

ISSN 2359-0386



6^a EDIÇÃO

ERI·GO

ESCOLA REGIONAL DE INFORMÁTICA

ANAIIS



6^a EDIÇÃO

ERI • **GO**

ESCOLA REGIONAL DE INFORMÁTICA

ANAIIS

2018

Goiânia - GO

Organização do Material

Vinicius da Cunha M. Borges (UFG/Goiânia)
Luciana de Oliveira Berretta (UFG/Goiânia)
Valdemar Vicente Graciano Neto (UFG/Goiânia)
Wellington Martins (UFG/Goiânia)

Editoração

Vinicius da Cunha M. Borges (UFG/Goiânia)

Dados Internacionais de Catalogação na Publicação (CIP) GPT/BC/UFG

E75 Escola Regional de Informática de Goiás (06. : 2018 : Goiânia, GO).
Anais da VI Escola Regional de Informática de Goiás [recurso eletrônico] / Organizadores, Antonio Carlos de Oliveira Júnior, e Marcelo Akira Inuzuka. – Dados eletrônicos. – Goiânia : Instituto de Informática, UFG, 2018.
362 p.

ISSN: 2359-0386

Disponível também em World Wide Web:

<<http://erigo.sbc.org.br/p/152-anais-eri-go>>

1. Informática – Estudo e ensino. 2. Informática – Aspectos sociais e econômicos. I. Oliveira Júnior, Antonio Carlos de. II. Inuzuka, Marcelo Akira. III. Título.

CDU: 004

COMITÊ

ORGANIZAÇÃO GERAL LOCAL

Antonio Carlos de Oliveira Júnior (UFG/Goiânia) – Coordenador
Marcelo Akira Inuzuka (UFG/Goiânia) – Vice-Coordenador

COORDENAÇÃO DO COMITÊ CIENTÍFICO

Vinicius da Cunha M. Borges (UFG/Goiânia) – Coordenador
Luciana de Oliveira Berretta (UFG/Goiânia) – Vice-Coordenadora
Valdemar Vicente Graciano Neto (UFG/Goiânia)
Wanderley Alencar (UFG/Goiânia)
Wellington Martins (UFG/Goiânia)

COMITÊ DE AVALIAÇÃO TÉCNICA E CIENTÍFICA

Adailton Araújo (UFG/Goiânia)
Adriano Braga (IFGoiano/Ceres)
Alessandro Silva (IFG/Anápolis)
Alba Melo (UnB)
Aleteia Araujo (UnB)
Alisson Cardoso (EMC/UFG)
Ana Claudia Loureiro Monção (UFG/Goiânia)
Anderson Corrêa de Lima (UFMS)
Antônio Oliveira Jr (UFG/Goiânia)
Bruno Silvestre (UFG/Goiânia)
Christiane Borges Santos (IFG/Luziânia)
Cleon Pereira Júnior (IFGoiano/Posse)
Cristiane Bastos Rocha Ferreira UFG/Goiânia)
Daniel Stefany Duarte Caetano (UFU)
Davi Viana (UFMA)
Deborah Fernandes (UFG/Goiânia)
Douglas Cordeiro (UFG)
Edison Morais (UFG/Goiânia)
Edmar Yokome (UEG/Santa Helena)
Eduardo Alchieri (UnB)
Eliomar Lima (UFG/Goiânia)
Fabian Cardoso (UniRV/Rio Verde)
Flávia Fernandes (UFCAT)
Flávio Vieira (EMC/UFG)
Flávio Geraldo Coelho Rocha (EMC/UFG)
George Teodoro (UnB)
Gilmar Arantes (UFG/Goiânia)
Hebert da Silva Coelho (INF/UFG)
Heitor Costa (UFLA)
Hugo do Nascimento (INF/UFG)
Ivairton Santos (UFMT)
Iwens Sene Jr (UFG/Goiânia)
José Maria David (UFJF)
Juliano Lopes de Oliveira (UFG/Goiânia)
Leandro Freitas (IFG/Inhumas)
Lenice Miranda Alves (UFG/Goiânia)
Leonardo Andrade Ribeiro (UFG/Goiânia)
Lina María Garcés Rodríguez (USP/SP)
Lívia Castro Degrossi (USP)
Lívia Ruback (UFRJ)
Luanna Lopes Lobato (UFCAT)
Luciana Berretta (UFG/Goiânia)

Luiz Henrique Rauber Rodrigues (Senac/Santa Cruz do Sul/RS)
Marcelo Akira Inuzuka (UFG/Goiânia)
Marcelo Ricardo Quinta (UFG/Goiânia)
Marcos Caetano (UnB)
Marcos De Sousa (PUC-GO)
Marcos Wagner de Souza Ribeiro (UFJ)
Marcus Vinícius Gonzaga Ferreira (UFG/Goiânia)
Márcio Dias (UFCAT)
Nádia Silva (UFG/Goiânia)
Núbia Silva (UFG)
Renato Ishii (UFMS)
Ricardo Franco (EMC/UFG)
Ricardo Santos (UFMS)
Robson Fidalgo (UFPE)
Rodrigo Cândido Borges (IFG/Inhumas)
Rodrigo Pereira dos Santos (UNIRIO/RJ)
Roney Lima (IFG/Jataí)
Rubens Pereira (UFG/Goiânia)
Scheila de Avila e Silva (Universidade de Caxias do Sul)
Sérgio Teixeira de Carvalho (UFG/Goiânia)
Sofia Paiva (UFSJ)
Taciana Novo Kudo (UFG/Goiânia)
Tércio Alberto Santos Filho (UFCAT)
Thales Nicolai Tavares (UFMS, RS)
Thiago Bittar (UFCAT)
Valdemar Vicente Graciano Neto (INF/UFG)
Vinicius da Cunha M. Borges (UFG/Goiânia)
Vinicius Garcia (UFMS, RS)
Vinicius Sebba Patto (UFG/Goiânia)
Wanderley Alencar (UFG/Goiânia)
Wendell Bento Geraldes (IFG/Luziânia)
Wellington Martins (UFG/Goiânia)
William Ferreira (UFG/Goiânia)

REVISORES EXTERNOS

Alex Pessoa (Agência Brasileira de Meteorologia)
Celso G. Camilo-Junior (UFG/Goiânia)
Fabrizio Soares Alves de Melo Nunes Soares (UFG/Goiânia)
Luis Pacheco (UnB)
Luiz Fernando Assis (USP)
Thierson Rosa (UFG/Goiânia)
Vinicius Medeiros (MediaLab/Labora/UFG)

E-mail de Contato: erigo@inf.ufg.br

DIRETOR DO INSTITUTO DE INFORMÁTICA – UFG
Sérgio Teixeira de Carvalho

Apresentação

A 6ª Edição da Escola Regional de Informática de Goiás (ERI-GO 2018) é um evento promovido pela UFG (Instituto de Informática). A ERI-GO 2018 foi realizada em consórcio com a 2ª Edição da Escola Regional de Sistemas de Informação de Goiás (ERSI-GO), 1ª Edição da Escola Regional de Alto Desempenho do Centro-Oeste (ERAD-CO) e com a 10ª etapa estadual de Goiás da Olimpíada Brasileira de Robótica, contribuindo assim para uma maior integração entre a comunidade goiana de pesquisadores e discussão de temas relevantes nas áreas de Informática em geral, aproximando estudantes, pesquisadores, profissionais e empresários. Esta edição do ERI-GO tem como objetivos: (a) estimular ações de pesquisa, mercado de trabalho, ensino e extensão sobre informática em geral nas instituições de ensino superior e técnico, como também centros de pesquisa no Estado de Goiás, baseado em trilhas dos principais cursos de Informática em Goiás, no Brasil e no mundo de acordo com ACM (Association for Computing Machinery), por exemplo, Sistemas de Informação, Engenharia de Software, Ciência da Computação e Engenharia e afins; (b) tornar público os projetos de informática desenvolvidos regionalmente; e (c) promover a cooperação entre produtores e usuários de tecnologia da informação e comunicação visando ampliar o desenvolvimento tecnológico e a inovação em soluções de informática no Estado. A ERI-GO 2018 ocorreu nos dias 14 e 15 de setembro de 2018, em Goiânia (Goiás). 15 (Quinze artigos) completos (apresentação oral), 5 (cinco) resumos expandidos (apresentação oral) e 4 (quatro) artigos curtos (apresentação de poster) foram incluídos nos anais selecionados pelo comitê de avaliação técnica e científica para a ERI 2018, os quais tratam de vários temas de pesquisa e desenvolvimento científico e tecnológico em Informática de forma geral. Além disso, os 3 (três) melhores artigos do evento foram selecionados e premiados, 1º Lugar com artigo intitulado "Reducing the Influence of Confounders on Predictive Models", 2º Lugar com artigo intitulado "Navegação indoor: um estudo de caso" e 3º Lugar com artigo intitulado "Intersecção de caminhos mais longos em produtos de grafos", a classificação levou em conta a avaliação da apresentação e a avaliação dos revisores. Os anais também estão disponibilizados eletronicamente nos sites: <http://www.inf.ufg.br/~erigo/4/o/anais-v-erigo-2018.pdf> e <http://erigo.sbc.org.br/p/152-anais-eri-go>.

SUMÁRIO

Artigos Completos

11

- 11 Representação semântica vetorial para análise de similaridade de documentos textuais
Kátia K. Cassiano
Douglas Cordeiro
- 25 Reducing the Influence of Confounders on Predictive Models
Ricardo Brito Alves
- 39 Seleção de Dispositivos para Caching Baseado no Interesse do Usuário através Comunicação D2D
Marcelo Almeida Silva
Antonio C. de Oliveira Jr.
Fausto da Silva Moraes
Kleber Vieira Cardoso
Vinicius da C. M. Borges
Waldir Moreira
- 53 Proposta de modelo de veículos elétricos para análise de consumo energético em redes veiculares
Leonardo Moraes Fleury
Divino Alves Ferreira Júnior
Vinicius da C. M. Borges
Antonio C. de Oliveira Jr.
- 65 Autenticação de dispositivos de Internet das Coisas baseada nas características do sinal eletromagnético
Marcos Felipe B. de Abreu
Pablo F. A. Sousa
Vinicius da C. M. Borges
Antonio C. de Oliveira Jr.
Kleber Vieira Cardoso
- 77 Otimização da trajetória de Drones como gateway de comunicação para sensores IoT
Lucas Soares Rodrigues
Kleber Vieira Cardoso
Vinicius da C. M. Borges
Leizer de Lima Pinto
Antonio C. de Oliveira Jr.
- 89 Análise Econômica-Financeira da Instalação de um Sistema de Energia Solar Fotovoltaica na Modalidade Geração Compartilhada em Goiás
Vitor Rezende Faria
Guilherme Moraes Spíndola
- 103 Análise e projeto de filtro FIR aplicado à tecnologia F-OFDM
Salatyel Fellipe da Silva
Leonardo L. de Assis Silva
Flávio Geraldo C. Rocha
- 117 Análise de Sentimentos de Conteúdos Textuais de Redes Sociais Por Meio de Modelos de Compressão de Dados
Jurandir J. D. Silva
Rogério Salvini

- 131 Utilização de Metaheurísticas na Resolução do Problema de Agendamento de Horários de Aulas numa Escola de Ensino Fundamental
Fábio Santanta de Oliveira
Wanderley de Souza Alencar
Hugo Alexandre Dantas do Nascimento
- 145 Intersecção de caminhos mais longos em produtos de grafos
Gabriel Matheus Faria de Almeida
Elisângela Silva Dias
- 159 Navegação Indoor – Um estudo de caso
Leandro da S. Taddeo
Paulo H. F. Alves
Francisco Klever A. Sobrinho
- 173 Protótipo de Monitoramento Remoto da Saúde do Idoso
Cícero Joasyo Mateus de Moura
Otávio Calaça Xavier
- 187 Experiências usando PERT-CPM para desenvolvimento ágil de software
Thiago Jabur Bittar
Patrícia Duarte Silva
Larissa P. C. dos Santos
Luanna Lopes Lobato
- 201 Parâmetros de Avaliação Multidimensionais: Aspiração para Modelos de Maturidade de Gestão do Conhecimento
Douglas de F. Alves
Eliomar A. de Lima

Resumos Expandidos

215

- 215 Caracterização de Mensagens Em Língua Portuguesa Com Traços de Racismo No Twitter
Rodolfo Costa Cezar da Silva
Deborah Silva Alves Fernandes
Marcio Giovane Cunha Fernandes
- 225 Extração de Regras de Associação de Dados Criminais no Município de Goiânia
Sílvio Passos Severino
Nádia Félix F. da Silva
- 237 Utilização de Sistemas Multiagentes em Rede de Energia Elétrica para Identificação de Irregularidades no Consumo
Gislainy Crisostomo Velasco
Luana Silva Carvalho
Vinícius Sebba Patto
- 249 Governança de TI no Setor Público: Um Estudo Sobre o Impacto na Efetividade Organizacional
Luciano da Costa Duarte
Eliomar A. de Lima
- 263 Avaliação de Metodologia para Implementação do BPM
Carlos Hernane de Oliveira
Nivaldo Pereira de Moraes Júnior
Eliomar Araújo de Lima

Artigos Curtos

277

- 277 Etnografia como base no desenvolvimento de um ERP voltado para o Hospital HAB
Pétersson S. Jesus
Alessandro R. Silva
- 287 Aplicabilidade da Norma ABNT NBR ISO/IEC 27002 em uma Empresa de Médio Porte
Yan S. Rocha
Eliomar A. de Lima
- 297 Aliança Defensivas Globais em Grafos Roda e em Prismas Complementares
Rafael de Castro Freitas
Elisângela Silva Dias
- 307 Análise de Desempenho de Sistemas MIMO Utilizando Codificação Alamouti e Diversidade na Recepção MRC
Vagner L. C. Silva
Flávio Henrique Teles Vieira
- 317 Sistema Integrado Educação Digital
Leidiane B. P. Rodrigues
Raimunda D. dos Santos

Fórum de Iniciação Científica e Pós-graduação da I ERAD-CO

327

- 327 Uma Plataforma Flexível de Computação de Alto Desempenho para Solução de Problemas de Sensoriamento Remoto
Leandro L. Parente
Roberto U. Paiva
Claudinei O. Santos
Evandro C. Taquary
Laerte G. Ferreira
- 333 FGSSJOIN: Um Algoritmo Paralelo Baseado em GPU para Problema da Junção por Similaridade em Bases de Dados Textuais
Rafael Quirino
Wellington Martins
Leonardo Andrade Ribeiro
- 339 A Fast Word2Vec implementation on manycore architectures for Text Representation and its applications
Leonardo A. Amorim
Mateus F. Freitas
Chayner C. Barros
Wellington S. Martins
- 345 Impulsionando Árvores Extremamente Aleatórias Ensacadas em Paralelo para Classificação de Textos
Julio C. B. Pires
Wellington S. Martins
Daniel X. de Sousa
- 351 Análise Custo x Desempenho em relação às Instâncias Preemptáveis em Provedores Públicos de Nuvem
Amanda L. Dantas
Jonas Soares
Matheus Vieira
Eduardo Costa
Aleteia Araujo

- 357 DStream: Plataforma de coleta, processamento e armazenamento de streams de dados de sensoriamento remoto
Sávio Oliveira
Vagner Rodrigues
Laerte Ferreira
Claudinei Santos
Wellington S. Martins

Index of Authors

363

Representação semântica vetorial para análise de similaridade de documentos textuais

Kátia Kelvis Cassiano¹, Douglas Farias Cordeiro¹

¹Faculdade de Informação e Comunicação – Universidade Federal de Goiás (UFG)
74.690-900 – Goiânia – GO – Brazil

{katiakelvis, cordeiro}@ufg.br

Abstract. *This paper is based on a Natural Language Processing tool called **Doc2Vec**, for the semantic representation of textual documents. The database of interest is composed of 44 (forty-four) undergraduate course final papers. Text mining techniques were used to process the digital archives of the monographs and generate the text. Each document is represented by word vectors and the model performs term inferences for semantic analysis. As a result, the similarity of the documents is in the form of a weighted graph, closeness between each element of the data sample.*

Resumo. *Este artigo descreve um modelo baseado em uma ferramenta de Processamento de Linguagem Natural denominada **Doc2Vec**, para representação semântica de documentos textuais. A base de dados de interesse é composta por 44 (quarenta e quatro) monografias de trabalhos de conclusão do curso Gestão da Informação da Universidade Federal de Goiás. Técnicas de mineração de texto foram utilizadas para processamento dos arquivos digitais das monografias e geração do corpus. Cada documento é representado por vetores de palavras e o modelo realiza inferência de termos para análise semântica. Como resultado, a similaridade dos documentos é apresentada na forma de um grafo ponderado, realçando a proximidade entre cada elemento da amostra de dados.*

1. Introdução

O registro de informações em documentos textuais é uma prática comum e cotidiana, sendo observada nos mais distintos cenários, desde o ambiente acadêmico, por meio dos registros das produções científicas em artigos, teses e dissertações, como também em ambientes organizacionais, em documentos, projetos, planejamentos, entre outros. O crescente volume deste tipo de registro de dados acaba por emergir desafios e obstáculos no que se refere à sua análise, demandando soluções que possam automatizar e fornecer informações de forma mais eficiente e eficaz, proporcionando a descoberta de conhecimento para o desenvolvimento de estratégias e tomadas de decisão mais assertivas.

Neste cenário, com os avanços alcançados no âmbito da mineração de dados, e neste caso mais específico, da mineração de textos, a aplicação de modelos provenientes desta área pode ser considerada como uma interessante alternativa, com resultados consideravelmente satisfatórios. Entretanto, para que seja possível alcançar soluções que sejam adequadas aos problemas a serem tratados, é necessário a realização de uma análise minuciosa acerca dos dados a serem considerados no problemas, assim como dos resultados desejados [Loh 2001]. Diante disso, através da

mineração de textos é possível explorar e desenvolver soluções voltadas à diferentes aspectos, tais como: classificação, segmentação, clusterização, associação, entre outros [Andrade 2015],[Castro and Ferrari 2016],[Hussein et al. 2015][Jurafsky and Martin 2009].

Diante deste contexto, a modelagem da similaridade semântica de documentos textuais pode ser destacada como uma área extremamente significativa para a ciência da informação, por possibilitar resultados nos campos da análise de sentimentos, recuperação e a geração de conhecimento para apoio à tomada de decisão, classificação de documentos, entre outros [Silva et al. 2016]. Entretanto, embora existam diversos modelos [Morais and Ambrósio 2007], tais como *bag-of-words*, *n-grama*, *skip-grama*, a subjetividade inerente ao conceito de similaridade acaba por tornar este um problema de complexa modelagem computacional. Neste sentido, é importante observar que a análise de documentos textuais requer uma modelagem consistente que considere aspectos relevantes como: ordenação, semântica e composicionalidade das palavras em uma sentença.

Neste sentido, este artigo traz um estudo de caso da aplicação de mineração de textos e representação semântica para o levantamento de similaridades em documentos textuais acadêmicos, como o propósito de fomentar dados e informações que possam ser posteriormente utilizadas em processos subsequentes de clusterização, agrupamento ou classificação. É importante ressaltar que o trabalho tem como propósito demonstrar a aplicação de uma das técnicas disponíveis para mineração de dados não estruturados, sem contudo desenvolver discussão comparativa com outras técnicas do estado da arte do PLN.

Primeiramente, são abordados os conceitos de Mineração de Texto e Processamento de Linguagem Natural. Em seguida, as características do algoritmo escolhido e a metodologia adotada para o desenvolvimento do trabalho, sendo apresentados detalhes da aplicação da técnica de Processamento de Linguagem Natural (PLN) denominada **Doc2Vec**, introduzido por [Le and Mikolov 2014], sobre uma base de dados textuais composta por um conjunto de monografias de trabalhos de conclusão do curso Gestão da Informação da Universidade Federal de Goiás. Por fim, para fins de análise, a matriz de similaridade dos documentos é apresentada na forma de um grafo ponderado, realçando a proximidade entre cada elemento da amostra de dados.

2. Mineração de Textos

A Mineração de Textos sempre se apresentou como um grande desafio no contexto da Ciência da Computação e da Ciência da Informação. Neste sentido, a constante evolução das técnicas de mineração de dados acabaram por proporcionar avanços consideráveis nesta área, permitindo a automatização de tarefas e processos que anteriormente demandavam um grande esforço humano. De forma geral, a mineração de documentos textuais pode ser vista como uma subárea da Recuperação de Informação (RI) [Salton and McGill 1983], na qual, através de um conjunto de rotinas de processamento e análise de padrões, a informação é recuperada a partir de dados textuais, gerando, consequentemente conhecimento. Neste sentido, é interessante observar que os fundamentos que regem esta área estão intrinsecamente ligados às definições de dado, informação e conhecimento.

Embora os conceitos de dado, informação e conhecimento possam ser considerados relativamente triviais, é comum encontrar interpretações que acabam por mesclar suas

definições. Diante disso, de modo a qualificar os conceitos relacionados à classificação de documentos, é importante fazer um parêntese e pontuar tais definições. De acordo com [Silva et al. 2016], dado pode ser descrito como algo bruto, sem contexto, ou seja, um símbolo ou um conjunto de símbolos quantificados ou quantificáveis. Por outro lado, a informação pode ser descrita como dados tratados, os quais possuem significado. É importante destacar que nem toda informação gerada é necessariamente útil e utilizada, e que nem todo dado processado é garantia de informação. Finalmente, conhecimento pode ser descrito como a informação explorada com algum propósito específico, ou seja, utilizada para, por exemplo, tomada de decisão, produção de cenários, entre outros.

Neste contexto, é possível concluir que a matéria prima essencial para o processo de análise são os dados. De forma geral, de acordo com [Silva et al. 2016], os dados podem ser classificados de duas formas: estruturados e não-estruturados. A identificação do tipo de dado é essencial para que o processo de mineração possa ser aplicado, uma vez que as peculiaridades de cada tipo de dado demandam rotinas específicas para seu processamento. Neste sentido, os dados estruturados podem ser descritos como aqueles que se referem ao resultado de processos transacionais, ou ainda de medição ou observação, sendo normalmente armazenados em uma tabela, ou em um formato que obedece um padrão pré-definido, facilmente compreensível por máquina. Por outro lado, os dados não-estruturados referem àqueles que não apresentam padrões pré-definidos, sendo necessárias rotinas adicionais para seu tratamento e processamento, tais como ocorre com textos, sons, vídeos e imagens.

A partir disso, no âmbito dos dados estruturados, a obtenção de informação e a geração de conhecimento podem ser alcançadas através do modelo KDD (do inglês, *Knowledge Discovery in Databases*) [Fayyad et al. 1996]. De forma geral, o KDD se refere a um conjunto de processos que vão desde a definição do problema até a geração dos resultados em si, ou seja, a geração de informação e conhecimento. Estes processos podem ser descritos como: Seleção, Pré-processamento, Transformação, Mineração de Dados, e Avaliação. Dentre tais atividades, a Mineração de Dados se destaca como a etapa mais importante na obtenção dos resultados, a qual pode ser definida como uma área multidisciplinar que proporciona, através de rotinas automatizadas, o reconhecimento de padrões, o levantamento de estatísticas, a visualização, e a extração de informações em grandes conjuntos de dados. Entretanto, devido às particularidades inerentes às bases de dados textuais, é necessário o emprego de técnicas e modelos especializados, proveniente da subárea denominada Mineração de Textos.

A Mineração de Textos pode ser descrita como um processo baseado na utilização de rotinas computacionais, para extração de padrões e conhecimento sobre conjuntos de dados textuais não-estruturados [Loh 2001]. É interessante observar que alguns autores consideram que a Mineração de Textos pode ser definida como a aplicação do modelo KDD sobre dados textuais [Morais and Ambrósio 2007], porém é importante destacar que a mineração sobre este tipo de dado demanda técnicas que extrapolam os processos tradicionais do KDD, compondo o que é denominado de KDT (do inglês, *Knowledge Discovery from Text*). Os avanços relacionados ao KDT incluem, entre outras coisas, contribuições que vão desde à exploração analítica em grandes bases de dados textuais, quantitativamente e qualitativamente, até a busca de informações em documentos, representadas através, por exemplo, de relacionamentos entre os termos mais relevan-

tes de um documento, análise de conteúdo, etc. Um exemplo deste tipo de aplicação é a análise de sentimentos em textos curtos aplicada no âmbito das mídias sociais [Silva 2016], ou ainda a exploração da mineração de textos para classificação de documentos [Hussein et al. 2015].

Neste sentido, é possível concluir que as técnicas provenientes do KDT podem trazer vantagens e benefícios tanto para problemas relacionados à volumes de dados extraídos em documentos eletrônicos da Internet, assim como em quaisquer outros cenários, como em documentos gerados por sistemas de informações gerenciais, ou similares. De acordo com [Beppler and Fernandes 2005], a maior parte dos dados de empresas estão contidos em documentos textuais, o que também acaba por potencializar a importância do KDT na geração de informação e conhecimento. Em termos práticos, pode-se descrever o KDT, por meio do conceito de Mineração de Textos, como um conjunto de processos que auxiliam na descoberta de conhecimento, ou seja, a realização de análises que transformam dados em informação, as quais devem então ser verificadas, analisadas e contextualizadas dentro de seus propósitos.

A Figura 1 apresenta as etapas do processo de Mineração de Textos. É importante destacar que o desenvolvimento da análise textual pode seguir, de forma geral, duas abordagens distintas: a Análise Estatística e Análise Semântica. A Análise Estatística trata aspectos mais relacionados à quantificação dos termos na base de dados, incluindo, por exemplo, estimativas, codificação e modelos de representação [Morais and Ambrósio 2007]. A Análise Semântica, por outro lado, explora aspectos mais ligados à representatividade de um termo em relação a outros, tendo sua base em PLN (Processamento de Linguagem Natural). A Análise Semântica é usada na proposta apresentada no presente trabalho, sendo descrita com maiores detalhes na Seção 3.

Entre as aplicações possíveis através da Mineração de Textos usando Análise Semântica, está a Classificação de Textos. De acordo com [Jurafsky and Martin 2009], a classificação de documentos textuais trata da identificação de um determinado documento d com relação a um conjunto de classes C , sendo $C = c_1, c_2, c_3, \dots, c_n$, em outras palavras, trata da determinação de qual classe c_i que o documento d pertence. Este tipo de aplicação pode ser utilizado em diversos contextos, tais como, definição de categorias, classificação de projetos, identificação de áreas de pesquisa, entre outros.

De forma geral, de acordo com [Andrade 2015], os modelos de classificação de documentos textuais podem ser classificados de duas formas: single-label, no qual cada documento está associado a apenas uma classe, ou multi-label, no qual um documento pode estar associado a uma ou mais classes. Neste sentido, uma das soluções para o processo de classificação é a utilização de algoritmos de aprendizagem baseados em dados previamente rotulados, o que é denominado de aprendizado supervisionado. Alguns exemplos de classificadores são: *Naives Bayes*, *Random Forest*, Árvores de Decisão e SVM (do inglês, *Support Vector Machine*).

Entretanto, em problemas em que não há disponibilidade de dados pré-rotulados, uma alternativa é a utilização de modelos não-supervisionados, como o caso do presente trabalho. Neste caso, pode ser realizada a clusterização de textos através de modelos de aprendizado não-supervisionado. De acordo com [Castro and Ferrari 2016], a

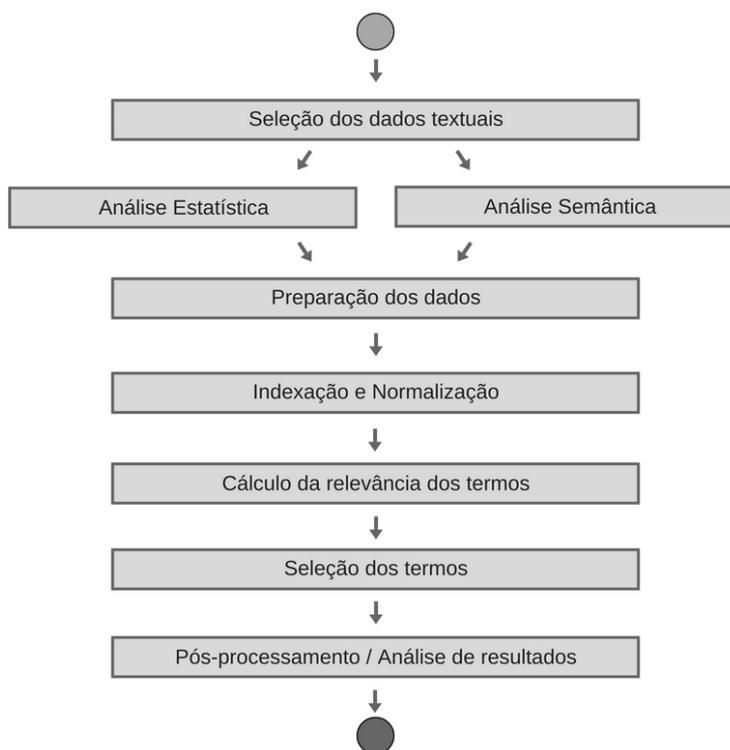


Figura 1. Processo de Mineração de Textos.

clusterização se refere a uma técnica de segmentação de dados com base na proximidade de padrões e tendências.

A clusterização é referenciada como um dos estudos iniciais realizados em processos de mineração, uma vez que através dela é possível realizar análise exploratória com a identificação de grupos ou classes que compartilham padrões e similaridades, os quais podem ser utilizados como entrada para outros métodos. No presente trabalho será explorada a técnica *Paragraph Vector* para levantamento de características relacionadas à similaridade de documentos textuais para propósitos de clusterização.

3. Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) constitui um conjunto de técnicas de inteligência artificial que possibilitam a comunicação entre seres humanos e computadores. No contexto do processamento de documentos textuais, PLN pode ser utilizado para diversos fins, tais como recuperação de informação, tradução, sumarização e classificação [Norvig and Russel 2011].

Em geral, um sistema PLN processa a linguagem em três níveis: morfológico, sintático e semântico. O sistema interpreta uma dada sentença por meio da análise de seu conteúdo léxico, das regras gramaticais e do significado dos termos, inicialmente armazenados em um dicionário. No contexto da inteligência artificial, o objetivo do PLN é expressar o conhecimento de forma tratável para os sistemas computacionais. Portanto, faz-se necessário e pertinente estudar formas de representação de sentenças e/ ou documentos em linguagem natural.

Segundo [Specia and Rino 2002], o processamento semântico é considerado um dos maiores desafios do PLN, dado a variabilidade morfológica e sintática das unidades lexicais e ainda, para o caso da língua portuguesa, ambiguidade de significados dos termos ou palavras. Neste sentido, um dos principais problemas sobre a interpretação semântica é referente ao processo composicional: as formas semânticas das entradas lexicais podem resultar em sentidos diversos para os termos ou palavras, dado que a interpretação depende de regras gramaticais.

A partir disso, o significado de um determinado termo de uma sentença depende dos significados dos demais termos e das regras gramaticais relacionadas. Tudo isso permite concluir que representação do conhecimento para o processamento de linguagem natural não é trivial, demandando a aplicação de técnicas e adequações específicas para cada tipo de problema tratado, levando-se em conta fatores como a representação semântica envolvida.

Neste contexto, no âmbito da linguística, o termo semântica caracteriza-se como um dos componentes do conhecimento, cuja função é representar o significado de uma sentença. No contexto do PLN, a extração do significado de expressões e a representação do mesmo por meio de estruturas semânticas constituem estágios do processamento [Specia and Rino 2002]. Quando se trata de documentos textuais, é importante observar que o significado depende não somente das informações semânticas dos termos isolados mas também da forma como essas informações são dispostas em um contexto. É importante destacar que existem uma série de técnicas e métodos baseados nestes conceitos e definições, entre as quais está a técnica **Doc2Vec**, utilizada no presente trabalho.

3.1. Paragraph Vector (Doc2Vec)

Paragraph Vector, também descrito como **Doc2Vec** e proposto por [Le and Mikolov 2014], é uma ferramenta de Processamento de Linguagem Natural para representar documentos e é uma generalização do método **Word2Vec**, introduzido por [Mikolov et al. 2013a]. De forma geral, esta técnica consiste de um modelo de aprendizado não-supervisionado que utiliza de representações vetoriais distribuídas dos termos ou palavras de um texto.

Neste sentido, os textos referentes à base de dados considerada podem ser de tamanho variável, de sentenças a documentos completos e, de uma forma geral, os vetores são treinados para prever palavras ou termos em um parágrafo e assim atribuir uma representação semântica. A partir disso, inicialmente, é realizado um mapeamento baseado em probabilidades, de forma que as palavras que possuem o mesmo sentido são distribuídas em um mesmo espaço vetorial, possibilitando realizar a distinção semântica entre as palavras de um parágrafo. Por exemplo, sejam os termos “poderoso”, “forte” e “Paris” em um segmento textual, esta representação vetorial pode descrever que o termo “poderoso” é semanticamente mais próximo de “forte” que de “Paris”.

Sequencialmente, o método realiza o mapeamento dos parágrafos para vetores distintos aos de palavras, concatenando o vetor do parágrafo com vários vetores de palavras presentes no parágrafo, com o objetivo de prever a próxima palavra no contexto considerado. Dessa forma, são levados em conta o tamanho variável das sentenças, a ordem das palavras e a semântica. Tanto os vetores de palavras, quanto os de parágrafo são treinados pela descida de gradiente estocástica e pós-propagação [Rumelhart et al. 1986]. É

importante destacar que enquanto os vetores de parágrafo são únicos entre os parágrafos, os vetores de palavras são compartilhados (o vetor de uma palavra é o mesmo para todos os parágrafos que possuem aquela palavra). No momento da predição, os vetores de parágrafo são inferidos corrigindo os vetores de palavra e treinando o novo vetor de parágrafo até a convergência. [Rumelhart et al. 1986] propuseram dois algoritmos para a geração de vetores de parágrafo:

- **PV-DM** (do inglês, *Distributed Memory Model of Paragraph Vectors*): neste modelo, cada parágrafo é mapeado para um vetor exclusivo, representado por uma coluna em uma matriz D. Cada palavra também é mapeada para um vetor exclusivo, representado por uma coluna em uma matriz W. A concatenação ou média do vetor de parágrafo com os vetores de palavras são utilizados para prever a próxima palavra em um contexto.

O vetor de parágrafo pode ser considerado uma pseudo-palavra e representa as informações que faltam no contexto atual, atuando como uma memória do tópico do parágrafo em questão. A Figura 2 apresenta a estrutura do modelo PV-DM, considerando a sentença “o gato senta no sofá”.

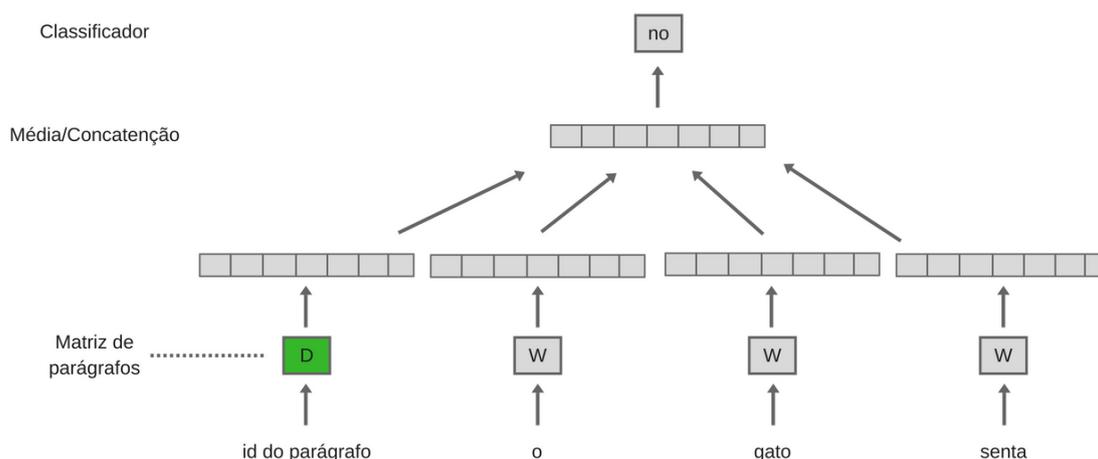


Figura 2. Estrutura do modelo PV-DM.

- **PV-DBOW** (do inglês, *Distributed Bag of Words version of Paragraph Vector*): neste modelo, as palavras de contexto são ignoradas na entrada e previstas aleatoriamente a partir do vetor de parágrafo. Na Figura 3 é apresentada a estrutura do modelo PV-DBOW.

4. Metodologia

O objetivo do presente trabalho é aplicar o método **Doc2Vec** com o propósito de construir uma representação da similaridade entre documentos textuais dentro uma base de dados textuais composta por documentos acadêmicos. Para tanto, a base de dados tratada neste trabalho é composta por 44 (quarenta e quatro) monografias do curso Gestão da Informação da Universidade Federal de Goiás (UFG), disponibilizadas em formato pdf (*Portable Document File*). No escopo deste trabalho, o termo documento pode fazer

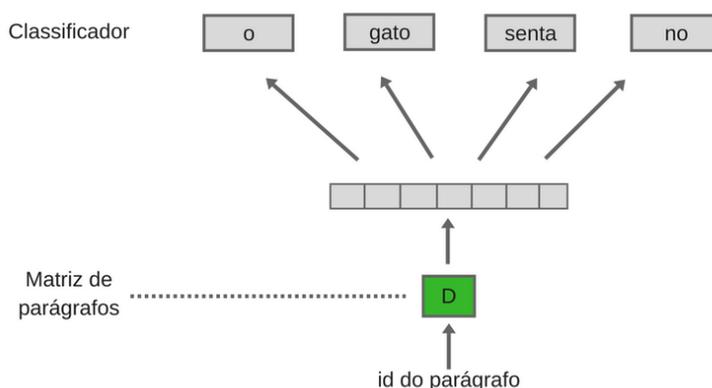


Figura 3. Estrutura do modelo PV-DBOW

referência a uma sentença, um artigo científico, uma monografia, um livro ou ainda um resumo destes. Em Processamento de Linguagem Natural (PLN) uma coleção ou conjunto de documentos é denominada *corpus*.

A partir disso, primeiramente, foi realizado o pré-processamento da base de dados para gerar, então, o *corpus* textual. Neste, cada documento é representado em uma linha e descrito por 5 (cinco) atributos, a saber:

- **doc-id**: identificador da monografia
- **autor**: nome do autor da monografia
- **orientador**: nome do professor orientador da monografia
- **título**: título da monografia
- **resumo**: resumo da monografia

Para reduzir a influência de termos ou palavras muito frequentes e irrelevantes no que tange à semântica, tais como preposições, pronomes e artigos, foram retiradas as *stopwords* de todo o *corpus* textual. Para processamento do *corpus* textual foi ainda adicionado um *label* para cada resumo, de forma a identificá-lo posteriormente (**doc-id**).

O algoritmo **Doc2Vec** foi utilizado para criação de um modelo de aprendizado de máquina do tipo não supervisionado. Neste sentido, inicialmente, uma função foi definida para ler e pré-processar o *corpus*. Este pré-processamento consiste, basicamente, em *tokenizar* o texto em termos ou palavras, gerando um vetor para cada documento. Ressalta-se que cada linha do *corpus* define um documento e o seu comprimento pode variar. Como resultado, tem-se a base de treinamento, denominada aqui de **train-corpus**, e a base de testes, denominada de **test-corpus**.

Para treinamento do modelo, a cada documento de **train-corpus** foi associada uma **tag**, sendo utilizado o número da linha que contém o documento, com base em zero. A parametrização do modelo foi realizada com base em trabalhos similares [Lee and Welsh 2005, Mikolov et al. 2013b, Le and Mikolov 2014]. A seguir são apresentados os valores para os principais parâmetros do modelo final validado:

- **size**: dimensionalidade dos vetores de palavras. Foi utilizado $size = 15$;
- **window**: quantidade de palavras anteriores e posteriores à palavra alvo. Este parâmetro é utilizado para a predição da palavra no contexto. Foi utilizado $window = 5$.

- **mincount**: define o valor mínimo de frequência, a partir do qual palavras ou termos serão consideradas, ou seja, atribui uma noção de relevância, descartando palavras com poucas ocorrências. Segundo revisão bibliográfica, utilizada para referência na definição dos parâmetros, a faixa de valores [10,20] é utilizada para *corpus* contendo dezenas de milhares a milhões de documentos. Empiricamente, tais trabalhos demonstraram que sem uma variedade de exemplos representativos de documentos (como é o caso do presente trabalho), a retenção de muitas palavras raras pode tornar o modelo pior. Portanto, foi definido empiricamente $mincount = 1$.
- **hs**: se 1, a função *softmax* hierárquico será utilizada para o treinamento do modelo. Foi utilizado $hs = 1$.
- **dm**: define o algoritmo de treinamento. Por padrão, o DBOW é usado ($dm = 0$). O outro é o DMPV ($dm = 1$). Foi utilizado $dm = 1$.
- **dm-concat**: se 1, usa a concatenação dos vetores de palavras e vetores de parágrafo para atribuir o contexto da representação. Portanto, foi utilizado $dm - concat = 1$.
- **iter**: número de iterações (épocas) de treinamento sobre **train-corpus**. Foi utilizado $iter = 100$. Tal valor foi definido empiricamente e considerando a dimensionalidade do conjunto de treinamento.

Para validar o modelo, a partir de **train-corpus** foram gerados vetores de termos para os documentos por meio de inferência. O algoritmo de inferência realiza a predição dos termos a partir dos vetores de palavras e então estes novos vetores podem ser comparados com os vetores do modelo treinado.

Basicamente, nesta abordagem, **train-corpus** é tratado como um dado desconhecido pelo modelo e, uma vez identificada semelhança entre os vetores (inferidos e modelados) obtém-se uma noção da consistência do modelo. Embora não seja um valor real de precisão, é uma forma de validar quão representativo é o modelo para as características dos documentos da base de dados. Neste sentido, a partir do modelo treinado e validado, foi realizada análise de similaridade semântica dos documentos. Um grafo ponderado foi gerado para ilustrar a relação entre os documentos. Para tanto, a matriz de similaridade foi convertida em um grafo direcional ponderado, no qual, para cada nó é associado um peso referente ao somatório dos valores de similaridade com os outros documentos

5. Resultados, discussões e conclusão

A representação semântica tratada no presente trabalho tem como objetivo a análise de similaridade dos documentos. Para melhor visualização dos resultados obtidos, é apresentado na Figura 4 o grafo ponderado da matriz de similaridade dos documentos. No referido grafo, cada nó se refere a um documento da base de dados, o qual por sua vez representa uma monografia de trabalho de conclusão do curso Gestão da Informação da UFG. Tendo como critério o tamanho do nó, pode-se ter uma ideia da distribuição dos documentos: nós de mesmo tamanho representam documentos similares entre si; quanto maior a diferença de tamanho entre um nó e outro, menor a similaridade entre os documentos representados pelos respectivos nós, ou seja, abordam conceitos de áreas de conhecimento distintas.

De uma forma geral é pertinente dizer que o grafo transmite uma ideia de agrupamento dos documentos de acordo com o tamanho do nó, e ainda que grande parte dos

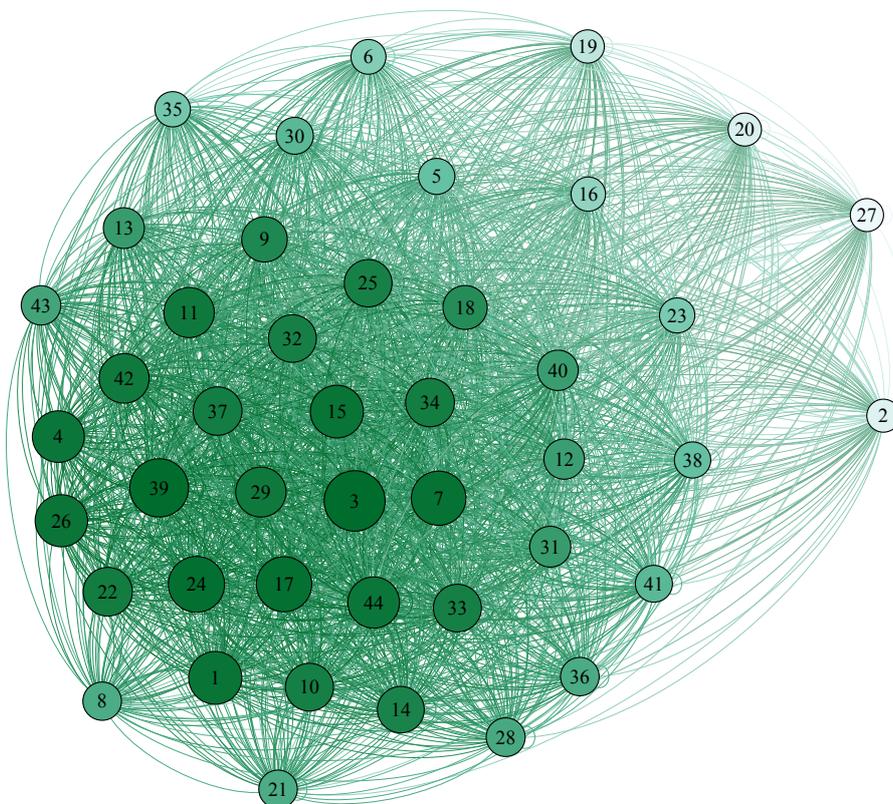


Figura 4. Grafo ponderado da similaridade dos documentos.

documentos estão concentrados em um dos grupos, que se difere dos demais pelos tamanhos dos nós. Ressalta-se ainda que tal observação é apenas uma inferência tomada a partir da análise visual, uma vez que não são objetivos do modelo proposto realizar a clusterização tampouco a classificação de documentos. No entanto, este tipo de hipótese é relevante e coerente quando da exploração de características e extração de padrões para subsidiar, posteriormente, processos de geração de conhecimento.

As Figuras 5 e 6, em Anexo, apresentam os resultados da análise de similaridade, bem como as características (**doc-id, título, resumo**) dos documentos representados pelos nós 42 e 5, respectivamente. Para cada documento alvo (**TARGET**), são apresentados seu identificador, o título e o resumo (atributos separados por ”) e, em seguida, estas mesmas características para os documentos de maior (**MOST**) e menor (**LEAST**) grau de similaridade.

As características do documento 42, por exemplo, validam a informação apresentada no grafo da 4. Os conceitos abordados no resumo do respectivo documento (definição de indicador/ gestão estratégica) são, de fato, significativamente distintos dos conceitos abordados no documento 2 (modelagem de processos) e coerentes com os conceitos abordados no documento 11 (análise de risco/ gestão estratégica).

O grau de similaridade é uma característica importante na presente análise. Observa-se na Figura 6 que os graus de similaridade para o documento 5 são em valor

absoluto relativamente menores que os apresentados para o documento 42 - vide grau apresentado para o documento 20, que é o mais similar ao documento 5. Tal resultado permite o seguinte apontamento: o fato de um documento ser similar ao outro não significa, necessariamente, que tratam do mesmo tema ou área de conhecimento do curso, mas pode ser uma característica relevante para identificação de grupos em processos de clusterização de documentos.

No que tange ao tamanho dos nós, observa-se no grafo que os nós referentes aos documentos 5 e 20 tem praticamente o mesmo tamanho e são relativamente menores que o nó referente ao documento 42. Tal fato possibilita definir uma hipótese de que os documentos 5 e 20 pertencem a um mesmo grupo e, ainda, por similaridade, o documento 42 é distinto do documento 20, assim como a análise da 5 demonstrou ser distinto do documento 2. A hipótese definida anteriormente é validada pelas informações apresentadas na Figura 6, conforme descrito anteriormente.

De uma forma geral, o modelo apresentado no presente trabalho é potencialmente relevante do ponto de vista de aplicação de técnicas e ferramentas de Processamento de Linguagem Natural para Recuperação da Informação. Foi realizada análise do conteúdo dos resumos das monografias do curso Gestão da Informação da UFG numa tentativa de representá-las semanticamente, atribuindo assim significado.

Uma abordagem potencial para trabalhos futuros é utilizar os resultados para a criação de coleções de dados científicos e, dessa forma, obter não apenas um estoque de dados mas uma organização analítica de dados focada na extração de características e geração de conhecimento. É possível, ainda, utilizar os resultados para classificação e correlação de áreas científicas. Nesse sentido, a solução abre um leque de oportunidades para análise de dados, recuperação da informação e geração de conhecimento no âmbito do Processamento de Linguagem Natural de textos científicos.

Referências

- Andrade, P. H. M. A. (2015). Aplicação de técnicas de mineração de textos para classificação de documentos: um estudo da automatização da triagem de denúncias na cgu. Dissertação de mestrado, Instituto de Ciências Exatas - Universidade de Brasília, Brasília.
- Beppler, M. D. and Fernandes, A. M. R. (2005). Aplicação de text mining para extração de conhecimento jurisprudencial. In *Anais do I Congresso Sul Catarinense de Computação*.
- Castro, L. N. and Ferrari, D. G. (2016). *Introdução à Mineração de Dados: conceitos básicos, algoritmos e aplicações*. Editora Saraiva, São Paulo.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases*. American Association for Artificial Intelligence.
- Hussein, H., Alaaeldin, H., and Hassan, M. (2015). Selection criteria for text mining approaches. *Computers in Human Behavior*, 51:729–733.
- Jurafsky, D. and Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall, Upper Saddle River, NJ, USA.

- Le, Q. V. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32, Beijing, China.
- Lee, M. D. and Welsh, M. (2005). An empirical evaluation of models of text document similarity. In *CogSci2005*, pages 1254–1259.
- Loh, S. (2001). *Abordagem baseada em conceitos para descoberta de conhecimento em textos*. Tese de doutorado, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *Proceedings of Workshop at ICLR*, abs/1301.3781.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, volume 2, pages 3111–3119.
- Morais, E. A. M. and Ambrósio, A. P. L. (2007). Mineração de textos. Technical report, Universidade Federal de Goiás, Goiânia.
- Norvig, P. and Russel, S. (2011). *Inteligência Artificial*. Elsevier, 3 edition.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back propagating errors. *Nature*, 323:533–536.
- Salton, G. and McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. John Wiley & Sons, New York.
- Silva, L. A., Peres, S. M., and Boscaroli, C. (2016). *Introdução à Mineração de Dados: com aplicações em R*. Elsevier, Rio de Janeiro.
- Silva, N. F. F. (2016). *Análise de sentimentos em textos curtos provenientes de redes sociais*. Tese de doutorado, Instituto de Ciências Matemáticas e de Computação, São Carlos.
- Specia, L. and Rino, L. H. (2002). Representação semântica: Alguns modelos ilustrativos. Technical report, NILC - ICMC-USP.

TARGET : 42 | Tornando tangível a participação social na Rede Humaniza SUS: proposta de indicador | Considerando que seria de interesse para as políticas públicas a adoção de mecanismos que explicitem níveis de interação horizontal, proporcionando ativação da inteligência coletiva e gerando propostas e soluções mais elaboradas, a investigação tem como objetivo propor um indicador que torne visível a participação social dos usuários de uma rede social online. Método: Baseia-se em análise de literatura pertinente relativa ao histórico de participação social no Brasil, no Sistema Único de Saúde e, particularmente à Rede Humaniza Sus (RHS). Apresentam-se características de indicadores que respaldam a discussão dos elementos que compõem o indicador proposto. Resultados: Foram identificadas as variáveis, considerando-se os elementos de postagem, comentários e votação e, com base nestes, criou-se um indicador ponderado de “participação social”. Conclusão: O indicador proposto apresentou quais são os usuários mais importantes da rede e também pode auxiliar em traçar o tipo de perfil que esse usuário possui

SIMILAR/DISSIMILAR DOCS PER MODEL Doc2Vec(dm/c,d15,n5,w5,s1e-05):

MOST (0.8468265533447266): 11 | Gestão da Informação e análise de risco: proposta metodológica baseada em assimetria de informação e rating | O presente trabalho trata da identificação de elementos que contribuem para a ocorrência de riscos no processo de aplicação do Modelo Processual de Administração da Informação proposto por Choo (2006). Mesmo com diversas análises dos processos que compõe o Modelo registradas na literatura, estas análises não são abordadas sob o ponto de vista dos elementos constituintes da Teoria Econômica da Informação (Assimetria de informação e Rating). Diante dessa ausência na literatura, propôs-se um estudo para levantar os elementos que podem gerar riscos para a implementação eficiente do Modelo Processual de Administração da Informação. Num segundo momento este estudo procurou caracterizar instrumentos que possam auxiliar de forma eficaz na implantação do Modelo em estudo. O estudo baseou-se, por um lado, essencialmente em pesquisa bibliográfica e, de outro lado, utilizou-se a chamada “pesquisa básica” que consiste, genericamente, no fato de gerar novos conhecimentos. Ao final pode-se inferir, que, diante da sociedade contemporânea, o Modelo proposto por Choo (2006) é flexível e útil para se tomar decisões quando são criadas novas respostas a situações não esperadas. O Modelo Processual de Administração da Informação é sujeito a vários fatores/elementos dinâmicos (interrupções, ciclos de compreensão, ciclos de fracasso) que podem inviabilizar sua aplicação ou até mesmo impossibilitar que ele tenha êxito ao ser utilizado para tomar uma determinada decisão na organização.

LEAST (-0.674375057220459): 2 | Modelagem de processos aplicada à gestão de organizações públicas: Um estudo de caso em uma IFES | Este trabalho tem como objetivo principal realizar um estudo sobre a aplicação de gestão de processos de negócios em instituições públicas, através de um estudo de caso em um problema específico de uma Instituição Federal de Ensino Superior. Neste sentido, durante o trabalho serão abordados os principais conceitos, ferramentas e metodologias relacionados à informação, gestão de processos e gestão de organizações públicas, abordando questões como a cultura organizacional características destas instituições. A partir disso, é apresentado um estudo de caso sobre um processo específico, a avaliação de estágio probatório de docentes na Universidade Federal de Goiás.

Figura 5. Análise de similaridade do documento 42.

TARGET : 5 | Modelagem de processos e serviços oferecidos em Bibliotecas Universitárias: aplicações em serviços de atendimento | A pesquisa tem por objetivo analisar os processos dos serviços oferecidos pelo Sistema de Bibliotecas/SIBI, da Universidade Federal de Goiás/UFG, especificamente, o serviço de quitação de multas e o de empréstimos entre bibliotecas. Faz-se revisão de literatura contextualizando a pesquisa e abordando os principais conceitos pertinentes. A metodologia adotada é classificada como de natureza aplicada, pois tem como objetivo gerar conhecimentos para auxiliar a solução de problemas específicos. A análise completa foi desenvolvida por meio da identificação, mapeamento e modelagem dos processos citados. Conclui-se, a partir destas considerações, que a gestão por processos é de fundamental importância na gestão de qualquer organização.

SIMILAR/DISSIMILAR DOCS PER MODEL Doc2Vec(dm/c,d15,n5,w5,s1e-05):

MOST (0.2787836194038391): 20 | Processamento de Sinais Digitais Aplicado à Transmissão de Mensagens de Áudio Criptografadas | Este trabalho tem como propósito demonstrar a importância da segurança da informação junto aos métodos de processamento de sinais vinculando a criptografia. Essa demonstração tem como efeito relatar o quanto a segurança da informação é necessária e eficiente para as organizações em suas tomadas de decisões. Sendo assim, este trabalho tem como proposta o envio e recebimento de mensagens criptografadas através de arquivos de áudio com vínculo à segurança informacional.

LEAST (-0.4536975622177124): 40 | Autopoiese do conhecimento em redes sociais de conversação: em busca de evidências das implicações da composição biológica na dinâmica social de relacionamento em rede | O advento da chamada Era do Conhecimento intensificou os estudos acerca da importância da produção cognitiva, sobretudo nos ambientes empresariais. No entanto, para que se possa compreender como ocorrem os processos de realização do real e de como se dá a construção de conhecimento, é necessário analisar a relação entre processos oriundos do modo de ser biológico do ser humano em suas vivências autopoieticas com a existência do ser em comunidade. Esta produção científica busca investigar a relação entre a composição biológica do ser humano e sua dinâmica relacional mundo contemporâneo e suas implicações na construção de conhecimento do indivíduo. Apontando que o modo de intensificação da comunicação em rede contemporânea tem profunda relação com a maneira de sustentação de visões individuais e coletivas na realização do ser.

Figura 6. Análise de similaridade do documento 5.

Reducing the Influence of Confounders on Predictive Models

Ricardo Brito Alves¹

¹Department of Electrical Engineering – Pontifícia Universidade Católica de Minas Gerais
– Belo Horizonte – MG – Brazil

rbalves@sga.pucminas.br

Abstract. *The analysis of Big Data has become so important with the progressive increase of the information stored in digital media. Extracting more value from diversified and unstructured data is really challenging. With the help of predictive models, it is possible to find new patterns and trends that could be innovation bases. Predictive models need to have a relevant reliability rate to aid us in decision-making processes. In this context, this article discusses the influence of confounding variables on predictive models and proposes techniques for identifying and minimizing their effect. Through a database with information collected in a hospital, it was possible to construct a predictive model, to identify possible confounding variables, to apply a technique to minimize its influences and to evaluate the accuracy of the model through machine learning techniques. The result was an efficient prediction model.*

Keywords: *Big Data, Predictive Model, Confounders, Multicollinearity, Machine Learning.*

1. Introduction

The objective of this study was to apply techniques that identify and minimize the influence of the confounding variables so that the predictive model obtain maximum efficiency in the prediction process. Confusion can arise when in an unbalanced sample an interfering variable distorts the association between an exposure variable and a response variable, changing the strength or even the direction. Build a predictive model is not simply to write an equation but rather to perform more consistent analyzes on the data and their relationships. As result in our study, through the use of Pearson and Spearman Correlation, VIF - Variance Inflation Factor, linear regression by the outcome it was possible to minimize the effect of a potential confounding variable and through classification techniques such as SVM and Logistic Regression, we verified the efficiency of the model after reducing the influence of the confounding factor.

The Data mining has been used to exploit large amounts of data in search of consistent patterns, such as association rules or time sequences. Once systematic relationships between variables have been detected, it is possible to obtain subsets of data [Han et al. 2011, Maurizio 2011]. Predicting or inferring about causality are two of the great scientific motivations of verifying the statistical association between variables. As part of this process, we have the predictive analysis that has been applied in many areas and is able to use data, algorithms, and machine learning techniques in an attempt to predict future situations. The increasing information digitization by society, ease of storage and the possibility of processing large volumes of data has made this analysis more accessible to those who seek to know it [Waller and Fawcett 2013].

Predictive modelling uses statistics to predict outcomes and in most often the event one wants to predict is in the future, but predictive modelling can be applied to any type of unknown event, regardless of when it occurred [Leatherman et al. 2018].

A predictive model is nothing more than a mathematical function that, applied to a set of data, can identify hidden patterns and can enable the prediction of scenarios that interest us with a relative margin of accuracy. Predictive models can either be used directly to estimate a response (output) given a defined set of characteristics (input), or indirectly to drive the choice of decision rules [Steyerberg 2009].

In many cases, the model is chosen on the basis of detection theory to try to guess the probability of an outcome given a set amount of input data. Models can use one or more classifiers. Most of the regression models can be used for prediction purposes. Broadly speaking, there are two classes of predictive models: parametric and non-parametric. A third class, semi-parametric models, includes features of both. Parametric models make "specific assumptions with regard to one or more of the population parameters that characterize the underlying distributions", while non-parametric regressions make fewer assumptions than their parametric counterparts [Sheskin 2011].

As part of predictive modelling, it is necessary to analyze variables and their correlation. A confounding variable or confounding factor is a variable that influences both a dependent variable and an independent variable, causing a spurious association. It is a situation where the effects of two variables are difficult to separate from each other [Austin 2011]. To be considered as confusion, the variable must be associated with the outcome and be associated with exposure, and not be part of the causal chain linking exposure to outcome. Variables with a bias of confusion are potential effect modifiers, that is when the effect of an exposure on an outcome varies according to the level of a third variable. Therefore, when between two exposures there is a potentiated of one to another, we can say that the effect has been changed. One factor alone has one effect, but in the presence of another, its effect is increased. The union of two factors results in a different risk than simply the effect of one plus the effect of the another. Unlike most biases, confusion bias can be controlled after data collection, provided they have been collected in a way that allows for such control. When the effect is suspected, a stratified analysis of the confounding factor should be done, showing that the risks between the different strata are similar to each other but different from the gross risk. By reducing the effects of bias, the possibility of misclassification is reduced.

In the scope of statistics, for a sample, we can have two types of errors, random or by chance, that we can say are associated with the precision of the measurements and systematic or bias, which are associated with the validity of the measurements. Bias is a random distortion as a result of some sampling process. Also known as "bias deviation", it consists of the difference between the average value of a statistical estimator and the value that it intends to estimate. Often, to correct the deviation, the estimator is changed. Systematic errors or bias can be classified as selection bias, information bias, and confounding.

- Selection bias: the measure of association estimated in the study is distorted because of the way in which individuals are selected to make up the study population.
- Information Bias: The measure of association estimated in the study is distorted due to errors in the way information on exposure and/or disease has been obtained.

- Confusion or confounding: part of the observed association arises from the existence of one or more variables, called confounding.

This article was structured in introduction, related works, contextualization, materials and methods, results and conclusion. In the contextualization, it was trying to clarify more the subjects approached. In materials and methods, there is a more detailed description of the methods used, divided into three stages: building a predictive model and reduction of the characteristics to be worked, identification of potential confounding variables and the application of a technique to correct the effect of confounding variables using linear regression. The results have been discussed obeying the same three stages.

2. Related Works

[Li et al. 2011] present us in this article a Support Vector Machine classifier that can correct the prediction for observed confounding factors. This is achieved by minimizing the statistical dependence between the classifier and the confounding factors.

[Li and Zhang 2015] analyzes the existence of confounders such as population structure in genome-wide association study makes it difficult to apply machine learning methods directly to solve biological problems.

[Low et al. 2016] show us a observational studies from EHR in real time, particularly in emergencies, rapid confounder control methods that can handle numerous variables and adjust for biases are imperative. This study compares the performance of 18 automatic confounder control methods.

[Schnitzer et al. 2016] investigates the appropriateness of the integration of flexible propensity score modeling (nonparametric or machine learning approaches) in semiparametric models for the estimation of a causal quantity, such as the mean outcome under treatment.

[Berk et al. 2018] provide an alternative methods draw on work in econometrics and statistics from several decades ago, updated with the most recent thinking to provide a way to properly work with misspecified models. They show how asymptotically, unbiased regression estimates can be obtained along with valid standard errors.

3. Materials and Methods

Data from patients were collected at a hospital in India. We attempted to minimize the effects of confounding factors, by the multivariate analysis method where the predictor variables are analyzed simultaneously so that the effect of each variable is adjusted for the effect of the others. Thus, it tried to identify the direct effect of each variable in the prediction of the outcome, an effect that is independent of other variables. This is called as an independent association. Confusion is as an intermediate variable with a false association between two other variables. Adjusting for this confounding variable, we will know about the relationship between predictor variable and outcome. This concept of a need for adjustment for confusion by multivariate analysis applies to the construction of a predictive model. The multivariate analysis makes the adjustment for the predictive variables, determining an independent association, a necessary condition for the variable to be part of a predictive model or considered one of the causes of the outcome.

The database was extracted from UCI - Machine Learning Repository [<http://mlr.cs.umass.edu/ml/> 2017], which aimed to classify two groups, carrier and non-

carrier to chronic kidney disease (CKD). CKD is a progressive pathology, with a high mortality rate. Discovering the disease before reaching the chronic phase and progressing to dialysis is fundamental to guarantee quality of life and to increase survival [Jena and Kamila 2015, Sinha and Sinha 2015, Kumar 2016]. According to the UCI - Machine Learning Repository data set extracted over a 2-month period in India, there are 400 patient records with 24 attributes to predict CKD or not CKD (age, blood pressure, relative density, albumin, sugar, red blood cells, pus cell, pus cell clusters, bacteria, random blood glucose, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, anemia).

As a statistical analysis tool, was used public domain software R [Team 2014]. R software was used mainly in multivariate analyses, which estimated the association of each independent variable with the dependent variable, after adjusting for the effects of all other variables.

As mentioned before, this study was divided into three stages: building a predictive model and selection of variables more adherent to the predictive model, identification of potential confounding variables and the application of a technique to correct the effect of confounding variables. However, before all, one of the important steps for model development was to eliminate records with undefined values, leaving 158 sets of data for use in this work without missing values.

Among the 25 variables in the database, the variable to be predicted is the class variable. The others are explanatory or independent variables: age, blood.pressure, specific.gravity, albumin, sugar, red.blood.cells, pus.cell, pus.cell.clumps, bacteria, blood.glucose.random, blood.urea, serum.creatinine, sodium, potassium, hemoglobin, packed.cell.volume, white.blood.cell.count,, red.blood.cell.count, hypertension, diabetes.mellitus, coronary.artery.disease, appetite, pedal.edema, anemia.

3.1. Stage 1

The first stage was about the predictive model construction. Linear regression was used in order to build an equation to estimate the conditional (expected value) of a variable y , given the values of some other variables. In a set of 25 variables, the "class" dichotomic variable was considered as the expected value that has CKD and not CKD information. The others were considered as independent or explicative variables. Thus, the model takes on the standard form (1) which describes a line with slope β_k and y-intercept β_0 involving the error term ϵ_i :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \epsilon_i \quad (1)$$

To optimize the predictive model it was necessary to analyze all independent variables in order to select the ones with the most relevant characteristics of the predictive model. It was an important step in the application of machine learning methods. Data sets are often described with many variables and some of these variables may be irrelevant to classification, and their use is a disadvantage for constructing models. Firstly, analyses of the set of variables were performed using the Pearson correlation coefficient and its respective p-value and the Spearman correlation coefficient in the search for the set of

characteristics more adherent to the predictive model. The correlation coefficient measures the degree to which two variables tend to change together. The coefficient describes the strength and direction of the relationship. Pearson’s correlation evaluates the linear relationship between two continuous variables while Spearman’s correlation evaluates the monotonic relationship between two continuous or ordinal variables. In a monotonic relationship, the variables tend to change together but not necessarily at a constant rate.

Then a new analysis was done using CFS - Correlation Feature Selection. This analysis could provide a numerical estimate of the importance of resource. The random forest classification algorithm could be run without setting parameters and provides a numerical estimate of the importance of the resource. It is a joint method in which classification is performed by voting on multiple nonindependent weak classifiers.

3.2. Stage 2

The second stage had the objective of identifying potentially confounding variable following the steps below.

- Identify variables with strong correlations, which tend to be redundant in the sense of adding little to the model and being potential confounding variables.
- Adjust relative risk analysis of more than 10 % relative to the gross relative risk, which suggests that it is a confounding variable. By the use of the linear regression, it is possible to analyze the influence of the confounding variables. The first linear regression model considers all variables, including those of confusion. The second linear regression model does not consider the confounding variables. If the estimates of the two models vary by more than 10 %, it is suggested that the variable is confusing [Austin 2011].
- Identify multicollinearity that is defined as the presence of a high degree of correlation between the independent variables [García et al. 2015]. Presence of multicollinearity means that there is a presence of collinearity between the variables, so any plane along the data dispersion axis will be unstable and results in the same sum of squares of the error. Variance Inflation Factor (VIF) was use to detect multicollinearity, assuming that the variables are centered and standardized, we have $R = (X^T X)^{-1}$ in which the diagonal elements of this matrix are called VIF and represent the increment of the variance due to the presence of multicollinearity [Montgomery et al. 2012]. The VIF can be calculated using the equation (2).

$$VIF_j = \frac{1}{1 - R_j^2} \text{ and } j = 1, 2, \dots, p \quad (2)$$

Where p is the number of predictor variables; R^2 is the multiple correlation coefficients resulting from the regression of X_j on the other $p - 1$ regressors.

3.3. Stage 3

The third stage was the application of a technique to correct the effect of the confounding variable by the use of linear regression by the outcome. The linear regression model was used to correct the effect of the confounding variable.

NSUB is the number of subjects, NVAR is the number of variables and NCOV is the number of covariables. For each variable y in Y and $nCovariables$ we have:

- y = observed value for the dependent variable Y i -nth level of the independent variable $nCOV$.
- b = regression constant, which represents the intercept of the line with the y -axis.
- a = regression coefficient, which is the variation of Y as a function of the variation of a unit of variable a .
- C = coefficient of regression covariates.

Y is a $NSUB \times NVAR$ matrix C is a $NSUB \times NCOV$ matrix

For each variable y in Y , with $NSUB$ samples:

Using the linear regression, we look for the coefficients of b , a and C (3).

$$y = b + a[1] C[1] + \dots + a[nCOV] C[nCOV] \quad (3)$$

Correction of the variable y to each line using the equation (4):

$$y = y + mean(y) - b - a[1] C[1] - \dots - a[nCOV] C[nCOV] \quad (4)$$

To check results at the corrected database, it was necessary to apply classification techniques to compare the accuracy of the predictive model. For this were used two technics, SVM and Logistic Regression. In technics, the corrected database was used, removing the predicted variable. New classifications were made that could be compared with the original classification. In the logistic regression, the database was separated on a training base and a test base, being the test base used for classification. Results of logistic regression was checked by confusion matrix.

In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

Computing the SVM classifier amounts to minimizing an expression of the form (5):

$$\left[\frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i(\vec{w} \cdot \vec{x}_i - b)) \right] + \lambda \|\vec{w}\|^2 \quad (5)$$

Where the parameter λ determines the tradeoff between increasing the margin-size and ensuring that the \vec{x}_i lie on the correct side of the margin. Thus, for sufficiently small values of λ , the second term in the loss function will become negligible, hence, it will behave similar to to the hard-margin SVM, if the input data are linearly classifiable, but will still learn if a classification rule is viable or not.

The logistic model or logit model is a statistical model that is usually taken to apply to a binary dependent variable. In regression analysis, logistic regression or logit regression is estimating the parameters of a logistic model. More formally, a logistic model is one where the log-odds of the probability of an event is a linear combination of independent or predictor variables. The two possible dependent variable values are often labelled as "0" and "1", which represent outcomes. The binary logistic regression model

can be generalized to more than two levels of the dependent variable: categorical outputs with more than two values are modelled by multinomial logistic regression, and if the multiple categories are ordered, by ordinal logistic regression.

The logistic regression can be understood simply as finding the β parameters (6):

$$\begin{cases} 1 & \beta_0 + \beta_1 x + \varepsilon > 0 \\ 0 & \text{else} \end{cases} \quad (6)$$

Where $A(x,t)$ and $d(x,t) =: D(x,t)$ are constant rank singular, possibly rectangular matrix functions which are in some sense well matched.

4. Results

As mentioned, the discussion of the results were divided according to the three stages of the study.

4.1. Stage 1

Firstly, analyses of the set of variables were performed using the Pearson correlation coefficient and its respective p-value and the Spearman correlation coefficient in the search for the set of characteristics more adherent to the predictive model. The correlation coefficient measures the degree to which two variables tend to change together.

The correlation between variables was evaluated using the Pearson correlation coefficient and the Spearman correlation coefficient, which measure the degree of correlation and the direction of this correlation. 17 of the 24 independent variables had significant correlations and significance with p-value < 0.05 as we can see in table 1.

Staying with many relevant variables, it was important to apply another method, the Correlation Feature Selection (CFS), which provides a numerical estimate of the importance of the resource. In this case, the result indicated the variables in order of importance for the predictive model. 23 attributes was confirmed important: albumin, anemia, appetite, bacteria, blood.glucose.random and 18 more. It is possible to verify all attributes by importance using the CFS method, which is represented in Figure 1.

Table 1. Pearson and Spearman Correlations

Class	Pearson Correlation	p-Value	Spearman Correlation
specific.gravity	0.790101503	2.20E-16	0.697503862
albumin	-0.925816188	2.20E-16	-0.97014838
sugar	-0.510615424	7.18E-09	-0.58507828
red.blood.cells	0.586390502	5.72E-13	0.586390502
pus.cell	0.775387573	2.20E-16	0.775387573
pus.cell.clumps	-0.50991462	7.75E-09	-0.50991462
blood.glucose.random	-0.591217296	2.88E-13	-0.44888816
blood.urea	-0.677610639	2.20E-16	-0.64139796
serum.creatinine	-0.820232963	2.20E-16	-0.7432218
sodium	0.640901871	2.20E-16	0.61444839
packed.cell.volume	-0.82798329	2.20E-16	0.740012593
red.blood.cell.count	0.736366841	2.20E-16	0.701921008
hypertension	-0.856334238	2.20E-16	-0.85633424
diabetes.mellitus	-0.758965495	2.20E-16	-0.75896549
appetite	-0.604622205	2.20E-16	-0.6046222
pedal.edema	-0.622572806	2.20E-16	-0.62257281
anemia	-0.548947121	8.16E-11	-0.54894712

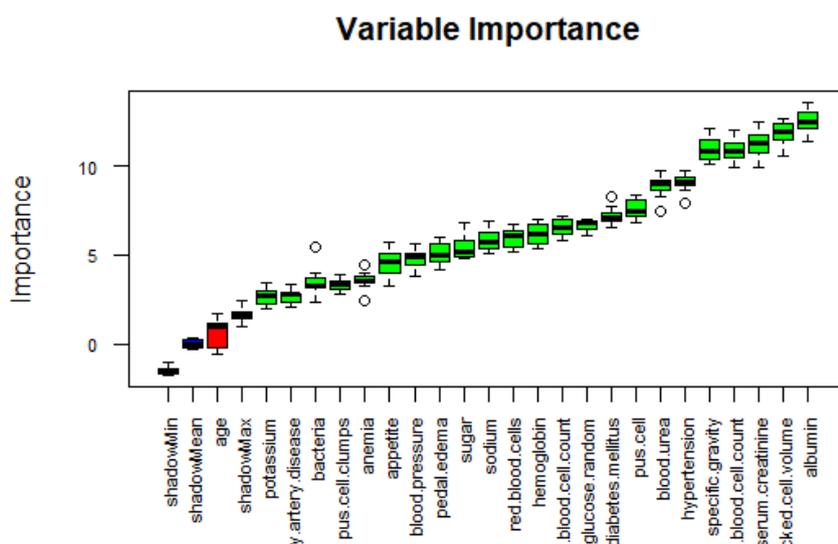


Figure 1. Correlation Feature Selection

Below in table 2 we have the list of the variables in order of significance and relevance. For the purpose of the study, was selected the variables considered most relevant to compose the predictive model.

Table 2. Order of Significance and Relevance

Significance	Relevance	Variable
1	-	blood.pressure
2	4	specific.gravity
3	1	albumin
4	15	sugar
5	14	red.blood.cells
6	8	pus.cell
7	-	pus.cell.clumps
8	-	bacteria
9	10	blood.glucose.random
10	7	blood.urea
11	3	serum.creatinine
12	13	sodium
13	-	potassium
14	11	hemoglobin
15	2	packed.cell.volume
16	12	white.blood.cell.count
17	5	red.blood.cell.count
18	6	hypertension
19	9	diabetes.mellitus
20	-	coronary.artery.disease
21	-	appetite
22	-	pedal.edema
23	-	anemia

4.2. Stage 2

The variables with strong correlations tend to be redundant, we first evaluate the albumin variable. According to the table 3, it can be seen albumin related to the dependent variable and other independent variables.

Table 3. Correlation Results

	albumin
class	-0.92582
specific.gravity	-0.71233
pus.cell	-0.75296
blood.urea	0.66194
serum.creatinine	0.802923
packed.cell.volume	-0.77553
hypertension	0,796876

The adjusted relative risk analysis showed that linear regression models considering and disregarding the variable albumin had large variations in the regression coefficients. Below we have the regression coefficients for the variables blood.pressure and specific.gravity.

- Blood.pressure varying from $-3.575e-04$ to $-6.333e-04$.
- Specific.gravity varying from $5.902e-02$ to $7.697e-02$.

To detect multicollinearity, VIF was used. A maximum VIF above 10 indicates that multicollinearity may be influencing least squares estimates and below, it is possible to see albumin variable with $VIF = 10.506670$. The highest VIF results is represented in table 4.

Table 4. VIF Results

Variable	VIF
albumin	10.506670
hypertension	7.423139
packed.cell.volume	5.221808
diabetes.mellitus	4.779383
pus.cell	4.612907
serum.creatinine	4.322035

In Figure 2, the presence of collinearity between the variables can be detected, so any plane along the data dispersion axis will be unstable.

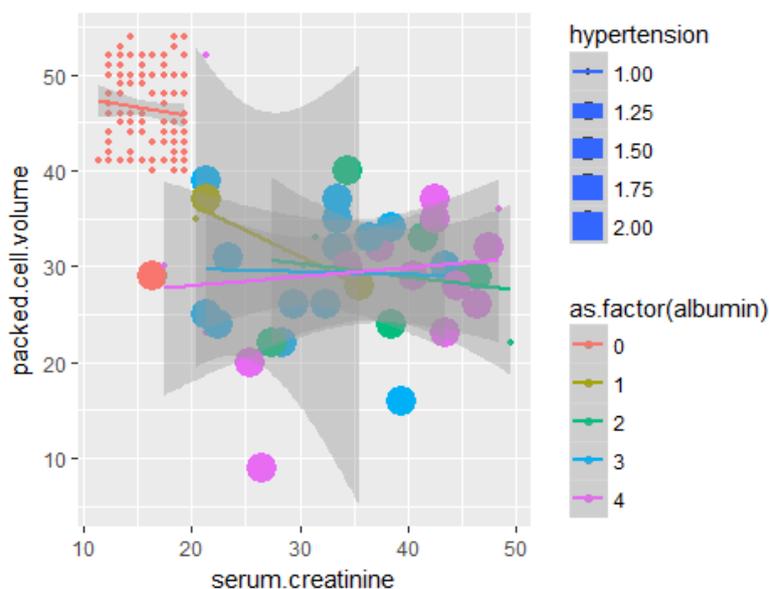


Figure 2. Comparing packed.cell.volume x serum.creatinine x hypertension x albumin

4.3. Stage 3

A technique of correction of the effect of the confounding variable was applied successful using the linear regression by the outcome to correct the effect of the confounding variable. With a set of corrected data, classifications techniques were applied, obtaining the results of the study.

Both technics SVM and logistic regression reached similar results. The accuracy of the predictive model was proved.

Applying the SVM technic in the original database, the classification in the presence of the confounder variable had 2 items incorrectly classified. In the absence of the confounder variable, 3 items were incorrectly classified. Both misclassifications cases were about CKD.

Applying the SVM technic in the corrected database, the result was 43 registers correctly classified as CKD and 115 registers correctly classified as NOT CKD in a total of 158 registers. None of the registers were incorrectly classified as we can see in table 5.

	CKD	NOT CKD
CKD	43	0
NOT CKD	0	115

Applying Logistic Regression, the result was 40 registers correctly classified as CKD and 28 registers correctly classified as NOT CKD in a total of 68 registers. None of the registers were incorrectly classified. Using the original database, we had misclassifications cases, as we can see in table 6.

Table 6. Logistic Regression Results

	CKD	NOT CKD
CKD	40	0
NOT CKD	0	28

5. Conclusion

The development of the present study made it possible to analyze techniques in order to identify and reduce the influence of confounders in predictive models by the use of a reduced set of data. In addition, the study also shows us that to build a predictive model is not simply to write an equation but rather to perform more consistent analyzes on the data and their relationships. In our study, through the use of Pearson and Spearman Correlation, VIF - Variance Inflation Factor, linear regression by the outcome it was possible to minimize the effect of a potential confounding variable and through classification techniques such as SVM and Logistic Regression, we verified the efficiency of the model after reducing the influence of the confounding factor. These techniques were used in three different databases of different sizes and the results were similar. For the other two databases tested, the McNemar test was used, which showed strong evidence of a statistically significant association. In general, it was possible to perceive that the accuracy of the model was maintained. Given the importance of the subject, it is necessary to deepen in other techniques of analysis of the relationship between variables. This analysis allows us to know more about the data, their relationships and the results we want to predict, so the resources described here are relevant in the discussion of the creation of predictive models. As a contribution, the techniques described in this study can be applied and different databases with a classification of unbalanced classes and for future work, it would be interesting to extend the studies in others in techniques that identify the relevance and the relation of the variables within the model.

References

- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424.
- Berk, R., Brown, L., Buja, A., George, E., and Zhao, L. (2018). Working with misspecified regression models. *Journal of Quantitative Criminology*, 34(3):633.
- García, C., García, J., López Martín, M., and Salmerón, R. (2015). Collinearity: Revisiting the variance inflation factor in ridge regression. *Journal of Applied Statistics*, 42(3):648–661.
- Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- <http://mlr.cs.umass.edu/ml/> (2017). Uci machine learning repository.
- Jena, L. and Kamila, N. K. (2015). Distributed data mining classification algorithms for prediction of chronic-kidney-disease. *International Journal of Emerging Research in Management & Technology*, 4(11):110–118.

- Kumar, M. (2016). Prediction of chronic kidney disease using random forest machine learning algorithm. *International Journal of Computer Science and Mobile Computing*, 5(2):24–33.
- Leatherman, E. R., Santner, T. J., and Dean, A. M. (2018). Computer experiment designs for accurate prediction. *Statistics and Computing*, 28(4):739.
- Li, L., Rakitsch, B., and Borgwardt, K. (2011). ccsvm: correcting support vector machines for confounding factors in biological data classification. *Bioinformatics (Oxford, England)*, 27:i342–i348.
- Li, L. and Zhang, S. (2015). Orthogonal projection correction for confounders in biological data classification. *International journal of data mining and bioinformatics*, 13:181–196.
- Low, Y. S., Gallego, B., and Shah, N. H. (2016). Comparing high-dimensional confounder control methods for rapid cohort studies from electronic health records. *Journal of comparative effectiveness research*, 5:179–192.
- Maurizio, M. (2011). Data mining concepts and techniques. *domenica*.
- Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). *Introduction to linear regression analysis*, volume 821. John Wiley & Sons.
- Schnitzer, M. E., Lok, J. J., and Gruber, S. (2016). Variable selection for confounder control, flexible modeling and collaborative targeted minimum loss-based estimation in causal inference. *The international journal of biostatistics*, 12:97–115.
- Sheskin, D. J. (2011). Parametric versus nonparametric tests. *International Encyclopedia of Statistical Science*.
- Sinha, P. and Sinha, P. (2015). Comparative study of chronic kidney disease prediction using knn and svm. *International Journal of Engineering Research and Technology*, 4(12):608–12.
- Steyerberg, E. (2009). Lessons from case studies. *Clinical Prediction Models*.
- Team, R. C. (2014). R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2013.
- Waller, M. A. and Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2):77–84.

Seleção de Dispositivos para *Caching* Baseado no Interesse do Usuário através de Comunicação D2D

Marcelo Almeida Silva¹, Antonio C. de Oliveira Jr.¹, Fausto da Silva Moraes¹,
Kleber Vieira Cardoso¹, Vinicius da C. M. Borges¹, Waldir Moreira²

¹ Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brazil

²Fraunhofer Portugal AICOS – Porto – Portugal

{marceloalmeidasilva, antonio, fausto, kleber, vinicius}@inf.ufg.br,
waldir.junior@fraunhofer.pt

Abstract. *The next-generation wireless networks (5G Networks) predict an expansion in the connected devices' number and network interfaces with higher transmission capacity, and consequently an exponential increase in the amount of traffic. In this context, solutions that do not overload the wireless communication infrastructure are needed to improve efficiency. This paper presents a cache's proposal based on interest and interaction between devices in networks with D2D communication, which seeks to settle some of these problems. The proposal is based on cache approaches' and solutions case studies aimed at improving the network's use through the efficient use of its resources. It was verified that the proposal presents improvement of 25% when compared to another cache's approach. From these results we can conclude that the use of this information can be beneficial for 5G networks strategies.*

Resumo. *As redes sem fio de próxima geração (Redes 5G) prevê um aumento na quantidade de dispositivos conectados e interfaces de redes com maior capacidade de transmissão e conseqüentemente, um aumento exponencial da quantidade de tráfego. Nesse contexto, soluções que não sobrecarreguem a infraestrutura de comunicação sem fio se faz necessária para melhorar a eficiência. Este trabalho apresenta uma proposta de cache baseada em interesse e interação entre os dispositivos em redes com comunicação D2D, que busca suavizar alguns desses problemas. A proposta é baseada nos estudos de casos de abordagens de cache e soluções voltadas para a melhoria do uso da rede através do uso eficiente dos seus recursos. Verificou-se que a proposta apresenta melhoria de 25% quando comparada com outra abordagem de cache. A partir desses resultados podemos concluir que o uso dessas informações pode ser benéfico para estratégias das redes 5G.*

1. Introdução

O crescente número de dispositivos conectados tem aumentado cada vez mais o tráfego na rede. Grande parte deste aumento se deve a massiva quantidade de dispositivos que estão conectados à rede enviando e recebendo dados a todo instante, por exemplo: sensores, câmeras, carros autônomos, drones entre outros. Um dos grandes desafios que as redes

de computadores vão enfrentar nas próximas gerações é atender às novas demandas sem deixar de lado as demandas dos usuários, ou seja, o uso de novos mecanismos e soluções junto às operadoras para suprir as necessidades geradas pelo crescimento da rede.

Alguns problemas surgem devido o aumento do tráfego: indisponibilidade de serviços, dificuldade na alocação de recursos e congestionamentos na rede. Uma consequência seria a redução drástica na qualidade do serviço. As mídias de vídeo são as principais responsáveis pelo uso do tráfego na rede, sendo responsáveis por mais de 73% do tráfego em redes móveis [Cisco 2018]. Os provedores de *Internet* já enfrentam problemas relacionados a desempenho e tráfego nas redes diante deste cenário.

Portanto, as redes 5G (Quinta Geração da rede móvel)[Marica et al. 2018], devem ser capazes de atender a todos os usuários utilizando melhor os seus recursos para aumentar a vazão, diminuir a latência e aumentar a disponibilidade da rede [Guan et al. 2016].

Para atender a essas demandas no 5G, temos como abordagens promissoras o uso de *Cache* Proativo [Moraes et al. 2017], NFV (*Network Function Virtualization*) [Chantre and d. Fonseca 2018], SDN (*Software Defined Network*) [Lee and Yoo 2017], D2D (*Device to Device*) [Feng et al. 2018] entre outras. O uso de *Cache* Proativo como alternativa, visa armazenar previamente conteúdos que podem ser utilizados pelo dispositivo em ocasiões futuras, dessa forma é possível explorar conexões com menor custo e maior velocidade de transmissão para realizar o *cache* de conteúdo.

O NFV procura realizar as principais funções de rede (por exemplo: *caches*, *firewalls*, etc.), além dos recursos virtualizados de computação, armazenamento e rede, prometendo a redução do CAPEX/OPEX (*CAPital EXpenditure e OPerational EXpenditure*) associado nas redes 5G. O SDN é uma solução que permite a programação controlada por *software* e o gerenciamento de recursos de rede de maneira dinâmica e escalonável. Estas soluções estão sendo amplamente adotadas por provedores de serviços, provedores de nuvem e empresas, a fim de acelerar a implantação de serviços de rede, economizando recursos e aumentando a confiabilidade.

As comunicações D2D são conexões diretas entre dispositivos próximos que exploram as interfaces de rede disponíveis nos mesmos para dar vazão aos dados. Dessa forma é possível transmitir dados sem passar pelas estações base da rede, evitando congestionamentos e aumento do fluxo. Por ser uma abordagem que explora a proximidade entre os dispositivos, a comunicação D2D pode ser realizada de forma oportunista, explorando os recursos da rede para melhorar o *offloading* de dados.

Uma possível abordagem nas redes 5G é a realização de *cache* de conteúdo nos dispositivos da borda de rede como os pontos de acesso e dispositivos dos usuários. A partir das características das comunicações D2D, as conexões oportunistas podem ser utilizadas como forma de distribuição de conteúdo armazenado nos dispositivos por meio de *cache*, assim, é possível obter o conteúdo dos dispositivos próximos sem a necessidade de passar por uma estação base [Moraes et al. 2017].

Devido ao seu impacto nas redes celulares, o conteúdo escolhido para ser armazenado são os conteúdos de vídeo. Por este motivo é importante analisar as características do consumo desse tipo de conteúdo, sobretudo os fatores que podem influenciar um (ou vários) usuário(s) a acessar um determinado vídeo. Um aspecto importante é a característica viral que alguns vídeos possuem fazendo com que a probabilidade de que um

conteúdo seja acessado aumente à medida que ele é compartilhado em redes sociais *online* (OSN - *Online Social Networks*) [Zhang et al. 2015].

Um segundo aspecto importante é o tipo do vídeo ou a qual categoria o mesmo pertence, pois o interesse do usuário nesse conteúdo pode aumentar a probabilidade de acesso a um vídeo o que impacta na forma como ele acessa e como os conteúdos são distribuídos [Bai et al. 2016]. Estudos apontam que os relacionamentos são formados com base nos interesses em comum entre as pessoas [Schiefele 1991] o que gera impacto de forma relevante nas abordagens de *cache*.

Outros dados importantes são: Como o usuário se comporta dentro desses cenários, a forma como se movimenta, quais dispositivos ele tem contato e quanto tempo permanece conectado aos outros dispositivos [Moreira et al. 2013]. Esses dados são fatores determinantes para a decisão de quais dispositivos podem ser utilizados para fazer *cache*.

Dessa forma, o melhor dispositivo para armazenar esses conteúdos poderia ser determinado, a partir da visão da vizinhança D2D e do histórico de comunicações entre os dispositivos, que as operadoras de redes celulares já possuem. Essas informações seriam combinadas para criar aplicações capazes de realizar o *cache* ao armazenar os conteúdos com maior probabilidade de serem acessados e utilizar a comunicação D2D para disseminá-las entre os dispositivos dos usuários, preservando os recursos da rede celular.

Existe uma carência de soluções de *cache* para redes 5G que façam uso de informações de interesse e de histórico de comunicações entre os dispositivos como informação principal. Dessa forma esse trabalho visa criar uma abordagem que utiliza essas informações com o intuito de melhorar a eficiência da rede.

Este artigo está organizado da seguinte forma. Após a introdução, a Seção 2 apresenta os trabalhos relacionados. A Seção 3 detalha as ferramentas de software e conceitos utilizados no trabalho. A Seção 4 apresenta a descrição da proposta. A Seção 5 mostra uma avaliação da proposta através de simulação, discutindo os resultados. Por fim, a Seção 6 faz conclusões desse trabalho e aponta alguns trabalhos futuros.

2. Trabalhos Relacionados

Trabalhos recentes têm mostrado como técnicas de *cache* de conteúdo em dispositivos da borda da rede podem utilizar a comunicação D2D para melhorar a eficiência do uso de recursos de uma rede celular.

No trabalho realizado por [Moreira et al. 2013] foi proposto o SCORP (*Social-aware Opportunistic Routing Protocol*). O SCORP explora a proximidade social e o conhecimento do conteúdo com o objetivo de melhorar a entrega de forma eficiente em cenários de rede com características urbanas e densas. Com base no grafo social *offline* (interações físicas entre os dispositivos) e interesses dos usuários foram utilizados os princípios de redes oportunistas para alcançar o que foi proposto.

Através do histórico de movimentação associado ao interesse, é possível prever qual o melhor dispositivo para a realização de *cache* de conteúdo, dessa forma consegue-se maior cobertura de dispositivos dos usuários. Isso se dá pelo fato de que o histórico de

movimentação é obtido por meio do tempo de contato entre os dispositivos com o mesmo interesse.

No entanto o SCORP não analisa a popularidade do conteúdo e o impacto de conteúdos com diferentes taxas de visualização nesses cenários. A consequência disso é que pode ocorrer sobrecarga na rede devido a conteúdos virais (que possuem uma alta taxa de visualização em um curto espaço de tempo) com características diversas.

A solução proposta por [Zhang et al. 2015] faz uso de uma abordagem *social-aware* para otimizar a comunicação D2D explorando a popularidade do conteúdo e o grafo social *offline*. Para modelar a distribuição de conteúdo nas redes sociais *offline* ele utiliza o IBP (*Indian Buffet Process*) para apresentar uma abordagem de *cache* reativo utilizando comunicação D2D. Os resultados da proposta apresentaram um aumento da eficácia de acordo com as configurações da rede e dos usuários.

Apesar do aumento da eficácia da rede utiliza-se apenas o grafo social *offline*, propagando conteúdo para dispositivos do usuário em que o mesmo pode não possuir interesse. Dessa forma, há o consumo de recursos da rede que poderiam ser melhor aproveitados por usuários que tenham interesse no conteúdo.

Bai et al. [Bai et al. 2016] utiliza um hipergrafo para representar as informações de três camadas. A primeira camada refere-se ao espectro e interferência em que o enlace D2D foi estabelecido. A segunda camada trata das relações sociais *online* (são as relações estabelecidas pelos usuários dentro de uma OSN - *Online Social Network*) assumindo que os enlaces entre os dispositivos são duráveis e confiáveis. A terceira camada lida com os diferentes interesses que os usuários podem demonstrar em relação aos diversos conteúdos. O uso das três camadas mostram melhorias nas simulações, nas estratégias de *cache* e na disseminação de conteúdo.

Entretanto essa abordagem deixa de atender os demais dispositivos que não possuem relação social *online* com o dispositivo que possui em *cache* o conteúdo de seu interesse, pois uma hiper-aresta (uma relação entre dois dispositivos) passa a existir quando: existe interesse em comum por algum conteúdo, o nó está dentro do alcance de transmissão do enlace, quando há pouca interferência e quando existe relação social *online* entre os dispositivos.

O ProSoCaD (*Probabilistic Social Cascade For caching in D2D communication*), proposto por [Moraes et al. 2017], combina popularidade do conteúdo, relação social *online* e o grafo social *offline*. Com essas informações é realizado o *cache* proativo no dispositivo selecionado com base no modelo de seleção de conteúdo social proposto também no artigo o SoCoSM (*Social-aware Content Selection Model*). O dispositivo selecionado é o que possui a maior quantidade de conexões D2D e o maior número de relações sociais (chamado de: dispositivo mais influente).

Nessa abordagem o uso das redes sociais *online* tornam o *cache* do conteúdo centralizado nos dispositivos mais influentes, consumindo de forma mais rápida os recursos dos dispositivos selecionados. O *cache* de um conteúdo no dispositivo em que o seu usuário não tem interesse pode acarretar no uso desnecessário do recurso caso não sejam realizados acessos a ele.

Apesar dos trabalhos descritos empregarem conceitos de interação social *online* e

offline entre os dispositivos e usuários, não são consideradas abordagens em que o interesse age como fator principal. A Tabela 1 representa a relação do que é abordado nos trabalhos descritos e na solução proposta.

Tabela 1. Configuração da Simulação

	Interesse do Usuário	Histórico de Movimentação	Modelo de Acesso ao Conteúdo	Popularidade do Conteúdo	Cache Proativo	Grafo Social Offline (D2D)	Grafo Social Online (OSN)
[Moreira et al. 2013]	X	X				X	
[Zhang et al. 2015]			X	X			
[Bai et al. 2016]	X				X	X	X
[Moraes et al. 2017]			X	X	X	X	X
Proposta	X	X	X	X	X	X	

Da mesma forma não foram exploradas as vantagens relacionadas a construção de uma solução de *cache* proativo nesse aspecto explorando comunicação D2D dando espaço para o surgimento de novas estratégias de *cache*, a fim de aumentar a eficiência da rede.

3. Ferramentas, Modelos e Conceitos Empregados

Nessa seção serão descritas as ferramentas, modelos e conceitos empregados nesse trabalho.

3.1. Grafo Social *Offline* (D2D)

O Grafo Social *Offline* (Grafo D2D) é obtido do uso de uma interface de comunicação diferente da rede celular. Pode-se obter assim pela interface utilizada, os registros dos encontros entre os dispositivos e os dispositivos que estejam em alcance de comunicação formando o Grafo D2D. Com essa informação é possível identificar os dispositivos melhor localizados para realização de *cache* de conteúdo.

3.2. Histórico de movimentação

As informações obtidas pelo Grafo D2D permitem estimar a probabilidade de que um determinado dispositivo, consiga atender outro através de uma comunicação mais estável, ou seja, um conexão pode ser mantida durante o tempo necessário para a transferência de conteúdo.

Para se obter essa informação é utilizado o histórico de movimentação dos dispositivos e o tempo de contato entre eles, determinando assim, qual dispositivo tem maior tempo de contato e maior quantidade de conexões com os demais. Dessa forma o dispositivo com essas características, considerado o mais popular, é o mais recomendado para a realização de *cache* de conteúdo, pois o mesmo pode atender uma maior quantidade de dispositivos em um intervalo de tempo maior.

3.3. Interesse

Muitos têm em mente que o comportamento dos usuários está associado apenas a contatos com amigos e familiares nas redes sociais *online* (OSN), porém, o comportamento está associado a mais que isso. Antes, as pessoas se comunicavam com base em relacionamentos restritos, mantendo laços as pessoas geograficamente mais próximas a elas.

Porém dado o avanço tecnológico as pessoas já não precisam sair de casa para encontrar alguém ou ir a um ambiente que possibilite conhecer e viver novas experiências, possibilitando ir além, como conhecer pessoas diferentes que tem afinidades e interesses em comum que estão além da geolocalização.

Neste trabalho abordamos o comportamento dos usuários baseado em seus interesses, ou seja, os usuários que acessam os mesmos conteúdos, compartilham dos mesmos tipos de interesses. As categorias de interesse são baseadas na classificação de cada conteúdo, podendo ser obtida através de *tags* e palavras chave encontradas no conteúdo.

A cada acesso a um conteúdo por um dispositivo, aquele interesse é associado a ele, dessa forma cada dispositivo contém uma tabela de interesses relacionada, que pode determinar quais conteúdos os usuários tem interesse e portanto tem maior probabilidade de acessar.

3.4. Modelo de Acesso ao Conteúdo

Os modelos de acesso a conteúdos são produzidos por processos estocásticos com o intuito de reproduzirem a distribuição de conteúdos do mundo real, através de funções matemáticas, observando séries de dados coletados de provedores de conteúdos.

O modelo SoCoSM (*Social-aware Content Selection Model*) proposto por [Moraes et al. 2017], gera rankings de visualização de conteúdos similares aos reais. Para isso o SoCoSM utiliza padrões de acesso à conteúdos ciente a interações sociais, fazendo o uso do IBP e do SI²RP (Figura 1).

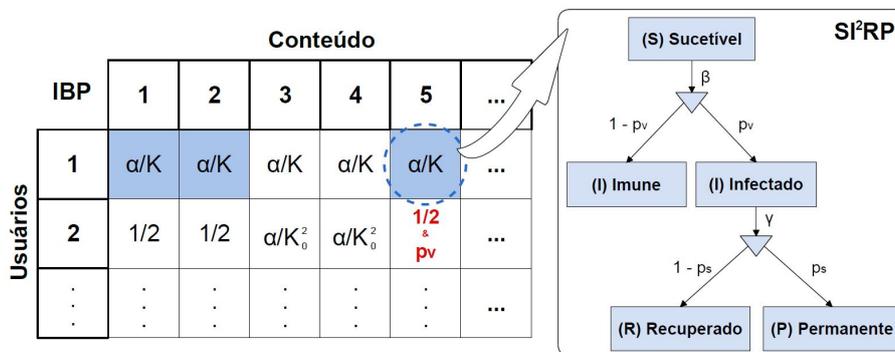


Figura 1. Modelo de seleção de conteúdo SoCoSM.

Zhang et al. [Zhang et al. 2015] utiliza o processo do Buffet Indiano (IBP) e valida o processo com séries de dados coletados do provedor de conteúdo YouTube, dessa forma demonstra que ele se aproxima dos resultados reais. O SI²RP é um modelo de propagação baseado em modelos epidêmicos utilizado para representar os tipos de usuários e seus comportamentos de forma probabilística.

Na literatura não foi encontrada uma base de dados pública que combine interesse, histórico de movimentação e conteúdos acessados, por esse motivo é importante que tenhamos modelos que representem esse comportamento e que combinados provêm todas as informações modeladas na rede.

Neste trabalho é utilizado o SoCoSM para selecionar os conteúdos acessados e distribuídos pelos dispositivos pois esse modelo gera dados capazes de representar o consumo de conteúdos dos usuários. A distribuição dos conteúdos é gerada a partir da saída do IBP, o SI²RP recebe essa saída como entrada a cada conteúdo selecionado pelos usuários, como mostrado na Figura 1. Dessa forma cada usuário poderá ter mais de uma oportunidade de acessar um mesmo conteúdo caso deseje assistir a algum conteúdo mais de uma vez.

3.5. Cache

O *cache* pode ser realizado de 2 maneiras: reativa e proativa. O *cache* reativo é o armazenamento de conteúdo resultado de uma ação, por exemplo, o *cache* de um vídeo que foi visualizado pelo usuário. Já o *cache* proativo não é resultado de uma ação direta do usuário, por exemplo, o usuário tem interesse em uma categoria de conteúdo e o dispositivo realizou o *download* de um conteúdo que é do interesse do usuário sem que seja necessário que ele realize alguma ação.

Devido os dispositivos móveis terem capacidade de armazenamento reduzida, são necessárias políticas de substituição para o *cache* do conteúdo. A política adotada neste trabalho é o LFU (*Least Frequently Used*), priorizando a remoção dos conteúdos que contêm a menor frequência de acesso. Assim é possível manter os conteúdos mais populares armazenados no *cache* visando atender a busca por conteúdos mais populares e auxiliar na substituição de conteúdo com baixa taxa de acesso.

3.6. Simulador

Para validar as propostas [Moraes et al. 2017], [Zhang et al. 2015] e [Moreira et al. 2013] utilizam simulações baseadas em traces reais e sintéticos, dessa forma conseguem avaliar o desempenho dentro dos cenários propostos. O NS3 (*Network Simulator*) [Consortium 2018] é o simulador eventos discreto mais utilizado pela comunidade acadêmica de redes que reproduz características de cenários reais em sistemas de *Internet*. Nesse trabalho utilizamos o NS3 para reprodução do cenário e validação da solução.

4. Proposta

O intuito deste trabalho é propor uma solução de *cache* proativo em redes com comunicação D2D que faz uso do interesse do usuário, histórico de movimentação, modelo de acesso ao conteúdo e popularidade do conteúdo, para apresentar as vantagens do uso do interesse dos usuários sobre as informações de interações sociais *online* em sistemas de distribuição de conteúdo, com intuito de melhorar a eficiência do uso dos recursos da rede celular.

Através dessas informações, buscamos como objetivos, reduzir o uso de recursos da *base-station*, fazendo uso de comunicação D2D e reduzir a quantidade de *cache* de conteúdo não utilizado que é realizado no dispositivo do usuário, através dos dados de interesse.

A partir dos dispositivos dos usuários obtemos o grafo D2D que ao longo de suas rotinas diárias geram o histórico de contatos entre os dispositivos. O histórico de contatos é associado ao interesse através do SCORP [Moreira et al. 2013] para geração do grafo de interesses. Além disso, a escolha dos conteúdos acessados por cada dispositivo é feita

conforme a popularidade dos mesmos através do modelo de seleção de conteúdo SoCoSM [Moraes et al. 2017].

A seleção do dispositivo para realização do *cache* proativo é feita combinando o grafo de interesses e o conjunto de conteúdos acessados por cada dispositivo, como apresentado na Figura 2. É importante ressaltar que cada conteúdo é previamente classificado em um tipo de interesse.

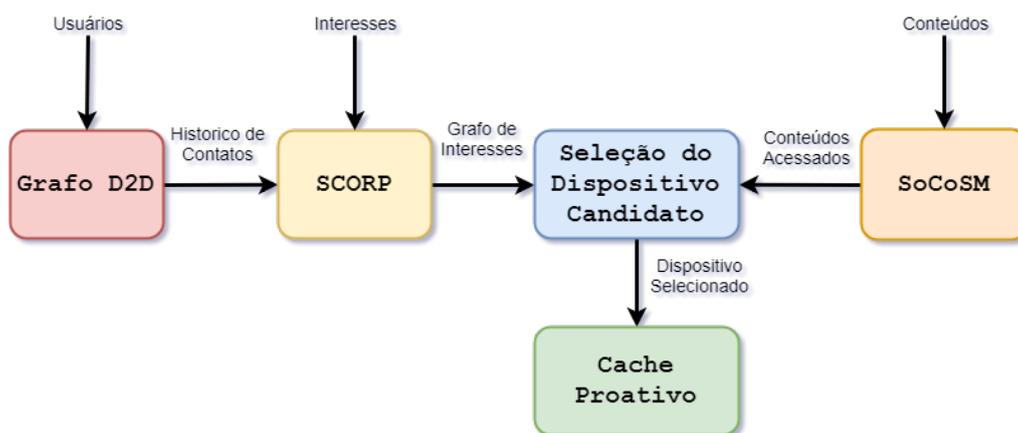


Figura 2. Arquitetura de distribuição de conteúdo baseado em interesse.

Buscando uma solução que atenda de forma eficiente os dispositivos da rede utilizando o *cache* proativo em dispositivos que tenham interesses em comum explorando o modelo de redes oportunistas.

Essa seção conta com duas subseções: Subseção 4.1 que descreve a forma de seleção do dispositivo candidato, e Subseção 4.2 que descreve o critério para a realização do *cache* proativo.

4.1. Escolha do Dispositivo

Na seleção dos dispositivos candidatos para a realização de *cache*, é utilizada a abordagem do SCORP que define o nível de proximidade social entre os dispositivos com certos interesses, independente dos dispositivos compartilharem de tais interesses, para determinar quais dispositivos são mais adequados a realizarem o *cache* proativo de acordo com a duração dos contatos que ocorreram no decorrer do tempo.

Para isso são escolhidos os dispositivos que possuem o maior $TECI(x)$ (*Time-Evolving Contact to Interest x*) do grafo D2D, visando atender todos os dispositivos que tem interesse em comum.

A partir do $TECI(x)$ é possível determinar o tempo de contato de cada usuário com os dispositivos próximos, classificando de acordo com o interesse, dessa forma podemos selecionar o dispositivo mais apropriado para realização de *cache* proativo.

Dessa forma, para cada interesse x obtemos o $Max(x)$ onde é maximizado o tempo de contato, resultando em uma maior probabilidade de sucesso na distribuição de conteúdo (Equação 1). Dessa forma obtemos o conjunto de dispositivos que mantém a propriedade temporal do $TECI(x)$ e atende os dispositivos no alcance de transmissão.

$$Max(x) = maximize(TECIx) \quad (1)$$

Portanto, como mostra a Figura 3, podemos representar a escolha do dispositivo da seguinte forma: i) tabelas de interesse de cada usuário gerado pelo SCORP; ii) o grafo D2D baseado no interesse; iii) o conjunto de subgrafos de interesse obtido a partir dos dispositivos selecionados para o *cache* proativo.

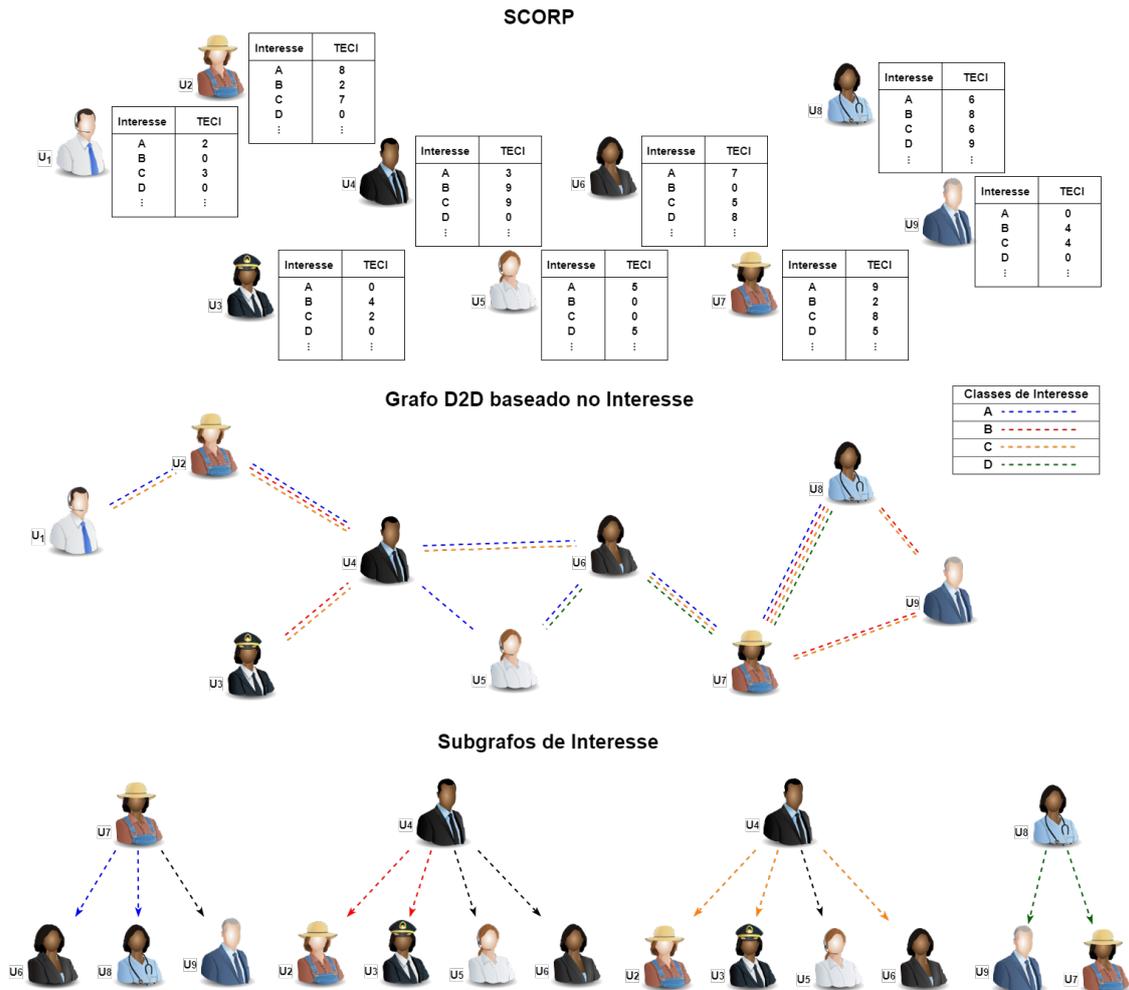


Figura 3. Escolha dos candidatos para armazenar um conteúdo.

Por exemplo, para uma distribuição de 9 usuários (U_i) e 4 tipos de interesses distintos (A, B, C e D) obtemos através do SCORP o valor do $TECI(x)$ associado a cada interesse, que inicia em zero para nenhum contato e aumenta de acordo com a duração dos contatos realizados.

Dessa forma obtemos o grafo D2D baseado no interesse, que representa a relação entre os usuários que possuem interesses em comum e que possuem possibilidade de conexão, representado por arestas de cores distintas para representar cada interesse. Os usuários que possuem mais que um interesse em comum estão representados com mais de uma aresta.

Ao realizar a seleção dos dispositivos candidatos encontramos Subgrafos de Interesse associados a cada classe de interesse, por exemplo, para o interesse B temos o dispositivo U_4 que possui $Max(B)$, sendo selecionado como dispositivo candidato. Dessa forma selecionamos o dispositivo que possui o maior nível de proximidade social entre os dispositivos com interesses em comum.

Cada subgrafo de interesse é representado pelo dispositivo candidato a realizar o *cache* proativo do conteúdo e os dispositivos que serão atendidos, ou seja, que estão dentro do alcance de transmissão no caso do dispositivo U_4 temos os dispositivos U_2 , U_3 , U_5 e U_6 . Os dispositivos que serão atendidos podem ter interesse em comum no conteúdo do dispositivo candidato, representado pelas arestas coloridas, ou não, representado pelas arestas pretas.

Como podemos observar nos Subgrafos de Interesse U_4 atende U_2 e U_3 que possuem o interesse em comum, e atende U_5 e U_6 mesmo sem que esses dispositivos tenham declarado interesse no mesmo conteúdo que U_4 . Dessa forma podemos atender os dispositivos que possuem classes de interesse em comum sem deixar de atender os demais dispositivos.

4.2. Cache Proativo

Para determinar quando um conteúdo deverá ser movido proativamente para o dispositivo escolhido pelo modelo de seleção de dispositivos candidatos para fazer *cache* deste conteúdo, é utilizado a quantidade de conexões que ele possui. Dessa forma temos o seguinte valor para o $p_{cv}(x)$ (probabilidade de realizar o *cache* proativo de um conteúdo de interesse x) descrito na Equação 2.

$$p_{cv}(x) = 1 - \prod_{k=1}^{n \forall P_k(a,x) > 0} (1 - p_v(k)) \quad (2)$$

A medida que a quantidade de conexões D2D aumenta, a probabilidade de realizar o *cache* proativo no dispositivo selecionado também, assim podemos garantir que os dispositivos presentes no momento do *cache* serão atendidos caso solicitem o conteúdo e as características temporais obtidas através do SCORP serão mantidas pois os dispositivos escolhidos são os que possuem um maior tempo de contato com os demais de mesmo interesse.

5. Estudo de Simulação

Essa seção é destinada a apresentação da avaliação da proposta descrita. Em seguida é realizada uma comparação com o ProSoCaD a solução proposta por [Moraes et al. 2017], que é uma abordagem de *cache* proativo de conteúdo que utiliza comunicação D2D. Essa seção é dividida em: Subseção 5.1 que descreve a configuração dos cenários e parâmetros de avaliação e Subseção 5.2 que discute os resultados de avaliação de desempenho.

5.1. Configuração do Cenário

Para a avaliação desse trabalho é considerado um cenário com poucos dispositivos distribuídos em uma grande área, dito cenários esparsos. Estes cenários são representados

por um modelo com diferentes tipos de usuários e movimentação, por exemplo, cidades pequenas e campus universitário.

Os usuários inseridos nesse contexto podem demonstrar interesses em comum, acessando os mesmos conteúdos ou de mesmo interesse, dessa forma, utilizamos um cenário que visa atender de forma justa as 2 abordagens. O cenário foi implementado utilizando a ferramenta de simulação de redes ns-3 (*Network Simulator 3*).

Para isso utilizamos os dados de uma base que contém dados reais de usuários móveis obtido do CRAWDAD (*Community Resource for Archiving Wireless Data At Dartmouth*) [Scott et al. 2009] contendo 36 dispositivos. Esses dispositivos foram entregues a alunos da Universidade de Cambridge e durante o período de 2 meses foram coletados dados ao longo de suas rotinas diárias.

Essa base de dados conta com informações de interesse dos usuários e a movimentação dos dispositivos, buscamos dessa forma mostrar as vantagens relacionadas ao uso do interesse nas propostas de *cache*.

O ProSoCaD faz uso das informações de relação social *online* entre os dispositivos, para representá-las, foram gerados de forma sintética as relações sociais *online*. Visando obter justiça na comparação utilizamos uma configuração em que todos os dispositivos estão conectados.

Para realização da comunicação D2D e percepção da proximidade entre os dispositivos utilizamos o *wireless* buscando maior velocidade de transmissão do conteúdo, assim obtemos uma configuração em que os dispositivos obtêm o conteúdo via dados móveis apenas quando não estão disponíveis em seus vizinhos próximos.

5.2. Avaliação da Proposta

O *Cache Local* ocorre quando o conteúdo que um dispositivo acessou é armazenado localmente no dispositivo, este conteúdo é computado quando obtido através dos dados da rede móvel. Esse parâmetro de avaliação é utilizado para aferir quanto do conteúdo acessado pelos dispositivos foi baixado da rede móvel.

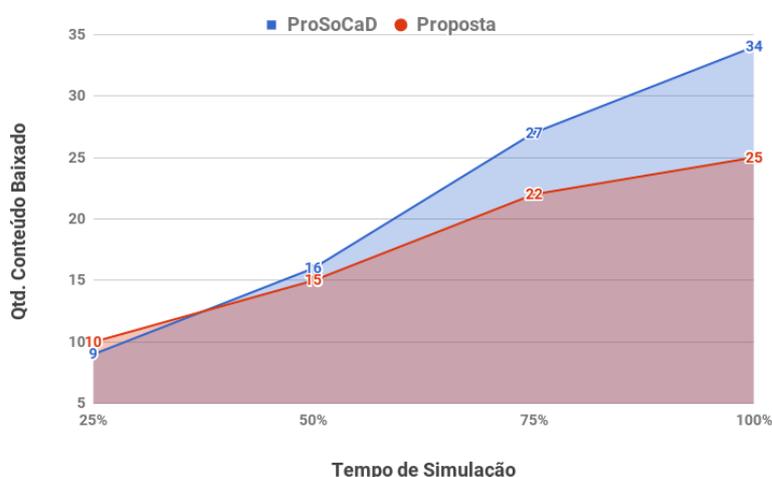


Figura 4. *Base Station Download*

Os valores apresentados pela Figura 4 representam a quantidade de conteúdo armazenado no dispositivo que foi obtido a partir das redes móveis a medida que a simulação acontece. Como podemos observar, quanto maior o valor obtido, maior o consumo da rede móvel.

A proposta leva em consideração o histórico de movimentação dos dispositivos para a realização de *cache*, portanto, no início do tempo de simulação nota-se uma similaridade entre os resultados, pois é necessário definir os interesses dos usuários e tempo de contato com cada interesse.

A medida que o tempo de simulação avança, a informação do histórico de movimentação passa a ser detectada pois necessita de mais tempo para ser percebida, logo percebe-se uma melhora nos resultados. Portanto, uma redução dos conteúdos baixados da rede móvel e armazenados localmente nos dispositivos, representa uma maior suavização na infraestrutura da rede móvel (*offloading*) uma vez que o dispositivo do usuário passa a requisitar menos conteúdo da estação base e os obtém através de comunicação D2D.

A Figura 5 apresenta os resultados obtidos do *Cache Miss*, ele ocorre quando o conteúdo que um dispositivo está solicitando, não está armazenado no *cache* de nenhum dos dispositivos que esta ao seu alcance, ou quando não consegue receber esse conteúdo com sucesso, então o dispositivo precisará solicitar o conteúdo desejado para a infraestrutura (estação base LTE) da rede móvel.

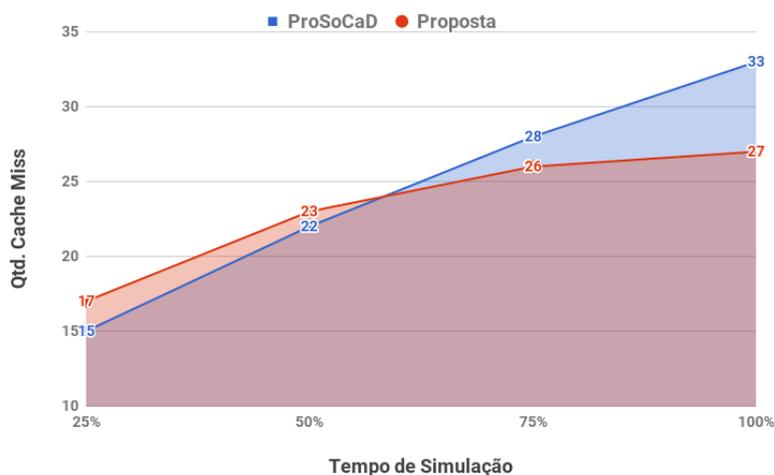


Figura 5. Cache Miss

Devido a distribuição dos conteúdos a medida que os dispositivos que solicitam novos conteúdos, é possível obter o conteúdo alvo dos dispositivos que estão próximos via comunicação D2D. Podemos notar que a redução do *Cache Miss* se dá pelo aumento de 25% da quantidade de conteúdos obtidos através da comunicação D2D, dessa forma conseguimos alcançar uma maior vazão reduzindo o consumo de recursos da *base-station*.

Podemos destacar que apesar do nível de relação social *online* ser alto (100%) os resultados mostram que o interesse do usuário e histórico de movimentação tem mais impacto na eficiência da *cache* que a relação social *online*.

6. Conclusões e Trabalhos Futuros

Neste artigo, foi proposta uma nova abordagem para reduzir o uso de recursos da rede móvel dos usuários e reduzir a quantidade de conteúdo baixado pelos dispositivos através da rede de dados móveis, explorando o interesse e as interações sociais *offline* entre os dispositivos.

Para isso propomos uma nova abordagem, que combina várias informações (interesse, popularidade do vídeo, histórico de movimentações, grafo D2D), para selecionar o dispositivo mais apropriado para fazer o *cache* do conteúdo de cada interesse. Ao selecionar o dispositivo é utilizado o SoCoSM para definir como o conteúdo deve ser distribuído.

Assim definimos uma solução de *cache* proativo em redes de comunicação D2D, que a partir dos resultados apresentados, mostra que o uso do interesse e das interações sociais entre os dispositivos apresenta melhora na distribuição e acesso ao conteúdo.

Como trabalhos futuros pretendemos avaliar a proposta em outros cenários com características diversas buscando melhorias na solução e analisar o impacto do interesse na distribuição do conteúdo.

Referências

- Bai, B., Wang, L., Han, Z., Chen, W., and Svensson, T. (2016). Caching based socially-aware d2d communications in wireless content delivery networks: a hypergraph framework. *IEEE Wireless Communications*, 23(4):74–81.
- Chantre, H. and d. Fonseca, N. L. S. (2018). Reliable broadcasting in 5g nfv-based networks. *IEEE Communications Magazine*, 56(3):218–224.
- Cisco (2018). The Zettabyte Era: Trends and Analysis.
- Consortium, N. (2018). Ns3 consortium.
- Feng, L., Zhao, P., Zhou, F., Yin, M., Yu, P., Li, W., and Qiu, X. (2018). Resource allocation for 5g d2d multicast content sharing in social-aware cellular networks. *IEEE Communications Magazine*, 56(3):112–118.
- Guan, P., Zhang, X., Ren, G., Tian, T., Benjebbour, A., Saito, Y., and Kishiyama, Y. (2016). Ultra-low latency for 5g-a lab trial. *arXiv preprint arXiv:1610.04362*.
- Lee, J. and Yoo, Y. (2017). Handover cell selection using user mobility information in a 5g sdn-based network. In *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 697–702.
- Marica, A., Claudia, C., and Antonella, M. (2018). Empowering 5g network softwarization through information centric networking. *Internet Technology Letters*, 1(2):e30.
- Moraes, F. d. S., Cardoso, K. V., and Borges, V. C. (2017). Improving video content access with proactive d2d caching and online social networking. In *Computers and Communications (ISCC), 2017 IEEE Symposium on*, pages 1043–1048. IEEE.
- Moreira, W., Mendes, P., and Sargento, S. (2013). Social-aware opportunistic routing protocol based on user’s interactions and interests. In *International Conference on Ad Hoc Networks*, pages 100–115. Springer.

- Schiefele, U. (1991). Interest, Learning, and Motivation. *Educational Psychologist*, 26(3-4):299–323.
- Scott, J., Gass, R., Crowcroft, J., Hui, P., Diot, C., and Chaintreau, A. (2009). CRAWDAD dataset cambridge/haggle (v. 2009-05-29). Downloaded from <https://crawdad.org/cambridge/haggle/20090529>.
- Zhang, Y., Pan, E., Song, L., Saad, W., Dawy, Z., and Han, Z. (2015). Social network aware device-to-device communication in wireless networks. *IEEE Transactions on Wireless Communications*, 14(1):177–190.

Proposta de modelo de veículos elétricos para análise de consumo energético em redes veiculares

Divino Alves Ferreira Júnior¹, Leonardo Moraes Fleury¹
Vinicius Cunha M Borges¹, Antônio Oliveira-Jr¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brazil

{profdivinojr, fleuryleomoraes}@gmail.com, {vinicius, antonio}@inf.ufg.br

Abstract. *Electric vehicles are rising in popularity because of the politicians, automakers, and citizens's interests. The same event happens with the number of mobile networks (mobile or wi-fi). With the growing of users each day, many researchers are concerned that the infrastructure can become insufficient, leading them to believe in opportunistic networks protocols as a solution. About the growing of electric cars, these protocols can impact in their autonomy. With that in mind, we propose a model of electric vehicles to analyze the effect of these protocols on the autonomy of the vehicle.*

Resumo. *Veículos elétricos estão se popularizando a cada dia, com interesses vindos dos Estados, das montadoras e dos cidadãos, assim como a comunicação por redes sem fio, seja ela móvel ou Wi-Fi. Com o crescente número de usuários, muitos pesquisadores se preocupam com o fato de que a infraestrutura tradicional possa se tornar insuficiente; logo apostam em redes oportunistas como uma solução. Quanto ao crescimento de carros elétricos, surge a preocupação do impacto que estes protocolos oportunistas podem causar em sua autonomia. Neste trabalho propomos um modelo de abstração e consumo de veículos elétricos para podermos analisar o efeito dos protocolos na autonomia do veículo.*

1. Introdução

As montadoras de veículos tm investido muito em produção de veículos elétricos. Algumas montadoras, como a Sueca Volvo anunciaram que em 2020 produzirão somente carros elétricos ou híbridos e não mais produzirão carros a combustão. Os carros elétricos apontam uma tecnologia promissora para o futuro. No escopo desta tecnologia não podemos considerar apenas os carros de passeios; há de considerar que os veículos de transporte de passageiros e de cargas seguem a mesma tendência [Verge 2017].

No início de 2018, a frota mundial chegou a 3,2 milhões de veículos elétricos – o que representa um aumento de 55% em relação ao mesmo período do ano anterior, de acordo com o Centro de Pesquisa em Energia Solar e Hidrogênio de *Baden-Württemberg* (ZSW), na Alemanha. Conforme o estudo, 1,2 milhões de carros elétricos já circulam pelas ruas chinesas. Somente em 2017, 579 mil unidades foram acrescentadas à frota do país [Globo 2018].

A cidade chinesa de *Shenzhen* migrou 100% de sua frota de ônibus do transporte público para veículos elétricos. São mais de 16 mil ônibus para atender uma população de 12 milhões de habitantes [Engenharia E 2018].

A grande motivação para a produção de carros elétricos está ligada às questões ambientais e isso é uma tendência mundial. Considerando a perspectiva futura sobre os carros elétricos, a evolução tecnológica no contexto dos componentes e acessórios eletrônicos para veículos, aliado ao conceito de cidades inteligentes que buscam a melhoria da qualidade de vida e segurança, as redes oportunistas VANETs se tornam um objeto de estudo à procura de soluções que viabilizem uma maior utilização e uma melhor eficiência.

O aumento na demanda de redes sem fio nos últimos anos cria um cenário de preocupação nas operadoras se considerarmos a infraestrutura das redes hoje. Esta preocupação levou pesquisadores a estudarem sobre tecnologia para redes oportunistas, em busca de soluções para: os gargalos nas infraestruturas de redes, a comunicação em locais onde a infraestrutura não alcança e a comunicação em locais de desastres onde a infraestrutura tenha sido danificada. Nas redes oportunistas, a informação é passada através de nós intermediários até que chegue ao destino, estratégia denominada *store-carry-and-forward* [Trifunovic et al. 2017]. Uma vantagem deste modelo é que a rede é baseada na necessidade, aproveitando uma oportunidade de modo que comunicação ocorra sem que seja preciso ter conhecimento prévio da topologia [Pond et al. 2013].

Dentro do contexto de redes oportunistas emprega-se a comunicação entre dispositivos móveis, como *Tablets e Smartphones* e veículos. O foco deste trabalho são as redes veiculares (VANET), um subtipo de MANET (*Mobile Ad hoc Network*). Nas MANET, os nós são móveis e a topologia é dinâmica. Já as VANET utilizam veículos como nós da rede. Esta última se destaca drasticamente em relação às MANET, principalmente, devido a alta velocidade e a restrição de mobilidade dos automóveis [Yousefi et al. 2006].

Por uma década os pesquisadores têm procurado meios para impulsionar a implantação global das Redes Oportunistas [Trifunovic et al. 2017]. Estas redes podem auxiliar a comunicação no descarregamento da rede de celular, a comunicação em áreas desafiadoras (área onde a infraestrutura está parcial ou totalmente destruída), a circunscisão de censura e as aplicações baseadas em proximidade. As redes veiculares vêm fortalecer os Sistemas de Transporte Inteligente – ITS (*Intelligent Transport System*), impactando diretamente no aumento da segurança veicular e na eficiência do tráfego. Logo, essas redes contribuem para o conceito de cidades inteligentes. Ao unirmos os dois conceitos, percebe-se que as redes oportunistas veiculares podem favorecer a implementação dos ITS e podem ainda ser uma solução para os problemas de infraestrutura de redes ao considerarmos as limitações já apresentadas.

As redes oportunistas ainda não foram amplamente implementadas e alguns fatores contribuem para isto, tais como o desinteresse comercial das operadoras de redes, visto que elas não teriam o controle sobre a infraestrutura e a preocupação com o consumo de bateria nos dispositivos conectados à rede. Em 10 anos de estudos, muitos aspectos foram analisados, mas há assuntos que ainda precisam ser explorados [Trifunovic et al. 2017].

A tecnologia das redes oportunistas aplicadas às redes veiculares vêm contribuir diretamente para com o conceito de cidades inteligentes, na perspectiva de não utilizar a infraestrutura de redes de telefonia. Há que se levar em consideração que os veículos possuem bateria e que estas são geradas/recarregadas enquanto os carros estão em funcionamento. Nos veículos com motores a combustão, o consumo de energia pela rede veicular

não interfere na autonomia. Já nos veículos elétricos, embora tenha uma recarga constante de bateria, o consumo desta energia pelas redes veiculares representa uma redução da autonomia do veículo, o que desperta então, uma necessidade de avaliação do consumo desta energia, na busca de uma melhor eficiência energética.

Os sistemas inteligentes de transportes contribuem diretamente para com o conceito de cidades inteligentes, pois melhoram a qualidade do transporte e a segurança das vias. Porém estes sistemas ainda atuam utilizando a infraestrutura de redes das companhias de telefonia.

O estudo sobre a eficiência energética em redes oportunistas veiculares se torna interessante, principalmente, quando identificamos que as montadoras e os governantes estão atuando diretamente para a redução da frota de veículos a combustão e migrando a produção de veículos totalmente ou parcialmente elétricos.

Os dados apontam um futuro para os carros elétricos, e neste contexto qualquer consumo de energia por circuitos eletrônicos nos carros implica em redução de autonomia. Considerando que as redes VANET vêm ao encontro desta evolução ao proporcionar maior segurança e maior mobilidade, entre outros benefícios, precisamos atentar também para o consumo energético. Destaca-se ainda que as trocas de informações que utilizam a infraestrutura de redes *Wi-Fi*, *bluetooth* ou outra tecnologia têm sido extremamente demandadas. Desta forma, ao pensar em uma alternativa de tráfego de dados, temos as redes oportunistas.

Este trabalho cria um modelo abstrato de veículos elétricos e analisa o consumo energético dos protocolos de encaminhamento *Epidemic*, *Spray and Wait*, *Dlife* e *Prophet* em VANETs. Estes protocolos possuem diferentes características como: 1. *Epidemic* é um protocolo de encaminhamento que dissemina mensagens aos nós vizinhos. Este ciclo é repetido até que a mensagem chegue ao destino [Mota et al. 2014]. 2. *Spray and Wait* propaga L cópias da mensagem, onde L indica as cópias máximas permitidas da mensagem [Khan and Rahim 2018]. 3. *Dlife* é um protocolo de roteamento com consciência social que considera o dinamismo do comportamento do usuário com base em sua vida diária [Khan and Rahim 2018]. 4. *PRoPHET* é um protocolo ciente do contexto. Este protocolo usa uma métrica probabilística chamada previsibilidade de entrega [Mota et al. 2014].

Ao considerar o cenário de redes veiculares oportunistas e a previsão de aumento da produção de veículos elétricos pela indústria automobilística é que este estudo analisa o consumo de energia em protocolos de encaminhamento. O objetivo deste trabalho é desenvolver um modelo de abstração de veículos elétricos e demonstrar, dentre os protocolos citados acima, qual dos quatro apresentam maior eficiência energética. Ao final da pesquisa comprova-se qual o protocolo que garante maior autonomia para os veículos elétricos. Isto porque quanto menor o consumo de energia dos carros elétricos, maior sua autonomia.

A avaliação do consumo energético nos protocolos de roteamento é feito a partir de simulações usando o ambiente de simulação *ONE*. O cenário é implementado considerando a cidade *Helsinki*. O objetivo é implementar testes no ambiente para cada protocolo e coletar as informações dos testes a partir de relatórios do simulador. Após a coleta de dados, os mesmos são tabulados para uma análise do consumo energético e para a

geração dos gráficos. A análise dos dados resulta na identificação do protocolo com maior eficiência energética dentre os quatro protocolos analisados. Serão analisados também os dados referentes à taxa de entrega e latência de cada protocolo.

Este artigo está organizado da seguinte forma: após a introdução, a seção 2 apresenta os trabalhos relacionados. A seção 3 aborda os protocolos de encaminhamento apresentando suas características. A seção 4 apresenta uma avaliação do ambiente de simulação, tais como as características do cenário de simulação, a configuração dos nós, o modelo proposto dos veículos elétricos, entre outras. Os resultados são apresentados e discutidos na seção 5. Por fim, na seção 6 concluímos o estudo.

2. Trabalhos relacionados

O consumo energético em redes *Ad Hoc* e Redes Oportunistas vêm sendo um tema bastante explorado nos últimos anos, o que resultou no desenvolvimento de vários protocolos que prometem reduzir o consumo de energia dos nós das redes. Muitos autores afirmam que VANETs não possuem limitações energéticas [Mota et al. 2014], o que é verdade em carros movidos a combustão. Porém, não podemos dizer o mesmo para veículos elétricos. Várias técnicas foram analisadas e pesquisadas para que se criassem algoritmos cada vez mais eficientes.

Encontramos na literatura alguns exemplos de discussões e algumas propostas de solução para o problema energético em nós de redes oportunistas [Mota et al. 2014, Landsiedel et al. 2012, Baji Baba and Rao 2016]. Em sua maioria, o foco das pesquisas são sensores ou dispositivos móveis, que possuem baixa capacidade energética e pouca ou nenhuma mobilidade. Eles exploram o fato de que, nesses cenários, a criação de uma conexão fim a fim entre dois dispositivos pode ser inviável, ou até mesmo impossível, para desenvolver algoritmos de roteamento oportunista.

[Mota et al. 2014] faz uma comparação entre vários protocolos em diferentes cenários e oferece uma boa compreensão sobre os mesmos. Embora não possua os mais recentes, serve como base para iniciarmos nossos estudos. [Landsiedel et al. 2012] introduz o conceito de ciclo de serviço, usando oportunisticamente os nós que estão acordados para fazer a transmissão dos pacotes. E [Baji Baba and Rao 2016] propõe um algoritmo que gerencia um escalonamento de nós, permitindo que todos durmam alternadamente por um tempo, sem que a rede seja prejudicada, assim economizando energia.

Devido à alta dinamicidade e previsibilidade das redes VANET, algoritmos de clusterização têm se tornado uma boa alternativa para aprimorar a conexão entre veículos. [Toutouh et al. 2013] explora uma metodologia que utiliza algoritmo evolucionário para buscar um OLSR (*Optimized Link State Routing Protocol*) de melhor eficiência energética. Nesta análise experimental, foi demonstrado uma redução considerável no consumo energético da rede, sem perda notável na qualidade do serviço. Já [Bali et al. 2017] desenvolveu um algoritmo de predição de variação média de veículos para superar o problema da curta duração do *cluster*, mostrando que o esquema proposto é eficiente em manter alta vazão, alta probabilidade de transmissão e menor atraso fim a fim.

A partir dos estudos acima, percebe-se que há uma grande preocupação com o consumo energético em redes oportunistas e em redes veiculares. Porém, ainda existem muitos casos em aberto. Um deles, e talvez o mais eminente, é o caso de veículos

elétricos, cujo consumo da bateria possui efeito direto na autonomia do carro. Diante do exposto, decide-se então, realizar uma simulação de alguns protocolos, a fim de obter dados a respeito do impacto dos mesmos no gasto da bateria dos automóveis. O simulador *The One* apresenta as características necessárias para realizar tal feito, sendo assim escolhido para esta pesquisa.

3. Protocolos analisados

Dentre os protocolos disponibilizados no simulador ONE, quatro foram escolhidos para serem analisados em redes oportunistas VANET: *Epidemic*, *Spray and Wait*, *dLife* e *Prophet*. Cada protocolo estudado possui características de roteamento distintas, como se segue.

Epidemic é um protocolo de encaminhamento que dissemina mensagens muito parecido com a disseminação de uma doença. Um nó que recebe uma mensagem encaminha para todos os seus vizinhos, que repetem esse processo. Este ciclo é repetido até que a mensagem chegue ao seu destino ou até que um número máximo de saltos seja atingido. O objetivo do *Epidemic* é maximizar a entrega de dados e minimizar o atraso [Mota et al. 2014].

Spray and Wait propaga L cópias da mensagem, onde L indica as cópias máximas permitidas da mensagem e esse parâmetro é selecionado com base na densidade da rede e no tempo médio desejado. *Spray* e *Wait* é composto por duas fases: Fase 1) L cópias da mensagem são encaminhadas para os primeiros L nós encontrados. Esta fase é chamada de fase “*Spray*”. Fase 2) no caso, o destino não pode ser identificado na fase “*Spray*”, todos os nós (individuais de nós L) que receberam uma versão da mensagem esperam para encontrar o nó de destino e entregar diretamente a mensagem. Isto é conhecido como “Aguarde” [Khan and Rahim 2018].

dLife é um protocolo de roteamento com consciência social que considera o dinamismo do comportamento do usuário com base em sua vida diária. Ele tira proveito de estruturas sociais que evoluem no tempo em um ambiente oportunista. No protocolo *dLife*, o nó de origem entrega uma cópia da mensagem para o nó encontrado, pois tem uma relação maior com o destino do que com a origem [Khan and Rahim 2018].

PRoPHET é um protocolo ciente do contexto. Este protocolo usa uma métrica probabilística chamada previsibilidade de entrega, que indica a probabilidade de um nó ser capaz de entregar uma mensagem para um determinado destino. Ele leva em conta o histórico de contatos anteriores. Além disso, assume que o movimento do nó não é aleatório e que é possível identificar padrões de mobilidade. Assim, se um nó visitou frequentemente uma parte da rede, existe uma grande probabilidade de que esse mesmo local seja visitado no futuro [Mota et al. 2014].

Alguns testes relativos ao desempenho dos protocolos foram apresentados por [Khan and Rahim 2018]. Este trabalho analisou três métricas destes protocolos: a taxa de entrega, a latência média e o custo de transmissão.

Em relação à taxa de entrega, considerando o aumento da densidade dos nós, observa-se que o protocolo baseado em redes sociais (*dLife*) é o com maior taxa de entrega, enquanto o *Epidemic* é o de menor taxa de entrega. Com relação ao tempo de vida, o *dLife* diminui um pouco a taxa de entrega, mas ainda mantém com maior taxa de entrega

e o *Epidemic* é o protocolo de menor taxa de entrega [Khan and Rahim 2018].

A latência média é verificada através do tempo gasto desde a geração no nó de origem até a recepção bem-sucedida no nó de destino. Com o aumento da densidade dos nós ou com o aumento do tempo de vida da mensagem (TTL) fica evidente que o protocolo *dLife* apresenta maior latência, e que o protocolo *Spray and Wait* (SNW) apresenta menor latência. Portanto, o SNW é o de melhor desempenho em ambos os casos, considerando a latência média [Khan and Rahim 2018].

O custo de transmissão é verificado considerando a retransmissão de pacotes não essenciais na rede com objetivo da entrega de um único pacote. O protocolo *Epidemic* é o de maior custo de transmissão, pois a sua característica de inundação da rede fez com que a curva do custo de transmissão tivesse um aumento constante com a variação da densidade dos nós ou mesmo do TTL. O menor custo de transmissão ficou para o protocolo SNW por ter a característica de uma réplica mínima de mensagens, tanto com o aumento da densidade dos nós como com o aumento do tempo de vida [Khan and Rahim 2018].

4. Avaliação

Nesta seção é apresentada e discutida as capacidades do simulador e os modelos de simulação, bem como as métricas usadas para comparar os protocolos de roteamento.

4.1. Ambiente de Simulação

O *ONE* foi projetado especificamente para avaliação de protocolos de roteamento e para a aplicação de redes oportunistas. Ele permite que os usuários criem cenários baseados em diferentes modelos de movimentos sintéticos e de rastreamentos reais e oferece uma estrutura para implementar o roteamento de protocolos e aplicativos.

As principais funções do *ONE* são a modelagem da movimentação dos nós, o contato entre nós, o roteamento e a manipulação de mensagens. A coleta e a análise de resultados são feitos a partir visualizações, relatórios e ferramentas de pós-processamentos. Uma visão geral do funcionamento do *ONE* pode ser vista na Figura 1 [Keränen et al. 2009].

O *ONE* nos permite especificar uma semente (seed) para a geração de valores aleatórios utilizados durante a simulação. Isto facilita a comparação de resultados pois garante que as mensagens enviadas, caminhos percorridos, etc. sejam os mesmos em diferentes execuções, possibilitando que isolemos as variáveis que desejamos comparar.

4.2. Modelo proposto para veículos elétricos

Nesta seção descreve-se o modelo proposto para simulação de veículos elétricos dentro do simulador *ONE* e como utilizá-lo. O objetivo deste modelo é permitir a análise do impacto energético dos protocolos em redes veiculares nos quais os carros movidos a bateria façam parte.

Este simulador já possui módulo de bateria e consumo de energia para os protocolos. Assim, foi necessário apenas uma adaptação para que o automóvel consumisse energia a medida em que andasse. Da mesma forma, o módulo de movimentação nos permite calcular de maneira suficientemente a precisa distância percorrida durante a movimentação, já que este utiliza de um vetor com pontos no mapa para descrever o

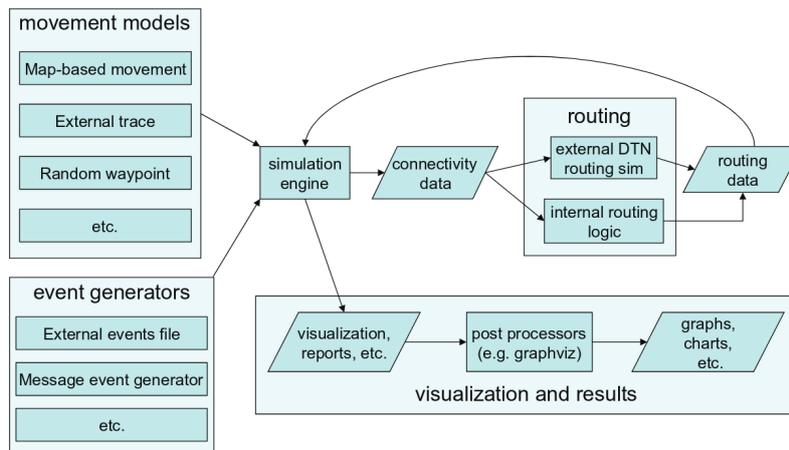


Figura 1. Visão geral do ambiente de simulação ONE

caminho escolhido, sendo necessário apenas somar a distância entre cada ponto do vetor a fim de obter o total percorrido. A equação pode ser vista em (1).

$$D = \sum_{i=1}^{n-1} \sqrt{(V_i x - V_{i+1} x)^2 + (V_i y - V_{i+1} y)^2} \quad (1)$$

Para se calcular o total gasto por um veículo elétrico durante um percurso precisamos entender sobre o funcionamento deste. Estes automóveis possuem uma bateria que alimenta um motor elétrico responsável pela movimentação do mesmo. [Miller 1993] descreve detalhadamente o funcionamento de um motor elétrico, que, de maneira simplificada, converte a energia elétrica em um campo magnético que faz um eixo girar e, assim, o faz movimentar.

Com estas informações, foi possível abstrair o funcionamento de um veículo elétrico para o simulador *ONE*. Notamos que as informações necessárias para o funcionamento do módulo seriam a capacidade da bateria e o consumo por quilômetro. Como já existe um módulo de bateria no simulador, implementamos apenas o consumo do automóvel.

Para calcular a energia consumida pelo veículo é necessário que sejam fornecido os valores das configurações citadas acima, sendo que para o valor de consumo deve-se utilizar a mesma unidade usada para definir a capacidade da bateria. A equação utilizada pode ser vista na Equação 2, onde G é o valor gasto de energia, C é o consumo fornecido nas configurações e D a distância percorrida, obtido a partir da Equação (1).

$$G = C * D \quad (2)$$

4.3. Cenário

Neste trabalho usou a modelagem da cidade de *Helsinki*, disponibilizado pelo *ONE*, para simular um ambiente urbano. Este cenário foi escolhido por ser uma representação realista de uma cidade, com diversas ruas e avenidas, e, portanto, serviu para todos os propósitos desta pesquisa.

A avaliação foi realizada em dois cenários distintos, variando a densidade dos nós. Em ambos os casos, os nós se movem nas ruas da cidade percorrendo o caminho mais curto entre a posição atual e outro ponto escolhido aleatoriamente. Estes cenários foram escolhidos com o propósito de se obter uma visão geral do comportamento dos protocolos em diferentes situações.

Cenário 1 – Baixa densidade: este cenário visa analisar o consumo dos protocolos em localidades de baixo tráfego, como bairros residenciais distantes dos grandes centros urbanos e fora do horário de *rush*. Neste caso temos poucos encontros e os mesmos são de curta duração, dificultando o envio de mensagens.

Cenário 2 – Alta densidade: neste caso, o objetivo do cenário é analisar o consumo energético nos grandes centros das cidades, especialmente durante o horário de *rush*, onde o número de veículos é alto e velocidade de tráfego é baixa, implicando um grande número de conexões possíveis que podem durar por bastante tempo.

Os parâmetros da simulação deste estudo podem ser vistas na Tabela 1, na qual foram detalhados os valores do tamanho da área e duração da simulação. Nesta tabela também especificamos as configurações da interface de rede utilizada pelos nós e os parâmetros para geração de mensagens pelo simulador.

Parâmetro	Valor
Dimensão do cenário	4500 x 3400 m
Tempo de simulação	86400 s
Velocidade de Transmissão	1 Mb
Alcance da Transmissão	50 m
Tempo de scan	32 s
Tamanho do Buffer	32 MB
Intervalo de geração de mensagens	25 a 35 s
Tamanho da mensagem	500 KB a 1 MB

Tabela 1. Configurações do cenário de simulação

Na tabela 2, temos as informações referentes às configurações dos nós que fazem parte da rede. Temos 4 grupos com características distintas comunicando entre si, pedestres, veículos elétricos, veículos à combustão e bonde.

Como características dos veículos elétricos temos os dados referente a capacidade da bateria e à autonomia do veículo, que foram obtido no site da Tesla, baseando no *Model S 75D*. Temos também o consumo de energia da placa de rede durante o *scan* da rede e a transmissão de mensagens, de acordo com os trabalhos de [Silva et al. 2012]. Consideramos a bateria com capacidade de 75 kWh e o veículo consumindo 188 W/kM [Tesla 2018], enquanto a placa de rede consome 0.93 mW durante o *scan* e 0.08 mW para responder e transmitir uma mensagem [Silva et al. 2012].

Temos ainda informações sobre o número de nós, sendo total de 192 nós para o cenário denso e 80 nós para o cenário esparsos. Cada nó possui um *buffer* que armazena 32MB. O cenário é a cidade de *Helsinki* e o movimento dos nós dos pedestres, dos

veículos elétricos e dos veículos à combustão são aleatórios no mapa. Já os nós dos bondes se movimentam obedecendo a uma rota definida.

Utilizando o modelo de veículo elétrico descrito, focamos as análises deste estudo no consumo energético dos protocolos de roteamento. Porém não descartamos outras métricas, como o atraso e a probabilidade de chegada da mensagem, que são muito importantes no contexto de redes oportunistas.

	Pedestre	Veículo Elétrico	Veículo à combustão	Bondes
Nº de nós (denso)	40	60	80	16
Nº de nós (esparso)	14	20	30	16
Velocidade (km/h)	2 a 5.5	10 a 50	10 a 50	5 a 37
Movimento	Aleatório	Menor caminho	Menor caminho	Ruas definidas

Tabela 2. Configurações dos nós

5. Resultados obtidos

As simulações foram executadas por um tempo equivalente a 24 horas. Os resultados foram coletados e são analisados nesta seção. Na Tabela 3, podemos ver um resumo do tráfego de mensagens pelos nós. A simulação considerou 20 veículos elétricos no cenário esparso (total 80 nós) e 60 veículos elétricos no cenário denso (total 196 nós).

O protocolo de encaminhamento *dLife* é o que resulta em maior energia residual, seguido pelo *Spray And Wait* (SnW). Neste caso, o veículo elétrico teria uma autonomia maior usando protocolo de encaminhamento *dLife* para a comunicação oportunista proposta. O protocolo *dLife* possui uma característica de encaminhamento com base no dinamismo do comportamento do usuário. O *Spray and Wait* encaminha mensagens considerando um limite máximo de mensagens a ser enviadas, calculado conforme a densidade da rede. Estas características permitem a redução do consumo de energia favorecendo a autonomia do veículo. O protocolo *Epidemic* é o de maior gasto energético. Isto se dá pela característica de disseminação, onde um nó recebe uma mensagem e encaminha para todos os vizinhos. Esta característica de encaminhamento leva ao consumo maior de energia do protocolo, reduzindo a energia que o veículo poderia utilizar para seu deslocamento. Considerando o número de mensagens entregues deparamos com a figura 3.

	Spray and Wait	dLife	Epidemic	Prophet
Mensagens criadas	2939	2939	2939	2939
Mensagens entregues	2909	2565	1439	1399
Probabilidade de entrega	0.99	0.87	0.49	0.48
Atraso médio	919.987	4579.889	1206.295	1336.715
Mediana do atraso	726.840	3317.700	955.480	1049.120

Tabela 3. Resultado das simulações

O *Spray and Wait* é o que tem maior taxa de entrega de mensagens, entregando 99% das mensagens durante as 24 horas, enquanto o *dLife* entrega 12% a menos, com uma taxa de entrega de 87% das mensagens. Os protocolos *Epidemic* e *Prophet* possuem taxa

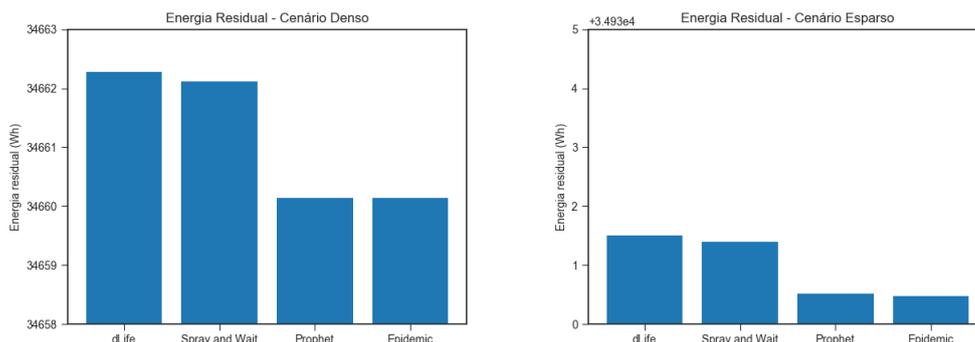


Figura 2. Energia residual de cada protocolo em cada cenário ao final da simulação

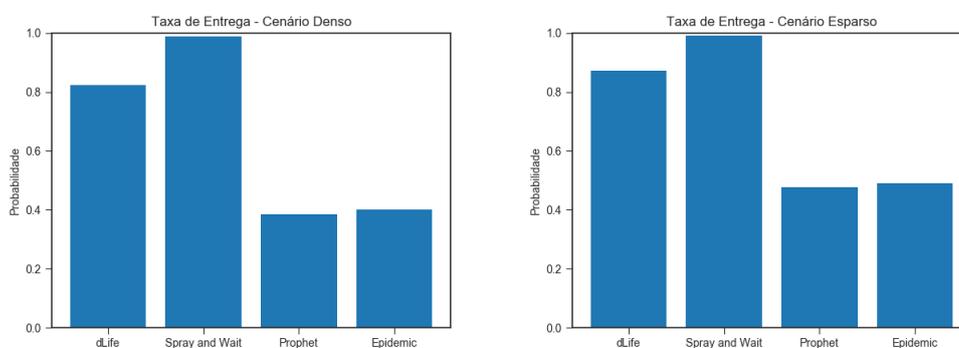


Figura 3. Taxa de entrega

de entrega de 49% e 48% respectivamente, uma diferença de apenas 1 ponto percentual, e consumo de energia elevado. Considerando que o movimento do veículo é similar em cada protocolo simulado, o baixo índice de entrega se justifica pelas características do próprio protocolo. O *Epidemic* dissemina as mensagens a todos os nós; isto leva ao *buffer overflow*, não permitindo uma taxa de entrega maior no prazo de 24 horas. Já o *Prophet* analisa primeiro o comportamento do nó e considera a probabilidade de fazer o mesmo caminho novamente e, assim, entregar a mensagem.

A latência média dos protocolos mostra uma diferença muito grande para o protocolo *dLife* e os demais protocolos. O *Spray and Wait* apresenta menor taxa de latência, seguido pelo *Epidemic* e depois pelo *Prophet*. O *dLife* apresenta uma latência quase 5 vezes superior à latência do *Spray and Wait*. Esta latência justifica a taxa de entrega menor do *dLife* em relação ao *Spray and Wait*.

A latência do protocolo *dLife* é alta em função de sua característica de consciência social, e sua taxa de entrega é menor que o *Spray And Wait*, mas ele ainda tem uma taxa de entrega de 87%. Ao analisarmos os gastos de energia nos protocolos temos os resultados apresentados na Figura 2.

Considerando os dados apresentados, a escolha pelo protocolo de menor consumo energético não prejudicaria a entrega de mensagem se levarmos em conta o benefício direto em relação à economia de energia, para fins de movimentação de veículo elétrico.

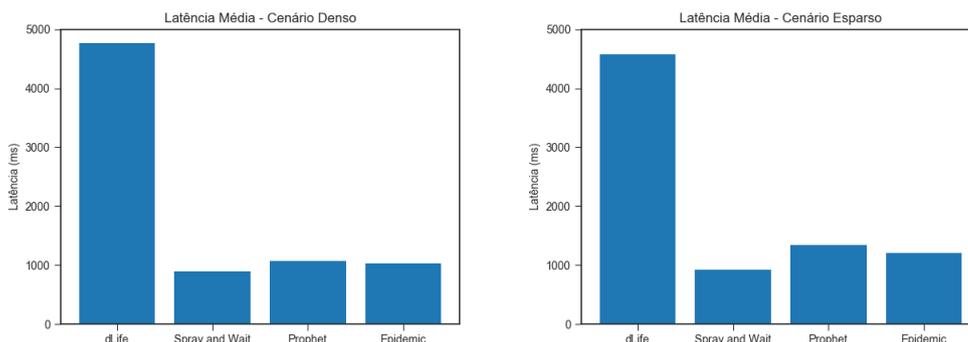


Figura 4. Latência Média

6. Conclusão

Neste trabalho propusemos um modelo de veículos elétricos para simulação e análise de consumo energético em simuladores. Esta abstração propõe que automóveis elétricos podem ser modelados a partir da capacidade de sua bateria e de seu gasto por quilômetro. Seu consumo pode ser calculado multiplicando-se a distancia percorrida pelo gasto. A autonomia destes veículos será maior se os mesmos tiverem seus componentes eletrônicos consumindo a menor energia possível.

Com este simples modelo, analisamos os protocolos almejados nesta pesquisa, *dLife*, *Epidemic*, *Prophet* e *Spray and Wait*, em dois cenários distintos, variando a densidade dos nós.

Nos cenários analisados, concluímos que a autonomia do veículo seria maior se utilizássemos o protocolo de encaminhamento *dLife*. O protocolo *Spray and Wait* viria como segunda opção do ponto de vista de consumo energético, embora possua uma taxa de entrega e uma latência média superior aos demais protocolos. Se o veículo elétrico trabalhasse com os protocolos *Epidemic* e *Prophet* a autonomia do veículo reduziria, visto que o consumo energéticos destes protocolos é maior, a taxa de entrega destes protocolos é baixa, o que tornaria mais negativo o uso destes protocolos na rede veicular.

Para trabalhos futuros, pretendemos melhorar o modelo de bateria utilizado e analisar outros protocolos e diferentes cenários, a fim de obter uma melhor perspectiva dos mesmos em veículos elétricos. Os cenários ainda podem ser variados conforme a velocidade de veículos; pode-se ainda aumentar a densidade e considerar apenas veículos elétricos na comunicação.

Referências

- Baji Baba, S. and Rao, K. M. (2016). Improving the network life time of a wireless sensor network using the integration of progressive sleep scheduling algorithm with opportunistic routing protocol. *Indian Journal of Science and Technology*, 9(17).
- Bali, R. S., Kumar, N., and Rodrigues, J. J. (2017). An efficient energy-aware predictive clustering approach for vehicular ad hoc networks. *International Journal of Communication Systems*, 30(2).

- Engenharia E (2018). Shenzhen anuncia sua nova frota de mais de 16.000 ônibus elétricos. online.
- Globo (2018). Frota mundial de carros eletrificados cresce 55% em um ano. online.
- Keränen, A., Ott, J., and Kärkkäinen, T. (2009). The ONE Simulator for DTN Protocol Evaluation. In *SIMUTools '09: Proceedings of the 2nd International Conference on Simulation Tools and Techniques*, New York, NY, USA. ICST.
- Khan, M. K. M. and Rahim, M. S. (2018). Performance analysis of social-aware routing protocols in delay tolerant networks. *International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering*.
- Landsiedel, O., Ghadimi, E., Duquenooy, S., and Johansson, M. (2012). Low power, low delay: opportunistic routing meets duty cycling. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks*, pages 185–196. ACM.
- Miller, T. (1993). *Switched Reluctance Motors and Their Control (Monographs in Electrical and Electronic Engineering)*. Clarendon Press.
- Mota, V. F., Cunha, F. D., Macedo, D. F., Nogueira, J. M., and Loureiro, A. A. (2014). Protocols, mobility models and tools in opportunistic networks: A survey. *Computer Communications*, 48:5–19.
- Pond, B. L., Júnior, A. C. O., and Moreira, W. (2013). Aspectos sociais como suporte ao roteamento em redes oportunistas. pages 1–6.
- Silva, D. R., Costa, A., and Macedo, J. (2012). Energy impact analysis on dtn routing protocols. *ExtremeCom 2012*, pages 1–6.
- Tesla (2018). Model s — tesla. online.
- Toutouh, J., Nesmachnow, S., and Alba, E. (2013). Fast energy-aware olsr routing in vanets by means of a parallel evolutionary algorithm. *Cluster computing*, 16(3):435–450.
- Trifunovic, S., Kouyoumdjieva, S. T., Distl, B., Pajevic, L., Karlsson, G., and Plattner, B. (2017). A decade of research in opportunistic networks: Challenges, relevance, and future directions. *IEEE Communications Magazine*, 55(1):168–173.
- Verge, T. (2017). Volvo to end gas-only cars by 2019. online.
- Yousefi, S., Mousavi, M. S., and Fathy, M. (2006). Vehicular ad hoc networks (vanets): challenges and perspectives. In *ITS Telecommunications Proceedings, 2006 6th International Conference on*, pages 761–766. IEEE.

Autenticação de dispositivos de Internet das Coisas baseada nas características do sinal eletromagnético

**Marcos Felipe B. de Abreu¹, Pablo F. A. Sousa¹,
Vinicius da C. M. Borges¹, Antonio C. de O. Júnior¹, Kleber V. Cardoso¹**

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Quadra D, Câmpus Samambaia CEP 74690-900 - Goiânia - GO – Brazil

{marcos,pablosousa,vinicius,antonio,kleber}@inf.ufg.br

Abstract. *Accessing unauthorized devices on a network is an security issue on Internet of Things. The unique identifiers of the devices used to authenticate them can be easily cloned, thus requiring another form of authentication. By monitoring the electromagnetic spectrum by means of a software defined radio it is possible to capture data transmitted by various wireless communication devices. These data allow us to extract unique characteristics of a device, since an electric circuit that generates the electromagnetic signal does not behave perfectly like another. These characteristics can then be used to create a unique signature, thus enabling device differentiation. There are several IoT technologies on the market and it is expected that the implementation of an authentication technique will be technology independent. For the validation of the technique are collected signals of nRF24L01+ technology devices, extracting characteristics of the magnitude of the signal. These characteristics are used in a classifier, which gives an accuracy of 94.7% in device differentiation.*

Resumo. *O acesso de dispositivos não autorizados em uma rede é um problema de segurança em Internet das Coisas. Os identificadores únicos dos dispositivos usados para autenticação dos mesmos podem ser facilmente clonados, sendo assim necessário outra forma de autenticação. Monitorando o espectro eletromagnético através de um rádio definido por software é possível capturar dados transmitidos por diversos dispositivos de comunicação sem fio. Esses dados nos permite extrair características únicas de um dispositivo, visto que, um circuito elétrico que gera o sinal eletromagnético não se comporta perfeitamente igual a outro. Essas características podem então ser usadas na criação de uma assinatura única, possibilitando assim a diferenciação de dispositivos. Há várias tecnologias IoT no mercado e espera-se que a implementação de uma técnica de autenticação seja independente de tecnologia. Para validação da técnica são coletados sinais de dispositivos da tecnologia nRF24L01+, extraindo características da magnitude do sinal. Essas características são usadas em um classificador, ao qual é obtida uma acurácia de 94,7% na diferenciação de dispositivos.*

1. Introdução

Internet das Coisas, do inglês *Internet of Things* (IoT), é um paradigma que está ganhando espaço na área de comunicações sem fio modernas devido ao seu grande impacto na vida da população. A quantidade de dispositivos conectados é enorme, com cerca de 9 bilhões de coisas conectadas à internet em 2013 e previsão de mais de 24 bilhões de coisas estejam conectadas até 2020 [J. Gubbi 2013]. E-saúde, assistentes pessoais, vida assistida são exemplos de aplicações presentes em um ambiente doméstico, enquanto no contexto corporativo temos, como exemplos, automação, logística, gerenciamento de negócios/pessoas e transporte inteligente [Gomez and Paradells 2010]

Padrões e especificações de segurança começam a ser pensados, a exemplo do Brasil com o Plano Nacional de Internet das Coisas [BNDES 2017] e nos Estados Unidos, onde foi elaborado um documento listando riscos e oferecendo sugestões para aplicação de segurança na Internet das Coisas [of Homeland Security 2016]. Observa-se, entretanto, um cenário diferente, dispositivos tem apresentado falhas graves que comprometem a segurança dos mesmos [Zhao and Ge 2013]. Um ataque bastante conhecido em redes de computadores, que se repete em IoT, é o roubo de identidade [Nawir et al. 2016]. Como a maior parte não provém mecanismos de segurança a implementação desses não ultrapassam o nível de enlace, a autenticação de dispositivos sem fio por padrão são baseados em identificadores únicos, como o endereço MAC (*Media Access Control*) no Bluetooth e no Wi-Fi. Entretanto, esses dispositivos podem ser forjados por alguém mal intencionado para se passar pelo dispositivo original, abrindo assim portas para execução de outros ataques.

Tradicionalmente o problema de autenticação é resolvido usando criptografia, mas isso se torna inviável quando falamos de IoT devido as características dos dispositivos, que contam com um baixo poder computacional. Os autores em [Xu et al. 2014] mostram que tratar ao nível de hardware é uma solução, mas demoraria para entrar no mercado incluir esse tipo de tecnologia em dispositivos que possuem como ideal serem baratos. Uma das soluções é trazer a responsabilidade de autenticação desses dispositivos para um terceiro que possui mais poder computacional.

Dispositivos sem fio podem ser diferenciados com base em sua assinatura eletromagnética [Danev et al. 2012]. A diferenciação de placas é possível, pois, são geradas imperfeições, não propositais, no processo de fabricação dos módulos geradores de sinais dos dispositivos. Essas imperfeições influenciam na geração do sinal a ser transmitido. Com base nisso é possível analisar o sinal transmitido por um dispositivo e criar uma assinatura para o mesmo.

A ideia de um *gateway* que conecte vários dispositivos IoT e ofereça, através de uma solução centralizada, segurança e processamento vem ganhando espaço em IoT [Zhu et al. 2010]. A proposta é analisar e identificar dispositivos IoT, sendo que para isso é necessário utilizar um SDR (*Software Defined Radio*). O SDR é um dispositivo capaz de capturar e analisar sinais independente da tecnologia usada pelas coisas, além de permitir implementar a solução no *gateway*, evitando aumento de latência na rede. Isso se torna viável, pois, a proposta é que a solução execute em um *gateway* que se comunique com múltiplas tecnologias de redes sem fio, e.g., SOFTWAY4IoT [LABORA 2017].

O objetivo do trabalho é investigar a viabilidade de autenticar dispositivos de IoT

com base no sinal eletromagnético das suas interfaces sem fio. É importante que a solução possa ser integrada a um *gateway* de comunicação IoT. O trabalho é dividido em duas etapas. Na primeira etapa, é apresentada uma prova de conceito da técnica de análise dos sinais eletromagnéticos que permite diferenciar dispositivos com base em um conjunto de características peculiares (*features*). Na segunda etapa, é feita uma análise sobre a aplicação da técnica avaliada a um *gateway* de comunicação IoT.

Esse trabalho está organizado conforme segue: Na Seção 2, são apresentados os trabalhos relacionados. Na seção 3, é detalhada a metodologia aplicada no trabalho. Na seção 4, são apresentados os experimentos e os resultados obtidos e, na seção 5, é discutida a implementação da técnica em um *gateway* IoT. A seção 6 apresenta a conclusão e os trabalhos futuros.

2. Trabalhos Relacionados

Identificar dispositivos com base nas suas características eletromagnéticas não é um tema novo. Na Segunda Guerra Mundial controladores de voo analisavam visualmente ondas emitidas por radares para identificar possíveis transmissões que não vinham de seus dispositivos. Os primeiros trabalhos na área de computação, que automatizaram esse processo se deu em 1995 com os trabalhos de [Toonstra and Kinsner 1995] e [Choe et al. 1995].

[Danev et al. 2012] faz um resumo da área de identificação de dispositivos sem fio e traz algumas métricas usadas para medir a viabilidade e a eficácia de novos métodos. A performance de um método é medida de acordo com a velocidade, acurácia, custo e segurança. A métrica velocidade é utilizada para medir o tempo gasto para o sistema tomar alguma decisão, por exemplo bloquear um dispositivo na rede. Essa métrica é de extrema importância, se o trabalho é executado de uma forma reativa, ou seja, autoriza-se qualquer dispositivo na rede enquanto a análise do sinal é feita, um dispositivo mal intencionado pode nesse meio tempo já ter obtido algum sucesso em seu ataque e ou exploração. Por outro lado, se o trabalho é executado de uma forma ativa, ou seja, o dispositivo é autorizado somente após a análise do seu sinal, é introduzido um atraso de transmissão na rede e torna-se necessária a implementação de políticas de cache. A segunda métrica, acurácia, trata da porcentagem de erros em uma técnica. Quer se minimizar duas taxas de erro, a FAR (*False Accept Rate*), que é quando há a autorização de um dispositivo impostor, e a FRR (*False Reject Rate*), que é a probabilidade de não autorizar um dispositivo genuíno. A métrica custo está relacionada ao hardware necessário para fazer a análise e o processamento desse sinal. Para se obter o sinal eletromagnético em um dispositivo é necessário que o hardware ofereça essa função.

A escolha das características é uma parte extremamente importante no trabalho, visto que essa terá total impacto na diferenciação dos dispositivos. A escolha está relacionada às imperfeições geradas na fabricação de cada placa. [Danev and Capkun 2009] trabalha com características coletadas do início da transmissão do sinal pelo dispositivo, essas chamadas na literatura de *transient features*. Baseia-se em parâmetros como distância, polarização da antena e voltagem, posteriormente verificam como esses parâmetros influenciam na acurácia da coleta e por fim verificam a viabilidade de certos ataques de personificação de dispositivo.

Várias tecnologias de redes sem fio já foram exploradas. Os trabalhos [Hall et al. 2005] e [Barbeau et al. 2006] diferenciam dispositivos Bluetooth usando

como características Amplitude, Frequência e Fase, coletadas do início da transmissão do sinal. [Bihl et al. 2016] aplica algoritmos de seleção de características para identificação de dispositivos Zigbee. Nenhum desses trabalhos considera múltiplas tecnologias, ou ainda, extração de características independentes de tecnologia, capazes de diferenciar dispositivos diversos, sejam eles pertencentes a qualquer padrão.

Há na literatura outras formas de abordar o problema de identificação de dispositivos com base no comportamento da transmissão. [Meidan et al. 2017] classifica dispositivos analisando pacotes TCP recebidos. [Bezawada et al. 2018] trabalha no problema de identificação de dispositivos IoT usando cabeçalho e carga útil dos pacotes TCP/IP. A desvantagem dessa abordagem é o tempo que leva para coletar, analisar e tomar uma decisão.

Outra abordagem, [Verma et al. 2015] propõe autenticar dispositivos com base em uma `tag`, sinal de baixa potência que o transmissor adiciona em sua transmissão. Uma prova de conceito é feita usando um par de SDRs, sendo demonstrada a viabilidade e a segurança oferecidas pela técnica. Entretanto para a aplicação do método é necessário que ambos dispositivos, transmissor e receptor ofereçam suporte, o que torna os dispositivos de Internet das Coisas atuais incompatíveis.

3. Metodologia

Nessa seção, é detalhada a metodologia usada neste artigo. A Figura 1 mostra o fluxo do experimentos, o qual é apresentado em detalhes a seguir.



Figura 1. Fluxo dos experimentos.

Transmissão

A Figura 2 apresenta o ambiente experimental utilizado para transmissão e captura do sinal.

Coleta do sinal

O equipamento utilizado para a captura do sinal é o SDR. Composto de todos os componentes de um rádio comum, o SDR possui o que é chamado de *RF Front End* programável. O *RF Front end* de um rádio é composto por filtros, amplificadores, oscilador local e mixer. Esses componentes podem ser acessados via softwares de configuração fornecidos pelas fabricantes, ou até de ferramentas de processamento de sinal como GNU-Radio [Blossom 2004]. Os componentes podem ser programados utilizando a linguagem C/C++, ou em alguns casos em Python. O SDR possui duas antenas, podendo ser configuradas para trabalharem ambas em modo de recepção ou, uma em modo de recepção e outra em modo de transmissão.

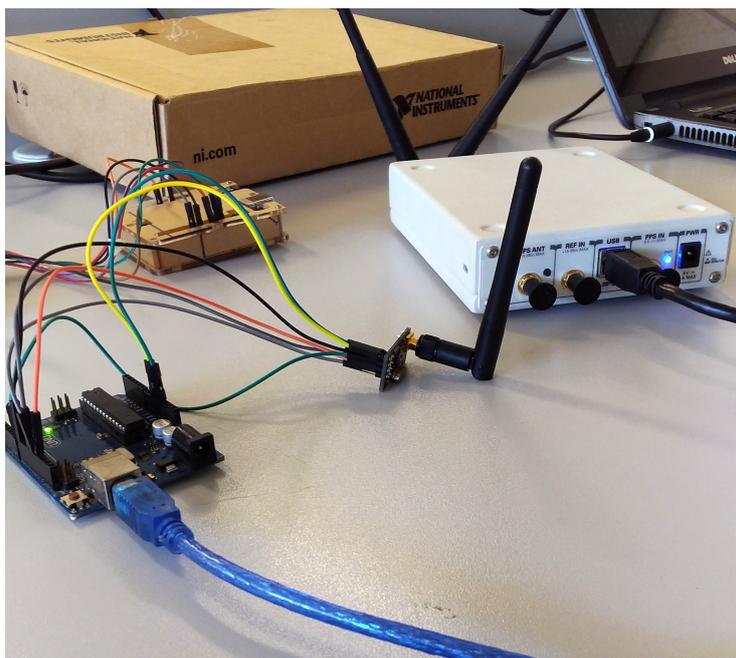


Figura 2. SDR e dispositivo de IoT.

Processamento do sinal

A ferramenta utilizada para realizar a análise e extração das características do sinal recebido é chamada GNURadio. GNURadio é um *toolkit* para implementação em SDR, que traz blocos como ferramentas de processamento de sinal. Esses blocos podem exercer as mais variadas funções, como filtros, moduladores, conversores de domínio do sinal, etc. Os blocos podem conter *buffers* de entrada, *buffers* de saída ou ambos, e são denominados *source blocks*, *sink blocks* e *general blocks* respectivamente.

Na Figura 3, é apresentado o conjunto de blocos usados no GNURadio para realizar o processamento do sinal eletromagnético e a extração das características peculiares. Para identificar um determinado dispositivo, é preciso primeiramente armazenar os dados capturados em um *buffer*, enviar essas amostras para um demodulador específico da tecnologia, identificar o preâmbulo do pacote transmitido pelo dispositivo, se há a existência de um preâmbulo, as amostras anteriormente guardadas no *buffer* são armazenadas em um arquivo.

Ao iniciar a captura, armazena-se os valores no *buffer* de entrada do bloco *Stream to vector*, que faz a decimação do sinal. Então esse bloco envia o sinal decimado para um outro bloco chamado FFT (*Fast Fourier Transform*) que converte os dados do domínio do tempo para o domínio da frequência. O *buffer* de saída do bloco FFT entrega um vetor de números complexos, a partir do qual é possível obter a magnitude do sinal (distância entre o símbolo e a origem) em cada frequência, onde cada valor do vetor é aplicado à seguinte expressão:

$$|X[k]| = \sqrt{X_{re}^2 + X_{im}^2}, \quad (1)$$

onde X_{re} representa a parte real do sinal e X_{im} representa a parte imaginária. A fase de

cada componente do sinal é obtida através da aplicação de cada valor do vetor na seguinte expressão:

$$\angle X[k] = \tan^{-1} \left(\frac{X_{im}}{X_{re}} \right). \quad (2)$$

Outra característica extraída é o EVM (*Error Vector Magnitude*), baseada nos erros relacionados à modulação. O EVM é a diferença entre o valor da magnitude em cada amostra do sinal capturado e o valor da magnitude ideal para aquela modulação. Colhidas as características necessárias, essas agora são armazenadas.

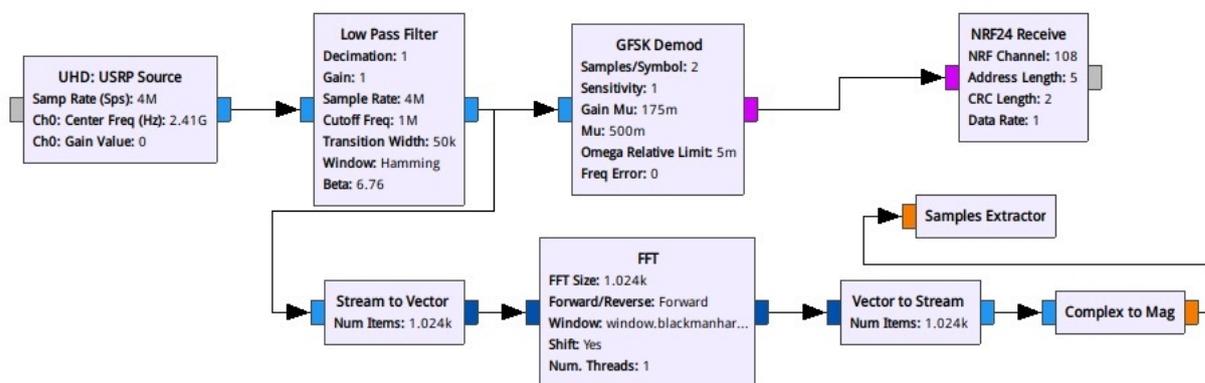


Figura 3. Blocos do GNURadio usados no processamento do sinal e extração das características.

Treinamento

Obtidas as características do transmissor é iniciada a classificação. Nessa fase é aplicado um classificador nas amostras. O problema se encaixa na categoria de aprendizado não supervisionado pois conhecemos as características dos dispositivos cadastrados mas não as dos intrusos. Dentre os algoritmos da categoria o que mais obteve desempenho em outros trabalhos [Brik et al. 2008], ou seja, menor taxa de erros, é o SVM (*Support Vector Machine*). O SVM é usado para detecção de anomalias, ou seja, dado um conjunto de amostras, ele detectará o limite flexível desse conjunto de modo a classificar novos pontos como pertencentes a esse conjunto ou não. Como resultado obtivemos um grau de similaridade entre o sinal que estamos analisando e o conjunto de sinais já cadastrados.

Avaliação

A avaliação consiste em comparar as amostras coletadas de um dispositivo com ele mesmo e com amostras de outros dispositivos, esses que seriam os possíveis intrusos. Quer se minimizar duas taxas de erro, a FAR (*False Accept Rate*), que é quando há a autorização de um dispositivo impostor, e a FRR (*False Reject Rate*), que é a probabilidade de não autorizar um dispositivo genuíno.

4. Experimentos

A captura do sinal é feita usando um SDR Ettus b200, que possui cobertura de sinal entre as frequências de 70 MHz a 6 GHz, operação full-duplex com largura de banda em

tempo real de 56 MHz, conectividade rápida e conveniente alimentada por barramento SuperSpeed USB 3.0. São usados 2 dispositivos da tecnologia nrf24L01+ transmitindo em intervalos pré-definidos em um mesmo canal de comunicação. Antes de realizar a captura do sinal é necessário definir a frequência e a taxa de amostragem em que o SDR vai trabalhar. Isso é feito na interface gráfica da ferramenta GNURadio. Para determinar que uma amostra é válida, primeiro ela é bufferizada, demodulada e decodificada. Assim que um quadro da tecnologia é identificado nessa amostra, os dados que estão no *buffer* são passados para o extrator de características. Foram coletadas 2000 amostras da transmissão de cada um dos dispositivos em diferentes posições no cenário. A mudança de local é realizada para ver a variação que há entre as magnitudes de um mesmo dispositivo, visto que essa característica é fortemente influenciada pela amplitude do sinal. Cada amostra é um conjunto de 1024 características que representam a magnitude do sinal.

4.1. Seleção de melhores características e análise dos dados

A seleção de características relevantes é uma fase importante para obtenção de melhores resultados na classificação. Para isso usa-se algoritmos que analisam os dados e apontam quais as características mais relevantes para a classificação. O algoritmo utilizado no conjunto de dados foi o *Principal Component Analysis* (PCA) [Jolliffe 2011].

4.2. Avaliação da solução

Nos experimentos, são realizados dois testes. Dos dois dispositivos, um é usado como referência e o outro faz o papel de um intruso que tenta se comunicar com o *gateway*. Um pequeno número de amostras é necessário para a fase de treinamento e de acordo com [Brik et al. 2008] a quantidade ideal é de 20 amostras. O conjunto de dados, com 2000 amostras de cada dispositivo, é dividido em 20 partes cada um contendo 100 amostras. São usadas 20 amostras para treinamento e as 80 restantes para teste. Cada parte é comparada com as 2000 amostras do dispositivo intruso.

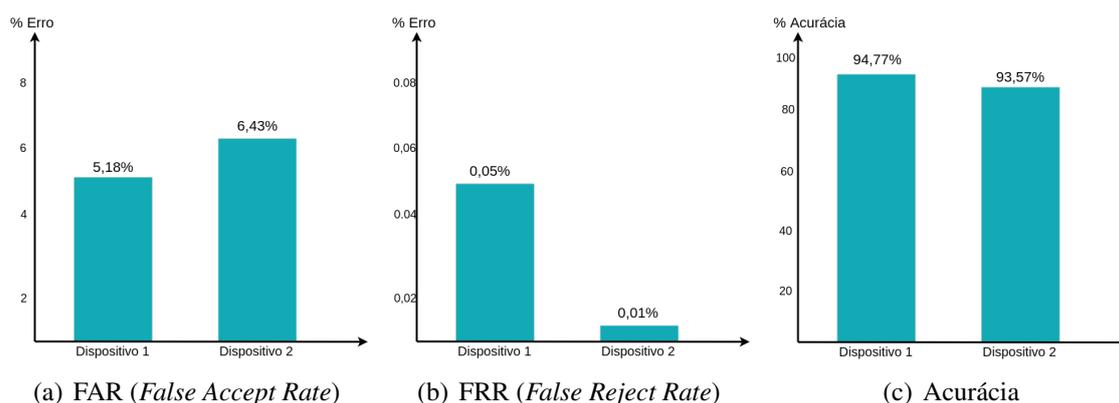


Figura 4. Erros e acurácia obtidas dos experimentos

Como visto na figura 4(c) foi obtida uma acurácia máxima de 94,77%. Usando o dispositivo 1 como referência é obtida uma taxa de erro FAR e FRR de 5,18% e 0,05%, respectivamente. Treinando o classificador com o dispositivo de referência 2 o erro FAR foi de 6,43% e o erro FRR foi de 0,01%.f As taxas de erro são inversamente proporcionais devido à característica do classificador. O SVM ajusta um limiar de variação das

características. Se esse limiar é alto, a taxa de falsos positivos (erro FRR) pode aumentar e conseqüentemente a taxa de verdadeiros negativos (erro FAR) diminuir. Tendo em vista essa característica, devemos então escolher um limiar para o classificador de acordo com o erro que se deseja minimizar.

5. Aplicação da técnica em um gateway IoT

O processo de autenticação em um gateway IoT é dividido em duas etapas. A primeira etapa é a de cadastro de uma novo dispositivo que precisa se comunicar com o gateway. Para isso, é necessário criar uma assinatura desse dispositivo que consiste basicamente em: 1) capturar o sinal eletromagnético emitido pelo dispositivo por um determinado período de tempo, 2) criar sua assinatura com base nas características do sinal e 3) associar essa assinatura com o identificador do dispositivo. A segunda etapa é dividida em cinco passos, os quais estão ilustrados na Figura 5 e que são detalhados a seguir. Esses passos são baseados na metodologia descrita na Seção 3.



Figura 5. Autenticação no gateway, dispositivo já cadastrado

O primeiro passo é a transmissão do dispositivo IoT. A vantagem da técnica apresentada no trabalho é que não é necessária nenhuma alteração nos dispositivos para o processo de autenticação. Enquanto o sinal está sendo processado pelo gateway para gerar a pilha TCP-IP um clone é enviado para o processo de identificação.

O segundo passo é o processamento do sinal. Como exposto anteriormente, o Rádio Definido por software nos dá a possibilidade de capturar e processar o sinal de

dispositivos em sua faixa de transmissão. A variedade de dispositivos depende do suporte do rádio às tecnologias.

Com o sinal já capturado pelo SDR, o terceiro passo é a extração das características que serão usadas na classificação. Como descrito anteriormente as características são obtidas a partir da aplicação da transformada de Fourier nos dados capturados. A partir disso é obtido o vetor de magnitude do sinal.

No quarto passo, as características extraídas da fase anterior são comparadas com a assinatura do dispositivo previamente cadastrado. O banco de dados de características pode ser atualizado conforme o tempo para se obter melhor acurácia na classificação. Diversos fatores, como a interferência, podem influenciar no resultado.

A última fase é a tomada de decisão pela aplicação. Ao se identificar um dispositivo intruso duas políticas podem ser aplicadas. Em uma política menos rígida um alerta pode ser gerado para o administrador do *gateway* cabendo a ele tomar a decisão. Aplicando uma política mais rígida, a comunicação do dispositivo com o *gateway* é interrompida e é gerado um alerta para administrador. Dispositivos intrusos estão na área de cobertura do *gateway*, sendo que a distância máxima depende da tecnologia de transmissão sem fio.

6. Conclusões e trabalhos futuros

Neste trabalho, foi apresentada uma técnica baseada no uso da magnitude do sinal como característica única para criação de assinaturas de dispositivos. A técnica apresentou resultados satisfatórios e se mostra viável na autenticação de dispositivos quando não há alteração de sua localização ao criar a sua assinatura. Vimos que a amplitude do sinal é um ponto importante quando utilizada somente a magnitude como característica para diferenciação dos dispositivos. Além disso, a técnica é simples e de baixo custo. Apresentamos também uma análise sobre a integração da técnica a um *gateway* de comunicação IoT.

Embora não tenha sido avaliado, a técnica apresentada é capaz de distinguir assinaturas de diferentes tecnologias de comunicações. Avaliar essa propriedade e sua acurácia são alguns trabalhos futuros planejados. Também pretendemos avaliar a técnica de maneira mais ampla, considerando diversos cenários que podem influenciar no processo de identificação, por exemplo, interferência com outros dispositivos, obstáculos que impactam na característica do sinal e incluir um número maior de características capazes de influenciar na diferenciação dos dispositivos como, largura de banda e frequência do pico de amplitude.

Referências

- Barbeau, M., Hall, J., and Kranakis, E. (2006). Detection of rogue devices in bluetooth networks using radio frequency fingerprinting. In *proceedings of the 3rd IASTED International Conference on Communications and Computer Networks, CCN*, pages 4–6. Citeseer.
- Bezawada, B., Bachani, M., Peterson, J., Shirazi, H., Ray, I., and Ray, I. (2018). IoTSense: Behavioral Fingerprinting of IoT Devices. *arXiv preprint arXiv:1804.03852*.

- Bihl, T. J., Bauer, K. W., and Temple, M. A. (2016). Feature Selection for RF Fingerprinting With Multiple Discriminant Analysis and Using ZigBee Device Emissions. *IEEE Transactions on Information Forensics and Security*, 11(8):1862–1874.
- Blossom, E. (2004). Gnu radio: tools for exploring the radio frequency spectrum. *Linux journal*, 2004(122):4.
- BNDES (2017). Produto 8: Relatório do plano de ação. Technical report, Banco Nacional do Desenvolvimento Social (BNDES).
- Brik, V., Banerjee, S., Gruteser, M., and Oh, S. (2008). Wireless device identification with radiometric signatures. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*, pages 116–127. ACM.
- Choe, H. C., Poole, C. E., Andrea, M. Y., and Szu, H. H. (1995). Novel identification of intercepted signals from unknown radio transmitters. In *Wavelet Applications II*, volume 2491, pages 504–518. International Society for Optics and Photonics.
- Danev, B. and Capkun, S. (2009). Transient-based identification of wireless sensor nodes. In *Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*, pages 25–36. IEEE Computer Society.
- Danev, B., Zanetti, D., and Capkun, S. (2012). On physical-layer identification of wireless devices. *ACM Computing Surveys (CSUR)*, 45(1):6.
- Gomez, C. and Paradells, J. (2010). Wireless home automation networks: A survey of architectures and technologies. volume 48, pages 92–101. IEEE.
- Hall, J., Barbeau, M., and Kranakis, E. (2005). Radio frequency fingerprinting for intrusion detection in wireless networks. *IEEE Transactions on Dependable and Secure Computing*, 12:1–35.
- J. Gubbi, R. Buyya, S. M. M. P. (2013). Internet of things (iot): A vision, architectural elements, and future directions in future generation computer systems. volume 29, pages 1645–1660.
- Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer.
- LABORA, T. (2017). SOFTware-defined gateWAY and fog computing for IoT.
- Meidan, Y., Bohadana, M., Shabtai, A., Guarnizo, J. D., Ochoa, M., Tippenhauer, N. O., and Elovici, Y. (2017). Profiliot: a machine learning approach for iot device identification based on network traffic analysis. In *Proceedings of the Symposium on Applied Computing*, pages 506–509. ACM.
- Nawir, M., Amir, A., Yaakob, N., and Lynn, O. B. (2016). Internet of things (iot): Taxonomy of security attacks. In *2016 3rd International Conference on Electronic Design (ICED)*, pages 321–326.
- of Homeland Security, U. D. (2016). Strategic principles for securing the internet of things (iot). pages 1–17.
- Toonstra, J. and Kinsner, W. (1995). Transient analysis and genetic algorithms for classification. In *WESCANEX 95. Communications, Power, and Computing. Conference Proceedings.*, IEEE, volume 2, pages 432–437. IEEE.

- Verma, G., Yu, P., and Sadler, B. M. (2015). Physical layer authentication via fingerprint embedding using software-defined radios. *IEEE Access*, 3:81–88.
- Xu, T., Wendt, J. B., and Potkonjak, M. (2014). Security of iot systems: Design challenges and opportunities.
- Zhao, K. and Ge, L. (2013). A survey on the internet of things security. In *Computational Intelligence and Security (CIS), 2013 9th International Conference on*, pages 663–667. IEEE.
- Zhu, Q., Wang, R., Chen, Q., Liu, Y., and Qin, W. (2010). Iot gateway: Bridging wireless sensor networks into internet of things. In *Embedded and Ubiquitous Computing (EUC), 2010 IEEE/IFIP 8th International Conference on*, pages 347–352. Ieee.

Otimização da trajetória de Drones como gateway de comunicação para sensores IoT

Lucas Soares Rodrigues¹, Kleber Vieira Cardoso¹, Vinicius da Cunha Martins Borges¹
Leizer de Lima Pinto¹, Antonio Carlos de Oliveira Júnior¹

¹Instituto de Informática, Universidade Federal de Goiás
Goiânia, Goiás, Brazil

{lucassoaresrodrigues, kleber, vinicius, leizer, antonio} @inf.ufg.br

Abstract. *This work proposes a flight path optimization model for drones that serve as a communication gateway or data mule for IoT (Internet of Things) sensors scattered in an Intelligent Campus. This drone has limited power autonomy (battery), and in many cases can not fly over the course required to serve all allocated sensors. Therefore, the paper proposes, using Linear Programming, a model that ensures the maximum of sensors are satisfied considering the autonomy found in the drone.*

Resumo. *Este trabalho propõe um modelo de otimização de trajetória de vôo para drones que servem como gateway de comunicação ou mula de dados para sensores IoT (Internet of Things) distribuídos em um Campus Inteligente. Este drone possui autonomia energética (bateria) limitada, e em muitos casos não pode sobrevoar o percurso necessário para atender todos os sensores existentes no conjunto alocado. Por isso o artigo propõe, utilizando Programação Linear, um modelo que assegure que o máximo de sensores sejam atendidos considerando a autonomia encontrada no drone.*

Palavras-chave— VANTs, IoT, Planejamento de Vôo, Otimização de Trajetória

1. Introdução

A importância de IoT (*Internet of Things*), a chamada Internet das Coisas, é praticamente inquestionável, em parte devido à enorme quantidade de aplicações disponíveis e em desenvolvimento. Apesar da implantação e uso de um grande número de dispositivos de IoT já ser uma realidade, o tema ainda é considerado extremamente importante tanto para a indústria quanto para a academia [Stankovic 2014].

No Brasil, por exemplo, o governo federal reconheceu a importância do tema através da iniciativa chamada Plano Nacional de Internet das Coisas [MCTIC 2017]. Dentre as aplicações para IoT, aparecem com destaque os ambientes inteligentes, e.g., Cidades e Campus Inteligentes [Zanella et al. 2014].

O campus pode ser visto como uma versão reduzida da cidade, portanto, o menor tamanho torna a gestão mais simples, e além disso, a possibilidade de acesso a toda a infraestrutura física do campus facilita a implantação e testes de soluções de maneira mais rápida. Por exemplo, o campus universitário é um laboratório adequado para testar soluções para ambientes inteligentes baseadas em IoT.

Inúmeros trabalhos aliando ambientes inteligentes e IoT vêm sendo desenvolvidos. A autonomia energética é um dos principais desafios encontrados acerca dos dispositivos e sensores IoT. Por causa de sua baixa autonomia, estes dispositivos possuem pouca mobilidade de conexão, ou seja, possuem limitação quanto à transmissão de longo alcance a partir de sua localização. [Mozaffari et al. 2017]. Os VANTs (Veículos Aéreos Não-Tripulados), e neste caso os *drones*, podem ser empregados para prover maior mobilidade a uma infraestrutura IoT.

Os VANTs podem ser usados como *relays* sem fio para aperfeiçoamento da conectividade e cobertura de dispositivos sem fio terrestres [Mozaffari et al. 2017]. Comparados com estações bases terrestres a vantagem do uso de estações bases usando VANTs é sua habilidade em prover comunicação *on-the-fly*, ou seja, no decorrer do voo. Além disso, a alta atitude dos VANTs faz com que eles estabeleçam um link de comunicação LoS (*line-of-sight*), mitigando bloqueios e sombreamentos.

Os VANTs são previstos como um importante componente de um avançado ecossistema ciber-físico da Internet das Coisas [Vermesan and Friess 2013]. Isso é devido ao fato que os VANTs possuem características únicas como serem dinâmicos, fáceis de implantar, fáceis de reprogramar durante o tempo de execução, capazes de medir qualquer coisa em qualquer lugar, e capazes de voarem em um controle de espaço aéreo com um alto grau de autonomia [Snow 2014].

Portanto, um VANT pode prover uma conectividade *uplink* para dispositivos IoT fixos numa localização em um campus inteligente, sendo que este VANT é empregado como um gateway de comunicação que faz associação com estes dispositivos durante um voo programável e autônomo. Para a realização deste voo e o estabelecimento dessas comunicações podem ser avaliados vários critérios e restrições tais como a autonomia energética (capacidade de bateria) dos VANTs e a trajetória ótima a ser percorrida para obter melhor performance considerando todos seus recursos.

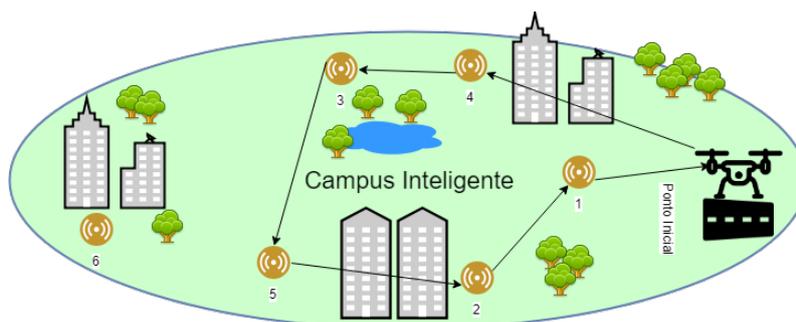


Figura 1. Exemplo de Campus Inteligente com sensores alocados

Um Campus Inteligente pode possuir um parque de sensores IoT implantados com a função de coletar os mais variados tipos de dados possíveis relativos ao funcionamento deste campus. Considerando um campus universitário, por exemplo, podemos implantar sensores para auxiliar a coleta de lixo, sensores para mapear vagas de estacionamento disponíveis no campus, sensores para monitorar funcionamento de recursos tais como ar condicionado em salas de aula, etc. A difícil conectividade móvel em lugares remotos onde foram ocasionados acidentes e catástrofes requerendo resgate imediato de pessoas também é um cenário onde se vê a utilidade do emprego de *drones* funcionando como

gateway de comunicações ou mula de dados. Essa comunicação é necessária para auxiliar tais ações, e a implantação eficiente de VANTs apoiando uma rede de sensores IoT pode ser crucial nessas situações extremas. Um exemplo deste cenário é demonstrado na Figura 1, mas no caso da figura a trajetória não foi necessariamente a mais eficiente possível.

Este trabalho visa avaliar um importante atributo quanto ao papel realizado por VANTs neste cenário: a trajetória. O caminho percorrido pelos *drones* que executam essa função de *gateway* ou mula de dados, é algo muito importante de ser avaliado, pois está relacionado diretamente com a autonomia e o gasto energético do *drone* durante o desempenho deste papel. Visualiza-se um conjunto de sensores IoT espalhados por uma determinada área geográfica e sabe-se que em muitos casos não é possível servir todos os nós devido a uma limitada carga de bateria encontrada nos *drones*.

Neste artigo, aborda-se o planejamento de voo autônomo de VANTs considerando a trajetória ótima para que seja atendido o máximo de sensores IoT possíveis dada uma baixíssima e limitada autonomia energética. Fazendo com que ao invés de realizar uma atualização da trajetória de voo enquanto estiver pairando no ar, esta atualização é feita antes do início de cada voo, e este *drone* somente realiza o voo focado em realizar a sincronização dos dados com cada um dos sensores.

Neste contexto, através de Programação Linear, é proposto um modelo matemático que esquematize o objetivo a ser atingido e as restrições que deva-se atender para que o *drone* sirva o máximo de sensores possíveis dentro de um Campus Inteligente, informando quais são os nós a serem sobrevoados e a sequência de nós a serem visitados.

Para além dessa introdução, as demais seções a seguir deste artigo são divididas em: A seção 2 apresenta os trabalhos relacionados enquanto que a seção 3 a proposta de um Modelo de Otimização de Trajetória de Voo. A seção 4 descreve a avaliação e a seção 5 as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Para possibilitar um sistema de transportes inteligente totalmente automatizado num contexto de Cidade Inteligente, [Menouar et al. 2017] usa VANTs como sensores conectados através de DSRC (*Dedicated Short Range Communications*) aos carros provendo comunicação V2V (*vehicle-to-vehicle*) e V2X (*vehicle-to-infrastructure*), capturando dados diretamente dos veículos e transmitindo essas informações até seu destinatário interessado na estação base via comunicação *multihop* (*relays* de rádios) numa abordagem D2D (*device-to-device*). O trabalho conclui que a operação verdadeiramente autônoma dos VANTs é um verdadeiro desafio, porque demanda a observação de seres humanos e outros obstáculos quaisquer.

[Seiber et al. 2018] usa *drones* como sensores IoT com a finalidade de detectar áreas perigosas contaminadas com agentes químicos sem que seja necessária uso de homens para fazer a detecção. A tecnologia de comunicação entre os *drones* foi Wi-Fi, e conexão entre *drone* e a estação base foi protocolo BLE (*Bluetooth Low Energy*). As limitações encontradas nos testes realizados foram em relação ao voo programado dos *drones* que limitou-se por estabelecer alinhamento horizontal no enxame de VANTs, o que não aconteceria numa situação real.

[Kirichek 2016] desenvolve um modelo de entrega de dados partindo de sensores

terrestres sendo transmitidos aos *drones* até a chegada em um servidor na nuvem. O modelo proposto possibilita o cálculo do número de VANTs necessários para coleta e entrega dos dados, levando em consideração a intensidade e volume do tráfego na rede, o número de unidades e o raio de sua conexão.

Em [Jiang and Swindlehurst 2012] os autores investigaram a trajetória ótima de VANTs equipados com múltiplas antenas para maximizar a somatória da taxa de transmissão em comunicações *uplink*. O trabalho [Zeng et al. 2016] maximiza a taxa de transferência de um sistema de relay baseado em VANT através da otimização da trajetória do VANT juntamente com a taxa de transmissão entre a fonte e o *relay*. Em [Mozaffari et al. 2016b] é investigado a implantação ótima e o movimento de um único VANT para apoiar comunicações sem fio *downlink*.

Para mitigar a interferência, em [Mozaffari et al. 2016a], a altitude dos VANTs são apropriadamente ajustadas baseadas no *beamwidth* da antena direcional bem como nos requisitos de cobertura. Além disso, o número mínimo de VANTs necessários para garantir a probabilidade de cobertura do alvo para uma área geográfica é determinado.

[Mozaffari et al. 2017] pretendeu encontrar uma implantação ótima para que os *drones* sirvam como estação base conseguindo estabelecer uma associação com os dispositivos ativos naquele instante. A quantidade de dispositivos ativos se altera no decorrer do tempo de acordo com que novos dispositivos são ativados e outros desativados, este padrão de ativação é conhecido pelo servidor na nuvem.

Para fazer esta atualização da trajetória de mobilidade dos VANTs é realizada uma parada aérea estacionária para uma pequena economia de energia antes que se continue o movimento do VANT sem saber qual será a nova trajetória. Este cenário provavelmente prevê o uso de um VANT de média autonomia energética pois considera uma manutenção do voo do *drone* por um período relativamente extenso, fazendo com que ele pare no ar por este tempo para servir como uma estação base para os dispositivos. Para o cenário determinado neste trabalho, é considerada situações com baixíssima autonomia.

3. Modelo de Otimização de Trajetória de Vôo

O modelo proposto visa reduzir o consumo de energia dos *drones* através do uso otimizado da autonomia restante de bateria de um *drone*, fazendo com que o máximo de sensores possíveis sejam visitados dentro dos nós alocados. Essa solução pode ser modelada usando PL (Programação Linear) [Ouahouah et al. 2017]. A escolha da técnica Programação Linear se dá pois as funções objetivo e as restrições são funções lineares. Mais especificamente foi aplicada a Programação Inteira Mista, pois algumas variáveis do problema assumem valor inteiro e outras assumem valor real.

Após a definição do modelo matemático usado para tratar as variáveis existentes dentro do contexto do problema é utilizado a ferramenta de otimização CPLEX [CPLEX 2009] para realização da implementação usando a linguagem de otimização OPL (Optimization Programming Language). Após coleta de dados extraídos das simulações e soluções encontradas no CPLEX, é utilizado o MATLAB [Guide 1998] para realizar a plotagem dos dados resultantes em gráficos condizentes.

Para otimização da trajetória de *drones* é aplicado a técnica de Programação Linear. Para cenários onde *drones* possuem baixa autonomia restante para vôo, é necessário

utilizar esta autonomia para sobrevôo e coleta de dados do máximo possível de sensores. Em alguns casos pode-se sobrevoar todos os sensores existentes em um campus inteligente, se a autonomia dos *drones* for utilizada otimizada.

Considere um conjunto de sensores $S = \{1, 2, \dots, N\}$, estes sensores estão espalhados em um campus de forma aleatória, estabelecendo entre cada um desses sensores um arco de distância. O conjunto de arestas de distâncias entre os pontos de localização dos sensores é dado por $D = \{(1, 1), (1, 2), \dots, (N - 1, N)\}$, onde os pares distância entre o mesmo sensor, por exemplo (1,1), é zero. A distância entre um par de sensores, por exemplo (1,2), é igual ao seu inverso (2,1). Este conjunto de arestas dá origem uma matriz de distâncias entre todos os sensores do conjunto S , que por sua vez gera um grafo completo bidirecional (Figura 2). Para fins de modelagem, o conjunto S possui o número de elementos igual a: Número de Sensores somado o Ponto Inicial. O grafo demonstra todas as possibilidades de percurso a ser sobrevoado por um *drone*, com finalidade de coletar dados em todos esses sensores. O *drone* que executa o vôo nessa trajetória possui autonomia limitada dada pela variável A . Os grafos não ilustram o tamanho das arestas real de acordo com os valores atribuídos a cada uma delas. As figuras apenas irão demonstrar a sequência de vértices visitados para cada conjunto de sensores e autonomia especificada.

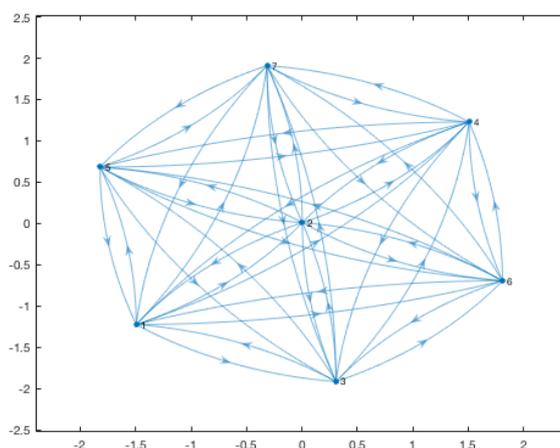


Figura 2. Grafo Completo Bidirecional

A partir da energia dada, em watt-hora, a autonomia é calculada pelo tempo restante $T = E \times P$, (P = Potência, em watt). Considerando uma velocidade constante V , em m/s , do *drone* se tem a distância máxima possível de ser sobrevoada, considerando a autonomia inicial, dada por $A = T \div V$. É importante frisar que para fins de simplificação não é considerada um outro vetor de localização que determinaria a altitude do drone num dado instante do percurso, bem como não são considerados outros fatores de influência na autonomia do *drone* como peso total do *drone*, empuxo líquido etc. Porém, para compreensão e solução de parte importante do problema as informações consideradas são suficientes para demonstração da otimização realizada e das simulações feitas.

Para cada uma das arestas que é computada como trajetória contida no percurso do vôo é necessário determinar uma variável booleana que especifica se a aresta é percorrida ou não para um par (i, j) sendo o $i, j \in S$ e $(i, j) \in D$. Esta variável de decisão é dada

em (1):

$$X_{ij} \in \{0, 1\}, \forall (i, j) \in D \quad (1)$$

Num cenário onde se tenha num conjunto de sensores espalhados por determinada área, estes sensores servem ao propósito de produção de dados que por sua vez serão utilizados para prover algum serviço ou aplicação. Portanto, para este contexto o maior interesse é que seja visitado o maior número de sensores possíveis. Dado um conjunto A de arestas percorridas, o número de sensores visitados é igual a $A - 1$, pois a última aresta (i, j) é relativa a volta do *drone* para o Ponto Inicial $(1, 1)$. O objetivo da otimização modelada é maximizar o número de sensores visitados dado pelo somatório de arestas percorridas. Este objetivo é demonstrado em (2):

$$MAX w = \left(\sum_{(i,j) \in D} X_{ij} \right) - 1 \quad (2)$$

Para atender o objetivo especificado têm-se várias restrições de comportamento na trajetória para se atender. O *drone* deve obrigatoriamente sair do Ponto Inicial e se dirigir para algum dos sensores somente uma vez. Para isso têm-se que o somatório das variáveis de decisão das arestas do Ponto Inicial para todos os N sensores deve ser igual a um, tal como demonstra em (3). Isso deve valer similarmente para o retorno ao Ponto Inicial, o somatório das variáveis de decisão também é um (4).

$$\sum_{(1,j) \in D} X_{1,j} = 1 \quad (3)$$

$$\sum_{(i,1) \in D} X_{i,1} = 1 \quad (4)$$

A restrição (5) garante que o somatório das variáveis de decisão de chegada em um sensor deve ser igual ao somatório das saídas no mesmo sensor. Essa restrição é importante por que em cada sensor que o *drone* entre sobrevoando por cima, percorrendo a trajetória de chegada neste sensor, o *drone* deve para esse sensor percorrer a trajetória de saída.

$$\sum_{(i,h) \in D} X_{i,h} = \sum_{(h,j) \in D} X_{h,j}, \forall h \in S \quad (5)$$

Dado um subconjunto de sensores dentro do conjunto total, neste subconjunto o total de arestas percorridas deve ser igual o total de vértices (sensores), menos um, para que dessa forma não se feche uma sub trajetória dentro daquele conjunto de sensores (6). Esta restrição atua para evitar qualquer desperdício na autonomia restante do *drone* durante o voo, uma vez que um sensor já tiver sido visitado não é necessário que se retorne a esse sensor.

$$\sum_{i \in R} \sum_{j \in R} X_{i,j} \leq |R| - 1, R \subset S, 2 \leq |R| \leq n \quad (6)$$

Para especificar que o *drone* não pode percorrer uma distância maior do que a autonomia que ele possui, deve-se multiplicar cada variável de decisão, zero ou um que seja, à distância daquela aresta, e o somatório de todas distâncias percorridas deve ser menor ou igual à autonomia total do *drone* (7).

$$\sum_{(i,j) \in D} D_{i,j} X_{i,j} \leq A \quad (7)$$

4. Avaliação

Foram simuladas dois conjuntos de sensores alocados, o primeiro com 6 sensores e o segundo com 14, cada conjunto com três instâncias de autonomies diferentes. Os sensores são alocados em uma área num raio de no máximo 100 metros de distância entre cada um dos sensores. As distâncias entre os sensores são estáticas, isto é, não mudam a cada voo. O nó inicial, que é o ponto de saída do *drone* deve ser levado em consideração para compor o total de nós do conjunto de sensores. O Ponto Inicial é um local onde será guardado o *drone* e eventualmente pode-se efetuar recargas, preferencialmente no período onde não há vôos programados, por isso neste cenário é considerado vários vôos executados sequencialmente por exemplo num campus universitário ao longo de um período que não houveram recargas no VANT.

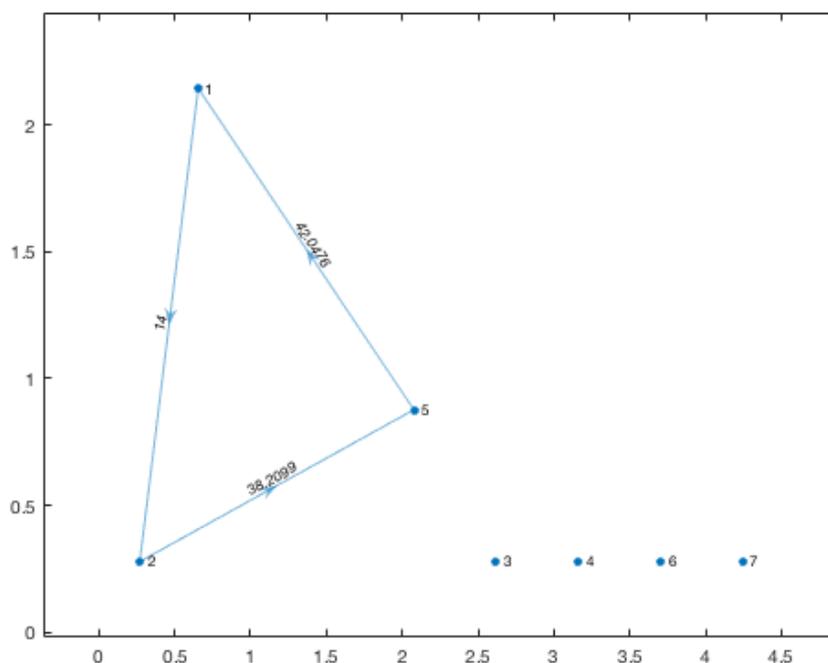


Figura 3. Trajetória gerada para o *drone* com autonomia de 150 metros

Na primeira instância (Figura 3) o *drone* tem uma autonomia total de 150 metros para ser percorrido, e considerando os 6 sensores alocados na área, o *drone* só consegue visitar 2 sensores, fazendo o percurso de três arestas para visitar estes sensores, sendo

eles: (1,5), (5,2) e (2,1). A distância total percorrida pelo *drone* é de 94,24 metros portanto restante 55,76 metros de autonomia que ainda pode ser percorridos pelo *drone*, porém não podem ser aproveitados pelo *drone* por que não seria suficiente para visitar mais sensores considerando não somente a distância para chegada no sensor como também a distância de volta até o ponto inicial. As distâncias entre os sensores utilizada na primeira instância é a mesma entre todas as demais instâncias simuladas dentro deste conjunto de sensores, mas as distâncias são diferentes no outro conjunto de sensores. Nos grafos ilustrados nas figuras é mostrado a sequência de pontos visitados porém não é marcada a localização geográfica dos pontos.

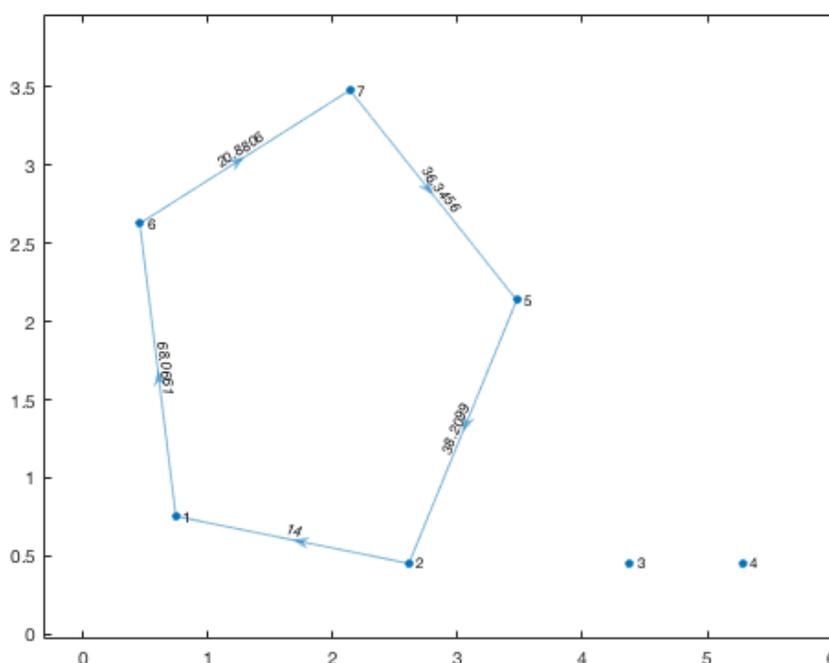


Figura 4. Trajetória gerada para o *drone* com autonomia de 200 metros

Quanto à segunda instância (Figura 4) analisada foi atribuído uma autonomia total do *drone* de 200 metros. Fazendo um aproveitamento desses 200 metros para o total de sensores alocados, o maior número possível de sensores visitados foi 4. O total da distância sobrevoada é de 125,28 metros com restante de autonomia de 74,72 metros. Veja que apesar da autonomia restante parecer alta a otimização realizada concentra-se na quantidade de sensores visitados, por isso é certo que independente da quantidade de autonomia restante o número de sensores visitados é o máximo que se conseguiria sobrevoar com a autonomia considerada. É possível que haja várias instâncias de distâncias totais percorridas para a mesma quantidade de sensores alcançadas com a autonomia limitada.

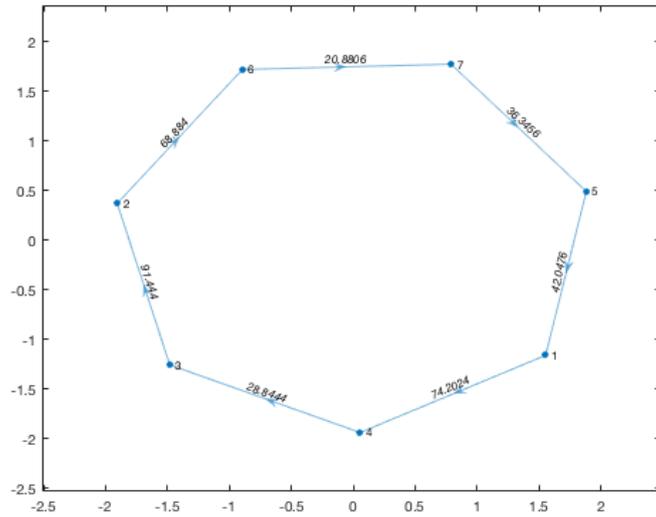


Figura 5. Trajetória gerada para o drone com autonomia de 500 metros

Uma terceira e última instância de autonomia é analisada, atribuindo 500 metros de autonomia para o *drone* (Figura 5). Para essa instância o *drone* consegue sobrevoar todos os 6 nós de sensores dentro da alocação estabelecida. Para efetuar toda essa trajetória ele gasta 362,62 metros de autonomia com restante de 137,38 metros. Nota-se que com apenas um pouco mais de carga no *drone*, já seria possível efetuar a trajetória da primeira instância. É importante enfatizar que não se pode afirmar que 362,62 metros é o mínimo de autonomia que o *drone* deve ter para sobrevoar todos os sensores. Por que as restrições não garantem que ele sobrevoe o mínimo de distância para sobrevoar o máximo de sensores. Essa distância mínima não é garantida pelo modelo de otimização proposto.

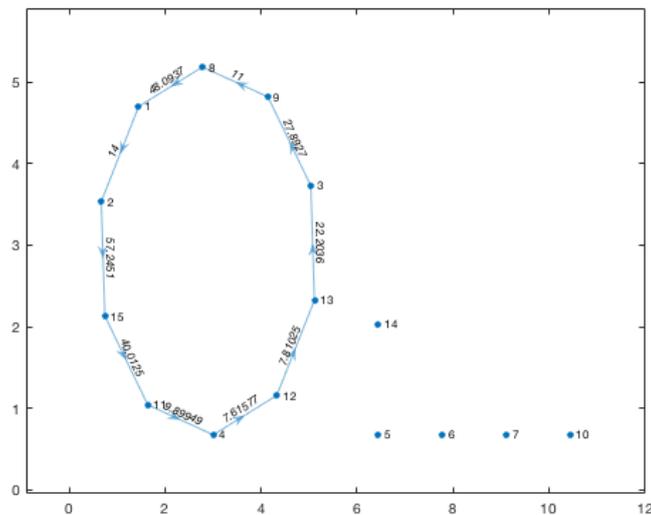


Figura 6. Trajetória possível para 15 sensores alocados e autonomia de 250 metros para o drone

Também é simulado um campus onde haja 14 sensores alocados, ou seja, mais que o dobro de sensores nos testes realizados no cenário anterior. Para este conjunto de sensores é executado três instâncias de testes. Na primeira instância, (Figura 6) para uma autonomia de 250 metros encontrada no VANT é percorrida a distância de 245,63 metros, e consegue-se sobrevoar 9 sensores, sem contar o Ponto Inicial de onde parte o *drone*. Na segunda instância (Figura 7) é alocada uma autonomia de 350 metros para o veículo, que assim percorre uma distância de 348,73 metros para se atingir 13 sensores. Quanto a terceira e última instância (Figura 8) com autonomia de 450 metros, são sobrevoados todos os 14 sensores e é percorrido 448.38 metros para chegar a este fim.

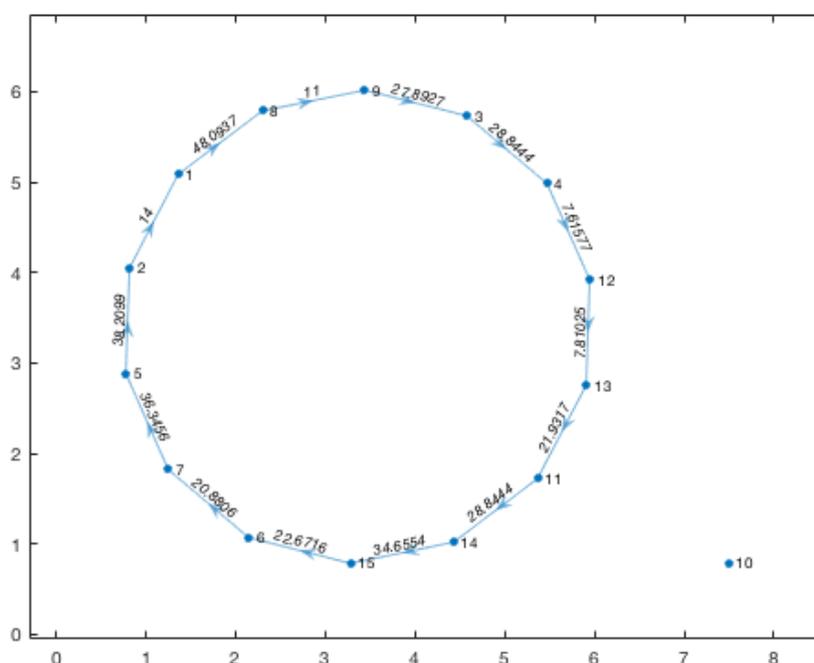


Figura 7. Trajetória possível para 15 sensores alocados e autonomia de 350 metros para o *drone*

Algo que pode ser destacado quanto a essas simulações se refere ao grande aproveitamento das autonomias encontradas no VANT, fazendo com que a autonomia restante seja muito pouca. Considerando que o limite de distância entre os sensores é o mesmo usado no conjunto anterior alocado (100 metros) nota-se que uma densidade maior de sensores num dado espaço é encontrado, e por isso há um maior aproveitamento das rotas, pois é possível fazer dentro dos mesmos deslocamentos referentes ao conjunto de sensores anteriores porém visitando um maior número de sensores.

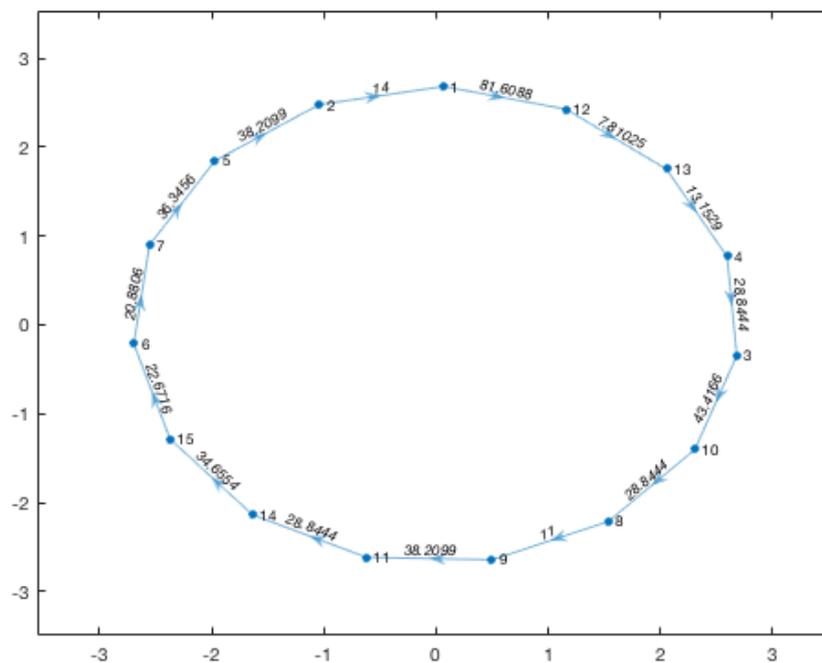


Figura 8. Trajetória possível para 15 sensores alocados e autonomia de 450 metros para o *drone*

5. Conclusão

Propomos neste trabalho uma modelagem de otimização combinatória muito semelhante ao problema de roteirização de veículo, porém, várias especificidades são adequadas neste cenário com um *drone* sobrevoando sensores IoT alocados numa determinada área. O modelo proposto visou atingir, com um *drone* de autonomia limitada, o máximo de visitas nos nós de sensores alocados. Este modelo foi elaborado considerando algumas simplificações quanto ao padrão de consumo dos *drones*, ignorando peso total do *drone* bem como o empuxo, fatores que deveriam ser considerados e podem ser agregados nos modelos de trabalhos futuros.

Outra simplificação empregada foi em relação a altitude do *drone* durante o sobrevôo. Esta localização do *drone* em três dimensões propiciaria um acompanhamento do VANT de forma que poderia ser calculado precisamente a distância do *drone* para o sensor e com isso determinar quais os alcances possíveis dentro das tecnologias de comunicação utilizadas entre o *drone* e os sensores. O trabalho se preocupou em delimitar algumas variáveis dentro deste contexto de uso dos *drones* como gateway de comunicação, ou mula de dados, e com essas variáveis obteve sucesso no objetivo do modelo e conquistou um terreno importante para tratamento deste tipo de problema de trajetória de *drones*.

Para trabalhos futuros, pretendemos agregar essas outras variáveis e simular utilizando as tecnologias de redes sem fio possíveis dentro da comunicação *drone* e sensor, também espera-se alocar os sensores usando coordenadas geográficas podendo calcular as distâncias num dado espaço geográfico real. Dessa forma pode-se comparar as simulações com um teste envolvendo *drones* reais.

Referências

- CPLEX, I. I. (2009). V12. 1: User's manual for cplex. *International Business Machines Corporation*, 46(53):157.
- Guide, M. U. (1998). The mathworks. *Inc., Natick, MA*, 5:333.
- Jiang, F. and Swindlehurst, A. L. (2012). Optimization of uav heading for the ground-to-air uplink. *IEEE Journal on Selected Areas in Communications*, 30(5):993–1005.
- Kirichek, R. (2016). The model of data delivery from the wireless body area network to the cloud server with the use of unmanned aerial vehicles. In *ECMS*, pages 603–606.
- MCTIC (2017). Plano nacional de iot. <http://www.mctic.gov.br/mctic/opencms/salaImprensa/noticias/?titulo=IoT>. Data acesso: Julho 2018.
- Menouar, H., Guvenc, I., Akkaya, K., Uluogac, A. S., Kadri, A., and Tuncer, A. (2017). Uav-enabled intelligent transportation systems for the smart city: Applications and challenges. *IEEE Communications Magazine*, 55(3):22–28.
- Mozaffari, M., Saad, W., Bennis, M., and Debbah, M. (2016a). Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage. *IEEE Communications Letters*, 20(8):1647–1650.
- Mozaffari, M., Saad, W., Bennis, M., and Debbah, M. (2016b). Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs. *IEEE Transactions on Wireless Communications*, 15(6):3949–3963.
- Mozaffari, M., Saad, W., Bennis, M., and Debbah, M. (2017). Mobile unmanned aerial vehicles (uavs) for energy-efficient internet of things communications. *IEEE Transactions on Wireless Communications*, 16(11):7574–7589.
- Ouahouah, S., Taleb, T., Song, J., and Benzaid, C. (2017). Efficient offloading mechanism for uavs-based value added services. In *Communications (ICC), 2017 IEEE International Conference on*, pages 1–6. IEEE.
- Seiber, C., Nowlin, D., Landowski, B., and Tolentino, M. E. (2018). Tracking hazardous aerial plumes using iot-enabled drone swarms. In *Internet of Things (WF-IoT), 2018 IEEE 4th World Forum on*, pages 377–382. IEEE.
- Snow (2014). Why drones are the future of the internet of things? <http://droneanalyst.com/2014/12/01/drones-are-the-future-of-iot/>. Data acesso: Julho 2018.
- Stankovic, J. A. (2014). Research directions for the internet of things. *IEEE Internet of Things Journal*, 1(1):3–9.
- Vermesan, O. and Friess, P. (2013). *Internet of things: converging technologies for smart environments and integrated ecosystems*. River Publishers.
- Zanella, A., Bui, N., Castellani, A., Vangelista, L., and Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things journal*, 1(1):22–32.
- Zeng, Y., Zhang, R., and Lim, T. J. (2016). Throughput maximization for uav-enabled mobile relaying systems. *IEEE Transactions on Communications*, 64(12):4983–4996.

Análise Econômica-Financeira da Instalação de um Sistema de Energia Solar Fotovoltaica na Modalidade Geração Compartilhada em Goiás

Vitor Rezende Faria¹, Guilherme Morais Spíndola²

¹Instituto Federal de Educação, Ciência e Tecnologia de Goiás (IFG)
Rua 75, nº 46, Centro, Goiânia – GO – Brasil

²Escola de Engenharia Civil e Ambiental
Universidade Federal de Goiás (UFG)
Av. Universitária, nº. 1488 – Goiânia – GO – Brasil

vitorocoro@gmail.com, eng.guilhermemorais@gmail.com

Abstract. *This work aims to evaluate the implementation of the photovoltaic solar energy system by shared generation in ten residences in the state of Goiás, from a economic-financial point of view. For the economic-financial analysis, the optimistic and pessimistic scenarios were economic indicators of the net present value, profitability index, internal rate of return and payback (simple and discounted), in addition to comparing with a case study reference. The results showed that the system is feasible, can be accepted in the two scenarios evaluated, it is superior to the reference case study and can attract participants with potential investors.*

Resumo. *Esse trabalho tem como objetivo avaliar do ponto de vista econômico-financeiro a implantação do sistema de energia solar fotovoltaica por geração compartilhada em dez residências do Estado de Goiás. Para a análise econômico-financeira foram avaliados nos cenários otimista e pessimista perante os indicadores econômicos do valor presente líquido, índice de lucratividade, taxa interna de retorno e payback (simples e descontado), além de comparar com um estudo de caso referência. Os resultados mostram que o sistema é viável, pode ser aceito nos dois cenários avaliados, se mostra superior ao estudo de caso referência em relação aos indicadores econômico-financeiros e pode atrair participantes com potencial de investidores.*

1. Introdução

Atualmente um dos grandes desafios de nossa sociedade é abastecer a população com a energia suficiente para o seu pleno desenvolvimento e fazer com que este abastecimento seja feito de forma a não degradar o meio ambiente, ou seja, de maneira sustentável.

O Brasil tem, em sua grande maioria, uma matriz energética constituída por energias renováveis. Em 2016 a oferta total de energia elétrica foi de 619,7 TWh, sendo 81,7% de fontes renováveis, com a seguinte composição: hidráulica (68,1%); gás natural (9,1%); biomassa (8,2%); eólica (5,4%); derivados de petróleo (3,7%); carvão e derivados (2,9%); nuclear (2,6%) [EPE 2017]. Mediante os dados apresentados, observa-se que a parcela da energia solar fotovoltaica é ínfima quando se compara com o restante.

Sobre a geração de energia limpa e sustentável, a Resolução Normativa (RN) 482/2012 de abril de 2012 [ANEEL 2012], da Agência Nacional de energia Elétrica (ANEEL), torna livre o acesso às redes de distribuição de energia elétrica para pequenas unidades geradoras, desde que sejam atendidos os requisitos básicos estabelecidos por ela. A última alteração da RN 482/2012 feita pela RN 687/2015 [ANEEL 2015], trouxe outras opções para a compensação de energia na intenção de fomentar a utilização de geração distribuída. Uma inovação importante foi a permissão da instalação de micro ou minigeração em locais com múltiplas unidades consumidoras (UCs), condomínios, por exemplo. Uma outra opção criada foi a geração compartilhada, que é a reunião de consumidores, que sejam atendidos pela mesma concessionária, por meio de consórcios ou associações, que possuam micro ou minigeração distribuída instalada em local diferente das UCs.

Apesar de haver embasamento legal, a opção de geração compartilhada ainda gera receio nos consumidores, pois os custos da instalação de uma geração fotovoltaica ainda são elevados. Neste contexto das inovações trazidas pela norma regulamentadora RN 687/2015 [ANEEL 2015], este trabalho pretende avaliar a viabilidade técnica e econômico-financeira da energia solar fotovoltaica na opção de compensação de energia elétrica por meio da geração compartilhada composto por uma mini usina localizada no estado de Goiás e com isso propor um modelo de negócios de um sistema fotovoltaico conectado à rede nesta modalidade.

Este trabalho está dividido em sete seções. A introdução que contextualiza o assunto no Brasil. Na seção 2 definimos termos técnicos e normativos, na seção 3 aprofundamos na geração compartilhada no Brasil, na seção 4 tratamos da questão econômica, depois, na seção 5 informamos o método de pesquisa utilizado para discutirmos os resultados na seção 6. Por último, fazemos algumas considerações finais e apresentamos as referências.

2. Sistemas Fotovoltaicos

A energia solar fotovoltaica (FV) é composta por um sistema que realiza a conversão da luz solar em corrente elétrica, através de placas construídas com fotocélulas produzidas a partir de um material semicondutor, como silício cristalino, arsenieto de gálio, dentre outros [Cabral and Vieira 2012]. Essa corrente elétrica gerada é contínua, para isso utiliza-se de um inversor que a transforma em corrente elétrica alternada, no qual é direcionada ao quadro geral e, posteriormente, distribuída para consumo.

De modo geral, a geração distribuída se caracteriza como uma forma de geração conectada ao sistema de distribuição e próxima ao ponto de consumo. Este trabalho adota os conceitos apresentados na RN nº 482 da ANEEL [ANEEL 2012], na qual estão definidos que micro e minigeração distribuída, são centrais geradoras de energia, que utilizam fontes com base em energia hidráulica, solar, eólica, biomassa ou cogeração qualificada, conectadas na rede de distribuição, por meio de instalações de unidades consumidoras, e especificamente o conceito da nova modalidade de Geração Distribuída, chamada Geração Compartilhada trazida pela RN 687 também da ANEEL [ANEEL 2015].

Na geração compartilhada, diversos consumidores, podendo ser pessoas físicas ou pessoas jurídicas, que estejam dentro da mesma área de concessão, ou seja, área de abrangência da concessionária de energia elétrica no estado, podem compartilhar energia

originada de uma única mini ou micro usina de energia solar fotovoltaica. Para isso, a localização da fonte geradora deve ser diferente de quaisquer endereços dessas unidades consumidoras e essas pessoas precisam se unir através de um consórcio ou cooperativa. A energia produzida será, então, compensada nas unidades consumidoras dos participantes através dos créditos gerados em quilowatt-hora (kWh), o excedente poderá ser consumido em até 60 meses [ANEEL 2015].

Para a criação de consórcios deve-se observar os dispostos na Lei nº6.404/76 [Nacional 1976] e na Instrução Normativa da Receita Federal nº 1.634/16 [do Brasil 2016]. Caso esse consórcio seja administrado por outro titular que não esteja entre os participantes, então deve-se seguir a Lei nº 11.795/08 [Nacional 2008]. Para a criação de cooperativas deve-se seguir os artigos 1.093 a 1.096 do Código Civil [Nacional 2002] e a Lei 5.764/71 [Nacional 1971].

3. Geração Compartilhada no Brasil

Nesta Seção são apresentados alguns dados relacionados à geração compartilhada no Brasil sobre as modalidades de geração distribuída. São apresentados na Figura 1 os números de conexões, a quantidade de UCs que recebem os créditos de energia e a capacidade instalada de cada modalidade no país.

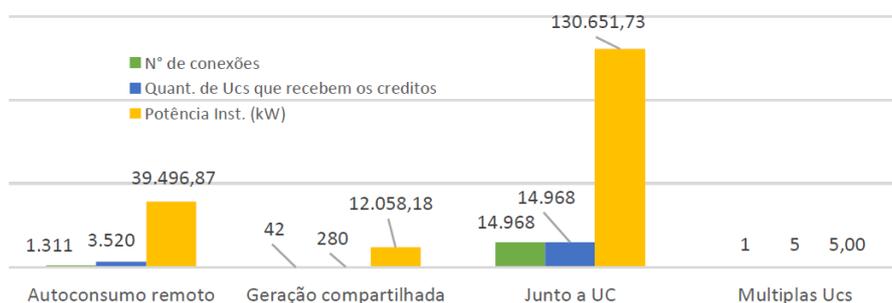


Figura 1. Dados relacionados com as modalidades de GD até 27/10/2017. Fonte: Adaptado de [Junior et al. 2018]

Como pode ser observado na Figura 1, o modelo que possui a maior quantidade de conexões é o modelo em que a UC gera a energia e a consome. De modo contrário, o modelo que possui a menor quantidade de conexões, apenas uma conexão, é o modelo de múltiplas UCs (ou condomínio). Em relação aos modelos que possibilitam o atendimento de mais do que uma UC (autoconsumo remoto e geração compartilhada), é visível que no autoconsumo a geração tem capacidade instalada normalmente menor do que o modelo de geração compartilhada. A principal razão é que na geração compartilhada o local da instalação é distinto da UC do proprietário, possibilitando uma geração de maior porte, respeitando os limites de potência para micro ou minigeração [Junior et al. 2018].

Atualmente, estão cadastrados 42 empreendimentos na ANEEL, como geração compartilhada. Cerca de 92% da capacidade instalada de geração compartilhada são de minigeradores de energia, que representam aproximadamente 24% dos empreendimentos. São ilustrados na Figura 2 os empreendimentos divididos pela sua fonte primária, representados por hídrica (CGH), solar fotovoltaica (UFV) e eólica (EOL).

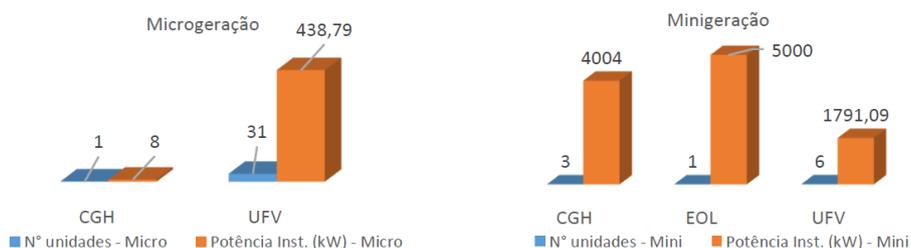


Figura 2. Dados de geração compartilhada até o dia 27/10/2017. Fonte: Adaptado de [Junior et al. 2018]

Da análise da Figura 2 se constata que quase a totalidade dos empreendimentos de microgeração compartilhada é de energia fotovoltaica e nos empreendimentos de minigeração a diversificação aumenta, uma vez que a capacidade de geração aumenta e, conseqüentemente, o retorno financeiro também. Distribuindo os empreendimentos no mapa do Brasil, conforme ilustrado na Figura 3, pode-se observar as principais regiões em que o modelo de geração compartilhada está sendo empregado.

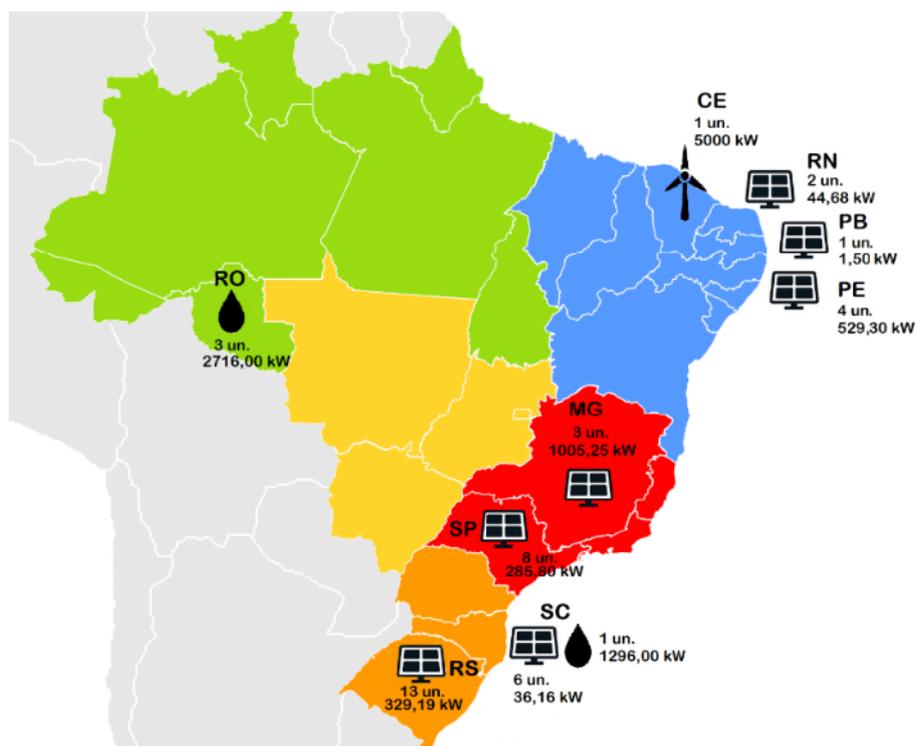


Figura 3. Distribuição dos empreendimentos de geração compartilhada. Fonte: Adaptado de [Junior et al. 2018]

Na Figura 3, pode-se observar a minigeração eólica de 5 MW no Estado do Ceará (denominada de Consórcio Prainha II), representa praticamente 44,48% da capacidade instalada de todos os empreendimentos de geração compartilhada e é responsável por gerar créditos de energia para 29 UCs. O empreendimento está localizado no Ceará sendo conectado na rede de distribuição da Enel Ceará. De maneira similar, pode-se observar na Figura 3, os empreendimentos que utilizam o potencial hidráulico para geração de energia. Todos os demais empreendimentos são de geradores fotovoltaicos, que representam

praticamente 88% dos empreendimentos de geração compartilhada, merecendo destaque a região sul e sudeste com 73,81% dos empreendimentos.

A seguir vamos apresentar alguns termos técnicos-econômicos necessários para uma melhor compreensão do método utilizado na análise dos cenários abordados.

4. Análise Econômico-Financeira

Para instaurar o sistema fotovoltaico em seu empreendimento, deve-se realizar um investimento inicial elevado e esses custos carecem de ser avaliados sob a perspectiva econômico-financeira, pois o investidor espera que esse bem adquirido apresente algum retorno econômico a médio/longo prazo.

A análise econômico-financeira de qualquer investimento auxilia na tomada de decisão e avalia a viabilidade desse negócio. [Casarotto Filho and Kopittke 2011] definiram diversas técnicas para os estudos de viabilidade:

- Taxa Mínima de Atratividade (TMA): É a taxa utilizada nos comparativos dos indicadores, e o seu valor é baseado no custo de oportunidade.
- Valor Presente Líquido (VPL): É a soma algébrica de todas as receitas e despesas transcorridas no tempo do projeto, descontadas à data inicial segundo a TMA;
- Taxa Interna de Retorno (TIR): É uma taxa de desconto que iguala o valor presente das receitas e o investimento inicial;
- Índice de Lucratividade (IL): É a relação do valor presente das receitas e os investimentos liquidados;
- *Payback*: Esse índice fornece o tempo necessário para que um investimento seja recuperado. Não há comparativos, a decisão de aceitação ou rejeição é subjetiva e depende dos objetivos do investidor.

5. Método de Pesquisa

Esse trabalho consiste em avaliar a viabilidade técnica e econômico-financeira de um sistema de energia solar FV por geração compartilhada. Para isso, foram selecionadas 10 UCs variadas e localizadas em todo o Estado de Goiás que compõem esse sistema de geração distribuída, sendo que todas UCs são propriedades de pessoas físicas. Essas UCs possuem consumos médios mensais de energia diversos variando de 420 a 127,5 kWh, então dividiu-se em 5 grupos distintos. Os grupos, o consumo mensal de cada grupo e a porcentagem de contribuição em relação ao sistema são apresentados na Tabela 1.

5.1. Indicadores Econômicos e Critérios

Para a análise econômico-financeira, utilizou-se os indicadores abordados anteriormente: TMA, VPL, TIR, IL e *payback* simples e descontado. Para isso, são apresentados na Tabela 2, resumidamente, os critérios de aceitação desse investimento para cada parâmetro escolhido. A aceitação ou rejeição foram baseadas conforme estabelecido por [Casarotto Filho and Kopittke 2011].

5.2. Cenários pessimista e otimista

Para a análise desse sistema, serão considerados dois cenários: pessimista e otimista. No cenário pessimista, o custo unitário de energia considerado foi de R\$0,78/kWh e a TMA

Tabela 1. Unidades consumidoras e suas contribuições no sistema

Grupos	UC	Consumo (kWh)	(%)
G1	A-1	420	15%
	A-2	420	15%
G2	A-3	365	13%
	A-4	365	13%
G3	A-5	300	11%
	A-6	300	11%
G4	A-7	200	7%
	A-8	200	7%
G5	A-9	127,5	5%
	A-10	127,5	5%
Total		2825	100%

Tabela 2. Indicadores econômicos e os seus critérios de aceitação

Indicador	Situação	Critério
VPL	$VPL > 0$	Projeto será aceito
	$VPL = 0$	Projeto poderá ser aceito
	$VPL < 0$	Projeto não será aceito
TIR	$TIR > TMA$	Projeto poderá ser aceito
	$TIR \leq TMA$	Projeto será rejeitado
IL	$IL > 0$	Viável
	$IL = 0$	Não gerará valor monetário
	$IL < 0$	Inviável
<i>Payback</i> Simples	Dependerá do tempo máximo aceitável de recuperação do investidor	
<i>Payback</i> Descontado		

igual a 10% ao ano (a.a.), sendo que para a TMA foram analisados investimentos atuais do mercado financeiro. No cenário otimista, o custo unitário de energia considerado foi de R\$0,85/kWh e a TMA igual a 6,5% ao ano (a.a.), sendo que para a TMA foi considerada a taxa do Sistema Especial de Liquidação e de Custódia (SELIC) imposta pelo Banco Central do Brasil (BC).

5.3. Estudo de caso referência

Utilizou-se o estudo de caso do [Spíndola 2018] para efeitos de comparação, no cenário pessimista, com o grupo 2 (G2) desse trabalho, que possuem os mesmos consumos médios de energia, iguais a 365 kWh. A Tabela 3 apresenta um resumo dos dados de referência, dados técnicos e dos resultados da análise econômico-financeira deste trabalho.

6. Resultados

6.1. Viabilidade Técnica

O terreno escolhido para a instalação do sistema de energia fotovoltaica está localizado na cidade de Dois Irmãos (GO). O consumo mensal médio total das 10 residências é de 2825 kWh, conforme contato com cada participante e exposto na Tabela 1.

Tabela 3. Dados gerais do estudo de caso de referência

Parâmetros	
Custo (R\$/kWh)	R\$ 0,78
TMA (%)	10%
Dados técnicos do sistema	
Consumo mensal (kWh)	365
Potência proposta (kW)	2,70
Geração estimada (kWh/ano)	3976
Área ocupada telhado (m ²)	17
Análise econômico-financeira	
Investimento inicial (R\$)	R\$16.650,00
VPL	R\$10.743,60
TIR	18%
IL	1,61
Payback simples	5 anos e 6 meses
Payback descontado	8 anos e 5 meses

A potência proposta é de 22,34 kW, com o objetivo de gerar cerca de 33.900 kWh por ano, ocupando cerca de 180 m² do terreno. Vale ressaltar que a empresa responsável pela distribuição de energia no estado de Goiás fixa uma tarifa mínima para um consumo de até 30 kWh em cada UC do tipo monofásica, ou seja, independente do investimento realizado nesse sistema, todas UCs terão que pagar uma tarifa mínima equivalente a um consumo de 30 kWh para a concessionária. Os painéis e seus componentes responsáveis pela geração de energia possuem uma vida útil de 25 anos, necessitando de manutenções bimestrais como limpeza dos módulos, troca de fiações elétricas e outros, que devem ser executadas por um electricista terceirizado da região que possua, pelo menos, um curso da Norma Regulamentadora 10 (NR-10).

Observa-se que o valor de produção de energia apresentado pela empresa é uma estimativa baseada em dados do clima coletados no passado. A produção real pode variar ao longo do tempo devido a uma série de fatores, como, por exemplo, incidência solar, períodos chuvosos fora da média, temperatura ambiente, sombra de árvores ou edifícios [Spíndola 2018].

6.2. Aspectos financeiros de orçamentação

Para uma melhor avaliação do benéfico financeiro, foi considerada uma análise de 25 anos, que é a vida útil dos painéis fotovoltaicos [Spíndola 2018]. Para a análise econômico-financeira desse investimento é apresentado de forma discriminada o orçamento do projeto na Tabela 4. Além disso, foi previsto um gasto anual de R\$1.000,00 por ano com um electricista terceirizado da região para realizar as manutenções preventivas do sistema.

Para esse custo total de R\$ 117.826,31 do sistema, cada UC contribui com uma parcela desse investimento inicial, considerando cada percentual de consumo médio mensal conforme disposto na Tabela 1. O investimento individual para cada UC é apresentado na Tabela 5.

Tabela 4. Orçamento para a implantação do sistema

Orçamento Energia Solar Fotovoltaica - Geração Compartilhada					
1	Custos com a Obra				
Item	Descrição	Unid.	Qtd.	P. Unitário (R\$)	P. Total(R\$)
1.1	Lote de 360 m2 em Dois Irmãos (GO)	vb	1	R\$ 8.500,00	R\$ 8.500,00
1.2	Custos com incorporação	vb	1	R\$ 850,00	R\$ 850,00
1.3	Alvenaria de tijolo furado 1/2 vez 14X29X9 - 6 furos - Arg. (1Calh:4Arml+100kg de CI/m3)	m2	196,25	R\$ 28,08	R\$ 5.510,62
1.4	Lastro de concreto magro, aplicado em pisos ou radiers, espessura de 5 cm. AF_07_2016	m2	8,62	R\$ 18,49	R\$ 159,44
1.5	Porta de madeira tipo veneziana, 80X210cm, espessura de 3cm, incluso dobradiças - fornecimento e instalação. AF_08/2015	uni	1	R\$ 519,75	R\$ 519,75
1.6	Portão de ferro em chapa galvanizada plana 14 GSG	m2	6,6	R\$ 198,52	R\$ 1.310,23
1.7	Trama de madeira composta por terças para telhados de até 2 águas para telha ondulada de fibrocimento, metálica, plástica ou termoacústica, incluso transporte vertical. AF_12/2015	m2	12,1	R\$ 10,25	R\$ 123,97
1.8	Telhamento com telha ondulada de fibrocimento e = 6 mm, com recobrimento lateral de 1 1/4 de onda para telhado com inclinação máxima de 10º, com até 2 águas, incluso çamento. AF_06/2016	m2	12,1	R\$ 32,79	R\$ 396,60
SUB-TOTAL (R\$)					R\$ 17.370,61
2	Energia Solar				
Item	Descrição	Unid.	Qtd.	P. Unitário (R\$)	P. Total(R\$)
2.1	Módulos Fotovoltaicos policristalinos de 265 W	Uni	85	R\$ 749,00	R\$ 63.665,00
2.2	Inversor de 20 kW de potência ou semelhantes	Uni	1	R\$ 23.240,70	R\$ 23.240,70
2.3	Estruturas de suporte em alumínio ou ferro galvanizado	vb	1	R\$ 5.150,00	R\$ 5.150,00
2.4	Dispositivos de proteção elétrica e materiais elétricos	vb	1	R\$ 3.550,00	R\$ 3.550,00
2.5	Serviços de instalação	vb	1	R\$ 1.500,00	R\$ 1.500,00
2.6	Projeto elétrico	vb	1	R\$ 2.500,00	R\$ 2.500,00
2.7	Homologação do gerador junto à CELG/ENEL	vb	1	R\$ 500,00	R\$ 500,00
2.8	Frete e deslocamento de pessoal qualificado	vb	1	R\$ 350,00	R\$ 350,00
SUB-TOTAL (R\$)					R\$ 100.455,70
TOTAL (R\$)					R\$ 117.826,31

6.3. Viabilidade econômico-financeira - Cenário otimista

Nesta seção são apresentados os resultados econômico-financeiros de todos os grupos do sistema considerando o cenário otimista e analisando a viabilidade perante essa conjuntura. Para todos os grupos, considera-se que a proposta da empresa garante a troca dos

Tabela 5. Investimento inicial de cada UC

Grupos	UC	Investimento (R\$)
G-1	A-1	R\$ 17.517,54
	A-2	R\$ 17.517,54
G-2	A-3	R\$ 15.223,58
	A-4	R\$ 15.223,58
G-3	A-5	R\$ 12.512,53
	A-6	R\$ 12.515,53
G-4	A-7	R\$ 8.341,69
	A-8	R\$ 8.341,69
G-5	A-9	R\$ 5.317,82
	A-10	R\$ 5.317,82

painéis e do inversor em caso de perda de potência ou defeitos de fabricação.

6.3.1. Grupo 1

Para o G1, o consumo médio anual é de 4.680 kWh, ou seja, o sistema propicia uma economia de R\$3.978,00, considerando o custo da tarifa de energia elétrica igual a R\$0,85 por kWh. Ao longo de 25 anos, o G1 consumirá cerca de 117.000 kWh, economizando cerca de R\$99.450,00, que seriam pagos à CELG/ENEL em preço público de energia elétrica. O cálculo do *payback* descontado considera uma TMA de 6,5% ao ano, assim, são necessários 5 anos e 6 meses para recuperar o investimento proposto de R\$17.517,54 acrescidos dos custos de manutenção. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 29.785,83. A TIR e o IL foram calculados e iguais a, respectivamente, 22% e 2,59. Salienta-se que considerando um *payback* simples, o fluxo de caixa acumulado durante um período de 4 anos e meio já supera o valor do investimento inicial.

6.3.2. Grupo 2

Para o G2, o consumo médio anual é de 4.020 kWh, ou seja, o sistema possibilita uma economia de R\$3.417,00, considerando o custo da tarifa de energia igual a R\$0,85 por kWh. No decorrer dos 25 anos, o G2 consumirá cerca de 100.500 kWh, economizando cerca de R\$85.425,00, que seriam pagos à concessionária.

O *payback* descontado considera uma TMA de 6,5% ao ano, assim, são necessários 5 anos e 8 meses para recuperar o investimento proposto de R\$15.223,58 acrescidos dos gastos com manutenção preventiva. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 25.236,78. A TIR e o IL foram calculados e iguais a, respectivamente, 22% e 2,53. A partir de 4 anos e 7 meses, considerando o *payback* simples, o investimento já pode ser considerado recuperado.

6.3.3. Grupo 3

Para o G3, o consumo médio anual é de 3.240 kWh, ou seja, o sistema possibilita uma economia de R\$2.754,00, considerando o custo da tarifa de energia igual a R\$0,85 por

kWh. Ao longo dos 25 anos, o G3 consumirá cerca de 81.000 kWh, economizando cerca de R\$68.850,00, que seriam pagos à empresa responsável pela distribuição de energia.

O cálculo do *payback* descontado considera uma TMA de 6,5% ao ano, assim, necessita-se de 5 anos e 10 meses para recuperar o investimento proposto de R\$12.512,53 acrescidos das despesas com manutenções bimestrais. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 19.860,64. A TIR e o IL foram calculados e iguais a, respectivamente, 21% e 2,45. Considerando o *payback* simples, a partir de 4 anos e 9 meses o investimento será considerado como recuperado.

6.3.4. Grupo 4

Para o G4, o consumo médio anual é de 2.040 kWh, portanto o sistema propicia uma economia de R\$1.734,00, considerando o custo da tarifa de energia igual a R\$0,85 por kWh. No 25º ano, o G4 terá consumido cerca de 51.000 kWh, economizando cerca de R\$43.350,00, que seriam pagos à CELG/ENEL.

Para o *payback* descontado a TMA considerada é de 6,5% ao ano, assim, após 6 anos e 5 meses o investimento proposto de R\$8.341,69 somado dos custos com manutenções será totalmente recuperado. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 11.589,65. A TIR e o IL foram calculados e iguais a, respectivamente, 19% e 2,21. Considerando o *payback* simples, partir de 5 anos e 1 mês o investimento inicial será recuperado.

6.3.5. Grupo 5

Para o G5, o consumo médio anual é de 1.170 kWh, portanto a economia é de R\$994,50, considerando o custo da tarifa de energia igual a R\$0,85 por kWh. No 25º ano, o G5 terá consumido aproximadamente 29.250 kWh, deixando de gastar cerca de R\$24.862,50, que seriam pagos à concessionária de energia elétrica da região.

No cálculo do *payback* descontado a TMA considerada é de 6,5% ao ano, assim, após 7 anos e 9 meses o investimento proposto de R\$5.317,82 somado dos custos com manutenções bimestrais será totalmente recuperado. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 5.593,18. A TIR e o IL foram calculados e iguais a, respectivamente, 16% e 1,86. Considerando o *payback* simples, a partir de 5 anos e 11 meses o valor investido será recuperado.

6.4. Viabilidade econômico-financeira - Cenário pessimista

Nesse item são apresentados os resultados econômico-financeiros de todos os grupos da geração compartilhada, analisando o cenário pessimista e averiguando a viabilidade diante desse cenário. Para todos os grupos, considera-se que a proposta da empresa garante a troca dos painéis e do inversor em caso de perda de potência ou defeitos de fabricação.

6.4.1. Grupo 1

Para o G1, o consumo médio anual é de 4.680 kWh, ou seja, o sistema propicia uma economia de R\$3.669,96, considerando o custo da tarifa de energia elétrica igual a R\$0,78 por kWh. Ao longo de 25 anos, o G1 consumirá cerca de 117.000 kWh, economizando

cerca de R\$91.749,06, que seriam pagos à CELG/ENEL em preço público de energia elétrica.

O cálculo do *payback* descontado considera uma TMA de 10% ao ano, assim, são necessários 7 anos e 1 mês para recuperar o investimento proposto de R\$17.517,54 acrescidos dos custos de manutenção. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 14.887,15. A TIR e o IL foram calculados e iguais a, respectivamente, 20% e 1,81. Considerando o *payback* simples, a partir de 4 anos e 11 meses o investimento já estará totalmente pago.

6.4.2. Grupo 2

Para o G2, o consumo médio anual é de 4.020 kWh, ou seja, o sistema possibilita uma economia de R\$3.152,40, considerando o custo da tarifa de energia igual a R\$0,78 por kWh. No decorrer dos 25 anos, o G2 consumirá cerca de 100.500 kWh, economizando cerca de R\$78.810,09, que seriam pagos à concessionária.

O *payback* descontado considera uma TMA de 10% ao ano, assim, são necessários 7 anos e 3 meses para recuperar o investimento proposto de R\$15.223,58 acrescidos dos gastos com manutenção preventiva. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 12.483,21. A TIR e o IL foram calculados e iguais a, respectivamente, 20% e 1,77. Considerando o *payback* simples, o fluxo de caixa decorrido durante um período de 5 anos já supera o valor do investimento inicial.

6.4.3. Grupo 3

Para o G3, o consumo médio anual é de 3.240 kWh, ou seja, o sistema possibilita uma economia de R\$2.540,74, considerando o custo da tarifa de energia igual a R\$0,78 por kWh. Ao longo dos 25 anos, o G3 consumirá cerca de 81.000 kWh, economizando cerca de R\$63.518,58, que seriam pagos à empresa responsável pela distribuição de energia.

O cálculo do *payback* descontado considera uma TMA de 10% ao ano, assim, necessita-se de 7 anos e 7 meses para recuperar o investimento proposto de R\$12.512,53 acrescidos das despesas com manutenções bimestrais. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 9.642,20. A TIR e o IL foram calculados e iguais a, respectivamente, 19% e 1,72. Considerando o *payback* simples, a partir de 5 anos e 2 meses o investimento será considerado como recuperado.

6.4.4. Grupo 4

Para o G4, o consumo médio anual é de 2.040 kWh, portanto o sistema propicia uma economia de R\$1.599,73, considerando o custo da tarifa de energia igual a R\$0,78 por kWh. No 25º ano, o G4 terá consumido cerca de 51.000 kWh, economizando cerca de R\$39.993,18, que seriam pagos à CELG/ENEL.

Para o *payback* descontado a TMA considerada é de 10% ao ano, assim, após 8 anos e 6 meses o investimento proposto de R\$8.341,69 somado dos custos com manutenções será totalmente recuperado. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 5.271,40. A TIR e o IL foram calculados e iguais a, respectivamente, 18% e

1,57. Considerando o *payback* simples, a partir de 5 anos e 7 meses o investimento inicial será recuperado.

6.4.5. Grupo 5

Para o G5, o consumo médio anual é de 1.170 kWh, portanto a economia é de R\$917,49, considerando o custo da tarifa de energia igual a R\$0,78 por kWh. No 25º ano, o G5 terá consumido aproximadamente 29.250 kWh, deixando de gastar cerca de R\$22.937,27, que seriam pagos à concessionária de energia elétrica da região.

No cálculo do *payback* descontado a TMA considerada é de 10% ao ano, assim, após 11 anos o investimento proposto de R\$5.317,82 somado dos custos com manutenções bimestrais será totalmente recuperado. O valor presente líquido ao final do vigésimo quinto ano será de R\$ 2.102,57. A TIR e o IL foram calculados e iguais a, respectivamente, 15% e 1,34. Considerando o *payback* simples, a partir de 6 anos e 6 meses o valor investido será recuperado.

6.5. Discussão com relação aos resultados

É possível ratificar que para todos os grupos, em ambos os cenários, o projeto será aceito, pois o VPL é maior do que zero, o IL é maior do que um, a TIR é maior do que a TMA proposta e o *payback* (simples ou descontado) é inferior à vida útil do sistema.

De forma geral, os resultados foram considerados satisfatórios para todos os grupos e cenários, visto que os indicadores econômicos VPL, IL, TIR e *payback* (simples e descontado) utilizados como variáveis de avaliação econômico-financeira, tiveram seus critérios atendidos, portanto o projeto foi aceito e conceituado como viável. No cenário otimista, o IL variou de 1,86 a 2,59, a TIR de 16% a 22% e o *payback* descontado variou de 5 anos e 6 meses a 7 anos e 9 meses. No cenário pessimista, o IL variou de 1,34 a 1,81, a TIR de 15% a 20% e o *payback* descontado variou de 7 anos e 1 mês a 11 anos. Os resultados do cenário pessimista foram piores do que o cenário otimista, e isso só corrobora as premissas adotadas para cada um. No pessimista, o custo da tarifa de energia elétrica é de R\$0,78 por kWh, ou seja, quando se compara com o otimista, a economia na fatura de energia ao longo da vida útil do sistema será menor. A TMA do cenário pessimista é de 10% ao ano, com isso pode-se considerar que a perda por não investir em outro negócio com essa remuneração é maior do que no cenário otimista, que tem a TMA igual a 6,5% ao ano. Vale ressaltar, que quanto menor o investimento, menor o desempenho dos indicadores econômicos, como no G5 do cenário pessimista que teve um *payback* descontado de 11 anos e o G1 de 7 anos e 1 mês, mostrando proporcionalidade econômica em análise de investimentos.

6.6. Comparação com o estudo de caso referência

A comparação foi realizada pelo G2 do cenário pessimista composto pela UCs nomeadas de A-9 e A-10, com o estudo de caso [Spíndola 2018], pois esses possuem os mesmos parâmetros e premissas adotadas. Um resumo das análises nos dois casos é apresentado na Tabela 6, para assim ser realizado uma comparação coerente.

Analisando a Tabela 6, observa-se que para todos os parâmetros, o G2 do estudo de energia solar FV por geração compartilhada, apresenta-se superior ao estudo de caso

Tabela 6. Comparação das análises econômico-financeiras

Item	G2	[Spíndola 2018]
Investimento inicial	R\$ 15.223,58	R\$ 16.650,00
VPL	R\$ 12.483,21	R\$ 10.743,60
TIR	20%	18%
IL	1,77	1,61
Payback Simples	5 anos	5 anos e 6 meses
Payback Descontado	7 anos e 3 meses	8 anos e 5 meses

de [Spíndola 2018], pois o investimento inicial do G2 foi inferior ao deste, o VPL de [Spíndola 2018] foi menor que o do G2, a TIR e o IL do G2 foram maiores, e o tempo de retorno do investimento mostrou-se mais favorável no G2.

6.7. Potencial de investimento

Uma das vantagens do sistema de geração compartilhada é que é possível desfazer da sua parte no futuro, pois a geração de energia pode acontecer distante da residência. Sendo a participação individual via consórcio ou cooperativa, o participante pode sair da parceria e outra ingressar. Essa dinâmica é favorável quando comparado com o sistema FV em residências, pois exige-se a permanência durante a vida útil do empreendimento.

7. Considerações Finais

O estudo analisou a viabilidade técnica e econômico-financeira do sistema FV de geração compartilhada como opção para diminuição de custos com energia elétrica para dez UCs localizadas no estado de Goiás.

Para a análise de viabilidade técnica foi contatada uma empresa do ramo que investigou a região para confirmar se haveria radiação solar suficiente para a produção de energia. Para a análise de viabilidade econômico-financeira de energia solar FV foram analisados o payback descontado, o VPL o IL e a TIR.

Os resultados mostraram que a implantação de energia solar FV por geração compartilhada é viável para os dois cenários. A análise de viabilidade técnica ratificou que o local de estudo possui incidência solar suficiente para geração de energia conforme o consumo médio mensal informado pelos moradores. Para a análise de viabilidade econômico-financeira, todos os indicadores econômicos tiveram seus critérios aceitos nas duas situações, tendo o projeto tido como aceite e viável. A comparação do estudo atual com o de referência apresentou melhores resultados no primeiro para todos os indicadores econômico-financeiros, mostrando-se que o sistema compartilhado precisa ser melhor aproveitado. Para futuros trabalhos sugerimos inserir a Análise de Risco para enriquecer a discussão.

Conclui-se que, além de diminuir os custos com energia elétrica e de apresentar viabilidade técnica e econômico-financeira para a região de análise, a energia solar FV por geração compartilhada, pode ser considerada uma mini usina energia renovável que propicia a várias residências parceiras uma economia mensal na fatura de energia elétrica, além de contribuir para a redução do impacto ambiental.

Referências

- ANEEL (2012). Resolução normativa 482. *Estabelece as condições gerais para o acesso de microgeração e minigeração distribuída aos sistemas de distribuição de energia elétrica, o sistema de compensação de energia elétrica, e dá outras providências*, 17. Disponível em <<http://www2.aneel.gov.br/cedoc/ren2012482.pdf>>.
- ANEEL (2015). Resolução normativa 687. *Altera a Resolução Normativa 482*. Disponível em <<http://www2.aneel.gov.br/cedoc/ren2015687.pdf>>.
- Cabral, I. and Vieira, R. (2012). Viabilidade econômica x viabilidade ambiental do uso de energia fotovoltaica no caso brasileiro: uma abordagem no período recente. In *III Congresso Brasileiro de Gestão Ambiental*, volume 19.
- Casarotto Filho, N. and Kopittke, B. H. (2011). *Análise de investimentos: matemática financeira, engenharia econômica, tomada de decisão, estratégia empresarial*. Atlas.
- do Brasil, R. F. (2016). Instrução normativa da receita federal do Brasil nº 1.634, de 06 de maio de 2016. *Dispõe sobre o Cadastro Nacional da Pessoa Jurídica (CNPJ)*. Diário Oficial da União, Brasília - DF.
- EPE (2017). *Balanco energético nacional 2017 - Ano base 2016*. EMPRESA DE PESQUISA ENERGÉTICA. Disponível em <<http://www.epe.gov.br>>.
- Junior, W. P., Takigawa, F., Neto, E. A. C. A., and Fernandes, R. C. (2018). Levantamento da geração distribuída compartilhada no Brasil. In *VII Congresso Brasileiro de Energia Solar-CBENS 2018*.
- Nacional, C. (1971). Lei nº 5.764, de 16 de dezembro de 1971. *Define a Política Nacional de Cooperativismo, institui o regime jurídico das sociedades cooperativas, e dá outras providências*. Diário Oficial da União, Brasília - DF.
- Nacional, C. (1976). Lei nº 6.404, de 15 de dezembro de 1976. *Dispõe sobre as Sociedades por Ações*. Diário Oficial da União, Brasília - DF.
- Nacional, C. (2002). Código civil brasileiro. *Capítulo VII - Da Sociedade Corporativa*. Disponível em <<https://www2.senado.leg.br/bdsf/bitstream/handle/id/70327/C%C3%B3digo%20Civil%202%20ed.pdf>>.
- Nacional, C. (2008). Lei nº 11.795, de 08 de outubro de 2008. *Dispõe sobre o Sistema de Consórcio*. Diário Oficial da União, Brasília - DF.
- Spíndola, G. M. (2018). Estudo de viabilidade técnica e econômico-financeira de energia solar fotovoltaica em uma residência de senador canedo (GO). *Short paper apresentado na disciplina de Introdução ao Gerenciamento da Construção do Programa de Pós-Graduação em Geotecnia, Estruturas e Construção Civil da UFG*.

Análise e projeto de filtro FIR aplicado à tecnologia F-OFDM

Salatyel Fellipe da Silva, Leonardo L. de Assis Silva, Flávio Geraldo C. Rocha

Escola de Engenharia Elétrica, Mecânica e de Computação (EMC) – Universidade Federal de Goiás (UFG) – Av. Universitária, n.º 1488 - quadra 86 - bloco A - 3º piso - Setor Leste Universitário – Goiânia – GO – Brasil

ssalson@hotmail.com, leonardolucas.a.s@gmail.com, flaviogcr@ufg.br

Resumo. *Uma tecnologia candidata a ser utilizada na nova geração de comunicação móvel, o 5G, é a modulação F-OFDM (Filtered-Orthogonal Frequency Division Multiplexing), que consiste basicamente na aplicação de filtros nas sub-bandas da modulação OFDM. Técnicas como o janelamento permitem o projeto de filtros do tipo FIR (Finite Impulse Response) com desempenho variável em função da janela escolhida, mas que em geral fornece maior eficiência ao processo de transmissão e recepção de sinais de dados quando comparada à modulação OFDM. Neste artigo são analisados diferentes tipos de janelas para o filtro FIR a ser utilizado na F-OFDM. Além disso, neste artigo é proposto o uso de técnicas de formatação de pulso para formar o protótipo do filtro FIR com o objetivo de aumentar o seu desempenho. Com o intuito de avaliar a proposta do artigo, são conduzidas simulações de transmissão de imagens utilizando a modulação F-OFDM em comparação à modulação OFDM em termos de BER (Bit Error Rate). Os resultados obtidos nas simulações indicam que o desempenho da F-OFDM é em geral superior ao obtido com a modulação OFDM, mas variável conforme o projeto de filtro utilizado. Nas simulações foram obtidos valores reduzidos de BER utilizando o filtro proposto: uma combinação da técnica de formatação de pulso aplicada ao protótipo do filtro FIR com a escolha adequada da janela a ser utilizada. Os resultados obtidos indicam a relevância do projeto do filtro da modulação F-OFDM para torná-la candidata a ser utilizada em novas arquiteturas de comunicações de dados, incluindo o 5G.*

Abstract. *A candidate technology to be used in the new generation of mobile communication, the 5G, is the F-OFDM (Filtered-Orthogonal Frequency Division Multiplexing) modulation, which basically consists of the application of filters in the OFDM modulation subbands. Techniques such as windowing allow the design of FIR (Finite Impulse Response) filters with variable performance depending on the chosen window, but which generally provides greater efficiency in the transmission and reception of data signals when compared to the OFDM modulation. In this article different types of windows are analyzed for the FIR filter design to be used in F-OFDM. In addition, this article proposes the use of pulse shaping techniques to design the FIR filter prototype in order to increase its performance. In order to evaluate the proposal of the article, simulations of image*

transmission using F-OFDM modulation are carried out in comparison to OFDM modulation in terms of BER (Bit Error Rate). The results obtained in the simulations indicate that the performance of F-OFDM is generally higher than that obtained with OFDM modulation, but variable according to the filter design. In simulations, reduced BER values were obtained using the proposed filter: an application of the pulse shaping technique to the FIR filter prototype combined to the appropriate choice of the window to be used in the filter. The results indicate the relevance of the F-OFDM modulation filter design to make F-OFDM a modulation candidate to be used in new data communications architectures, including 5G.

1. Introdução

As tecnologias evoluem continuamente, e recentemente tem sido propostas alternativas a serem utilizadas na próxima geração de sistemas de comunicação móvel, o 5G. Para esse novo padrão de comunicação, espera-se um aumento significativo na taxa de dados em relação a geração antecessora (com a nova tecnologia espera-se taxas acima de 1 Gbps). Além disso, a tendência é que a quantidade de equipamentos usuários (*User Equipments* – UEs) aumente muito, o que não poderá ser suportado pelas tecnologias existentes. Visando responder aos desafios do 5G, tecnologias promissoras, tais como MIMO (*Multiple-Input and Multiple-Output*) de larga escala (*massive MIMO* – *ma-MIMO*), protocolos para dispositivos inteligentes e autônomos, comunicações utilizando ondas milimétricas (*mmWave*), entre outras têm sido analisadas de modo combinado [Zhang et al. 2015], [Abdoli, Jia e Ma 2015], [Guerra 2016].

Adicionalmente às tecnologias supracitadas, é preciso citar tecnologias de modulação e multiplexação de dados, indispensáveis para se obter a taxa de dados solicitada pelo 5G, assim como a baixa latência requerida. A OFDM (*Orthogonal Frequency-Division Multiplexing*) tem sido a principal tecnologia de modulação e multiplexação de dados para padrões de comunicações móveis sem fio, incluindo o padrão IEEE 802.11 e também o 4G.

Entretanto, para viabilizar a próxima geração de comunicação móvel, são necessárias formas de ondas alternativas, as quais não apresentem as limitações existentes em sistemas OFDM ou que pelo menos não apresentem custos ou limitações tecnológicas proibitivas para sua utilização em sistemas 5G [Guerra 2016]. Em geral, os novos requisitos de desempenho enfrentados pela forma de onda 5G podem ser atendidos pela F-OFDM (*Filtered-Orthogonal Frequency-Division Multiplexing*) e a eficiência geral do espectro pode ser melhorada. Entre todas as candidatas à forma de onda 5G a F-OFDM aparece como a mais promissora, em termos de desempenho geral, complexidade associada ao custo e viabilidade no caminho de evolução do atual 4G para a geração futura quando comparada a outras candidatas, tais como GFDM (*Generalized Frequency Division Multiplexing*), FBMC (*Filter Bank Multi-Carrier*) e UFMC (*Universal Filtered Multi-Carrier*) [Zhang et al. 2015].

A principal diferença entre a modulação F-OFDM e a modulação OFDM é que a F-OFDM utiliza um filtro FIR (*Finite Impulse Response*) para filtrar a parte do espectro desejada, potencializando as frequências de interesse e mitigando as demais. O filtro FIR

utilizado na F-OFDM emprega uma técnica chamada de janelamento. De fato, as funções de janelamento são amplamente utilizadas no processamento digital de sinais para aplicações em análise e estimativa de sinais e projeto de filtros digitais. As janelas utilizadas para calcular a transformada rápida de Fourier (*Fast Fourier Transform – FFT*) reduzem os efeitos de vazamento de sinal, mas não conseguem eliminar totalmente este efeito. Além disso, cada tipo de janela afeta o espectro de uma maneira diferente. Muitas janelas diferentes foram propostas ao longo do tempo, cada uma com suas vantagens e desvantagens em relação às outras. Algumas são mais adequadas para tipos específicos de sinais (por exemplo, sinais senoidais), outras melhoram a resolução de frequência (isto é, facilitam a detecção da frequência exata de um pico no espectro), já algumas melhoram a precisão da amplitude (ou seja, indicam com maior precisão o nível do pico). Portanto, o tipo de janela deve ser escolhido para cada aplicação específica objetivando os melhores resultados [Patil 2015].

Duas das principais aplicações das janelas no processamento digital de sinais são: análise de dados baseada em Transformada Rápida de Fourier (FFT) e projeto de filtros, tais como filtros de Resposta de Impulso Finita (FIR – *Finite Impulse Response*) e filtros de Resposta de Impulso Infinita (IIR – *Infinite Impulse Response*). Para análise de FFT, as janelas são empregadas para suprimir o chamado “efeito de vazamento” e, para o projeto de filtros FIR utilizando o “método de janelas”, as oscilações de Gibbs são atenuadas (o chamado fenômeno de Gibbs ocorre na forma de oscilações indesejáveis perto das bordas da largura de banda do filtro) [Patil 2015]. O filtro FIR, que é o foco deste artigo, é um tipo de filtro digital caracterizado por uma resposta ao impulso que se torna nula após um tempo finito, diferentemente do filtro IIR que possui duração infinita.

A resposta ao impulso dos filtros ideais é de duração infinita, porém, não é possível avaliar a resposta de frequência correspondente e implementar o filtro por hardware ou software. Assim, a resposta ao impulso deve ser truncada em ambas as extremidades em relação à faixa central, todavia um corte súbito tenderá a causar efeitos indesejáveis. A escolha adequada de janelas para o filtro poderá reduzir tais efeitos indesejados.

Neste artigo, serão analisadas algumas janelas já conhecidas nos estudos de processamento de sinais. Essa análise de diferentes tipos de janelas tem por objetivo encontrar aquelas com melhores características a fim de, conciliadas a técnicas de formatação de pulso, melhorar o desempenho do projeto de filtro FIR a ser utilizado na F-OFDM, com o intuito de torná-la uma tecnologia candidata a ser utilizada na tecnologia de comunicação móvel de 5ª geração (5G).

O restante do artigo está organizado da seguinte forma: na seção 2 é apresentada a tecnologia F-OFDM; a seção 3 discorre sobre o projeto do filtro FIR utilizado na F-OFDM; a seção 4 descreve diferentes tipos de janelas empregada no filtro FIR; a seção 5 apresenta uma análise comparativa das janelas descritas na seção 4; a seção 6 apresenta a proposta de uso e projeto de filtro FIR combinando a escolha apropriada de janela a uma técnica de formatação de pulso, ainda na seção 6 são apresentados os resultados obtidos em termos de BER na transmissão de imagens, comparando o F-OFDM com o OFDM; por fim, na seção 7, são apresentadas as conclusões obtidas neste trabalho.

2. F-OFDM: Aplicação de um Filtro FIR para melhoria de desempenho do OFDM

A técnica de modulação OFDM consiste em um esquema de transmissão e de recepção baseado em multiportadoras. Na OFDM, os dados são sempre multiplexados utilizando-se primeiramente algum esquema de modulação digital (QAM (*Quadrature Amplitude Modulation*) ou PSK (*Phase Shift Keying*), por exemplo), por isso a OFDM é referida de forma intercambiável na literatura como modulação e multiplexação. Em seguida, esses dados modulados são colocados em paralelo, sendo realizado o processamento desses dados através da IFFT (*Inverse Fast Fourier Transform*), transformando assim o sinal no domínio do tempo. Posteriormente, é adicionado o pre-echo cíclico (*Cyclic Prefix* – CP) para combater a ISI (*Inter Symbol Interference*). Os dados modulados no domínio do tempo são convertidos de digitais para analógicos e então serializados. Por fim, o sinal é trasladado de banda-base para banda passante via multiplicação pela frequência da portadora para, então, ser transmitido. Do lado do receptor é feito o mesmo procedimento, basicamente, mas na ordem contrária e utilizando-se a FFT no lugar da IFFT, visto que a informação original no transmissor encontra-se no domínio da frequência [Guerra 2016].

O sistema OFDM apresenta diversas vantagens, tais como: a capacidade de transformar canais de comunicação com seletividade em frequência em canais aproximadamente planos, robustez contra desvanecimento multipercurso, fácil implementação através do uso de algoritmos baseados na transformada rápida de Fourier (*Fast Fourier Transform* – FFT). Apesar dessas vantagens, a modulação OFDM também possui certas características indesejáveis pelo fato de ter uma grande sensibilidade na sincronização de frequência. Como já mencionado, sistemas de comunicações baseados no OFDM podem utilizar o pre-echo cíclico para combater os efeitos da Interferência Intersimbólica (ISI), o qual funciona como uma redundância para o sinal transmitido e que está sujeito à dispersão temporal do canal [Guerra 2016].

Para evitar as limitações da OFDM acima mencionadas e para enfrentar os novos desafios que surgem com as novas gerações de comunicações de dados, foi desenvolvido a F-OFDM (OFDM filtrado). Com divisão e filtragem baseadas em sub-bandas, sistemas F-OFDM independentes estão estreitamente contidos na largura de banda atribuída. Desta forma, a F-OFDM é capaz de superar as desvantagens da OFDM enquanto mantém as suas vantagens [Zhang et al. 2015; Abdoli, Jia e Ma 2015]. Para habilitar a filtragem baseada em sub-banda e, assim, aproveitar os benefícios sugeridos pela F-OFDM, são necessários projetos e sistemas de filtros. Em geral, o projeto do filtro envolve a negociação (*trade-off*) entre as características do domínio do tempo e da frequência, e também é fundamentado pela complexidade da implementação. O vazamento de energia do sinal no domínio do tempo deve ser controlada para mitigar a interferência entre símbolos (ISI), também é interessante melhorar a nitidez da região de transição no domínio da frequência, com o intuito de reduzir a interferência entre canais vizinhos (ICI - *Inter Carrier Interference*).

3. Projeto e implementação do filtro FIR para o F-OFDM

Para projetar o filtro FIR com janelamento, dois parâmetros devem ser selecionados: a frequência de corte, f_c , e o comprimento do kernel do filtro, M . A frequência de corte é expressa como uma fração da taxa de amostragem e, portanto, deve estar entre 0 e 0,5. O valor para M depende do fator de *roll-off* (o quão rápida a amplitude da resposta em frequência de uma função decai), o valor de M está relacionado com a largura de banda de transição B por meio da seguinte aproximação [Smith 1999]:

$$M \approx \frac{4}{B} \quad (1)$$

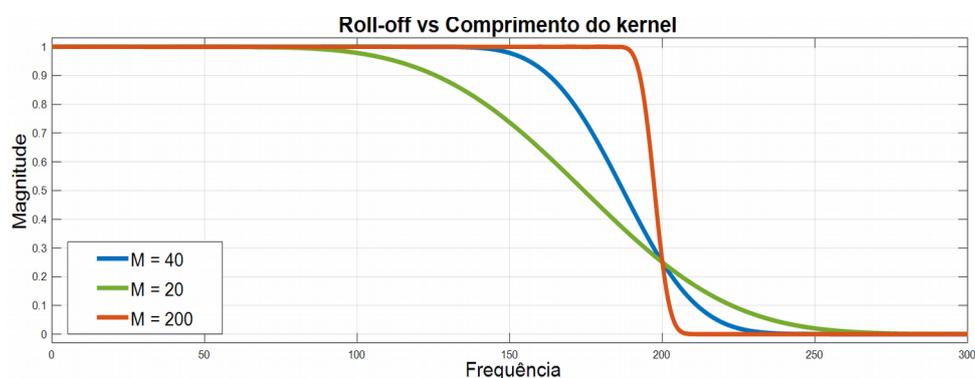


Figura 1. Comparação do *roll-off* de acordo com o valor de M

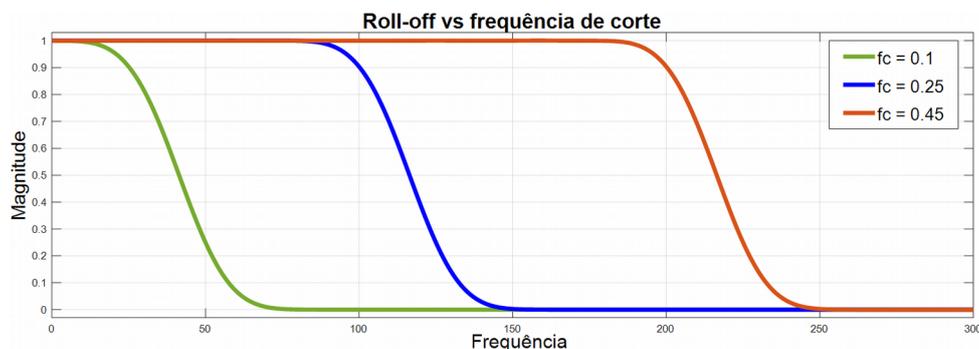


Figura 2. Comparação do *roll-off* de acordo com o valor da f_c

Na Figura 1, B pode ser visto como a distância entre as frequências em que a função sai da amplitude 1 para 0. É possível perceber que quanto maior o valor de M , mais rápida será a taxa de decaimento da amplitude em função da frequência (fator de *roll-off*). A largura de banda de transição também é expressa como uma fração da frequência de amostragem, e deve estar entre 0 e 0,5. A forma da resposta em frequência não depende do valor da frequência de corte, o formato do espectro será o mesmo, todavia deslocado na frequência (Figura 2)[Smith 1999]. Todas as figuras deste trabalho foram obtidas por meio de simulações utilizando o software MATLAB da MathWorks®.

No domínio do tempo, o janelamento signi ca multiplicar a função $h_D(n)$ por uma função de janela de duração finita $w(n)$ para se obter uma itragem do sinal. Lembre-se que multiplicação de duas funções no domínio do tempo (Figura 3e) é equivalente a convolução destas funções no domínio da frequência (Figura 3f). A função utilizada para a janela $w(n)$ pode resultar em um ltro com resposta truncada, como aquela obtida utilizando-se uma janela retangular (Figura 3e), ou pode apresentar transições suaves para zero em suas duas extremidades, de modo a evitar saltos abruptos no início e no final da resposta do ltro.

Além disso, o janelamento fornece uma localização de tempo razoável para o símbolo na resposta ao impulso do ltro e, assim, mantém a interferência intersimbólica (ISI) do sinal F-OFDM dentro de um limite aceitável [Zhang et al. 2015].

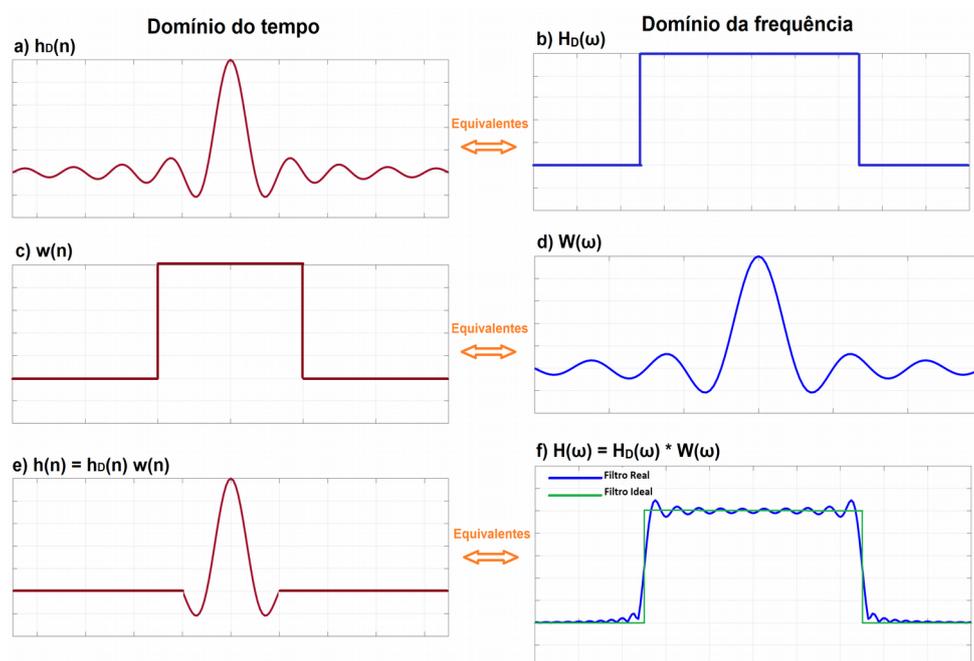


Figura 3. Passo a passo do funcionamento do ltro FIR

Como o tempo requerido para uma convolução é proporcional ao comprimento dos sinais, a equação (1) expressa uma negociação (*trade-o*) entre o tempo de computação (tamanho de M) e a qualidade do ltro (o valor de B). Quanto maior o valor de M , mais recursos computacionais são exigidos, porém a qualidade do ltro é melhor, já que B será menor (*roll-o* mais rápido), ou seja, o ltro se aproximará de sua resposta ideal (Figura 4). Após f_c e M terem sido selecionados, a resposta do ltro é calculada a partir da seguinte equação:

$$h(n) = K h_D(n) w(n) \quad (2)$$

Na equação (2) tem-se uma constante K , um formato de pulso $h_D(n)$ (sinc, por exemplo) com deslocamento $M/2$ e a janela $w(n)$ a ser utilizada. Para o ltro ter ganho unitário em DC, a constante K deve ser escolhido de tal forma que a soma de todas as

amostras seja igual a um. Alguns cuidados devem ser tomados na escolha da função $h_D(n)$, por exemplo, se for utilizada a função sinc a equação (3) é obtida de (2):

$$h(n) = K \frac{\sin[2\pi n f_c (n - M/2)]}{n - M/2} w(n) \quad (3)$$

onde, para $n = M/2$, tem-se:

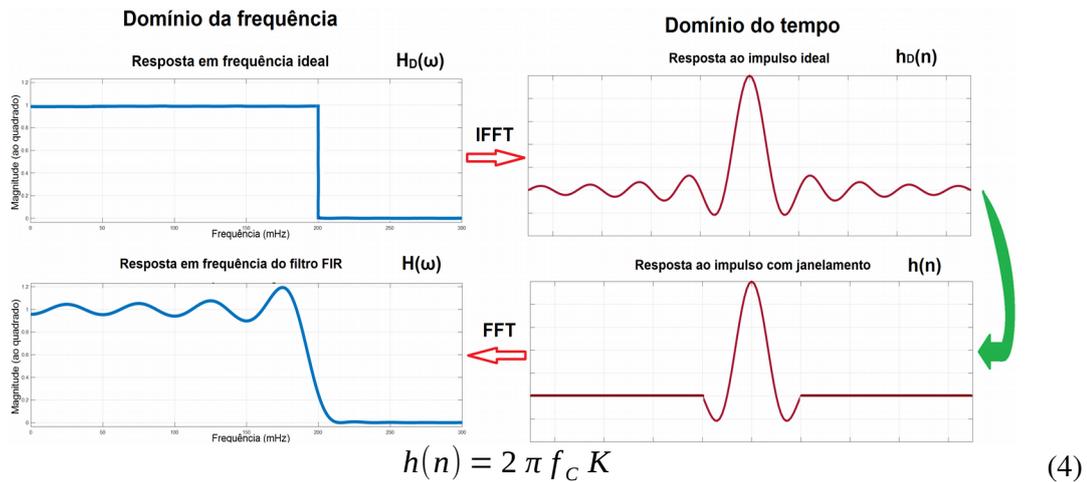


Figura 4. Respostas ideais (em cima) e após o janelamento (embaixo).

No projeto do filtro para o F-OFDM, avalia-se a resposta ao impulso do filtro $h_D(n)$ desejada a partir da resposta em frequência $H_D(\omega)$ desejada e aplica-se a janela apropriada. A resposta em frequência do filtro $H_D(\omega)$ e a correspondente resposta ao impulso $h_D(n)$ estão relacionadas pela transformada inversa de Fourier. O subscrito D é utilizado para distinguir entre a resposta ao impulso ideal e a resposta obtida na prática, onde $h_D(n)$ é definido por:

$$h_D(n) = \int_{-\pi}^{\pi} H_D(\omega) e^{j\omega n} d\omega \quad (5)$$

O filtro passa-baixa padrão é aquele com uma resposta de frequência retangular, isto é, cuja resposta ao impulso é a função sinc, com uma largura de banda apropriada. Este filtro é ideal no sentido de que não causa distorção na banda passante, enquanto fornece total rejeição fora de banda. Para implementação prática, a função sinc é truncada suavemente (*soft-truncated*) com diferentes funções de janela [Abdoli, Jia e Ma 2015]. Diferentes funções para a janela $w(n)$ são apresentadas e discutidas na próxima seção.

4. Funções para a janela do filtro FIR

Existem diferentes janelas utilizadas em filtros FIR, desde a mais simples (janela retangular) até as mais complexas (janela Kaiser, por exemplo). O principal objetivo dessas janelas é truncar suavemente a resposta ao impulso do filtro para gerar um filtro de comprimento fixo com amplitude máxima dentro da largura de banda de interesse do filtro e amplitude mínima fora dela. A escolha apropriada da janela tenta conciliar os melhores resultados de ISI no domínio do tempo e de ICI (*Inter Carrier Interference*) no domínio da frequência,

aumentado assim o desempenho do F-OFDM em relação ao OFDM. Nesta seção são detalhadas algumas janelas frequentemente utilizadas em filtros FIR.

4.1. Retangular

A função retangular é a janela mais simples utilizada por um filtro FIR, ela é definida pela seguinte equação:

$$w(n) = 1, \quad 0 \leq n \leq M \quad (6)$$

onde M é o comprimento do filtro (ordem do filtro) e n assume $M+1$ valores no intervalo $[0, M]$.

4.2. Hamming

A janela de Hamming é projetada para minimizar os picos dos lóbulos laterais. A janela de Hamming é calculada a partir da seguinte equação:

$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{M}\right), \quad 0 \leq n \leq M \quad (7)$$

4.3. Hanning (Cosseno levantado)

A janela de Hanning também é conhecida como janela Hann, ela é calculada a partir da seguinte equação:

$$w(n) = 0,5 - 0,5 \cos\left(\frac{2\pi n}{M}\right), \quad 0 \leq n \leq M \quad (8)$$

A janela de Hanning é um caso especial da janela cosseno levantado, quando $\alpha = 1$. A janela cosseno levantado é calculada por meio de:

$$w(n) = 0,5 - 0,5 \cos\left[\left(\frac{2\pi n}{\alpha}\right) \left(\frac{1}{M - \frac{(1-\alpha)}{2}}\right)\right], \quad 0 \leq n \leq M \quad (9)$$

4.4. Blackman

As janelas de Blackman têm lóbulos centrais ligeiramente mais largos e menos vazamento de banda lateral do que as janelas de Hamming e Hanning de comprimentos equivalentes. A janela Blackman é definida por:

$$w(n) = 0,42 - 0,5 \cos\left(\frac{2\pi n}{M-1}\right) + 0,08 \cos\left(\frac{4\pi n}{M-1}\right), \quad 0 \leq n \leq M-1 \quad (10)$$

onde M é igual a $N/2$ para N par e igual a $(N+1)/2$ para N ímpar.

4.5. Nuttall

A janela de Nuttall é formada por uma janela simétrica de Blackman-Harris de 4 termos. O

cálculo dos coeficientes para esta janela difere do cálculo dos coeficientes de janela de Blackman-Harris e produz lóbulos secundários ligeiramente menores. A janela de Nuttall é obtida a partir da seguinte equação:

$$w(n) = a_0 - a_1 \cos\left(\frac{2\pi n}{M-1}\right) + a_2 \cos\left(\frac{4\pi n}{M-1}\right) - a_3 \cos\left(\frac{6\pi n}{M-1}\right), \quad 0 \leq n \leq M-1 \quad (11)$$

onde os valores dos coeficientes são: $a_0 = 0,3635819$; $a_1 = 0,4891775$; $a_2 = 0,1365995$; $a_3 = 0,0106411$.

4.6. Flat-top (topo plano)

Janelas de topo planas (*Flat-top*) têm baixo *ripple* na faixa de banda passante (<0,01 dB). Sua largura de banda é aproximadamente 2,5 vezes maior que a de uma janela Hann. A função da janela *Flat-top* é semelhante à de Nuttall, porém com um termo a mais:

$$w(n) = a_0 - a_1 \cos\left(\frac{2\pi n}{M-1}\right) + a_2 \cos\left(\frac{4\pi n}{M-1}\right) + a_3 \cos\left(\frac{6\pi n}{M-1}\right) + a_4 \cos\left(\frac{8\pi n}{M-1}\right), \quad 0 \leq n \leq M-1 \quad (12)$$

onde os valores dos coeficientes são: $a_0 = 0,21557895$; $a_1 = 0,41663158$; $a_2 = 0,277263158$; $a_3 = 0,083578947$; $a_4 = 0,006947368$.

4.7. Kaiser

A janela do Kaiser é definida pela seguinte equação:

$$w(n) = \frac{I_0 \left[\beta \sqrt{1 - \left(\frac{n - M/2}{M/2} \right)^2} \right]}{I_0 \beta}, \quad 0 \leq n \leq M \quad (13)$$

onde β é o parâmetro da janela de Kaiser que afeta a atenuação dos lóbulos laterais da transformada de Fourier da janela, um valor comum para β é 0,5 e I_0 é a função de Bessel modificada de primeiro tipo e de ordem zero.

4.8. Patil

O objetivo da janela proposta por Patil é modificar a janela de Hamming para reduzir seu pico máximo de lóbulo secundário, mantendo inalterada a largura do lóbulo principal. A janela de Patil é implementada através da seguinte equação:

$$w(n) = a_0 - a_1 \cos\left(\frac{6\pi n}{M}\right) - a_2 \cos\left(\frac{6\pi n}{M}\right), \quad 0 \leq n \leq M \quad (14)$$

para ordens suficientemente grandes, os seguintes valores podem ser utilizados para os coeficientes:

$$w(n) = 0,536 - 0,46 \cos\left(\frac{2\pi n}{M}\right) - 0,003 \cos\left(\frac{6\pi n}{M}\right), \quad 0 \leq n \leq M \quad (15)$$

Assim como a janela de Hamming, a resposta em frequência da janela de Patil é degradada para ordens baixas. Portanto, dependendo da ordem da janela, os parâmetros acima são modificados para manter a eficiência, isso mostra a dependência de a_0 e a_1 em relação à M . Tais coeficientes são compostos de uma função monotônica e um termo DC [Patil 2015]. Estes parâmetros são regidos conforme as seguintes equações:

$$a_0 = 0,537 - \frac{0,3}{M+15}; \quad a_1 = 0,46 + \frac{0,25}{M+15}; \quad a_3 = 1 - a_0 - a_1 \quad (16)$$

5. Análise comparativa das janelas

Na Figura 5 são mostradas as respostas tempo-magnitude e na Figura 6 as respostas frequência-magnitude de todas as funções de janela descritas na seção 4.

As funções de janela mais interessantes são aquelas que conseguem conciliar maior largura de lóbulo principal com um decaimento mais suave da função, minimizando assim a altura máxima dos lóbulos secundários após a aplicação do filtro e consequentemente reduzindo a ICI [Gandhi 2013].

A Figura 5 apresenta as curvas de várias janelas no domínio do tempo, onde é possível observar comportamentos bem distintos, como a janela retangular, que ao ser utilizada, fornecerá um truncamento abrupto para o filtro e consequente resposta em frequência com muito *ripple*. Por outro lado, a janela *Flat-Top* é a mais estreita no domínio do tempo e por isso, ao ser utilizada no filtro, resultará em uma largura de banda passante mais susceptível à ICI, mas com *ripple* praticamente nulo. Outras janelas apresentam comportamentos muito similares no domínio do tempo, é o caso das janelas de Patil e Hamming.

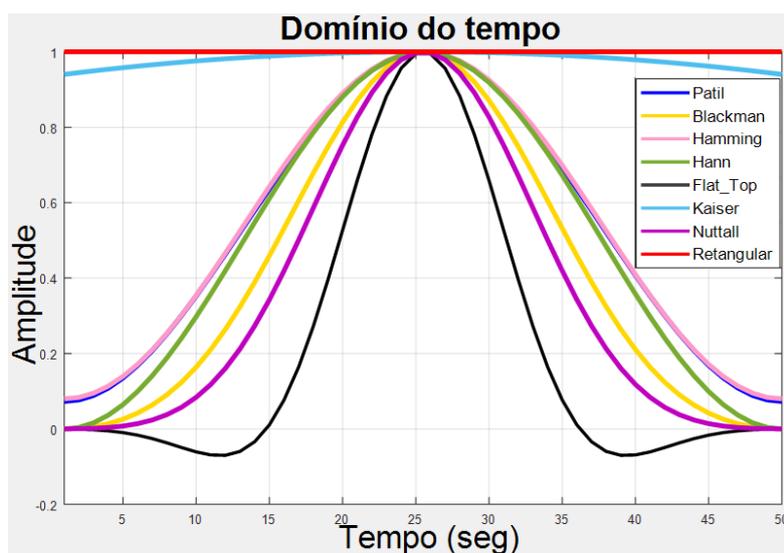


Figura 5. Resposta tempo-magnitude de diferentes janelas

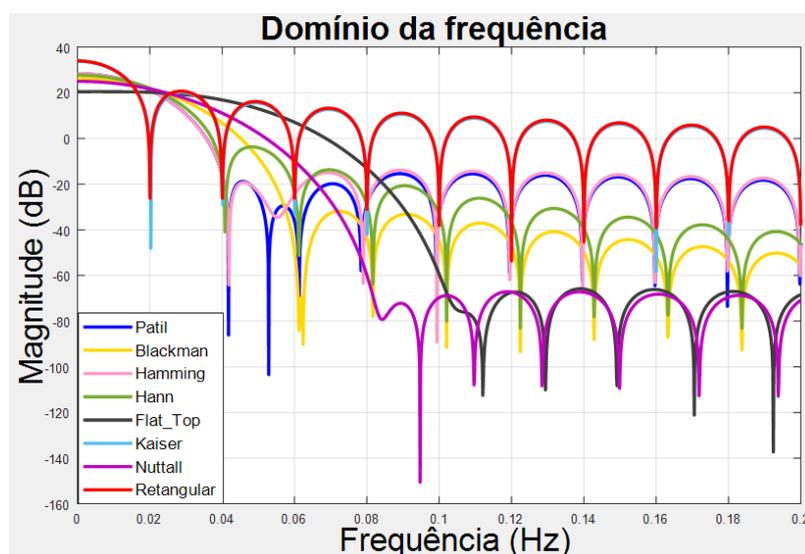


Figura 6. Resposta frequência-magnitude de diferentes janelas

No domínio da frequência serão analisadas com maior atenção as janelas com decaimento mais rápido e com maior atenuação de frequência de parada (maior distância entre pico e início do primeiro lóbulo lateral, apesar de isso geralmente implicar em menor *roll-off*) e cujos lóbulos secundários possuem menor altura de pico em relação ao principal (atenuação alta). As janelas que melhor preenchem estes critérios supracitados são: Blackman, Hann, Hamming e Patil.

Na Figura 5 observa-se a resposta tempo-magnitude e na Figura 6 a resposta frequência-magnitude das quatro janelas supracitadas. No domínio do tempo é evidenciado que as janelas Hamming e Patil possuem formatos de onda praticamente iguais, todavia no domínio da frequência (Figura 6) é possível perceber que Patil possui lóbulos secundários com menor magnitude. Enquanto Hamming possui atenuação da banda de parada (*stopband attenuation*) de aproximadamente -64dB, a atenuação da banda de parada de Patil é de aproximadamente -110dB (Figura 6), além de ter pico máximo de lóbulos secundários cerca de 3dB menor que o observado em Hamming. Estas duas janelas possuem as maiores larguras de lóbulo principal.

Apesar de, dentre estas 4 janelas analisadas, Blackman possuir largura de lóbulo principal no tempo mais estreito, esta janela possui aproximadamente a mesma atenuação de banda de parada que Patil (cerca de -110dB). Conforme observado na literatura, algumas características mais podem ser destacadas: Blackman tem um *ripple* de banda de passagem de apenas 0,02%, enquanto Hamming possui tipicamente 0,2%; Hamming e Patil possuem fator de *roll-off* cerca de 20% mais rápido que Blackman [Smith 1999].

Portanto, a partir das observações supracitadas, percebe-se que a janela Patil possui vantagens em ambos os domínios, destacando-se em relação às demais janelas analisadas neste artigo. A janela Blackman também possui qualidades interessantes e, junto a Patil, deve ser preferida em relação às demais funções de janelamento, mesmo aquelas com maior

largura de lóbulo principal no domínio do tempo, já que o *roll-off* lento de Blackman é mais fácil de lidar do que a fraca atenuação de banda de parada das demais janelas.

Com o intuito de melhorar a eficiência do filtro FIR para o F-OFDM, além da análise da janela adequada a ser utilizada no filtro FIR, este artigo propõe que a modificação na função *kernel* do filtro (protótipo do filtro) pode melhorar a resposta do filtro com o objetivo de redução da BER na transmissão de dados.

O filtro FIR do F-OFDM utiliza a função sinc como função base para o filtro. Essa função pode ser substituída por outros formatos de pulso, com potencial tanto no combate à ISI quanto no combate à interferência entre portadoras (*Inter-carrier Interference* – ICI). Neste artigo, propõe-se que melhores filtros para o F-OFDM podem ser obtidos utilizando-se as janelas conciliadas a pulsos formatados. Alguns exemplos de formatos de pulsos presentes na literatura são: cosseno levantado (*Raised Cosine* – RC), cosseno levantado melhorado (*Better Than Raised Cosine* – BTRC) [Tan e Beaulieu 2004], *Sinc Power* (SP) [Mourad 2006] e *Improved Sinc Power* (ISP) [Kumbasar e Kucur 2007; Gandhi e Dalal 2013].

6. Aplicações da Modulação F-OFDM

São diversas as aplicações possíveis para a F-OFDM. É suficiente dizer que a F-OFDM pode ser utilizada como substituta natural da modulação OFDM. A F-OFDM tem sido estudado em aplicações de sistemas que utilizam rádios cognitivos [Bala, Li e Yang, 2014], mas é como tecnologia de modulação candidata a ser utilizada no 5G que se encontram os maiores esforços de desenvolvimento da modulação F-OFDM, para os quais estão disponíveis trabalhos promissores [Abdoli, Jia e Ma 2015], [Zhao *et al.*, 2016].

Neste artigo, foi simulada uma aplicação prática da F-OFDM com a janela proposta por Patil (2015) combinada à formatação de pulso ISP – que segundo Gandhi, Ghupta e Dalal (2013) é a melhor modelagem de pulso para uma ampla gama de aplicações – incluindo o processamento de imagem. Essa conclusão também foi obtida pelos autores deste artigo, após diversas simulações com diversas janelas e diversas formas de pulsos.

A função ISP é calculada conforme a seguir:

$$ISP(\alpha) = e^{-\alpha(\omega)^2} \text{sinc}^2(\omega) \quad (17)$$

onde α é o fator de *roll-off*.

A imagem processada possui formato .jpg com dimensões 200x181 pixels e foram utilizados os seguintes parâmetros para o processamento: 1024 pontos para a FFT; 50 blocos de recurso; 12 subportadoras por bloco de recurso; período cíclico de comprimento 72; filtro de comprimento 513 e modulação QPSK.

Antes da utilização da F-OFDM realizou-se o processamento da imagem via OFDM, sem filtros, para se ter um valor base da taxa de erros de bit (BER – *Bit Error Rate*). Com os parâmetros supracitados e sem filtros o resultado obtido foi BER = 8%. Ao ser utilizado o filtro de sinal (pulso formatado e truncamento) a taxa caiu para 0,12%, isto equivale a uma redução percentual de 98,5% da BER, ou seja, uma melhora de 98,5% na

qualidade da imagem processada. Estes resultados corroboram que a F-OFDM pode ser superior à OFDM convencional. Resultados similares foram obtidos para outras imagens e para outros parâmetros do cenário de simulação.

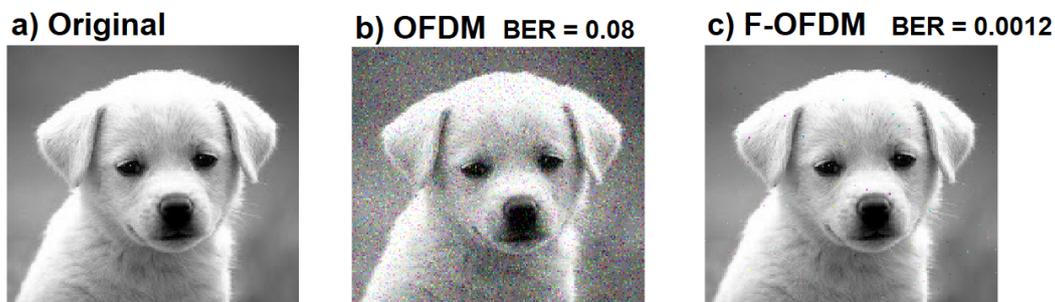


Figura 7. Processamento de imagem com F-OFDM

7. Conclusões

Neste artigo, foram feitas análises comparativas entre diversas funções de janela comumente utilizadas na literatura de processamento de sinais para o projeto de filtros FIR.

Com o objetivo de reduzir ISI no domínio do tempo, ICI no domínio da frequência e o *ripple* da banda passante, as funções das janelas foram comparadas e as que mais se destacaram foram: Hamming, Patil, Blackman e Hanning.

Ao escolher a janela, com base nas negociações (*trade-offs*) de parâmetros desejados, várias compensações podem ser gerenciadas de modo a maximizar a qualidade do projeto do filtro em uma determinada aplicação.

O método de janela para o projeto de filtro digital é bem conhecido na literatura [Patil 2015] e foi utilizado com obtenção de resultados interessantes neste trabalho. Escolhendo uma das janelas que se destacou nas análises, neste caso Patil, foi proposta uma implementação de um filtro FIR utilizando a janela de Patil aliada à formatação do pulso da função utilizada como *kernel* para o protótipo do filtro, sendo essa a maior contribuição deste trabalho. Para tanto, foi utilizada a função ISP (*Improved Sinc Power*). Por meio do projeto de filtro proposto, foram conduzidas simulações de transmissão de imagens comparando os resultados de BER obtidos para o OFDM e para o F-OFDM. Observou-se que os resultados foram interessantes, com redução de BER no processamento da imagem em 98,5% em relação a transmissão utilizando OFDM. Isto corrobora que o filtro utilizado é eficiente para redução da BER no processo de transmissão de dados, tornando a tecnologia F-OFDM, com projeto de filtro apropriado, candidata a ser utilizada no 5G.

Como trabalho futuro, pretende-se comparar os resultados obtidos com outras modulações candidatas a serem utilizadas no 5G, como por exemplo: FBMC e UFMC. Além disso, deseja-se conduzir simulações em um ambiente que reúna maior número de características de uma futura rede 5G.

Referências

- Abdoli, J., Jia, M. and Ma, J. (2015) "Filtered OFDM: A new waveform for future wireless systems", In: 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), p. 66-70, Stockholm, Sweden.
- Bala, E., Li, J. and Yang, R. (2014) "Pulse shaped OFDM for Cognitive Radio", In: InterDigital Communications, Melville, USA.
- Gandhi, D., Gupta, S. S. and Dalal, S. U. (2013) "Implementation of Pulse Shaping Techniques in OFDM System", In: International Journal of Computer Applications, vol. 68, no.10, India.
- Guerra, D. W. M. (2016) "FBMC × OFDM: Análise de Desempenho e Implementação de Sistemas de Comunicação sem Fio Multiportadoras de Quarta e Quinta Geração", In: Universidade Estadual de Londrina, Brasil.
- Heydt, G. T., Fjeld, P. S., Liu, C. C., Pierce, D., Tu, L. Tu and Hensley G. (1999) "Applications of the windowed FFT to electric power quality assessment", In: IEEE Transactions on Power Delivery, vol. 14, no. 4, p. 1411-1416.
- Harris, F. J. (1978) "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", In: Proceedings of the IEEE, vol. 66, no. 1.
- Kumbasar, V. and Kucur, O. (2007) "ICI reduction in OFDM systems by using improved sinc power pulse", In: Digital Signal Processing 17, p. 997– 1006.
- Mourad, H. M. (2006) "Reducing ICI in OFDM systems using a proposed pulse shape", In: Wireless Personal Communications 40, p. 41–48.
- Patil, A. K. (2015) "A new window function for filter design and spectral analysis", In: International journal of advanced research in science and engineering vol. no.4, Issue 09.
- Smith, S. W. (1999) "The Scientist and Engineer's Guide to Digital Signal Processing", USA, California Technical Publishing, 2nd edition, p. 285–296.
- Tan, P. and Beaulieu, N. C. (2004) "Reduced ICI in OFDM systems using the better than raised-cosine pulse", In: IEEE Communications Letters, vol. 8, no. 3, p. 135–137.
- Zhang, X., Jia, M., Chen, L., Ma, J. and Qiu, J. (2015) "Filtered-OFDM – Enabler for Flexible Waveform in the 5th Generation Cellular Networks", In: 2015 IEEE Global Communications Conference (GLOBECOM). San Diego, USA.
- Zhao, Z., Schellmann, M., Gong, X., Wang, Q., Bohnke, R. and Guo, Y. (2016) "Pulse Shaped OFDM for 5G Systems", In: German Research Center, Huawei Technologies, Munich, Germany.

Análise de Sentimentos de Conteúdos Textuais de Redes Sociais Por Meio de Modelos de Compressão de Dados

Jurandir J. D. Silva¹, Rogerio Salvini¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG) – Goiânia, GO – Brazil

{jurandirsilva, rogeriosalvini}@inf.ufg.br

***Resumo.** O conjunto das técnicas que são exploradas no tratamento de opiniões é abordado pela área de pesquisa da Análise de Sentimentos (AS), que combina conceitos de diversas áreas como, Inteligência Artificial, Reconhecimento de Padrões, Análise Textual etc. Por outro lado, técnicas baseadas em compressão de dados podem ser úteis para achar padrões em dados não estruturados, como textos com opiniões encontrados na internet. Neste trabalho foi testado o método DAMICORE, que utiliza estas técnicas, para verificar sua eficácia no problema de AS. Os resultados ficaram aquém dos esperados, entretanto abrindo novas oportunidades de pesquisa na área.*

1. Introdução

Atualmente, sistemas web como blogs, redes sociais, sites de compras etc., permitem cada vez mais a interação dos usuários, e a utilização de textos para comunicação e/ou expressão de opinião tornou-se relativamente comum. De forma natural, o conteúdo destes textos se tornou objeto de atenção para os maiores interessados no “ponto de vista” de seu público: as empresas. De acordo com Giachanou [Giachanou and Crestani 2016], a informação gerada pelos usuários da internet é uma boa fonte de opiniões e pode ser valiosa para uma variedade de aplicações que requerem uma compreensão da opinião pública sobre um conceito. Um exemplo típico que ilustra a importância da opinião pública refere-se às empresas que podem capturar as opiniões dos clientes sobre seus produtos ou seus competidores. Por outro lado, a organização desses dados para se inferir correlações em geral não tem sido um trabalho trivial. A dificuldade em se automatizar esse tipo de processamento ocorre, principalmente, porque a quantidade de formas diferentes de se expressar uma opinião, ou mesmo um sentimento, é relativamente grande.

O conjunto das técnicas que são exploradas no tratamento de opiniões é abordado pela área de pesquisa de Análise de Sentimento, que combina conceitos de diversas áreas como, Inteligência Artificial, Reconhecimento de Padrões, Aprendizado de Máquina, Processamento de Linguagem Natural e Análise Textual; objetivando analisar fragmentos textuais e determinar a atitude, emoção, opinião, avaliação ou sentimento do usuário com relação a algum tópico ou entidade [B. Liu 2012]. Dentre as técnicas de

aprendizado de máquina que vêm sendo empregadas na Análise de Sentimento estão: *Support Vector Machines* (SVM), *Naïve Bayes* (NB) e *Multinomial Naïve Bayes* (MNB) [Giachanou and Crestani 2016]. Além dessas técnicas, algoritmos evolutivos (*Evolutionary Algorithms*) [Dufourq and Bassett 2017], e o algoritmo de propagação (*Propagation Algorithm*) [Che et al. 2015], ambos utilizando compressão de dados, já foram empregados para análise de sentimento, através de uma abordagem supervisionada. No entanto, abordagens não supervisionadas, como a Análise de Agrupamentos (ou *Clustering*), ainda não estão muito exploradas nessa área.

A Análise de Agrupamentos é o conjunto de técnicas computacionais cujo propósito consiste em separar objetos em grupos, baseando-se nas características que estes objetos possuem. O agrupamento torna-se um mecanismo útil quando se tem a necessidade de analisar dados de diferentes tipos e tamanhos, assim como é possível extrair características não conhecidas anteriormente de um determinado grupo.

Modelos para análise de agrupamentos utilizando compressão de dados já se mostraram eficazes quando utilizados no agrupamento de textos [Sanches et al. 2011]. Neste contexto, os compressores geram arquivos compactados que são usados como padrões para determinar grupos semelhantes, formando *clusters* de objetos textuais.

O objetivo deste trabalho é aplicar e verificar se estas técnicas também podem ser eficazes no problema de análise de sentimento agrupando mensagens provenientes de websites como Twitter (*tweets*) e Amazon (comentários sobre produtos) em função dos sentimentos positivo ou negativo expressos através das mensagens dos usuários. Para isto, será utilizado um método chamado DAMICORE (*DAta MIning of COde REpositories*) que será apresentado em maiores detalhes na Seção 2 a seguir.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta o método DAMICORE para geração de agrupamentos baseado em compressão de dados; a Seção 3 aborda a metodologia empregada nos experimentos deste trabalho; a Seção 4 mostra os resultados alcançados; a Seção 5 faz uma discussão sobre estes resultados e levanta algumas questões sobre sua utilização para o problema abordado; e finalmente a Seção 6 faz uma conclusão deste estudo e apresenta alguns trabalhos futuros.

2. DAMICORE (*DAta MIning of COde REpositories*)

O DAMICORE¹ é um método proposto por Soares [Soares and Delbem] para análise de agrupamentos (*Clustering*) que se baseia em procurar similaridades nos dados utilizando compactação de dados, que são técnicas utilizadas para diminuir o volume da informação [Sanches et al. 2011].

A geração dos agrupamentos por meio do DAMICORE segue a combinação dos

¹Disponível em: <https://github.com/sidgleyandrade/damicore-python>

seguintes algoritmos: *Normalized Compression Distance* (NCD), proveniente da Teoria da Informação [Cilibrasi and Vitányi 2005], o *Neighbor Joining* (NJ), proveniente da Filogenética [Felsenstein J. 2003], e o *Fast Algorithm* (FA), proveniente de Redes Complexas [Newman 2004].

A NCD (do inglês, Distância por Compressão Normalizada) é uma métrica que possibilita quantificar a semelhança entre diferentes dados a partir da quantidade de informação do arquivo compactado (tamanho do arquivo). Esta abordagem não requer qualquer conhecimento específico do domínio de aplicação. De acordo com Cilibrasi [Cilibrasi and Vitányi 2005], a NCD é uma métrica universal e robusta que tem sido aplicada com sucesso em áreas como a genética, literatura, música e astronomia. A NCD é dada pela seguinte equação:

$$NCD(X, Y) = \frac{C(XY) - \min\{C(X), C(Y)\}}{\max\{C(x), C(y)\}} \quad (1)$$

onde, $C(XY)$ é o tamanho dos arquivos X e Y concatenados e, em seguida, compactados; $C(X)$ e $C(Y)$ são os tamanhos dos arquivos X e Y compactados, respectivamente.

A NCD é usada no DAMICORE para gerar matrizes de distância para um conjunto de variáveis, possibilitando verificar níveis de semelhança entre elas e propriedades em comum desses códigos que podem ser explorados.

O algoritmo NJ possibilita agrupar arquivos, a partir de sua semelhança, formando clusters de arquivos que compartilham alguma característica em comum. O NJ constrói uma árvore a partir de uma matriz de distâncias evolutivas, adaptando o critério de evolução mínima, em que o algoritmo tenta minimizar a soma dos tamanhos de todos os nós da árvore. A ideia central da técnica é identificar pares de objetos mais próximos.

O *Fast Algorithm* (FA), proposto em [Newman 2004], é um algoritmo de Detecção de Comunidades, cujo objetivo é encontrar a estrutura que maximize a fração de arestas que conectam vértices de uma mesma comunidade [Crocomo 2012].

O DAMICORE funciona seguindo uma sequência de etapas abaixo:

- Etapa 1: O conjunto de objetos é reorganizado de forma a compor uma amostra;
- Etapa 2: É gerada uma matriz que relaciona cada objeto com os demais através da NCD. Nesta etapa, a NCD necessita de um compressor, que é indicado pelo usuário;
- Etapa 3: É gerada uma árvore filogenética usando a heurística NJ a partir da matriz de distâncias construída na Etapa 2. O DAMICORE faz uso da técnica de árvores aditivas, sem raiz, em que a distância entre dois objetos é dada pela soma dos

ramos que os unem;

- Etapa 4: O DAMICORE utiliza a saída do algoritmo NJ, que será um arquivo no formato Newick [Felsenstein J. 2003], para formar uma Matriz de Adyacências.
- Etapa 5: É determinado um Particionamento Final a partir da Matriz de Adyacências da Etapa 4. Para isso, é utilizado o algoritmo FA;

3. METODOLOGIA

3.1. Bases de Dados

As bases de dados utilizadas nos resultados expostos neste trabalho foram as bases *Sanders*² e a *Sentiment Labelled Sentences Data Set*³, ambas compostas de dados reais e disponíveis publicamente.

3.1.1. Sanders

A base *Sanders* é composta por mensagens provenientes do Twitter (*tweets*), e foi escolhida por já ser utilizada em diversos artigos, o que possibilita uma melhor comparação com os resultados aqui obtidos.

Esta é formada por mensagens obtidas de forma manual por especialistas, e foram coletadas a partir de quatro termos de buscas: *@apple*, *#google*, *#microsoft* e *#twitter*. A Tabela 1 apresenta o número de *tweets* de cada tópico e classe:

Tabela 1. Especificação da base de dados *Sanders*

Tópico	Classe Positiva	Classe Neutra	Classe Negativa	Termo de busca no Twitter
Apple	191	581	377	<i>@apple</i>
Google	218	604	61	<i>#google</i>
Microsoft	93	671	138	<i>#microsoft</i>
Twitter	68	647	78	<i>#twitter</i>

Por se tratar de um estudo que tem por objetivo verificar se o DAMICORE pode ser eficaz no problema de análise de sentimento, e também visando uma melhor avaliação dos resultados, os exemplos da classe neutra da *Sanders* foram ignorados, permanecendo apenas as classes positiva e negativa.

3.1.2. Sentiment Labelled Sentences Data Set (SLSDS)

A base *Sentiment Labelled Sentences Data Set* é formada por comentários/avaliações de usuário nos *websites*: *amazon.com*, *imdb.com* e *yelp.com*; sobre produtos, filmes e restaurantes, respectivamente [Kotzias et al. 2015].

²Disponível em <http://www.sananalytics.com>

³Disponível em <https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences>

Algumas diferenças desta base com relação à base *Sanders* são:

- Os exemplos não são sobre termos/temas específicos;
- Os exemplos foram selecionados de outras bases maiores, de forma que esta se manteve balanceada (mesma quantidade de exemplos de cada classe);
- Das bases maiores, foram selecionados apenas as sentenças com conotação clara de “positivo” (indicados na base por 1) ou “negativo” (indicados na base por 0);

A Tabela 2 a seguir apresenta o número de sentenças de cada *website* e classe:

Tabela 2. Especificação da base de dados *Sentiment Labelled Sentences Data Set*

Website	Classe Positiva	Classe Negativa
<i>amazon.com</i>	500	500
<i>imdb.com</i>	500	500
<i>yelp.com</i>	500	500

3.2. Pré-processamento

Para a realização deste trabalho foi desenvolvido um pacote de *scripts*, chamado *py-DataPreparation*⁴, desenvolvido na linguagem Python 3, que é responsável por toda a preparação e o pré-processamento dos dados.

O pré-processamento foi realizado em etapas, onde a cada experimento realizado, quando feito o pré-processamento, foi adicionado pelo menos um item de pré-processamento dos dados. Os itens de pré-processamento foram os seguintes:

1. Passagem de todas as letras maiúsculas para minúsculas;
2. Remoção de todos os símbolos não alfanuméricos (que não são letras ou números), mantendo os espaços;
3. Remoção das *stopwords*, utilizando o conjunto de palavras presentes na *corpora stopwords* da biblioteca NLTK (*Natural Language Toolkit*);
4. Substituição de cada palavra, quando possível, por um sinônimo “mais relevante”. Por exemplo, caso as palavras “*happy*” e “*glad*” apareçam nas mensagens, mas a palavra “*happy*” aparece com mais frequência, todas as ocorrências da palavra “*glad*” são substituídas pela palavra “*happy*”;

É importante frisar que o conteúdo das sentenças após o pré-processamento é composto apenas pelas mensagens propriamente ditas, ou seja, a classe (ou algo que a indique) não é inserida nos arquivos finais, com exceção do nome apenas para identificação.

⁴Disponível em: <https://github.com/jurandirjdsilva/pyDataPreparation>

3.3. Compressor de Dados

Como abordado por Cilibrasi em [Cilibrasi and Vitányi 2005], diversos compressores de dados podem ser utilizados no cálculo da NCD. Neste trabalho, os experimentos foram realizados utilizando o compressor GZIP⁵, que utiliza o algoritmo de compressão *Lempel–Ziv*. Em [Cilibrasi and Vitányi 2005], o GZIP foi escolhido para comprimir sequências de textos cujo tamanho não excede a janela deslizante do compressor (32 *kilobytes*), que é justamente o caso dos experimentos deste trabalho, em que os maiores arquivos têm tamanhos que se aproximam de 200 *bytes*.

4. RESULTADOS

Diversos experimentos, em diferentes situações, foram executados. No entanto, para fim de uma melhor discussão posteriormente, os resultados aqui relatados correspondem a apenas os resultados mais relevantes alcançados. Por resultados mais relevantes entende-se os que tiveram mais impacto, seja sob os resultados esperados, seja na comparação com outros trabalhos.

Os resultados apresentados nesta seção representam as filogenias resultantes da execução do algoritmo NJ. Tais filogenias são exibidas neste trabalho na forma de Cladogramas Diagonais, cuja interpretação é como segue: O NJ identifica pares de objetos mais próximos, chamados de nós vizinhos, que são conectados por um nó mais interno formando uma subárvore bifurcada [Valdivia 2007]. A relação de vizinhança é ilustrada nos gráficos aqui apresentados na forma de linhas que conectam as folhas (os objetos dos experimentos) formando nós, e linhas que unem esses nós a outras folhas ou outros nós; de maneira que quanto menos nós houver entre dois objetos mais próximos (semelhantes) estes são.

O cladograma diagonal é semelhante a um dendrograma, bastante utilizado na análise de agrupamentos, estando a diferença na representação das distâncias, onde no dendrograma estas ficam explícitas no tamanho das arestas enquanto que no cladograma as distâncias ficam representadas apenas nos níveis. Devido à quantidade de exemplos nas bases utilizadas, nos cladogramas gerados não é possível ler claramente os rótulos dos exemplos (devido à restrição de espaço). Assim, para facilitar a interpretação dos resultados, optamos por representar os exemplos por diferentes cores conforme a sua classe. Os rótulos dos exemplos foram coloridas manualmente a fim de facilitar a identificação dos subgrupos gerados pelo algoritmo. As correspondências entre cor e classe, no Experimento 1, encontra-se na legenda da figura. Já nos Experimentos 2 ao 8, as sentenças que expressam sentimento positivo estão destacados com a cor verde e negativo com a cor vermelha.

⁵Disponível em: <https://www.gnu.org/software/gzip/>

A fim de possibilitar uma melhor avaliação dos resultados, o primeiro experimento abordado nesta sessão foi realizado utilizando-se a base de dados que acompanha o DAMICORE, e servirá de *baseline* para a comparação com os resultados obtidos nas bases de dados de análise de sentimentos utilizadas neste trabalho.

4.1. Experimento 1: Base de dados de textos em diferentes idiomas

A base de dados utilizada neste experimento é composta por arquivos textuais de cinco idiomas diferentes, sendo eles: alemão, francês, inglês, italiano e português; contendo ao todo 150 exemplos, 30 de cada classe/idioma.

Cada texto é armazenado em um arquivo de forma que as três primeiras letras desses arquivos representam o idioma do texto, da seguinte maneira: “GER” para o idioma alemão; “FRA” para o idioma francês; “ENG” para o idioma inglês; “ITA” para o idioma italiano; e “POR” para o idioma português.

O cladograma diagonal apresentado na Figura 1, obtido a partir do arquivo *newick* que é gerado, resume o resultado do DAMICORE, exibindo os *clusters* na forma de cladogramas. Pode-se notar nesta figura que os agrupamentos dos textos por idioma ficaram perfeitos.

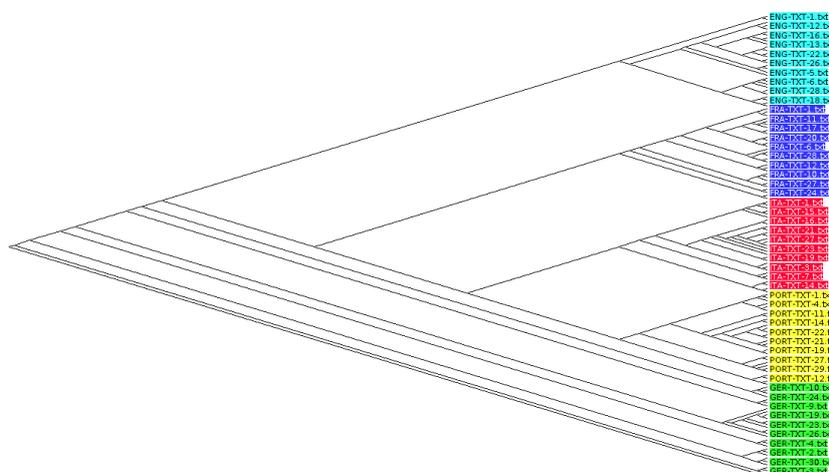


Figura 1. Cladograma diagonal mostrando o resultado do DAMICORE executado sobre a base de dados de textos em diferentes idiomas. Na figura, textos de idioma inglês estão destacados com a cor azul claro, francês na cor azul escuro, italiano na cor vermelho, português na cor amarelo e alemão na cor verde.

4.2. Experimento 2: Base Sanders completa sem pré-processamento

Tabela 3. Descrição do Experimento 2

Base de Dados	#Exemplos	#Pos.	#Neg.	Pré-processamento
Sanders completa	1219	565	654	Sem pré-processamento

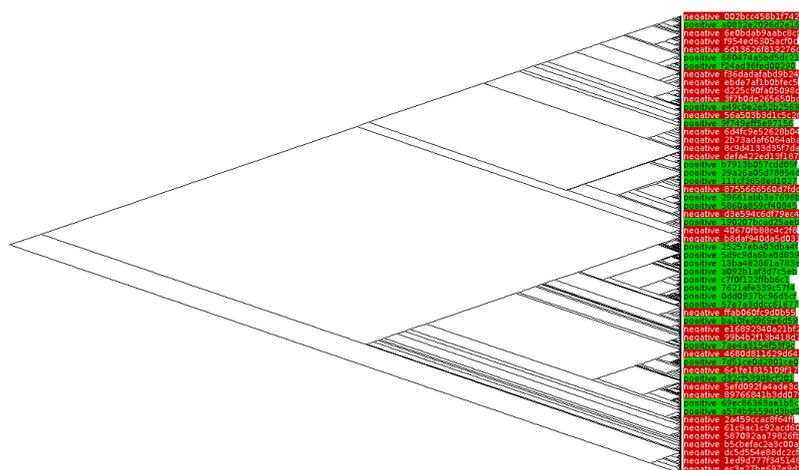


Figura 2. Cladograma diagonal mostrando o resultado do DAMICORE executado sobre a base de dados Sanders completa sem nenhum pré-processamento.

4.3. Experimento 3: Base Sanders completa com remoção de símbolos e stopwords

Tabela 4. Descrição do Experimento 3

Base de Dados	#Exemplos	#Pos.	#Neg.	Pré-processamento
Sanders	1217	563	654	Letras minúsculas; Remoção de símbolos não alfanuméricos; Remoção de <i>stopwords</i>

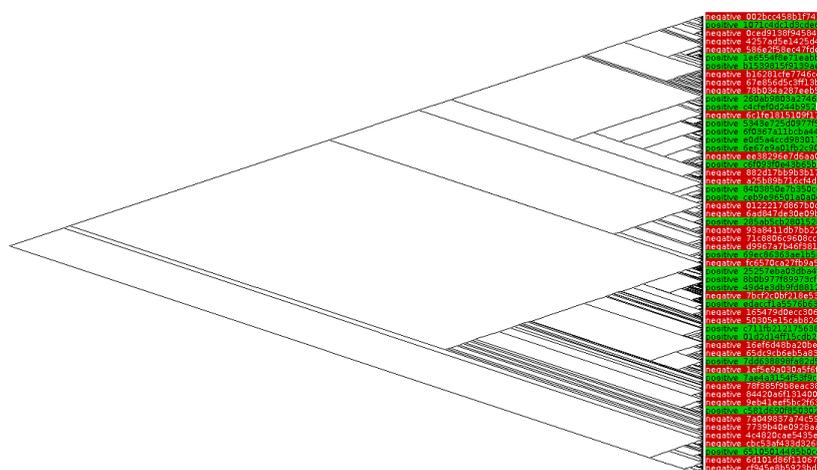


Figura 3. Cladograma diagonal mostrando o resultado do DAMICORE executado sobre a base Sanders completa e pré-processamento descrito na Tabela 4.

4.4. Experimento 4: Base Sanders completa com substituição de sinônimos

Tabela 5. Descrição do Experimento 4

Base de Dados	#Exemplos	#Pos.	#Neg.	Pré-processamento
Sanders	1217	563	654	Letras minúsculas; Remoção de símbolos ; Remoção de <i>stopwords</i> ; Substituição de palavras por sinônimos

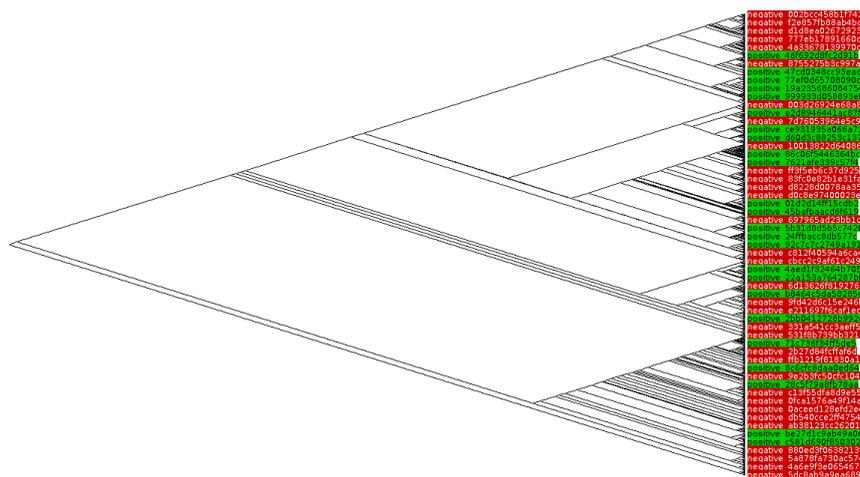


Figura 4. Cladograma diagonal mostrando o resultado do DAMICORE executado sobre a base Sanders completa e pré-processamento descrito na Tabela 5.

4.5. Experimento 5: Subconjunto referente ao termo #apple da base Sanders sem pré-processamento

Tabela 6. Descrição do Experimento 5

Base de Dados	#Exemplos	#Pos.	#Neg.	Pré-processamento
Sanders: apple	568	191	377	Sem pré-processamento

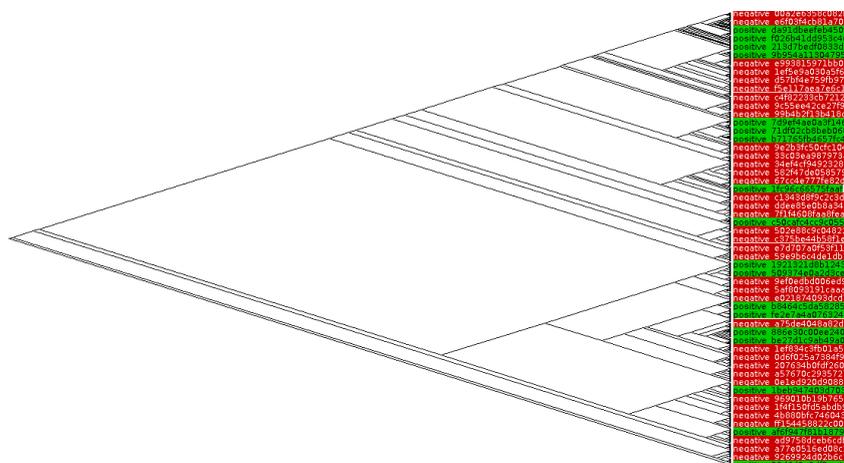


Figura 5. Cladograma diagonal mostrando o resultado do DAMICORE executado sobre o subconjunto da base Sanders referente ao termo #apple e sem pré-processamento.

4.6. Experimento 6: Subconjunto balanceado referente ao termo #apple da base Sanders sem pré-processamento

Tabela 7. Descrição do Experimento 6

Base de Dados	#Exemplos	#Pos.	#Neg.	Pré-processamento
Sanders: apple	380 (selec. aleat.)	190	190	Sem pré-processamento

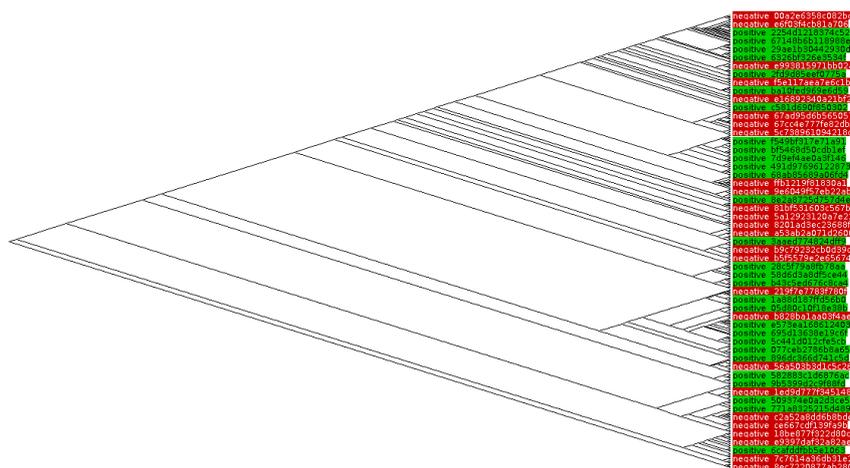


Figura 6. Cladograma diagonal mostrando o resultado do DAMICORE executado sobre o subconjunto da base Sanders referente ao termo #apple e sem pré-processamento.

4.7. Experimento 7: Subconjunto balanceado referente ao termo #apple da base Sanders com remoção de símbolos e stopwords

Tabela 8. Descrição do Experimento 7

Base de Dados	#Exemplos	#Pos.	#Neg.	Pré-processamento
Sanders: apple	380 (selec. aleat.)	190	190	Letras minúsculas; Remoção de símbolos não alfanuméricos; Remoção de <i>stopwords</i>

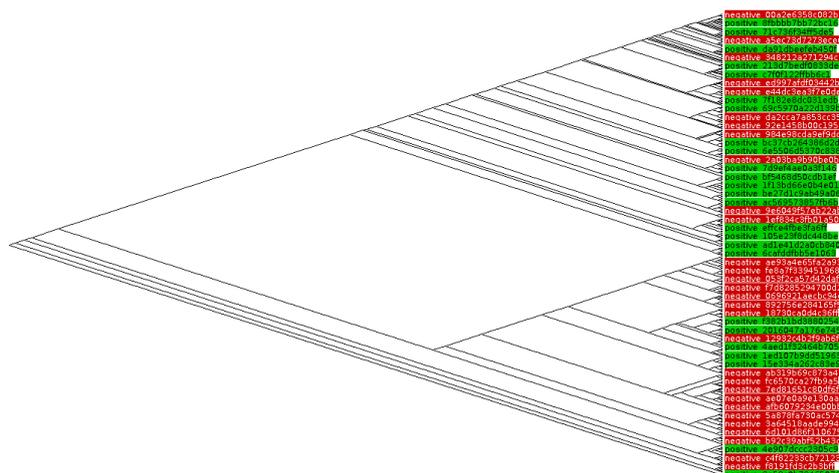


Figura 7. Cladograma diagonal mostrando o resultado do DAMICORE executado sobre o subconjunto da base Sanders referente ao termo #apple e pré-processamento descrito da Tabela 8.

4.8. Experimento 8: Subconjunto balanceado proveniente do *website* Amazon da base *Sentiment Labelled Sentences Data Set* com remoção de símbolos e *stopwords*

Tabela 9. Descrição do Experimento 8

Base de Dados	#Exemplos	#Pos.	#Neg.	Pré-processamento
SLSDS: Amazon	200 (selec. aleat.)	100	100	Letras minúsculas; Remoção de símbolos não alfanuméricos; Remoção de <i>stopwords</i>

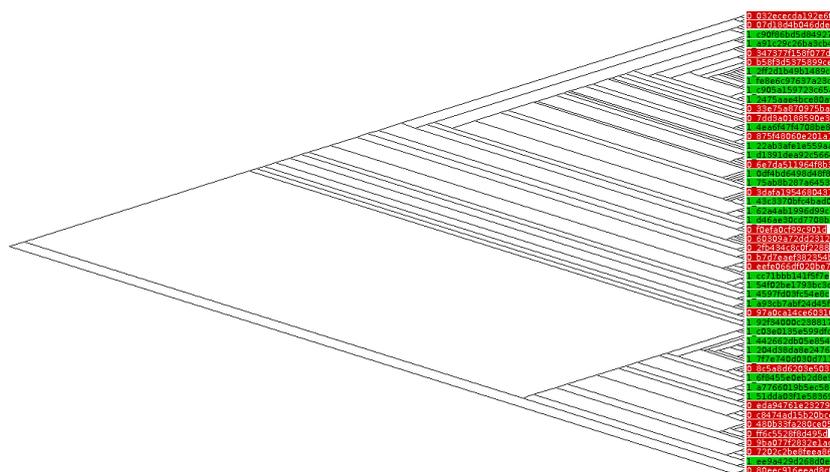


Figura 8. Cladograma diagonal mostrando o resultado do DAMICORE executado sobre o subconjunto da base *Sentiment Labelled Sentences Data Set* provenientes do *website* Amazon e pré-processamento descrito da Tabela 9

5. DISCUSSÃO

O Experimento 1 usou o DAMICORE para achar agrupamentos em uma base de textos em diferentes idiomas. Esta base é obtida junto com o código do DAMICORE e foi utilizada como *baseline* deste trabalho, ou seja, o cladograma do resultado obtido nesta base de dados é utilizado como modelo do que se esperava dos resultados nas bases escolhidas para a análise de sentimentos. Foram realizadas verificações minuciosas, tanto no cladograma em tamanho original, quanto no arquivo com os *clusters* gerado pelo DAMICORE, e constatou-se que todos os *clusters* gerados continham textos de apenas 1 idioma. Ou seja, o DAMICORE se mostrou realmente eficaz na tarefa de agrupamento de textos por idiomas.

Nos experimentos 2, 5 e 6, o DAMICORE foi executado sobre as bases de dados voltadas para análise de sentimentos sem nenhum pré-processamento. Conforme abordado em [Ziegelmayr and Schrader 2012], a classificação de textos baseada em compactação praticamente não requer pré-processamento, pois é capaz de capturar *non-words* (grupos de letras que se parecem com determinadas palavras) e *metawords* (carac-

terísticas que abrangem mais de uma palavra). No entanto, nestes experimentos realizados, o DAMICORE não foi capaz de agrupar os exemplos segundo sua classe, como é possível verificar através dos cladogramas em cada subseção, em que os exemplos das duas classes (destacados pelas cores vermelho e verde) ficam “misturados” entre si. Em um resultado positivo, seria possível observar dois grandes agrupamentos, sendo cada um formado, em sua maioria, por exemplos de uma mesma classe.

Nos experimentos 3, 4, 7 e 8, os dados passaram por pré-processamento a fim auxiliar o método. No entanto, o DAMICORE também não obteve sucesso nesses casos.

No experimento 4, um dos pré-processamentos adotados nas sentenças foi a substituição de cada palavra, quando possível, por um sinônimo mais “relevante” (com maior probabilidade de ocorrência nas demais sentenças). Essa medida foi adotada com o objetivo de aumentar a similaridade entre as sentenças (quanto maior a similaridade, menor o valor da NCD), sem alterar o significado destas. Porém, nem mesmo isso foi suficiente para melhorar a eficiência do DAMICORE.

É interessante notar que, em alguns experimentos, como o 2 e o 7, ocorrem alguns subagrupamento com uma quantidade relevante de exemplos de mesma classe. Diante disso, poderíamos concluir que o DAMICORE funciona parcialmente na análise de sentimentos. No entanto, a ocorrência desses subagrupamentos ocorre de forma desordenada, não sendo possível prever quando e como esses subagrupamentos surgirão.

Como os resultados não foram os esperados, estudos prévios foram realizados e, pelo menos, duas razões para a ineficiência do método foram descobertas:

- De acordo com Cilibrasi [Cilibrasi and Vitányi 2005], certas características dominantes que regem a similaridade são automaticamente descobertas pelo NCD. Além disso, essas características podem não ser explicitamente conhecidas para nós. Isso significa que o NCD pode agrupar dados de acordo com características que não são de interesse do problema em estudo e, desta forma, o resultado final (os *clusters*) não necessariamente estará incorreto.
- A NCD, apresentada em [Cilibrasi and Vitányi 2005], é definida como uma medida de similaridade que expressa uma distância entre dois objetos quaisquer por meio da compressão dos dados. No entanto, alguns testes prévios foram realizados e constatou-se que, em alguns casos, a distância entre dois objetos ($NCD(A, B)$, para $A \neq B$) é diferente ao se inverter a ordem dos objetos ($NCD(B, A)$). Isso pode ser um motivo real da ineficiência da NCD para o agrupamento de objetos, uma vez que, por se tratar de uma medida de similaridade, as distâncias $NCD(A, B)$ e $NCD(B, A)$ tinham que ser exatamente as mesmas. Esse fato coloca em risco a própria validade desta medida como uma métrica de distância.

6. CONCLUSÃO

Devido à eficácia do método DAMICORE no agrupamento de textos por idiomas, no presente trabalho, este mesmo método foi aplicado no problema de análise de sentimentos, para verificar se técnicas baseadas em compressão podem ser também eficazes agrupando mensagens de texto em função do sentimento expresso através destas.

Constatou-se porém, por meio de vários experimentos, e utilizando diversos tipos de pré-processamento dos dados a fim de ajudar o método, que o DAMICORE não foi capaz de formar agrupamentos de mensagens que tenham como característica em comum o sentimento (positivo ou negativo) expresso.

Dentre as razões que podem ter contribuído para este insucesso do DAMICORE está a possibilidade do método estar formando grupos a partir de características que não estão relacionadas às palavras referentes aos sentimentos nas mensagens. Além disso, a métrica utilizada para fazer o agrupamento dos objetos, a NCD, não é simétrica ($NCD(A, B) \neq NCD(B, A)$) o que interfere diretamente no agrupamento dos objetos.

Como trabalhos futuros, serão realizados novos estudos com o objetivo de entender melhor as limitações do DAMICORE, bem como realizar modificações no método a fim de aperfeiçoá-lo. Dentre as modificações possíveis, será considerada a implementação de um novo compressor de dados, além de possíveis modificações na métrica NCD.

Referências

- B. Liu (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Che, W., Zhao, Y., Guo, H., Su, Z., and Liu, T. (2015). Sentence compression for aspect-based sentiment analysis. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 23(12):2111–2124.
- Cilibrasi, R. and Vitányi, P. M. B. (2005). Clustering by compression. *IEEE Transactions on Information Theory*, 5:1523–1545.
- Crocomo, M. K. (2012). *Algoritmo de otimização bayesiano com detecção de comunidades*. PhD thesis, Universidade de São Paulo.
- Dufourq, E. and Bassett, B. A. (2017). Text compression for sentiment analysis via evolutionary algorithms. In *Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech), 2017*, pages 116–121. IEEE.
- Felsenstein J. (2003). *Inferring phylogenies*. Sinauer Associates.
- Giachanou, A. and Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, 49(2):28.

- Kotzias, D., Denil, M., De Freitas, N., and Smyth, P. (2015). From group to individual labels using deep features. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 597–606. ACM.
- Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6). arXiv: cond-mat/0309508.
- Sanches, A., Cardoso, J. M. P., and Delbem, A. C. B. (2011). Identifying merge-beneficial software kernels for hardware implementation. In *International Conference on Reconfigurable Computing and FPGAs (ReConFig)*, pages 74–79, Cancun. IEEE.
- Soares, A. H. M. and Delbem, A. C. B. Detecção de correlação em dados complexos usando ncd.
- Valdivia, A. M. C. (2007). *Mapeamento de dados multidimensionais usando árvores filogenéticas: foco em mapeamento de textos*. PhD thesis, Universidade de São Paulo.
- Ziegelmayr, D. and Schrader, R. (2012). Sentiment polarity classification using statistical data compression models. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pages 731–738. IEEE.

Utilização de Metaheurísticas na Resolução do Problema de Agendamento de Horários de Aulas numa Escola de Ensino Fundamental

Fábio Santanta de Oliveira¹, Wanderley de Souza Alencar¹,
Hugo Alexandre Dantas do Nascimento¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Quadra D, Câmpus Samambaia, CEP: 74690-900
Goiânia – Goiás – Brasil

{fabiosantanaoliveira, wanderley, hadn}@inf.ufg.br

Abstract. *This article reports the use of metaheuristic techniques for the resolution of the so-called Teacher-Classroom Timetabling Problem, one of subcategories of the Educational Timetabling Problem. The three metaheuristic techniques employed are: In Vitro Fertilization Genetic Algorithm, Tabu Search and Greedy Randomized Adaptive Search Procedure. The experimentation was carried out with one real instance of the problem coming from a Brazilian school dedicated to elementary education and, in addition, with instances from widely known datasets. It turns out that the methods are able to produce solutions suitable for practical use in their contexts, being that method In Vitro has overcome the others for the established optimization goal.*

Resumo. *O artigo relata o emprego de metaheurísticas na resolução do Problema do Agendamento de Horário de Aulas, um exemplar da classe Educational Timetabling Problem. As três técnicas metaheurísticas empregadas foram: (1ª) In Vitro Fertilization Genetic Algorithm; (2ª) Tabu Search; e (3ª) Greedy Randomized Adaptive Search Procedure. A experimentação foi realizada com instâncias reais do problema advindas de uma escola brasileira dedicada à oferta do ensino fundamental e, adicionalmente, de datasets amplamente conhecidos. Verifica-se que os métodos são capazes de produzir soluções adequadas para utilização prática em seus contextos, sendo que o método In Vitro superou os demais para o objetivo de otimização estabelecido.*

1. Introdução

Antes do início de cada período letivo, nas mais variadas instituições de ensino – escolas, colégios, faculdades e universidades, dentre outras – há um desafiador problema a ser resolvido: o da elaboração do horário de aulas para aquele período. A literatura científica internacional nomeia, de maneira abrangente, este problema de *Educational Timetabling Problem* (ETTP), como também concebe nomes particulares para suas diferentes especificações, gerando subclasses dele, como se delineará adiante neste texto.

Por sua vez, o ETTP é um exemplar de uma categoria mais ampla de problemas denominados por *Scheduling Problems* (SPs)¹, nos quais o objetivo é alocar todos os recursos necessários para a realização de cada uma das tarefas constantes num conjunto

¹Por isso, é também referenciado por *Academic Scheduling Problem*, como em [Teoh et al. 2015].

predefinido delas. Em situações reais, é comum a existência de conflitos entre a *disponibilidade* e a *demanda* de recursos para a realização destas tarefas e, por isso, se busca uma solução viável, ou seja, uma que possa ser aplicada na prática e que maximize uma *função de utilidade*² definida. Esta função deve refletir a preferência do usuário por uma certa solução em relação às demais para a instância particular do problema sendo solucionada.

Nos ETTPs, o *tempo* necessário para a execução das atividades acadêmicas – nomeadas, genericamente, de *aulas* – é o principal recurso a ser contingenciado. O foco é, em verdade, no *intervalo de tempo* exigido para a realização de uma aula em relação às demais. Assim, a resolução de uma instância do ETTP consiste em associar cada aula a um intervalo de tempo (ou *horário*) em que ela deverá ocorrer num certo conjunto de horários disponíveis. Adicionalmente à condicionante temporal, existe a exigência de alocação de professores, geralmente de maneira exclusiva a uma aula, e de recursos de infraestrutura (como salas, laboratórios, equipamentos tecnológicos, etc.).

Ainda hoje, em muitas instituições de pequeno e médio portes, este problema é resolvido *manualmente*, o que pode exigir substancial esforço de um profissional especializado ou mesmo de uma equipe deles, sendo comum que a obtenção de uma solução consuma de algumas horas a diversos dias, de acordo com as exigências impostas à aceitação de uma candidata à solução viável, ou seja, aplicável na prática, como enfatiza van Staereling [van Staereling 2012]: “*School timetable construction can be an extremely difficult task and usually consumes a large amount of time. Although some have a better intuition for the problem than others, it is most of the time practically impossible to construct a timetable that satisfies the wishes of every teacher and student.*”.

Desde meados da década de 1960, tem surgido propostas de métodos semi-automáticos/automáticos que visam a resolução do ETTP, como os trabalhos de Appleby, Gotlieb, Berghuis e Lions ([Appleby 1961], [Gotlieb 1962], [Berghuis et al. 1964] e [Lions 1966, Lions 1967]). Transcorrido mais de meio século, este problema ainda se encontra aberto à investigação científica, devido a diversos fatores, dentre os quais se podem citar:

- (1) apesar de haver um conjunto identificável de características comuns entre as diferentes instituições de ensino que atuam num mesmo nível de ensino e/ou localidade, cada instituição possui particularidades que a tornam única, impossibilitando o uso de soluções gerais e/ou genéricas, de aplicação ampla;
- (2) a constatação de que o problema é um exemplar típico das classes dos denominados problemas *\mathcal{NP} -Completo* e *\mathcal{NP} -Difícil* em suas modelagens sob a forma de problema de *decisão* e de *otimização*, respectivamente, como demonstrado em diversos trabalhos, dentre eles o elaborado por de Werra [de Werra 1997].
- (3) o grande número de variáveis de decisão envolvidas em exemplares reais e as complexas interdependências entre elas, muitas vezes antagônicas entre si.

Tendo em vista a grande diversidade de categorizações para os ETTPs existentes, neste texto adota-se a taxonomia proposta por Carter e Tovey [Carter and Tovey 1992]:

- (1) *Course Timetabling* – Cada estudante é considerado o elemento básico para o agendamento. Deve-se associar horários para a realização de aulas de cada uma das disciplinas constantes no currículo a ser cumprido pelo estudante. Podem

²Função de qualidade, de avaliação ou objetivo.

existir várias restrições aplicadas com respeito à estrutura do currículo, salas de aulas, professores, etc. Muitas universidades adotam esse sistema de associação, que é conhecido, no Brasil, pelo nome genérico de “Sistema de Créditos”: cada aluno tem o seu horário de aulas próprio, não pertencendo a uma “turma”.

- (2) *Class-Teacher Timetabling* – A unidade utilizada para o agendamento é uma “turma”. Normalmente há uma associação prévia entre professores e cursos/turmas nos quais estes atuarão, bem como as disciplinas pelas quais cada um será responsável. O problema é fixar as aulas a serem realizadas para as turmas existentes, sem gerar nenhum conflito temporal e, complementarmente, satisfazer as outras restrições impostas pelas diversas entidades envolvidas. A factibilidade é, frequentemente, a maior dificuldade apresentada nessa subclasse de problemas e, muitas vezes, se faz necessário reconsiderar associações previamente estabelecidas de maneira a atingir uma solução. Esse tipo de problema está presente na maioria das escolas da educação infantil, ensino fundamental e médio brasileiras;
- (3) *Student Scheduling* – Essa categoria ocorre quando as disciplinas são oferecidas em múltiplas vezes, com aulas simultâneas ou não. Assim, uma vez que cada estudante tenha fixado quais disciplinas irá cumprir no período, deve-se elaborar um horário válido para ele. A existência de múltiplos horários de oferta de uma disciplina se constitui um facilitador na obtenção de horários de aulas de melhor qualidade para os estudantes e auxilia no balanceamento da quantidade de alunos;
- (4) *Teacher Assignment* – O problema é associar professores às disciplinas de maneira a maximizar uma função de preferência destes com relação a cada uma das disciplinas. Normalmente a resolução desse problema é feita como uma fase anterior à elaboração de um horário de aulas propriamente dito, pois não considera as restrições impostas pelos demais entes envolvidos (estudantes, salas, etc.);
- (5) *Classroom Assignment* – As aulas entre estudantes e professores para o cumprimento das exigências estabelecidas no currículo devem estar associados às salas de aulas específicas. Uma sala de aula é associada de maneira a satisfazer condicionantes: sua capacidade (número máximo de alunos), sua localização na instituição (câmpus/blocos), os recursos materiais nela disponíveis, os horários em que ela está disponível, etc. Idealmente, nessa situação, os horários de ocorrência das aulas e as salas de aulas deverão ser associados de maneira simultânea.

O problema aqui tratado está na categoria *Class-Teacher Timetabling* (CT-TTP) para a qual técnicas metaheurísticas, que não precisam se relacionar intimamente com o domínio do problema sendo resolvido, têm representado, nos últimos anos, uma abordagem promissora para a descoberta de soluções eficazes e eficientes. Por esse motivo, este trabalho explora três técnicas metaheurísticas para um problema desta categoria: (1^a) *Genetic Algorithms* (GA), com o aprimoramento proposto por Camilo-Junior & Yamanaka [Camilo-Junior and Yamanaka 2011], que mimetiza a fertilização *in vitro*; (2^a) *Tabu Search* (TS); e (3^a) *Greedy Randomized Adaptive Search Procedure* (GRASP).

O restante deste texto está organizado em cinco seções. A seguinte, Seção 2, exhibe uma sucinta revisão da literatura científica dedicada ao *Educational Timetabling Problem* e afins. A Seção 3 especifica o problema-alvo a ser aqui tratado para, ato contínuo, a Seção 4 apresentar os métodos empregados em sua resolução. A experimentação computacional é delineada na Seção 5, sendo as conclusões dela extraídas e uma análise geral do trabalho registradas na Seção 6.

2. Revisão da Literatura

O tratamento científico-computacional para o *Problema de Agendamento de Horários de Aulas*, ou *Educational Timetabling Problem* (ETTP), teve início na década de 1960, sendo o trabalho de Appleby [Appleby 1961] um dos pioneiros. Nele, técnicas para a “construção” de horários por meio do uso de computadores são apresentadas e o problema é comparado com outros problemas de escalonamento já conhecidos na época. Gotlieb [Gotlieb 1962] propõe um método iterativo que emprega matrizes de variáveis lógicas, demonstrando ser possível gerar horários viáveis. Este método é aprimorado por Csimá & Gotlieb [Csimá and Gotlieb 1964], que apresentam fundamentação teórica de seu funcionamento. Um método que emprega uma função recursiva de “troca de horários” é proposto por Berghuis *et al.* [Berghuis *et al.* 1964] para resolver a geração de horários infactíveis, sendo ele aplicado em diversos problemas reais. Lions, num par de trabalhos, [Lions 1966, Lions 1967], utiliza técnicas para a redução de matrizes como abordagem para a geração de horários para escolas da cidade de Ontário – Canadá.

Nas duas décadas seguintes – 1970 e 1980 – houve grande esforço para a identificação/resolução de diferentes especificações do ETTP. Tendo em vista a grande variedade existente, de Werra [de Werra 1985] se propõe a apresentar uma “introdução” a esta nova categoria de problemas, com destaque para a modelagem baseada na Teoria de Grafos que, à época, começa a se mostrar útil para sua formulação e resolução.

A partir de 1990, ganham destaque as propostas baseadas em heurísticas e metaheurísticas. Os trabalhos de Hertz [Hertz 1991, Hertz 1992] empregam *Tabu Search* (TS) – concebida por Glover [Glover 1989, Glover 1990] – para resolução de problemas de “grande escala”, onde o número de conflitos entre os horários das aulas que possuem estudantes e/ou professores comuns é, normalmente, elevado. Na abordagem, TS é usada para a resolução de subproblemas (*assignment problems*) que são combinados para a síntese de uma solução global. Num experimento se demonstra a capacidade para resolver um problema com 143 professores, 288 disciplinas, 1.729 estudantes e 67 salas em 15 prédios.

O trabalho de Colorni *et al.* [Colorni *et al.* 1992] anuncia os *Genetic Algorithms* (GAs) como uma “nova abordagem” para a resolução de TTPs, empregando-os, com sucesso, em escolas italianas. Em sequência, ele e outros [Colorni *et al.* 1998] confrontam três abordagens para solucionar o mesmo problema: TS, *Simulated Annealing* (SA)³ e GA. Este último possui duas versões: *sem* e *com* o uso de uma técnica de *local search* (LS). A experimentação, em duas escolas italianas, evidenciou que GA+LS e TS são capazes de superar soluções sintetizadas pelo TS ou aquelas confeccionadas manualmente empregando um *software* comercial como apoio à tarefa. Os problemas considerados possuíam de 20 a 24 professores, 10 turmas e 30 aulas semanais por turma.

Em seu trabalho, Ferland & Lavoie [Ferland and Lavoie 1992] usam um método heurístico iterativo, realizando experimentos com problemas artificiais sistematicamente concebidos. O método se mostra adequado para obtenção de soluções que são aceitas por usuários reais. No início dos anos 2000, a interatividade com o usuário é empregada por Müller & Barták [Müller and Barták 2001] como paradigma de resolução colaborativa

³Baseada na versão original concebida por Metropolis [Metropolis *et al.* 1953] e como empregada por Kirkpatrick *et al.* [Kirkpatrick *et al.* 1983].

usuário/computador. Os experimentos mostram a capacidade desta combinação para resolver ETTPs de uma universidade de médio porte. A mesma abordagem é empregada por Geske & Goltz [Geske and Goltz 2004], obtendo resultados similares aos primeiros.

Um GA é proposto por dos Santos & Santos [dos Santos and Santos 2004] para a resolução do problema numa escola que oferta ensino fundamental e médio. Nesse GA, cada cromossomo representa uma solução possível, mas não necessariamente viável, para o horário de aulas. As restrições são hierarquizadas quanto ao grau da penalidade imposta por sua violação (*grave*, *médio* e *baixo*). A experimentação com 16 professores, 03 turmas, 45 disciplinas e 90 aulas semanais foi capaz de identificar soluções avaliadas como “*subótimas*” pelos autores.

O trabalho de Santos [Santos et al. 2004] utiliza um processo *tabu search* (TS) para resolver o problema de escalonar professores/turmas, incluindo-se duas estratégias de diversificação. Os resultados experimentais, quando comparados a outras propostas de TS da literatura, se mostraram melhores para todas as instâncias testadas pelos autores.

De Haan *et al.* [de Haan et al. 2007] apresentam estudo de caso numa instituição de ensino médio. O objetivo é associar individualmente *estudante–disciplina–sala* para tratar o fato de que os estudantes não são vinculados a turmas e, sim, considerados isoladamente nesta escola. O processo de elaboração do horário é, por isso, dividido em três passos: (1º) as disciplinas são distribuídas em turnos; (2º) aulas – das disciplinas – associadas a horários; e (3º) otimiza-se o horário por meio da aplicação de um processo *tabu search*.

Moura & Scaraficci [Moura and Scaraficci 2010] apresentam a combinação de GRASP com um processo de *path relinking improvement* para resolver um horário associado a período de tempo prefixado e envolvendo aulas/professores que devem satisfazer restrições operacionais, algumas específicas das instituições brasileiras. A combinação, de acordo com os autores, apresenta bons resultados na resolução de instâncias reais.

Mais recentemente, a integração de técnicas tem merecido destaque, como demonstra o trabalho de Fonseca *et al.* [Fonseca et al. 2016], no qual se combina a metaheurística *Variable Neighbourhood Search Algorithm*, também usada por Fonseca & Santos [Fonseca and Santos 2014], com um modelo de programação matemática (matheurística) para solucionar um conjunto de ETTPs. Apesar disso, técnicas clássicas, como programação matemática inteira, não são desconsideradas como evidência Friedman [Friedman 2016], sendo também comparadas com outras, como mostrado por Veenstra & Vis [Veenstra and Vis 2016].

3. Especificação do Problema

O problema-alvo deste trabalho, o *School Timetabling Problem* (STTP), tem sido especificado de variadas maneiras e resolvido por diferentes métodos, como delineado na seção anterior. No sistema educacional brasileiro, nas instituições de ensino fundamental e médio – normalmente denominadas de *escolas* e *colégios*, respectivamente – cada estudante está associado apenas a uma turma que, por sua vez, possui seu horário semanal de aulas fixado antes do início do período letivo. É comum, mas não obrigatório, que uma turma utilize um conjunto bem limitado de salas de aula no transcorrer da semana. Outro costume é que há prioridade, nos horários concebidos, ao atendimento das restrições im-

postas pelos professores. Como resultado o problema se insere, na taxonomia apresentada na Seção 1, como um exemplar da categoria *Class-Teacher Timetabling* (CT-TTP).

Uma restrição pode ser de atendimento obrigatório (restrição *forte* ou *hard constraint*) ou opcional, mas desejável (*fraca*, ou *soft constraint*). A violação, por um agendamento de horários, de qualquer restrição forte (HC), o torna de adoção inviável e, por isso, ele deve ser descartado. Por outro lado, um agendamento de horários que não atenda a todas as restrições fracas (SC), continua sendo uma solução válida, mas penalizada na proporção ponderada das SCs violadas.

Nesta especificação do CT-TTP, considera-se que há um conjunto finito de:

- (1) professores $\mathcal{P} = \{p_1, p_2, p_3, \dots, p_n, \text{ com } n \in \mathbb{N}^*\}$, capaz, cada um, de ministrar certas disciplinas;
- (2) disciplinas $\mathcal{D} = \{d_1, d_2, d_3, \dots, d_m, \text{ com } m \in \mathbb{N}^*\}$, sendo um conjunto comum expresso por $\mathcal{D} = \{ \text{Biologia, Ciências, Cultura surda, Dança, Educação física, Ensino religioso, Filosofia, Física, Geografia, História, Inglês, Língua portuguesa, Libras, Literatura, Matemática aplicada, Matemática, Química e Sociologia} \}$;
- (3) turmas $\mathcal{T} = \{t_1, t_2, t_3, \dots, t_y, \text{ com } y \in \mathbb{N}^*\}$;
- (4) ofertas $\mathcal{R} = \{r_1, r_2, r_3, \dots, r_x, \text{ com } x \in \mathbb{N}^*\}$. Uma oferta é a designação de um professor p_i para poder ministrar as aulas de uma disciplina d_j , ou seja, é um par ordenado $(p_i, d_j) \in \mathcal{P} \times \mathcal{D}$.

As restrições usadas neste texto são as seguintes:

- (1) hc_1 – Um professor somente pode ministrar, num determinado horário, uma única aula para uma certa turma (não conflito de horário do professor);
- (2) hc_2 – Um professor somente pode ministrar aulas em horários nos quais está disponível, tendo em vista que ele pode registrar indisponibilidade para um subconjunto horários (obediência à disponibilidade do professor);
- (3) hc_3 – Uma turma, num determinado horário, somente pode assistir a uma única aula (não conflito de horário da turma);
- (4) sc_1 – Deve-se evitar a existência de “aula vaga” no agendamento de horários semanal de uma turma;
- (5) sc_2 – Deve-se evitar a existência de “aula vaga” no agendamento de horários de um professor;
- (6) sc_3 – Deve-se evitar a existência de intervalo entre duas aulas de uma mesma disciplina (ou professor), num mesmo dia de agendamento de horário de uma turma.

Considera-se que a semana letiva possui $d \in \mathcal{N}^*$ dias úteis, com $1 \leq d \leq 6$. Cada dia letivo, por sua vez, possui h intervalos de tempo⁴ disponíveis para abrigar uma aula. Por consequência, a semana letiva possui $q = d \times h$ intervalos. Uma solução Q para o problema é a associação estabelecida entre todos os q intervalos de tempo de uma semana letiva e as ofertas \mathcal{R} efetivamente alocadas nestes intervalos.

Para realizar a comparação qualitativa entre duas diferentes soluções propostas foi

⁴A literatura científica denomina de *timeslot* a cada um dos intervalos de tempo no qual um dia é dividido de maneira discreta. Por exemplo, das 08h às 09h, um *timeslot*. Das 09h às 10h, outro.

concebida uma *função objetivo* (FO) expressa por:

$$\mathcal{F}(Q) = \sum_{i=1}^3 w_i \times \mathcal{F}_i(Q) \quad (1)$$

onde cada \mathcal{F}_i é uma função concebida para quantificar a penalidade, na FO, advinda da violação das restrições fracas (SCs) sc_1 , sc_2 e sc_3 , respectivamente:

$$\mathcal{F}_1(Q) = \sum_{t=1}^{|\mathcal{T}|} horariovago(t) \quad (2)$$

$$\mathcal{F}_2(Q) = \sum_{t=1}^{|\mathcal{T}|} \sum_{d=1}^{|\mathcal{D}|} horariovago(d) \quad (3)$$

$$\mathcal{F}_3(Q) = \sum_{t=1}^{|\mathcal{T}|} \sum_{p=1}^{|\mathcal{P}|} intervalo(p) \quad (4)$$

os valores w_i , $w_i \in \mathbb{R}^*$, representam a importância relativa entre cada uma das diferentes violações das SCs.

Por exemplo, um conjunto de disciplinas é representado na Tabela 1:

Tabela 1. Conjunto de disciplinas ofertadas em determinado período letivo.

Nome da Disciplina
Biologia
Ciências
Cultura surda
Dança
Educação física
Ensino Religioso
Filosofia
Física
Geografia
História
Inglês
Língua portuguesa
Libras
Literatura
Matemática aplicada
Química
Sociologia

As ofertas, que relacionam professores e disciplinas, são destacadas na Tabela 2, onde cada linha representa um par (professor, disciplina) que será associado ao agendamento de horários.

Tabela 2. Conjunto de ofertas – professor/disciplina

Nome do Professor	Nome da Disciplina
Alessandra	Biologia
Alessandra	Ciências
Anderson	Física
Elaine	Dança
Elaine	Educação física
Fernanda	Cultura surda
Fernanda	Libras
Graça	Geografia
Graça	História
Graça	Ensino Religioso
Graça	Sociologia
Kelly	Matemática aplicada
Mariana	Língua portuguesa
Mariana	Literatura

Por fim, um exemplo de agendamento de horários para a instituição de ensino é mostrado na Tabela 3.

Tabela 3. Exemplo de agendamento de horários de aulas.

Horário	Segunda	Terça	Quarta	Quinta	Sexta
07:00 – 07:50	Ed. Física <i>Fernanda</i>	Inglês <i>Elisângela</i>	Inglês <i>Elisângela</i>	Dança <i>Alessandra</i>	Física <i>Fernanda</i>
07:50 – 08:40	L. Port. <i>Elisângela</i>	Biologia <i>Elisângela</i>	Matemát. <i>Kenia</i>	Dança <i>Alessandra</i>	Libras <i>José</i>
08:40 – 09:30	L. Port. <i>Elisângela</i>	Matemát. <i>Kenia</i>	Inglês <i>Elisângela</i>	Ciências <i>Agda</i>	Física Anderson
09:45 – 10:35	Matemát. <i>Kenia</i>	Ciências <i>Agda</i>	Biologia <i>Elisângela</i>	Ciências <i>Agda</i>	Biologia <i>Kenia</i>
10:35 – 11:15	Ens. Rel. <i>Graça</i>	Ciências <i>Agda</i>	Libras <i>Josiane</i>	Biologia <i>Agda</i>	História <i>Graça</i>
11:15 – 12:00	Geografia <i>Graça</i>	História <i>Graça</i>	Libras <i>Josiane</i>	História Anderson	Geografia <i>Graça</i>

4. Métodos Propostos

Esta seção discute, de maneira sintética, o *Genetic Algorithm* (GA) por meio de sua variante *In Vitro Fertilization* (IVFm), o *Tabu Search* (TS) e, por fim, *Greedy Randomized Adaptive Search Procedures* (GRASP).

4.1. *In Vitro Fertilization Genetic Algorithm* (IVFm)

O algoritmo genético (*Genetic Algorithm* – GA) se baseia na melhoria de soluções através de mecanismos inspirados em genética evolucionária e seleção natural. Os principais

mecanismos são a *Seleção* e a *Variação*. Na *Seleção*, os indivíduos mais adaptados ao meio em que vivem têm mais chances de sobreviver e deixar descendentes. Na *Variação*, pode ocorrer a recombinação genética oriunda do processo reprodutivo ou de mutações de origem aleatória e imprevisível.

Cada indivíduo representa uma solução para o problema e a aptidão é dada pelo valor da *função objetivo*. O algoritmo gera repetidamente uma população de tamanho n , sendo que a cada população é obtida a partir da população anterior, através da seleção dos melhores indivíduos. Na população gerada, é aplicada recombinação e/ou mutação para gerar novos indivíduos, que serão adicionados à população.

Um dos possíveis comportamentos dos GAs são a perda e o baixo aproveitamento das informações presentes nas populações de indivíduos, cada um representando uma possível solução para o problema-alvo. Para minimizá-los houve a proposta um algoritmo auxiliar, paralelo, de forma análoga à fertilização *in vitro* biológica, sendo este denominado de *In Vitro Fertilization Genetic Algorithm* (IVFm) por Camilo-Junior & Yamanaka [Camilo-Junior and Yamanaka 2011] e, por isso, esta abordagem foi utilizada no presente texto e parametrizada conforme mostra a Seção 5.

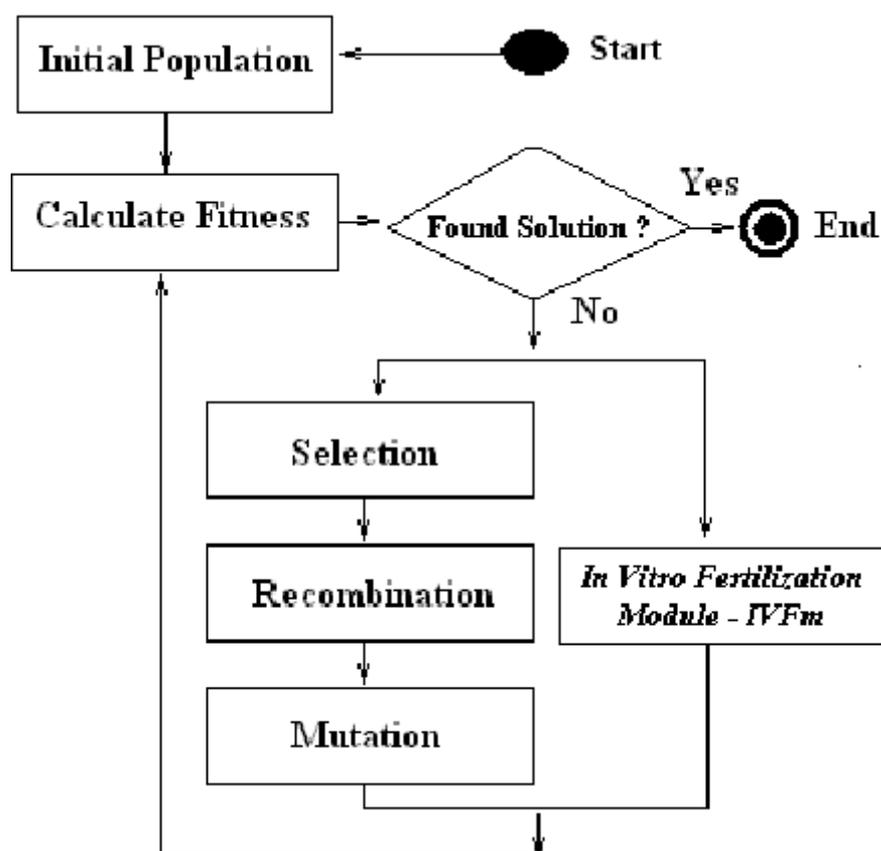


Figura 1. Fluxo do *In Vitro Fertilization Genetic Algorithm*, concebida por Camilo-Junior & Yamanaka [Camilo-Junior and Yamanaka 2011].

Em particular, o IVFm recebe como entrada parcelas da população atual e gera um ou mais indivíduos, melhores que o melhor indivíduo da população atual. Desta forma, ele recebe e analisa uma boa quantidade de informações que poderia ser perdida,

recombina e abastece a população com bons indivíduos, otimizando o processo evolucionário. O fluxo do algoritmo é apresentado na Figura 1, conforme Camilo-Junior & Yamanaka [Camilo-Junior and Yamanaka 2011].

O IVFm avalia várias possibilidades e seleciona os genes do pai e das mães (*father*, *mothers*) que irão compor o novo indivíduo. São selecionados os genes do melhor indivíduo (*father*) e dos outros bons indivíduos (*mothers*), sendo o novo indivíduo é avaliado para verificar a sua aptidão (ou *fitness*). Caso seja um bom indivíduo, ele é introduzido na população. O processo de recombinação é apresentado na Figura 2, extraída de Camilo-Junior & Yamanaka [Camilo-Junior and Yamanaka 2011].

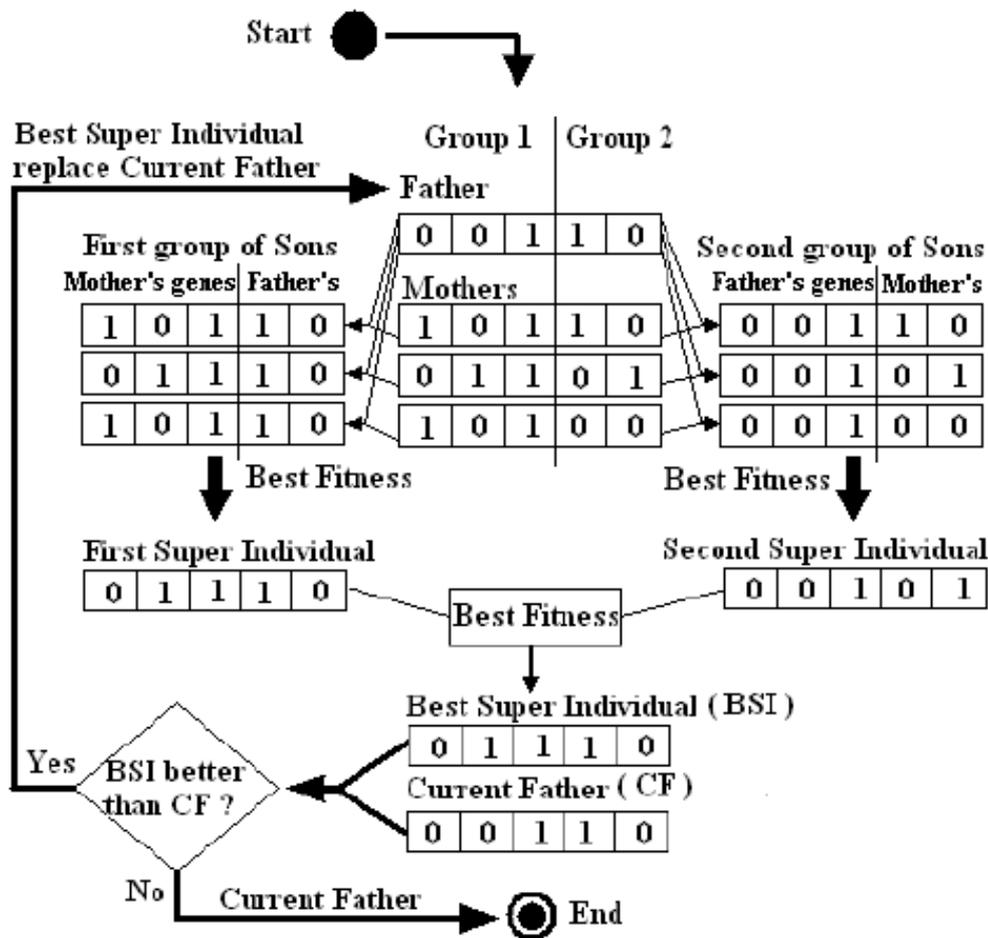


Figura 2. Processo de recombinação proposto por Camilo-Junior & Yamanaka [Camilo-Junior and Yamanaka 2011].

4.2. Tabu Search (TS)

A abordagem *tabu search* (TS), concebida por Glover [Glover 1989, Glover 1990], é uma metaheurística baseada em técnicas de busca local e que emprega um processo de seleção na vizinhança de uma certa solução x , dita *atual*, para eleger outra solução, dita x' , para substituí-la. A esperança é que x' seja melhor que x , quando avaliada por uma certa função de aptidão previamente definida. O termo *vizinhança* se refere ao fato de que x' pouco difere de x . Um dos problemas potenciais da ideia de vizinhança é que pode ocorrer um processo cíclico de busca e, para evitá-lo, é utilizada uma “lista tabu”, ou seja, uma

lista com as soluções visitadas recentemente durante uma certa quantidade de iterações (prefixada) no processo de busca.

A escolha considera a existência de uma lista tabu \mathcal{T} , onde a melhor solução x' , não tabu na vizinhança de x , é a escolhida. Em geral, a lista tabu é representada por uma fila circular de tamanho máximo prefixado.

4.3. Greedy Randomized Adaptive Search Procedures (GRASP)

O *Greedy Randomized Adaptive Search Procedure* (GRASP), amplamente descrito por Feo & Resende [Feo and Resende 1995], é uma metaheurística de múltiplas inicializações e, por isso, tende à obtenção de ótimos locais de melhor qualidade que métodos de inicialização única.

O processo é dividido em duas fases: criação e busca local. Em cada iteração do algoritmo, é gerada uma solução inicial aleatória utilizando um algoritmo guloso e sobre ela é efetuada uma *Busca Local*. O histórico das soluções não é guardado. O processo de repete, até que algum critério de parada seja satisfeito como registrado no trabalho de J. Hirsch *et al.* [J. Hirsch et al. 2010].

5. Experimentação Computacional

A experimentação computacional foi conduzida utilizando-se de um *notebook* DELL, modelo Inspiron 7560, Intel I7 com velocidade de processamento de 2.70GHz, memória principal de 16GiB e tecnologia DDR4, unidade de disco rígido com capacidade de 916 GiB e contando com uma placa de vídeo da marca NVIDIA, modelo GeForce 940MX. O sistema operacional instalado é o Windows 10, versão 1803 (compilação 17134.165) com todas as atualizações publicadas pelo produtor. A aplicação foi desenvolvida utilizando-se a linguagem C++ e compilada para o ambiente supramencionado por meio do compilador C++ Builder 10.2, versão 25.0.29899.2631.

Foram usadas as informações de uma escola de ensino fundamental e médio, em Goiânia, e instâncias do Brasil publicadas pela ITC2011, disponível em <https://www.utwente.nl/en/eemcs/dmmp/hstt/archives/XHSTT-2012/>. A Tabela 4 apresenta o número de horários disponíveis para alocação, o número de professores, o número de turmas e o número de aulas.

Tabela 4. Características das instâncias.

Instância	Quantidade de Horários	Professores	Turmas	Aulas
Escola em Goiânia	30	10	8	240
BrazilInstance1	25	8	3	75
BrazilInstance4	25	23	12	300
BrazilInstance5	25	31	13	325
BrazilInstance6	25	20	14	350
BrazilInstance7	25	33	20	500

As metaheurísticas foram parametrizadas, para todas as instâncias, da seguinte maneira:

(1) **GRASP**

GRASP Max = 100
Candidate List Size= 10

(2) **Tabu Search**

BTMax = 100
Tabu Size= 10

(3) **In Vitro Fertilization Genetic Algorithm (IVFm)**

Population Size = 3000
Quantidade de indivíduos avaliados = 20% do Population Size.

Os parâmetros $w_i, i = 1, 2, 3$ e $w_i \in \mathbb{R}^*$, que refletem a importância relativa entre as violações das restrições fracas (SCs), foram todos fixados em 1.

5.1. Resultados e Análise

Os resultados dos experimentos são apresentados na Tabela 5. A metaheurística *In Vitro Fertilization Genetic Algorithm* (IVFm) obteve, sempre, os melhores valores para função objetivo – que deve ser minimizada. A *Tabu Search* apresentou-se como a segunda melhor técnica aplicada aos problemas resolvidos. Os valores calculados para as instâncias *BrazilInstance4*, *BrazilInstance5* e *BrazilInstance6* foram muito próximos, mesmo com diferenças na quantidades de professores, turmas e aulas (Tabela 4).

Tabela 5. Resultado da função objetivo.

Instância	<i>Tabu Search</i>	GRASP	<i>IVFm</i>
Escola em Goiânia	280	284	268
<i>BrazilInstance1</i>	156	168	142
<i>BrazilInstance4</i>	368	372	340
<i>BrazilInstance5</i>	388	396	370
<i>BrazilInstance6</i>	368	388	362
<i>BrazilInstance7</i>	456	500	428

6. Conclusões

As soluções do problema foram feitas utilizando as metaheurísticas: *In Vitro Fertilization Genetic Algorithm* (IVFm), *Tabu Search* e *Greedy Randomized Adaptive Search Procedures* (GRASP). O estudo de caso foi o agendamento de horários de escola de ensino fundamental (6º ano, 7º ano, 8º ano, 9º ano) e ensino médio (1º ano, 2º ano e 3º ano) em Goiânia. Também, foram utilizadas instâncias do Brasil publicadas pela ITC2011. A oferta é representada por uma tupla de professor e disciplina. A grade de horários é representada por uma estrutura de dados baseada em matriz com os dias e os horários da escola.

As três metaheurísticas encontraram soluções viáveis para as instâncias, sendo que a metaheurística *In Vitro Fertilization Genetic Algorithm* apresentou, sempre, melhor valor da função objetivo, sendo seguida pela *Tabu Search*.

Referências

- Appleby, J. S. (1961). Techniques for producing school timetables on a computer and their application to other scheduling problems. *The Computer Journal*, 3(4):237–245.
- Berghuis, J., van der Heiden, A. J., and Bakker, R. (1964). The preparation of school timetables by electronic computer. *BIT Numerical Mathematics*, 4(2):106–114.
- Camilo-Junior, C. G. and Yamanaka, K. (2011). In vitro fertilization genetic algorithm. In Kita, E., editor, *Evolutionary Algorithms*, pages 1–14. InTech Open Access Publisher.
- Carter, M. W. and Tovey, C. A. (1992). When is the classroom assignment problem hard? *Operations Research*, 40(S1):28–29.
- Coloni, A., Dorigo, M., and Maniezzo, V. (1992). Genetic algorithms: a new approach to the timetable problem. *Combinatorial Optimization*, pages 235–239.
- Coloni, A., Dorigo, M., and Maniezzo, V. (1998). Metaheuristics for high school timetabling. *Computational Optimization and Applications*, 9(3):275–298.
- Csima, J. and Gotlieb, C. C. (1964). Test on a computer method for constructing school timetables. *Communications of the ACM*, 7:160.
- de Haan, P., Landman, R., Post, G., and Ruizenaar, H. (2007). A case study for timetabling in a dutch secondary school. In *Practice and Theory of Automated Timetabling VI*, volume 3867, pages 267–279.
- de Werra, D. (1985). An introduction to timetabling. *European Journal of Operational Research*, 19(2):151–162.
- de Werra, D. (1997). The combinatorics of timetabling. *European Journal of Operational Research*, 96:504–513.
- dos Santos, C. N. and Santos, R. J. d. S. (2004). Implementação de um algoritmo genético para a construção automática de horários em uma escola de ensino fundamental e médio. In *Anais do VIII Simpósio Brasileiro de Pesquisa Operacional e Logística da Marinha*, pages 1–8, Rio de Janeiro, Brasil.
- Feo, T. A. and Resende, M. G. C. (1995). Greedy Randomized Adaptive Search Procedures. *Journal of Global Optimization*, 6:109–133.
- Ferland, J. A. and Lavoie, A. (1992). Exchanges procedures for timetabling problems. *Discrete Applied Mathematics*, 35:237–253.
- Fonseca, G. H. G. and Santos, H. G. (2014). Variable neighborhood search based algorithms for high school timetabling. *Computers & Operations Research*, 52:203–208.
- Fonseca, G. H. G., Santos, H. G., and Carrano, E. G. (2016). Integrating matheuristics and metaheuristics for timetabling. *Computers & Operations Research*, 74:108–117.
- Friedman, J. S. (2016). Automated timetabling for small colleges and high schools using huge integer programs. *CoRR*, abs/1612.08777.
- Geske, U. and Goltz, H.-J. (2004). Automatische und interaktive Stundenplanung. *Informatik Forsch. Entw.*, 19:65–73.
- Glover, F. W. (1989). Tabu search (Part I). *ORSA Journal on Computing*, 1:190–206.

- Glover, F. W. (1990). Tabu search (Part II). *ORSA Journal on Computing*, 2:4–32.
- Gotlieb, C. C. (1962). The construction of class-teacher timetabling. In *October*, volume 13, pages 73–77.
- Hertz, A. (1991). Tabu search for large scale timetabling problems. *European Journal of Operational Research*, 54:39–47.
- Hertz, A. (1992). Finding a feasible course schedule using tabu search. *Discrete Applied Mathematics*, 35:255–270.
- J. Hirsch, M., Pardalos, P., and Resende, M. (2010). Speeding up continuous GRASP. *European Journal of Operational Research*, 205:507–521.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220:671–680.
- Lions, J. (1966). Matrix reduction using the Hungarian method for the generation of school timetables. *Communications of the ACM*, 9(5):349–354.
- Lions, J. (1967). The Ontario school timetabling problem. *The Computer Journal*, 10(1):14–21.
- Metropolis, N., Rosenbluth, A., Resenbluth, N., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.
- Moura, A. V. and Scaraficci, R. A. (2010). A GRASP strategy for a more constrained School Timetabling Problem. *International Journal of Operational Research*, 7(2):152.
- Müller, T. and Barták, R. (2001). Interactive timetabling: concepts, techniques and practical results. Technical report, Charles University, Prague, Czech Republic.
- Santos, H. G., Ochi, L. S., and Souza, M. J. F. (2004). An Efficient Tabu Search Heuristic for the School Timetabling Problem. In Ribeiro, C. C. and Martins, S. L., editors, *Experimental and Efficient Algorithms (WEA 2004)*, volume 3059 of *Lecture Notes in Computer Science (LNCS)*, pages 468–481, Berlin. Springer.
- Teoh, C. H., Wibowo, A., and Ngadiman, M. S. (2015). Review of state of art for metaheuristic techniques in academic scheduling problems. *Artificial Intelligence Review*, 44:1–21.
- van Staereling, I. v. H. (2012). School timetabling in theory and practice. Technical report, VU University, Amsterdam, Holland.
- Veenstra, M. and Vis, I. F. A. (2016). School timetabling problem under disturbances. *Computers & Industrial Engineering*, 65:175–186.

Intersecção de Caminhos mais Longos em Produtos de Grafos

Gabriel Matheus Faria de Almeida¹, Elisângela Silva Dias¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Câmpus Samambaia – 74690-900 – Goiânia – GO – Brasil

{gabrielmatheus, elisangela}@inf.ufg.br

Abstract. *The problem of longest paths in graphs, began to be discussed in 1966 from a question raised by Gallai, which questions whether any connected graph has at least one vertex that appears in all the longest paths. In this work, we studied the longest paths in chordal graphs, weighted graphs, trees and complementary prisms graphs and we obtained theoretical results and some factorial time algorithms, trying to understand the problem and the behavior of the longest paths in some types of graphs.*

Resumo. *O problema de caminhos mais longos começou a ser discutido por volta de 1966 a partir de uma questão levantada por Gallai, que questiona se todo grafo conexo tem pelo menos um vértice que aparece em todos os caminhos mais longos. Neste trabalho, foram estudados os caminhos mais longos em produtos de grafos, grafos cordais, grafos ponderados, árvores e grafos prismas complementares e foram obtidos resultados teóricos e alguns algoritmos de tempo fatorial buscando entender o problema e o comportamento dos caminhos mais longos nos diferentes tipos de grafos.*

1. Introdução

Um grafo $G = (V(G), E(G))$ consiste em $V(G)$, um conjunto não vazio de *vértices*, e $E(G)$, um conjunto de *arestas*, que são pares não ordenados de vértices. Cada aresta contém um ou mais vértices associados a ela, chamados de extremidades. Sejam v_1, v_2 vértices de G e e uma aresta que os associa; diz-se que e liga ou conecta v_1 e v_2 . Os vértices v_1, v_2 são ditos *adjacentes* em G , pois existe $e = v_1v_2 \in E(G)$. Diz-se também que e é incidente a v_1 e v_2 .

Um grafo não orientado é dito *conexo* se existir um caminho entre qualquer par de vértices distintos no grafo, caso contrário diz-se que o grafo é *desconexo*. Um vértice é dito *isolado* quando não tem outros vértices adjacentes a ele. Uma aresta é chamada de *laço* quando tem o vértice de saída igual ao vértice de chegada, isto é, uma aresta que liga o vértice a ele mesmo. *Arestas múltiplas* são arestas distintas incidentes ao mesmo par de vértices e um grafo é dito *simples* quando não contém laços em seus vértices nem arestas múltiplas.

Um grafo orientado é o grafo em que as arestas possuem um vértice de partida e um vértice de chegada, denotadas por uma seta e a passagem pela aresta deve obedecer a direção e o sentido que a aresta representa. No grafo não orientado as arestas não possuem orientação e são denotadas por segmentos de retas, ou seja, pode-se ir e voltar na aresta sem a necessidade de obedecer uma orientação.

O grau de um vértice $v \in V(G)$, denotado por $g(v)$, é o número de arestas incidentes a ele. Observe que um laço em um vértice contribui duas vezes para o seu grau. Por exemplo, na Figura 1, $g(a) = 2$, $g(d) = 4$ e $g(e) = 3$.

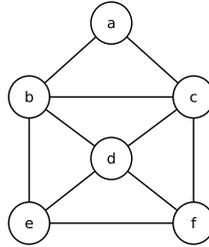


Figura 1. Exemplo de grafo simples não orientado.

Um *caminho simples* é uma sequência finita de vértices $\{v_1, v_2, \dots, v_n\}$ denotado por P_n , tal que $v_i v_{(i+1)} \in E(G)$ para todo $i \in \{1, \dots, n - 1\}$ e sem vértices repetidos na sequência, isto é, $v_i \neq v_j$ para todo $j \in \{1, \dots, n\}$ com $j \neq i$. Um ciclo simples em um grafo G , denotado por C_n , é um caminho simples $\{v_1, v_2, \dots, v_n\}$ tal que $v_n v_1 \in E(G)$ e $n \geq 3$. Em um ciclo simples, o número de vértices é igual ao número de arestas e para todo vértice $v_i \in C_n$, $g(v_i) = 2$. Como exemplo, na Figura 2 temos um ciclo C_3 e um ciclo C_4 .

O produto cartesiano $G_1 \times G_2$ de dois grafos G_1 e G_2 , é o grafo que contém o conjunto de vértices $V(G_1 \times G_2) = V(G_1) \times V(G_2)$ e dois vértices $v = v_1, v_2$ e $u = u_1, u_2$ de $G_1 \times G_2$ são adjacentes se ou $[v_1 u_1 \in E(G_1) \text{ e } v_2 = u_2]$ ou $[v_2 u_2 \in E(G_2) \text{ e } v_1 = u_1]$.

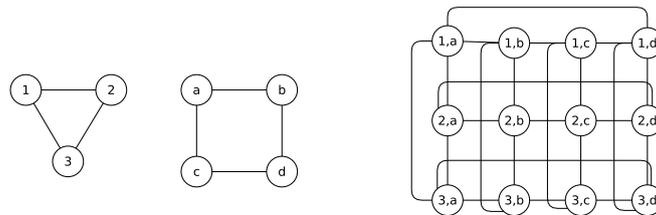


Figura 2. Exemplo de produto cartesiano dos grafos $C_3 \times C_4$.

Sejam G_1 e G_2 dois grafos, o *produto lexicográfico*, denotado por $G_1 \circ G_2$, é formado pelo conjunto de vértices de $V(G_1) \times V(G_2)$, sendo que os vértices $u = u_1, u_2$ e $v = v_1, v_2$ são adjacentes se $u_1 v_1 \in E(G_1)$ ou $u_1 = v_1$ e $u_2 v_2 \in E(G_2)$.

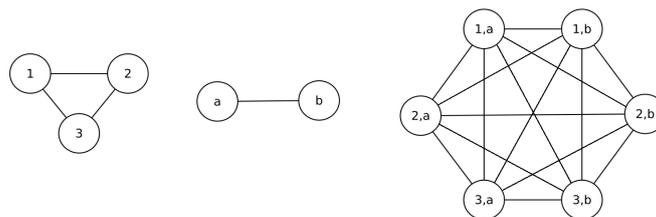


Figura 3. Exemplo de produto lexicográfico dos grafos $C_3 \times P_2$.

2. Algoritmos para caminhos

Em teoria dos grafos são estudados diversos problemas, e entre eles está o problema de caminhos em grafos. Existem inúmeros tipos de caminhos em grafos e dois deles são os caminhos mais longos e os caminhos mais curtos. Em uma definição breve, os caminhos mais longos em grafos não ponderados são caminhos entre pares de vértices que passam pela maior quantidade de vértices possíveis do grafo, e no caso de grafos ponderados são os caminhos entre pares de vértices onde a soma dos pesos é a maior possível. Os caminhos mais curtos em grafos não ponderados são os caminhos entre pares de vértices que passam pela menor quantidade possível de vértices, e no caso de grafos ponderados são os caminhos entre pares de vértices onde a soma dos pesos é a menor possível.

2.1. Matriz de adjacências

Uma *matriz de adjacências* é uma estrutura de dados do tipo matriz usada para armazenar um grafo a partir de suas adjacências. É usada para representar um grafo de maneira que os índices de linhas e colunas da matriz represente os vértices e a posição linha \times coluna da matriz armazena se há ou não aresta incidente aos vértices. Assim, pode-se imaginar que uma matriz de adjacências tem um tamanho $n \times n$, onde n é o número de vértices. Porém, é facilmente detectado que a matriz é idêntica em suas posições acima e abaixo da *diagonal principal*, que é configurada pelas posições onde o índice da linha é igual ao índice da coluna, ou seja, a posição (1, 2) da matriz é idêntica à posição (2, 1). No caso de um grafo simples, verifica-se também que a diagonal principal tem todas suas posições iguais a zero, ou seja, a diagonal principal é nula. Logo, pode ser representada por uma matriz menor de tamanho $\frac{n \cdot (n-1)}{2}$ para grafos simples ou $\frac{n \cdot (n+1)}{2}$ para grafos que contêm laço, pois a diagonal principal não será nula.

Em grafos ponderados, a matriz armazena em suas posições o peso da aresta entre os vértices representados pelo índice da linha e da coluna, diferentemente da matriz para um grafo sem peso, que armazena apenas se há ou não uma aresta incidente aos vértices, como pode ser visto na Figura 4.

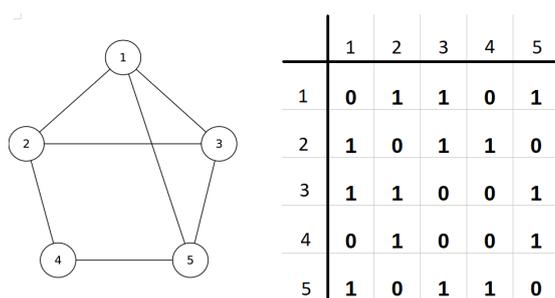


Figura 4. Exemplo de grafo simples e matriz de adjacências $n \times n$.

2.2. Algoritmo de Dijkstra

Um dos algoritmos usados para calcular o caminho mais curto em um grafo é o algoritmo de *Dijkstra*, que foi desenvolvido pelo matemático holandês Edsger Dijkstra, em 1959. O algoritmo é usado para encontrar o caminho mais curto entre todos os vértices de um grafo de arestas com pesos positivos, orientado ou não orientado. A grosso modo, o algoritmo calcula o menor caminho do vértice inicial até seus vértices adjacentes, e dos vértices

adjacentes até os próximos vértices sempre escolhendo a aresta de menor soma entre seu peso e sua distância até o vértice final do grafo.

De maneira mais formal, o algoritmo funciona da seguinte maneira em um grafo G com vértice inicial v_i , onde $d[a]$ é a distância até o vértice a e $\pi[a]$ é o vértice que antecede a . A *distância* armazena o peso calculado para ir até o vértice e o π armazena qual o vértice é seu antecessor no caminho. O algoritmo inicia atribuindo ao vértice inicial, chamado v_i , $d[v_i] = 0$ e $\pi[v_i] = Nil$, e os demais vértices $v \in V(G)$ recebem $d[v] = \infty$, $\pi[v] = Nil$, até o seu vértice final v_f . Assim que estiver preparada a estrutura que irá armazenar o caminho mais curto, dá-se início ao processo de cálculo do caminho mais curto no grafo G representado por uma matriz de adjacências com pesos.

Algoritmo 1: *Dijkstra*(G, s)

Entrada: Um grafo ponderado G e um vértice $s \in V(G)$ inicial.

Saída: O caminho mais curto no grafo G .

```

1 para vértice  $v \in V(G)$  faça
2    $d[v] \leftarrow \infty$ 
3    $\pi[v] \leftarrow NIL$ 
4 fim
5  $d[s] \leftarrow 0$ 
6  $\pi[s] \leftarrow NIL$ 
7  $S \leftarrow \emptyset$ 
8  $Q = V(G)$ 
9 enquanto ( $Q \neq \emptyset$ ) faça
10   $u \leftarrow EXTRACT\_MIN(Q)$ 
11   $S \leftarrow S \cup \{u\}$ 
12  para vértice  $v \in Adj[u]$  faça
13    se ( $d[v] > d[u] + w(u, v)$ ) então
14       $d[v] \leftarrow d[u] + w(u, v)$ 
15       $\pi[v] \leftarrow u$ 
16    fim
17  fim
18 fim
```

O Algoritmo 1 descrito é uma adaptação do algoritmo apresentado por [Cormen et al. 2009]. Nas linhas 1 a 6, o algoritmo inicializa a lista encadeada que armazenará o caminho mais curto através da distância representada por $d[v]$ e o vértice antecessor representado por $\pi[v]$.

A estrutura S utilizada na Linha 7 é uma lista encadeada que armazena os vértices do grafo que já passaram pelo algoritmo; sendo assim, inicia vazia e no fim do algoritmo terá todos os vértices do grafo armazenados. A estrutura Q é também uma lista encadeada que armazena os vértices do grafo que ainda não foram analisados, iniciada com todos os vértices do grafo e no fim fica vazia, pois o algoritmo testa todos os vértices do grafo.

Nas linhas 9 a 18 acontece a iteração que modifica o caminho durante o al-

goritmo, suas funções são: controlar as estruturas S e Q descritas através da função $EXTRACT_MIN(Q)$ que retira e retorna o vértice de menor distância na estrutura passada por parâmetro e verificar se o caminho analisado no momento é menor que o caminho visto anteriormente. Caso seja, é feita a modificação nos campos de distância e vértice antecedente na estrutura que armazena o caminho mais curto. O método $w(u, v)$ calcula e retorna o peso das arestas incidentes aos vértices recebidos por parâmetro, no caso de uma matriz de adjacências retorna o valor da posição da aresta, o vértice u é utilizado como variável e $Adj[v]$ é o conjunto de vértices adjacentes ao vértice v passado por parâmetro.

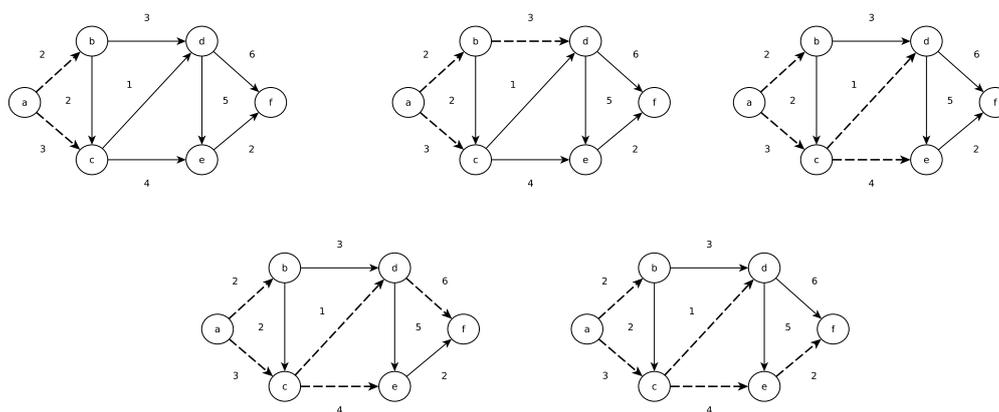


Figura 5. Exemplo de aplicação do algoritmo de Dijkstra.

Na Figura 5, temos uma ilustração do Algoritmo 1 passo a passo. Assim, pode-se perceber como é feita a escolha e a mudança das arestas que compõem o caminho em sua execução e, por fim, apresenta qual seria o caminho mais curto, representado por arestas pontilhadas, iniciando o algoritmo no vértice a do grafo.

Tabela 1. Estrutura que armazena o caminho mais curto do grafo na Figura 5.

<i>Vértices</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
$d[v]$	0	2	3	4	7	9
$\pi[v]$	<i>Nil</i>	<i>a</i>	<i>a</i>	<i>c</i>	<i>c</i>	<i>e</i>

A Tabela 1 apresenta a estrutura de dados usada para armazenar os dados dos vértices no caminho. A estrutura Q descrita no Algoritmo 1 tem exatamente a configuração ilustrada e a função $EXTRACT_MIN(Q)$ vista na Linha 10 do Algoritmo 1, retorna o vértice v que possui o maior $d[v]$ na estrutura Q , ilustrada na Tabela 1. Os valores apresentados na tabela correspondem aos valores do caminho calculado na Figura 5.

2.3. Caminho mais longo

Um grupo de turistas planejando a próxima viagem pelo Brasil, se reuniram para decidir quais cidades do país visitariam. Todos estavam de acordo que queriam visitar o maior número possível de cidades, com a única condição de não ter que repetir nenhuma. Coletaram todos os dados que necessitavam e após dias de estudo, alguns observaram que

era impossível percorrer todas as cidades sem repetir nenhuma e outros notaram que havia diversos percursos diferentes que passavam pelo maior número possível de cidades. A questão agora a ser observada é a mesma levantada por Gallai, Karger, Motwani e Ramkumar [Erdős and Katona 1966]. No exemplo dos turistas, se cada turista escolhesse um tal percurso distinto, teria alguma cidade que todos visitariam? No caso de Gallai, todo grafo conexo tem um vértice que aparece em todos os caminhos mais longos? E a questão respondida por Karger, Motwani e Ramkumar, [Erdős and Katona 1966], visto que é difícil traçar um caminho mais longo que passasse pelo maior número de cidades, será que ao menos poderiam encontrar um caminho mais longo onde visitariam uma fração do maior número possível de cidades?

Gallai [Erdős and Katona 1966] perguntou em 1966, em um colóquio em Tihany, se todo grafo conexo tem um vértice que aparece em todos os caminhos mais longos. A pergunta de Gallai é natural, pois é bem conhecido o fato de que em um grafo conexo, quaisquer dois caminhos mais longos sempre têm um vértice em comum. Porém, pouco tempo depois Walther [Zamfirescu 1976] deu uma resposta a essa questão construindo um grafo conexo provando que nem sempre é verdade que existe um vértice comum a todos os caminhos mais longos. O grafo de Walther está ilustrado na Figura 6, tem 25 vértices e 13 caminhos mais longos de tamanho 21 com intersecção vazia.

Com o surgimento do grafo de Walther, surgiu também a pergunta se o grafo de Walther era o menor possível. No começo dos anos 70, Walther e Voss [Walther and Voss 1974], e Zamfirescu [Zamfirescu 1976], responderam essa questão apresentando o grafo ilustrado na Figura 6 que tem 12 vértices, 9 caminhos de tamanho 9 (veja Tabela 2) que também responde negativamente à pergunta de Gallai.

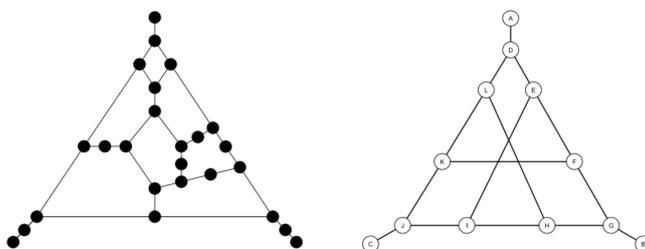


Figura 6. Grafo de Walther e grafo de Walther, Voss [Walther and Voss 1974] e Zamfirescu [Zamfirescu 1976].

2.3.1. Árvores

Uma árvore é um grafo conexo (existe caminho entre quaisquer de seus vértices) que não possui ciclos. Em uma árvore, cada vértice terá 1, ..., n vértices filhos e apenas um vértice pai, com exceção do vértice raiz da árvore que não possui vértice pai. Por volta de 1960, Dijkstra [Rezende 2014] propõe o primeiro algoritmo em tempo polinomial que encontra os caminhos mais longos em uma determinada classe de grafos, o algoritmo calcula os caminhos mais longos em uma árvore em tempo linear. Os Algoritmos 2, 3 e 4, formam uma adaptação do algoritmo de Dijkstra, utilizando para calcular o caminho mais longo, o algoritmo de busca em largura (*BFS*).

Tabela 2. Caminhos mais longos no grafo de Walther, Voss [Walther and Voss 1974] e Zamfirescu [Zamfirescu 1976].

Nº	Caminhos mais longos
01	$\langle A, D, E, I, H, G, F, K, J, C \rangle$
02	$\langle A, D, E, I, H, L, K, F, G, B \rangle$
03	$\langle A, D, E, I, J, K, L, H, G, B \rangle$
04	$\langle A, D, L, H, G, F, E, I, J, C \rangle$
05	$\langle A, D, L, H, I, E, F, K, J, C \rangle$
06	$\langle A, D, L, H, I, J, K, F, G, B \rangle$
07	$\langle B, G, F, K, L, D, E, I, J, C \rangle$
08	$\langle B, G, H, L, D, E, F, K, J, C \rangle$
09	$\langle B, G, H, L, K, F, E, I, J, C \rangle$

Algoritmo 2: $BFS(G, s)$

Entrada: Um grafo árvore G e o vértice inicial s .

Saída: A árvore resultante da busca em largura de G .

```

1 para cada  $u \in V(G) - \{s\}$  faça
2    $cor[u] \leftarrow branco$ 
3    $d[u] \leftarrow \infty$ 
4    $\pi[u] \leftarrow Nil$ 
5 fim
6  $cor[s] \leftarrow cinza$ 
7  $d[s] \leftarrow 0$ 
8  $\pi[s] \leftarrow Nil$ 
9  $Q \leftarrow s$ 
10 enquanto  $Q \neq \emptyset$  faça
11    $u \leftarrow pop(Q)$ 
12   para cada  $v \in Adj[u]$  faça
13     se  $cor[v] = branco$  então
14        $cor[v] \leftarrow cinza$ 
15        $d[v] \leftarrow d[u] + 1$ 
16        $\pi[v] \leftarrow u$ 
17        $push(Q, v)$ 
18     fim
19   fim
20    $cor[u] \leftarrow preto$ 
21 fim
```

O Algoritmo 2 calcula a altura de todos os vértices de uma árvore G a partir da busca em largura, onde para todo $v \in V(G)$, $d[v]$ corresponde à sua altura na árvore resultante da busca em largura. O algoritmo é gerenciado por uma fila Q que armazena a sequência em que os vértices serão analisados.

As cores também são importantes para o processo, onde a cor branca significa um vértice ainda não descoberto pelo algoritmo, a cor cinza sinaliza um vértice que foi

descoberto, porém não finalizado e a cor preta um vértice já finalizado.

Algoritmo 3: *Muda_raiz*(G, u)

Entrada: Um grafo árvore G e o vértice folha a ser raiz de G .

Saída: A árvore resultante da mudança de raiz.

```

1  $v \leftarrow \pi[u]$ 
2  $direita[u] \leftarrow v$ 
3 enquanto  $\pi[v] \neq Nil$  faça
4   se  $direita[v] = Nil$  então
5      $direita[v] \leftarrow \pi[v]$ 
6   fim
7   senão
8      $esquerda[v] \leftarrow \pi[v]$ 
9   fim
10   $v \leftarrow \pi[v]$ 
11 fim

```

O Algoritmo 3 muda a raiz do grafo árvore G pelo vértice folha u , onde u é o vértice de maior altura na árvore resultante da busca em largura. Como u sempre será folha, logo o algoritmo não trata quando u tem filhos. O parâmetro $\pi[v]$ utilizado é resultante da busca em largura feita previamente.

Algoritmo 4: *CML_arvore*(v)

Entrada: Uma árvore G e raiz da árvore v

Saída: O caminho mais longo da árvore G .

```

1  $BFS(G, v)$ 
2  $muda\_raiz(G, u)$  -  $u$  é o vértice de maior distância resultante do
    $BFS$ 
3  $BFS(G, u)$ 
4  $t$  é o vértice de maior distância, resultante do  $BFS$ 
5 enquanto  $t \neq Nil$  faça
6    $F \leftarrow t$ 
7    $t \leftarrow \pi[t]$ 
8 fim
9 retorna  $F$ 

```

O Algoritmo 4 calcula o caminho mais longo na árvore G . O algoritmo faz a chamada dos algoritmos 2 e 3 vistos anteriormente e faz uso de uma fila F para armazenar o caminho mais longo. Os parâmetros $\pi[v]$ e distância utilizados são calculados no algoritmo de busca em largura. Na Figura 7 temos o exemplo de funcionamento do Algoritmo 4 para uma árvore com 8 vértices. O vértice pontilhado é o vértice calculado pelo Algoritmo 2 de maior altura e o caminho mais longo retornado pelo algoritmo está também representado por arestas pontilhadas.

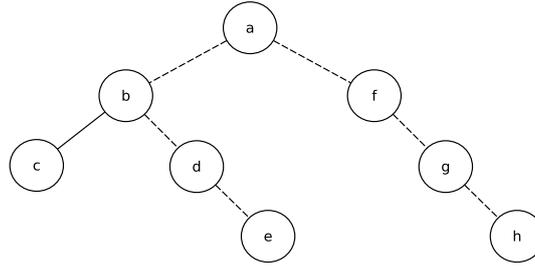


Figura 7. Exemplo de árvore e seu caminho mais longo.

2.3.2. Prisma complementar

Seja G um grafo simples e \overline{G} seu complemento. O prisma complementar de G denotado por $G\overline{G}$ é o grafo formado a partir da união disjunta de G e \overline{G} , adicionando as arestas para um emparelhamento perfeito entre os vértices correspondentes de G e \overline{G} . O complemento \overline{G} é o grafo cujo conjunto de vértices é $V(G)$ e cujas arestas são os pares de vértices não adjacentes de G . Considere n é o número de vértices de G , denotado por $|V(G)|$.

Proposição 1. *Se $G\overline{G}$ é um grafo prisma complementar, então $G\overline{G}$ tem $(\frac{n \cdot (n-1)}{2} + n)$ arestas.*

Demonstração. Seja $G\overline{G}$ um grafo prisma complementar. Pela definição de prisma complementar, temos que G é um grafo com n vértices e \overline{G} seu complemento. Iremos provar por indução que o número de arestas de um grafo G e seu complemento \overline{G} é $(\frac{n \cdot (n-1)}{2})$. Assumimos que G é um grafo com n vértices isolados. Temos, pela definição de complementar, que \overline{G} será um grafo completo K_n com $(\frac{n \cdot (n-1)}{2})$ arestas. Seja m o número de arestas de G . Por hipótese, o complemento \overline{G} terá $(\frac{n \cdot (n-1)}{2}) - m$ arestas. Portanto, temos que se o número de arestas de G é $m+1$, logo o número de arestas de \overline{G} é $(\frac{n \cdot (n-1)}{2}) - m - 1$, pois a cada aresta e adicionada em G , e deixa de ser uma aresta de \overline{G} , iterando até que G passe a ser um K_n e \overline{G} um grafo com n vértices isolados, com $(\frac{n \cdot (n-1)}{2})$ arestas. Pela definição de prisma complementar, para todo vértice $v \in G$ e $\bar{v} \in \overline{G}$, existe uma aresta $v\bar{v}$ adicionando assim n ao número de arestas de $G\overline{G}$. Logo, o número de arestas de um grafo prisma complementar $G\overline{G}$ é $(\frac{n \cdot (n-1)}{2} + n)$. \square

Proposição 2. *Se $P_n\overline{P_n}$ é um grafo prisma complementar com $n > 5$, então o maior ciclo C_i em $P_n\overline{P_n}$ tem tamanho $i = 2n$ e $P_n\overline{P_n}$ é hamiltoniano.*

Demonstração. Seja P_n um caminho de u até v com $n > 5$. Pela definição de complemento, $\overline{P_n}$ é um grafo completo K_n , onde $P_n \not\subseteq K_n$. Assim, para cada vértice $\bar{t} \in \overline{P_n}$, com exceção de \bar{u} e \bar{v} , $g(\bar{t}) = n - 3$ e $g(\bar{u}) = g(\bar{v}) = n - 2$. Logo, temos que $\overline{P_n}$ é um grafo de grau n e para todo vértice $\bar{s} \in \overline{P_n}$, $g(\bar{s}) \geq \frac{n}{2}$. Sendo assim, pelo teorema de Dirac [Dirac 1952], $\overline{P_n}$ é hamiltoniano, possuindo assim um ciclo hamiltoniano. Desse modo, existe um caminho hamiltoniano $P_i \in \overline{P_n}$ de \bar{u} até \bar{v} . Pela definição de grafo prisma complementar para todo $v \in P_n$ e $\bar{v} \in \overline{P_n}$ a aresta $v\bar{v} \in P_n\overline{P_n}$. Assim temos que $P_n \cup v\bar{v} \cup P_i \cup u\bar{u}$ é um ciclo C_i que passa por todos os vértices de $P_n\overline{P_n}$. Como

$P_n\overline{P_n}$ tem $2n$ vértices, C_i tem $i = 2n$. Logo, C_{2n} é um ciclo hamiltoniano e $P_n\overline{P_n}$ é hamiltoniano. \square

Algoritmo 5: $MC_Prisma(G)$

Entrada: Um grafo $P_n\overline{P_n}$ dos vértices de a até n com $n > 5$.

Saída: O maior ciclo em $P_n\overline{P_n}$.

```

1  Insira os vértices em um vetor de tamanho  $n$  de  $\bar{a}$  até  $\bar{n}$ 
2  enquanto  $i \neq n$  faça
3       $u \leftarrow vet[i]$ 
4      se  $(i + 1) = n$  então
5           $F \leftarrow u; cor[u] \leftarrow preto$ 
6           $i \leftarrow 2; u \leftarrow vet[2]$ 
7      fim
8      se  $i \bmod 2 \neq 0$  e  $(i + 2) = n$  então
9           $F \leftarrow u; cor[u] \leftarrow preto$ 
10          $t \leftarrow vet[2]$ 
11          $v \leftarrow vet[n - 1]$ 
12          $F \leftarrow t$ 
13          $cor[t] \leftarrow preto$ 
14          $F \leftarrow v$ 
15          $cor[v] \leftarrow preto$ 
16          $i \leftarrow 4$ 
17          $u \leftarrow vet[4]$ 
18     fim
19      $t \leftarrow vet[i + 2]$ 
20     se  $i \bmod 2 = 0$  e  $cor[t] \neq branco$  então
21          $F \leftarrow u; cor[u] \leftarrow preto$ 
22          $v \leftarrow vet[n]$ 
23          $F \leftarrow v$ 
24          $cor[v] \leftarrow preto$ 
25          $i \leftarrow n$ 
26     fim
27     senão
28          $F \leftarrow u$ 
29          $cor[u] \leftarrow preto$ 
30          $i \leftarrow i + 2$ 
31         se  $i = n$  então
32              $v \leftarrow vet[n]$ 
33              $F \leftarrow v$ 
34              $cor[v] \leftarrow preto$ 
35         fim
36     fim
37 fim
38 retorna  $P_n \cup u\bar{u} \cup F \cup v\bar{v}$ 

```

O Algoritmo 5 inicia inserindo todos os vértices de \overline{P}_n em um vetor de vértices com tamanho n , iniciando do vértice \bar{a} até \bar{n} . Este vetor é usado para gerenciar o andamento do algoritmo armazenando e atualizando as características dos vértices. A partir da Linha 2 temos a iteração que calcula o caminho entre \bar{a} e \bar{n} e todos os casos possíveis. A iteração é finalizada quando o caminho calculado chega a \bar{n} . Por fim, é feita a concatenação dos caminhos de P_n e \overline{P}_n , e das arestas de ligação $a\bar{a}$ e $n\bar{n}$.

Como visto, o Algoritmo 5 calcula o maior ciclo C_{2n} em grafos $P_n\overline{P}_n$, com $n > 5$. Temos pela proposição 2 que este C_{2n} é um ciclo hamiltoniano, pois passa por todos os vértices de $P_n\overline{P}_n$. Isto posto, é facilmente visto que se retirarmos qualquer aresta entre dois vértices u e $v \in C_{2n}$, tem-se deste modo um caminho mais longo entre u e v .

2.3.3. Grafos simples com pesos não negativos

Algoritmo 6: *Dijkstra_2(G, s)*.

Entrada: Um grafo simples G com pesos não negativos.

Saída: Todos os caminhos mais longos do grafo G .

```

1  para vértice  $v \in V(G)$  faça
2  |    $d[v] \leftarrow 0$ 
3  |    $\pi[v] \leftarrow NIL$ 
4  |    $cor[v] \leftarrow branco$ 
5  fim
6   $S \leftarrow \emptyset$ 
7   $Q = V(G)$ 
8  enquanto ( $Q \neq \emptyset$ ) faça
9  |    $u \leftarrow EXTRACT\_MAX(Q)$ 
10 |    $S \leftarrow S \cup \{u\}$ 
11 |   para vértice  $v \in Adj[u]$  faça
12 |   |   se ( $d[v] > d[u] + w(u, v)$  e  $cor[v] \neq preto$ ) então
13 |   |   |    $d[v] \leftarrow d[u] + w(u, v)$ 
14 |   |   |    $\pi[v] \leftarrow u$ 
15 |   |   |    $cor[v] \leftarrow cinza$ 
16 |   |   fim
17 |   fim
18 |    $cor[u] \leftarrow preto$ 
19 fim

```

O Algoritmo 6 foi desenvolvido para calcular e listar todos os caminhos mais longos presentes em grafos simples com pesos não negativos. O algoritmo é uma adaptação do algoritmo de Dijkstra descrito anteriormente. A função $EXTRACT_MAX(Q)$, vista na Linha 9 do Algoritmo 6, é uma função que retorna o índice na matriz de adjacências, do vértice de maior distância da lista encadeada Q . Por fim, são listados todos os caminhos que são armazenados na fila encadeada S ao fim de cada iteração.

Na Figura 8, temos um exemplo de grafo simples, não orientado e com arestas de peso não negativos. Na Tabela 3, temos os maiores caminhos do grafo ilustrado na figura.

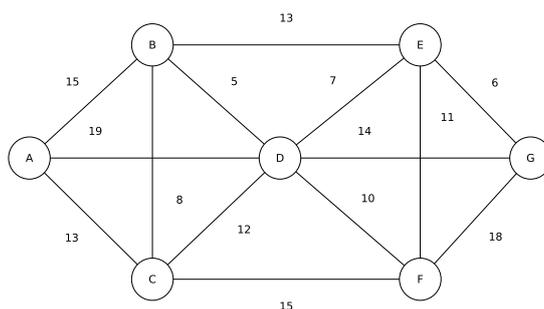


Figura 8. Exemplo de grafo simples com pesos não negativos.

Tabela 3. Caminhos mais longos presentes no grafo da Figura 8.

Nº	Caminhos mais longos
01	< A, D, G, F, C, B, E >
02	< B, E, D, A, C, F, G >
03	< C, F, G, D, A, B, E >

2.3.4. Grafos cordais formados por grafos completos K_n

Um grafo, G é dito *cordal* quando todo ciclo $C_i \in G$, com $i > 3$, tem pelo menos uma corda. Ou seja, se todo ciclo sem corda (*csc*) pertencente a G é um ciclo C_3 , então G é um grafo cordal.

Seja G um grafo simples formado por um grafo completo K_n , e uma árvore de raiz v para todo vértice $v \in K_n$, diz-se que G é um grafo cordal K_n com n subárvores. Na Figura 9, temos um exemplo de grafo cordal e de um grafo cordal com 4 subárvores, formado por um K_4 com as subárvores de seus vértices e um de seus caminhos mais longos destacados pelas arestas pontilhadas.

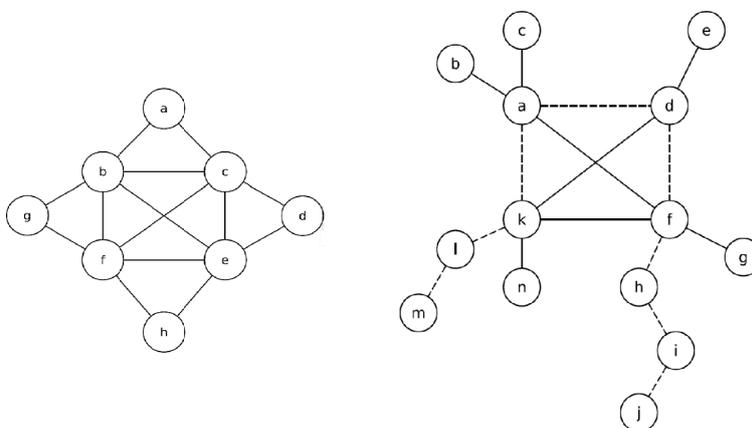


Figura 9. Exemplo de grafos cordais.

Algoritmo 7: $CML_Kn(G, u, v)$

Entrada: Um grafo K_n , o vértice inicial u e final v .

Saída: Um caminho mais longo em K_n de u até v .

```

1 Todo vértice  $t \in K_n$  inicia com cor branca.
2  $t \leftarrow u$ 
3 enquanto  $t \neq v$  faça
4     se  $cor(Adj[t]) = branco$  e  $Adj[t] \neq v$  então
5          $F \leftarrow t; cor(t) \leftarrow preto$ 
6          $t \leftarrow Adj[t]$ 
7     fim
8     senão
9          $F \leftarrow t; cor(t) \leftarrow preto$ 
10         $F \leftarrow v; cor(v) \leftarrow preto$ 
11         $t \leftarrow v$ 
12    fim
13 fim
14 retorna F

```

O Algoritmo 7 calcula o caminho mais longo entre dois vértices u e v em grafos completos K_n utilizando uma fila F para armazenar o caminho e a propriedade de grafos completos, que diz que há uma aresta entre quaisquer pares de vértices.

Algoritmo 8: $CML_Cordal(G)$

Entrada: Um grafo K_n com n subárvores.

Saída: Um caminho mais longo.

```

1 para todo  $u \in K_n$  faça
2      $BFS(S_v, v)$ 
3 fim
4  $u$  - vértice em  $K_n$  de maior distância no  $BFS$ .
5  $v$  - vértice em  $K_n$  de segundo maior distância no  $BFS$ .
6  $F \leftarrow CML\_arvore(S_u, u)$ 
7  $F \leftarrow CML\_Kn(K_n, u, v)$ 
8  $F \leftarrow CML\_arvore(S_v, v)$ 
9 retorna F

```

O Algoritmo 8 usa o Algoritmo 2 em toda subárvore S_v iniciada pelos vértices $v \in K_n$. Desta forma, os vértices adjacentes a v que também pertencem a K_n não pertencem a S_v . Os vértices u e v nas linhas 3 e 4, são vértices de K_n onde o BFS calculou os maiores valores de d , localizando assim quais são as maiores subárvores do grafo. Feito isso, o algoritmo concatena os caminhos mais longos das subárvores e o caminho mais longo do grafo completo fazendo as chamadas dos algoritmos 4 e 7 vistos anteriormente.

3. CONCLUSÕES

No decorrer do trabalho, vimos que há uma grande dificuldade de encontrar os caminhos mais longos em grafos, dado que é um problema NP-completo. Esse fato dificulta o

desenvolvimento de algoritmos em tempo polinomial dado à complexidade do problema.

Visando determinar a dificuldade de problemas menores, os estudos foram iniciados com análises de classes mais específicas de grafos e desenvolvendo alguns algoritmos que nos auxiliaram a entender o comportamento dos caminhos mais longos nestas classes.

A princípio foram estudados, grafos ponderados e grafos prismas complementares onde conseguimos resultados teóricos sobre os tamanhos dos caminhos e o tamanho dos grafos e algoritmos que listam os caminhos mais longos. Em sequência, foram estudados os grafos árvores, grafos completos e grafos cordais cujos algoritmos calculam um caminho mais longo nos diferentes tipos de grafos.

Como trabalhos futuros, seria interessante alterar os algoritmos para que funcionem em tempo menor e encontrar resultados teóricos e algoritmos para outros produtos de grafos e outras classes de grafos, como grafos cordais e k -árvores.

Referências

- Cormen, T. H., Leiserson, C. E., Rivest R. L., and Stein, C. (2009). *Introduction to algorithms*. MIT press, 3th edition.
- Bondy, J. A. and Murty U. R. (2008). *Graph Theory, Graduate Texts in Mathematics*. Springer.
- Dirac, G. A. (1952). “Some theorems on abstract graphs.” Proceedings of the London Mathematical Society 3.1, pages 69–81.
- Erdős, P. and Katona G. (1966). Theory of Graphs. Proceedings of the Colloquium held at Tihany, Hungary. Academic Press, New York. Problem 4 (Gallai T.), page 362.
- Rezende, S. F. (2014). “Caminhos mais longos em grafos”. Dissertação de mestrado, USP.
- Walther, H. and Voss H. J. (1974). *Über Kreise in Graphen*. VEB Deutscher Verlag der Wissenschaften, Berlin.
- Zamfirescu, T. (1976). *On longest paths and circuits in graphs*. Mathematica Scandinavica, Vol. 38, pages 211–239.

Navegação Indoor - Um estudo de caso

Leandro da S. Taddeo, Paulo H. F. Alves, Francisco Klever A. Sobrinho

Centro Universitário 7 de Setembro - Fortaleza, CE - Brasil

{leandrotaddeo, fk.a316}@gmail.com, paulo_he14@hotmail.com

Abstract. *This work presents a mobile mobility project that aims to find and find the best path to the destination indoors. In this prototype it is shown that a user can find a QR Codes usage code to determine their current position. The long term of the work should present a theoretical basis that will serve as the basis for the development of the prototype, as well as the detailed development of the same, such as its architecture and particularities. A final chapter will be presented in the case of the study, where the prototype will be applied to a real context, using the Centro Universitário 7 de Setembro and the Clóvis Beviláqua Forum as study areas.*

Resumo. *Este trabalho apresenta um estudo de caso sobre um projeto de aplicação mobile que tem por principal objetivo auxiliar pessoas a se localizarem e encontrarem o melhor caminho para determinado destino em ambientes fechados. Neste protótipo, é demonstrado como um usuário pode localizar um destino utilizando-se de Códigos de Resposta Rápida (QR Codes) para determinar sua posição atual. Ao longo do trabalho será apresentado a fundamentação teórica que servirá de base para o desenvolvimento do protótipo, bem como o detalhamento do desenvolvimento do mesmo, como sua arquitetura e particularidades. No último capítulo será apresentado o estudo de caso, onde o protótipo será aplicado a um contexto real, se utilizando do Centro Universitário 7 de Setembro e do Fórum Clóvis Beviláqua como áreas de estudo.*

1. Introdução

Muitas vezes, encontrar um destino em um ambiente interno pode ser uma tarefa de difícil execução, mais ainda quando não se conhece o lugar ou o setor que se quer ir. Muitas pessoas pedem informações para pessoas próximas ou procuram informações em placas e letreiros informativos. Contudo, estes podem ser difíceis de interpretar, não serem uma representação fiel da realidade e na maioria das vezes não acompanharem com a rapidez necessária as alterações que sofre a estrutura.

As pessoas precisam muitas vezes se localizar rapidamente em locais com estrutura física complexa, mas esse não é o principal foco principal das aplicações de rotas atuais, como o Google Maps ou Waze, onde só é disponibilizada a rota para veículos e transportes públicos. Essa situação pode acontecer com qualquer um, desde alunos novatos em uma universidade até consumidores fazendo compras em um shopping recém-inaugurado.

O fato é que o estilo de vida moderno está cada vez mais sedentário, e com isso, muitas pessoas acabam passando boa parte do seu tempo em espaços internos, onde não chega o sinal do Sistema de Posicionamento Global (GPS), nem câmeras fotográficas

dos aviões e onde não entram cartógrafos. Essa é também uma das razões para a necessidade de um sistema de posicionamento direcionado para ambientes internos.

Segundo um estudo da *Strategy Analytics (2018)*, as pessoas passam 80-90% do seu tempo em ambientes internos. O estudo ainda indica que cerca de 70% das chamadas efetuadas e 80% das transmissões de dados são efetuados a partir de ambientes internos.

Para contornar esses problemas, surgiram estudos sobre a navegação *indoor* que utilizam técnicas de navegação para localizar pessoas ou objetos em locais fechados. De acordo com as necessidades específicas de localização *indoor*, a abordagem clássica do GPS é inadequada, uma vez que este é inoperacional em espaços internos.

A navegação *indoor* levanta assim desafios adicionais quando comparada com a navegação *outdoor*. O interesse no *indoor* tem-se expandido significativamente, e como tal, diversos métodos têm vindo a ser estudados e apresentados, com resultados, custos e contextos bastante diferentes entre si. Fora do País já se pode encontrar alguns produtos implantados em shoppings, museus, e aeroportos, mas aqui no Brasil essa área ainda está sendo muito estudada, encontrando-se apenas protótipos em algumas universidades em teses de mestrado e doutorado.

É sobre uma dessas ferramentas que este trabalho trata, onde será apresentado um protótipo de aplicativo baseado em navegação *indoor* que auxilie pessoas em ambientes fechados a acharem o caminho mais curto para o seu destino, com a utilização de *QR Codes* para determinar a sua localização e guiá-las através de uma rota.

Ao final do trabalho será apresentado um estudo de caso onde o protótipo será utilizado no Centro Universitário 7 de Setembro e no Fórum Clóvis Beviláqua. Neste estudo será apresentado o passo a passo da implantação desde aplicativo nesses locais, bem como os resultados obtidos com o seu uso, por meio de feedback de usuários entrevistados.

2. Navegação Indoor

2.1 Representações Indoor

Um ambiente *indoor* é um espaço composto de injunções complexas – como corredores, salas, escadas, elevadores – dentro de um ou vários edifícios. A representação adequada dessas injunções é o objetivo do mapa *indoor*, pois os edifícios apresentam propósitos diferentes [OGC et al, 2014].

Existem três formas de representar um ambiente *indoor*: plantas arquitetônicas, plantas baixas e mapas esquemáticos [Nossum, 2013]. Plantas arquitetônicas são mapas que apresentam informações sobre a estrutura do edifício, como detalhes de infraestrutura, como podemos observar na figura 1 (a). Plantas baixas são uma generalização das plantas arquitetônicas, com menos detalhes. Estas plantas são voltadas a um público geral, que tem necessidade apenas de conhecer o ambiente [Sarat, 2015]. A Planta Arquitetônica e a Planta Baixa permitem que o usuário visualize mais de uma estrutura da edificação por vez, quando os pisos podem ser apresentados em perspectiva para que um andar sobreponha o outro no mapa [Nossum, 2013], conforme figura 1 (b).

As plantas esquemáticas são definidas pelo *Intergovernmental Information Systems Advisory Council* como “um mapa simplificado preocupado com a precisão

topológica” [Sarot, 2015], ou seja, o objetivo desse mapa é abstrair as informações para uma rota onde as linhas representam os corredores e os símbolos pontuais são os compartimentos (salas, refeitórios, auditórios, etc.). Para criar mapas esquemáticos as rotas que ligam as diferentes posições de um ambiente são selecionadas, esquematizadas e recebem uma simbologia. Essa esquematização pode incluir a simplificação, o deslocamento e mudança de direção das linhas [Farias e Delazari, 2017], conforme podemos observar na 1 (c), onde as linhas representam os corredores de um prédio e os símbolos pontuais são as salas.



Figura 1. Representação indoor.

2.2 Sistemas de Posicionamento

Sistemas de posicionamento determinam a localização de um objeto ou pessoa em um determinado local. Existem diversos sistemas de posicionamento que nos ajudam a nos localizar e o mais conhecido deles é o GPS.

Existem três tipos de sistemas: satélite, celular e os utilizados em ambientes fechados [Kupper, 2005]. O primeiro se caracteriza pela sua ampla área de cobertura e sua precisão e disponibilidade, contudo, sua utilização não é viável em locais fechados como prédios, estacionamentos cobertos ou áreas com pouca visibilidade do céu [Lei Fang et al, 2005].

O posicionamento por celular corresponde às tecnologias de redes móveis, onde a mais conhecida é o Sistema Global para comunicação Móvel (GSP). Este geralmente funciona bem em ambientes fechados, mas não é muito preciso.

Por último, o posicionamento em ambientes fechados ocorre predominantemente em edifícios, universidades, *shoppings*, museus e centros comerciais e existem várias maneiras de se obter a posição de um objeto nesses locais. Os meios mais conhecidos serão apresentados na próxima sessão deste trabalho.

2.3 Formatos de Dados Geográficos

Serão abordados nessa seção os principais formatos de dados disponibilizados para uma área de visualização geográfica. Por ser um trabalho voltado para o mobile, somente formatos que tem suporte para tal serão apresentados.

O crescimento acelerado do uso massivo de dispositivos móveis, nos dias de hoje, leva-nos a enfrentar desafios de disponibilização da informação nos mesmos. Estes desafios se juntam com a dificuldade de encontrar um equilíbrio entre aquilo que é possível mostrar na tela de um dispositivo móvel e aquilo que é essencial para que a informação fique o mais perceptível possível [Simões, 2015]. A informação geográfica não foge à regra e do ponto de vista da sua visualização, é importante que os dados

geográficos sejam mostrados de uma forma legível, em diferentes níveis de zoom, independentemente do tamanho das telas dos dispositivos.

Toda informação que é trocada, sobretudo através da Web, é formatada de maneira a que todas as aplicações, com tecnologias diferentes que as constituem, as possam ler e decodificar. Este princípio levou ao aparecimento de alguns formatos bem conhecidos como *eXtensible Markup Language* (XML) [W3, 2018] ou *JavaScript Object Notation* (JSON) [JSON, 2018]. A partir das normatizações sobre troca e disponibilização de informação, e com uma maior procura de informação centrada na localização geográfica, foram desenvolvidos formatos de dados normalizados, baseados nos que já existiam anteriormente. Muita da informação codificada nestes formatos, trata-se de mapas ou interações com os mesmos.

2.4 Aquisição

A aquisição se refere às técnicas de obtenção da posição de um objeto em ambientes fechados [Zlatanova et al, 2013]. Dentre as diversas técnicas de aquisição, três se destacam por ter baixo custo de implantação: *Pedestrian Dead Reckoning* (PDR), *Radio Frequency based* (Wi-fi e o *Bluetooth*) e *Image Analysis based* (neste trabalho representado pelos *QR Codes*).

O PDR é uma das técnicas com maior relevância na localização *indoor*, sendo baseado nos populares Sistemas de Localização Inercial (INS). Esta técnica tem como base principal a estimação acumulada do deslocamento (norma e direção) da pessoa/objeto para localizar sua posição atual. Sua precisão varia de acordo com a precisão dos sensores utilizados, as características físicas dos passos do sujeito, o trajeto, o local onde o utilizador leva o dispositivo de medição e variáveis do ambiente (por exemplo, a existência ou não de interferências magnéticas).

A metodologia pode ser dividida em determinação da orientação, detecção do passo, filtragem e estimativa de comprimento do passo [Pratama, Widyawan, e Hidayat, 2012]. São utilizados a bússola ou giroscópio para determinar a orientação e o acelerômetro para detectar o passo. O comprimento do passo é obtido a partir da diferença de dois passos detectados.

Existem diversos sensores que podem ser utilizados para detectar a ocorrência de passos e proporcionar um meio de determinar a distância e direção em que estes foram dados. Alguns exemplos são os acelerômetros, bússolas/magnetômetros e giroscópios.

O segundo método de aquisição é por Rádio Frequência, que é uma família de métodos de localização *indoor* que utilizam uma ou mais propriedades das ondas eletromagnéticas para identificar o posicionamento de uma pessoa ou objeto em um ambiente fechado. Dentre as tecnologias existentes neste grupo, a mais comum e já muito estudada é a WLAN (também conhecida como WI-FI). Contudo, outras técnicas também estão se tornando muito populares, como o *Bluetooth*, a banda ultralarga (UWB), o RFID e o NFC.

O *Wireless Fidelity* (WI-FI) é amplamente utilizado em vários estabelecimentos, e é integrado em diversos dispositivos, como celulares, computadores, e etc. Um ponto de acesso WI-FI transmite continuamente seu sinal e cada dispositivo na vizinhança imediata recebe este sinal. O sinal inclui o endereço *MAC* que identifica o ponto de acesso proveniente daquele sinal e, além disso, todos os dispositivos próximos têm a

capacidade de medir a intensidade deste sinal. Quanto mais forte for a potência do sinal, mais precisa vai ser a posição estimada.

O *Bluetooth* caracteriza-se como uma simples e segura tecnologia de comunicação de curto alcance que pode ser encontrada em toda parte. Semelhante ao WI-FI, essa tecnologia está presente em diversos dispositivos, como smartphones, computadores, *Wearables* (computação vestível) e até mesmo em dispositivos de Internet das coisas (IoT). As principais vantagens dessa tecnologia são a robustez, baixo custo e baixo consumo de energia. A desvantagem dos sistemas de posicionamento baseados em *Bluetooth* é o fato do sistema poder proporcionar uma precisão de apenas 2 a 3 metros, e com um atraso de cerca de 20 segundos.

Por fim, o último método de aquisição que será citado nesse trabalho é o Análise de Imagem. Este método se baseia em determinar a posição através da detecção, no espaço visual a sua volta, de referências visuais diferentes das demais que nos são conhecidas. Geralmente são utilizadas marcas fiduciais para essa detecção, que são imagens artificiais adicionadas ao ambiente real, de modo que auxilie um software a identificar padrões e localizar a pessoa, como etiquetas RFID por exemplo.

As marcas fiduciais mais famosas na detecção da localização do usuário são os *QR Codes*. Estes são um tipo de código de barras 2D utilizados para codificar e decodificar dados a grande velocidade. Surgido no Japão em 1994, esta tecnologia já é amplamente utilizada em diversos setores, chegando ao ponto de ser cada vez mais comum encontrá-los em revistas, jornais, cartões de empresas, produtos de consumo do dia a dia como uma simples embalagem de shampoo.

Para se entender melhor o que é um *QR Code*, se faz necessário apresentar os códigos de barras, que podem ser divididos em dois grandes grupos, os lineares ou 1D e os de matriz ou 2D.

Os primeiros caracterizam-se pela quantidade de informação reduzida que contêm e pelo tipo de informação da mesma, podendo ser apenas do tipo numérico e alfanumérico. O seu propósito é a rotulação de objetos e indexação, sendo por isso bastante utilizados em produtos disponíveis na venda a consumidores [Pereira, 2016].

Em comparação, os códigos 2D podem conter uma quantidade de informação muitíssimo superior, sendo que essa é suportada também no formato binário. O seu propósito também é outro, divergindo da rotulação de objetos, procuram fornecer informação associada aos mesmos. Deste tipo de códigos destacam-se os *QR Codes*, por terem uma maior capacidade e leitura mais rápida em relação aos restantes.

Um *QR Code* pode ser facilmente interpretado pela câmera de um smartphone, por exemplo, e contém informações tanto na vertical quanto na horizontal. Eles armazenam um número consideravelmente maior de informações comparado aos códigos de barras convencionais e são capazes de lidar com vários tipos de dados, ou seja, podem ser utilizados para direcionar o usuário para um link, um cartão de visita, ou até mesmo uma rede social.

A sua capacidade de dados é bastante elevada podendo conter 7.089 caracteres numéricos, 4.296 alfanuméricos e 2.953 *bytes* de informação binária. Além disso, a informação pode ser recuperada mesmo que este se encontre danificado. Existem quatro níveis de recuperação possíveis, sendo eles de 7, 15, 25 e 30 por cento da informação contida no QR Code. [Pereira, 2016].

No processo de leitura os cantos são marcados e estimados, para a informação poder ser decodificada. Primeiramente o algoritmo de leitura faz a detecção dos contornos, detecção da forma, identificação da barra de controle, identificação da orientação, dimensão e densidade dos bits e por último calcula o valor do mesmo.

Atualmente qualquer smartphone equipado com câmara pode fazer leitura de *QR Codes*, necessitando apenas de software adequado. Para a criação desses programas de leitura existem vários *Software Development Kits*, como por exemplo o *Google ZXing* que se encontra disponível num vasto leque de linguagens de programação [Kim et al, 2010].

Mesmo tendo algumas desvantagens, como a necessidade de alterações no ambiente para a inserção dos QR Codes, a iluminação, que pode interferir nos resultados, a necessidade de deslocamento até o código para realizar o escaneamento, dentre outras, em comparação com as outras tecnologias, constatou-se que os *QR Codes* são uma solução simples, de baixo custo de implantação e fácil de ser implementada.

2.5 Modelagem e Visualização

Novas formas de edição e mapeamento de dados geográficos estão se tornando cada vez mais populares, principalmente as soluções colaborativas e voluntárias. Para esta tendência existem diferentes nomes, porém um dos termos mais conhecidos é *Volunteered Geographic Information (VGI)*. Este compreende um conceito que um grupo de utilizadores cria, mantém, agrupa e divulga dados geográficos e espaciais de forma voluntária e colaborativa.

O mais popular e bem-sucedido projeto VGI é provavelmente o *Open Street Maps (OSM)*, uma plataforma de mapeamento online onde utilizadores voluntários podem contribuir e editar informações geográficas de forma colaborativa, ajudando a manter os mapas atualizados.

O OSM é a principal plataforma colaborativa para dados geográficos para mapas outdoor. Esta recebe contribuições de uma comunidade de voluntários que mantêm atualizados dados sobre estradas, edifícios e outros elementos existentes no mundo exterior. As informações contidas nessa plataforma são públicas e estão disponíveis sob a licença *Open Data Commons Open Database License* [Pereira, 2016].

3. Sistema Desenvolvido

Neste capítulo será apresentado todo o processo de desenvolvimento do protótipo de aplicativo mobile desenvolvido para servir de prova de conceito para este trabalho. Serão abordadas as ferramentas utilizadas no processo de desenvolvimento, a arquitetura do software, bem como a infraestrutura geral da aplicação, os componentes que a compõem e por fim as etapas do desenvolvimento, com o passo a passo desde o início do levantamento de requisitos até a implantação e preparação para ele ser utilizado no estudo de caso.

3.1 Ferramentas Utilizadas

Para desenvolver a aplicação foi utilizada a plataforma *Mapbox*. Esta é uma plataforma de criação de mapas personalizados para aplicações mobile e web. Nessa plataforma é possível criar mapas, sistema de navegação e conta com APIs e SDKs para

desenvolvedores implementarem seus aplicativos mobile ou inserir seus mapas em sites e sistemas.

A plataforma *Mapbox* conta com um produto chamado *Studio*, onde desenvolvedores podem estilizar mapas com controle completo sobre o design destes, permitindo criar mapas personalizados. Esta ferramenta foi utilizada para criar o design do mapa e cadastrar todos os pontos de interesse dos locais de estudo, para servirem de possíveis destinos para os usuários.

A *Mapbox* se utiliza do OSM como plataforma de mapeamento *online*, sendo uma alternativa, gratuita e *open-source*, às plataformas comerciais de mapeamento como a *Google Maps* e a *Bing Maps*. A OSM possui a sua própria plataforma de edição dos seus mapas, mas por uma questão didática, optou-se por utilizar o software JOSM para realizar a edição dos mapas comunitários e possibilitar assim, o desenvolvimento do protótipo.

As principais características do software JOSM são: A comunidade que a desenvolve, e nível de documentação existente; O número elevado de *plugins* existentes, que permitem acrescentar funcionalidades ao programa base; O esquema livre de etiquetas, utilizadas no OSM, que se podem atribuir aos objetos criados no programa, podendo assim criar uma nomenclatura própria para os nomes a atribuir aos elementos do edifício; O formato que o programa utiliza para guardar o trabalho desenvolvido. Para além de ter a vantagem de não ser obrigatório enviar para a plataforma os mapas criados, é possível guardar os mesmos em XML [Pereira, 2016].

Para ser feito o georreferenciamento das plantas dos locais de estudo, foi utilizado o *software* QGIS. QGIS é um *software* livre com código-fonte aberto multiplataforma, licenciado segundo a Licença Pública Geral GNU, de sistema de informação geográfica (SIG). Ele permite a visualização, edição e análise de dados georreferenciados, possibilitando também criar mapas para impressão. O QGIS é um projeto oficial da *Open Source Geospatial Foundation* (OSGeo), funciona em Linux, Unix, Mac OSX, Windows e Android e suporta inúmeros formatos de vetores, *rasters*, bases de dados e funcionalidades.

Para o desenvolvimento do protótipo de aplicação mobile para dispositivos Android, foi utilizada a IDE *Android Studio*, que é a IDE oficial da *Google* para criação de aplicações mobile nativas para Android. O *Android Studio* conta com uma série de funcionalidades que auxiliam desenvolvedores a criarem aplicativos cada vez melhores. Dentre elas pode-se destacar um editor visual de layout, analisador de APK, emulador de aplicações, editor de códigos inteligente, sistema de compilação flexível, e perfis de teste em tempo real [Android, 2018].

Como será utilizado *QR Codes* para se obter a localização do usuário na aplicação, é necessário utilizar uma biblioteca de desenvolvimento para realizar o escaneamento desses *QR Codes* dentro da aplicação. Para tal, foi utilizada a ZXING, uma biblioteca de processamento de imagens e códigos de barra 1D e 2D. A biblioteca possui código aberto e foi implementada em JAVA, mas possui suporte para outras linguagens de programação.

Para criar e gerar os *QR Codes* mencionados anteriormente, foi utilizado um site onde é possível criar vários tipos de *QR Codes*. Nele, pode-se escolher o tipo de informação que será decodificada, o seu design e resolução. Este site pode ser encontrado no endereço www.qrcode-monkey.com. Nesta ferramenta foi gerado um *QR*

Code para cada ponto de interesse, para serem espalhados pelos locais de estudo e cadastrados no aplicativo.

3.2 Arquitetura do Sistema

A base da infraestrutura da aplicação é a plataforma *Mapbox*. Segundo o site da plataforma, sua infraestrutura é executada em nove regiões da *Amazon Web Services* (AWS) em cinco continentes.

A AWS é uma plataforma de serviços em nuvem segura, oferecendo poder computacional, armazenamento de banco de dados, distribuição de conteúdo e outras funcionalidades para ajudar as empresas em seu dimensionamento e crescimento [Amazon, 2018]. Essa rede globalmente distribuída é importante para a entrega de baixa latência de mapas em cache e protege contra cenários de falha, como desastres naturais.

O tráfego de serviços da *web* varia de milhares a vários milhares de solicitações por segundo. O *Mapbox* pode lidar com essas solicitações porque cada origem regional é dimensionada de forma independente com base no tráfego em tempo real. O sistema desenvolvido baseia-se na arquitetura apresentada na figura 2.

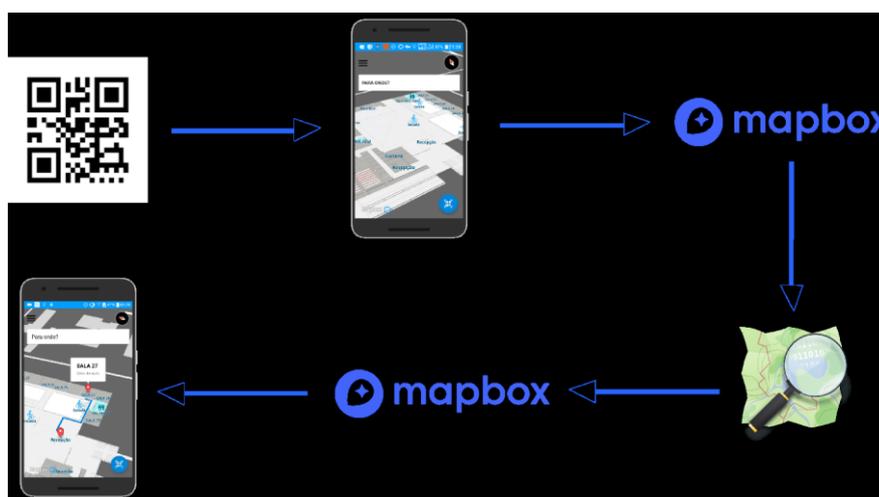


Figura 2 Ilustração da arquitetura do aplicativo.

3.3 Etapas do Desenvolvimento

Para qualquer sistema de navegação os mapas são sempre o ponto de partida. Sem um mapa, o utilizador não sabe em que posição se encontra e muito menos consegue navegar para algum lugar.

Na navegação por GPS indica-se ao sistema de navegação uma localidade ou ponto de destino e o mesmo, com base nos mapas que tem ao seu dispor, gera uma rota que permite satisfazer essa necessidade.

Deste modo, para o mesmo princípio ser aplicado à navegação interior é necessário mapear as divisões dos edifícios. Como a generalidade das construções possui uma planta, em formato digital ou físico, é essa a base de trabalho.

O desenvolvimento do protótipo pode ser dividido em seis etapas a saber, seguindo essa ordem: Georreferenciamento da planta do local; Inserção da planta georreferenciada no *Mapbox Studio*; Cadastro dos pontos de interesse (destinos) no *Mapbox Studio*; Edição do mapa no JOSM; Implementação do código Java; Geração

dos *QR Codes*. Esse desenvolvimento se deu por cerca de 6 meses e foi utilizada a planta do <omitido para revisão cega> como prova de conceito. A seguir será apresentada cada etapa em mais detalhes.

3.3.1 Georreferenciamento da planta

A primeira etapa para se fazer um aplicativo de navegação indoor é georreferenciar a planta, ou seja, tornar suas coordenadas conhecidas num dado sistema de referência. Este processo inicia-se com a obtenção das coordenadas (pertencentes ao sistema no qual se pretende georreferenciar) de pontos da imagem ou do mapa a serem georreferenciados, conhecidos como pontos de controle. Os pontos de controle são locais que oferecem uma feição física perfeitamente identificável, tais como intersecções de estradas e de rios, represas, pistas de aeroportos, edifícios proeminentes, topos de montanha, entre outros.

Para realizar o georreferenciamento da planta da UNI7, foi utilizado o software QGIS, como foi mencionado no subcapítulo Ferramentas utilizadas. Nele, foi inserida a planta em formato PNG do 2º andar do prédio da referida organização.

Em seguida foram escolhidos os principais pontos de referência para possibilitar o georreferenciamento. Normalmente são escolhidas as extremidades da construção, para facilitar o processo posteriormente. Após a escolha das coordenadas (Figura 3), o mapa já está pronto para ser georreferenciado. A ferramenta realiza todo o processo de relacionamento dos pontos geográficos da planta com os mesmos pontos geográficos do mapa real. Ao final do processo, a planta pode ser exportada no formato que será utilizada no *Mapbox Studio*, na próxima etapa.

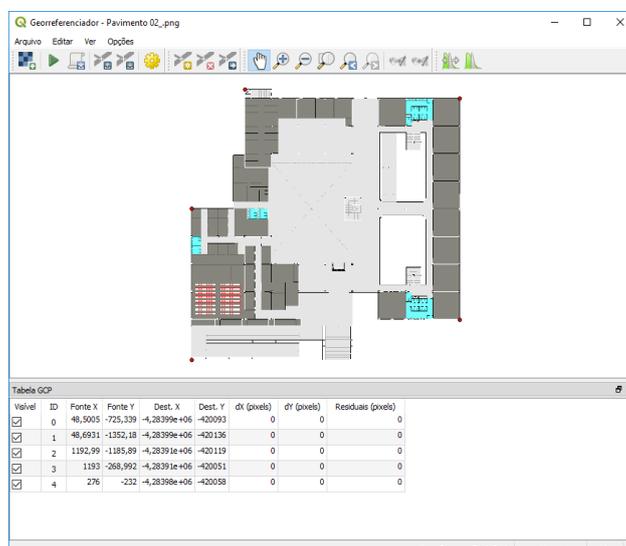


Figura 3. Planta com as coordenadas de georeferenciamento.

3.3.2 Inserção da planta no Mapbox Studio

Após georreferenciar e exportar a planta para o formato TIFF, devemos importar esse arquivo no Mapbox Studio como um novo Tileset. Um tileset é uma coleção de dados vetoriais ou raster divididos em uma grade uniforme de ladrilhos quadrados com 22 níveis de zoom predefinidos. O Tilesets é usado em bibliotecas e SDKs do Mapbox como peça central para tornar os mapas visíveis em dispositivos móveis ou no

navegador. Eles também são o principal mecanismo que usamos para determinar as visualizações de mapa.

Como a imagem já está georreferenciada, ela estará habilitada para visualização, quando seu utilizador estiver nas coordenadas (GPS) do prédio. Em seguida, selecionamos o estilo de visualização do mapa, no nosso caso, um dos estilos disponibilizados pelo Mapbox (*Light*).

Dentro do Editor, pode-se observar que já é possível ver o mapa com o local do edifício da UNI7, mas ainda não é possível ver a planta interna. O motivo é que ainda não foi criada a camada onde será inserido o *Tileset* criado anteriormente.

O próximo passo é criar a camada que receberá o *Tileset* com a planta georreferenciada. Depois de algumas configurações, como o nível de *zoom* que a planta será visualizada na tela, dentre outras, já é possível ver a planta no mapa, com todas as suas dependências.

3.3.3 Cadastro dos pontos de interesse

A próxima etapa no processo de desenvolvimento é cadastrar na plataforma todos os destinos do local de estudo. Para tal, será necessário criar um *Dataset*, que é um conjunto de dados em formato GeoJSON, que armazenará as coordenadas de cada ponto do mapa. Após criado o *Dataset*, se faz necessário selecionar o *Tileset* criado anteriormente para servir de referencial para a criação dos pontos, ficando entre o fundo do mapa e os dados que serão criados a seguir.

Para o referido protótipo, o ponto possui duas propriedades que serão utilizadas posteriormente no código, o nome e a descrição do local. No arquivo GEOJSON é possível visualizar as coordenadas de latitude e longitude do ponto criado. Estas coordenadas também serão salvas para depois serem inseridas no código. Após todos os pontos serem criados, o *Dataset* será exportado para o *Tileset* criado anteriormente. A Figura 4 (a) apresenta a planta com todos os pontos de interesse cadastrados.

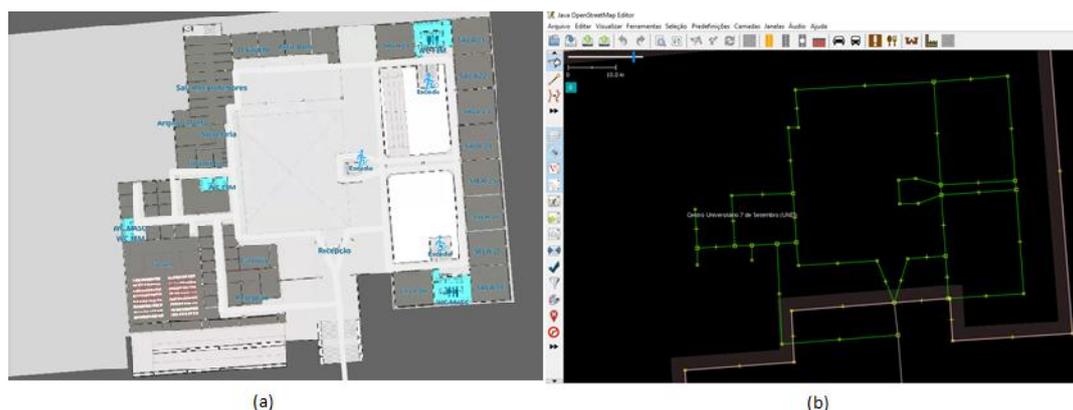


Figura 4 – Mapa com todos os pontos de interesse cadastrados e rotas no JOSM

3.3.4 Edição do Mapa no JOSM

Como o *Mapbox* utiliza o *OpenStreetMap* como plataforma de mapeamento online, se faz necessário editar o mapa nessa plataforma para ser possível criar uma aplicação de localização indoor, visto que o OSM é voltado para navegação *outdoor*. Como foi dito anteriormente, a *OpenStreetMap* possui um site onde é possível aditar os seus mapas

online, mas por questões didáticas foi utilizado o software JOSM, por ser uma ferramenta muito prática de edição de mapas do OSM.

Ao abrir a ferramenta, o primeiro passo é baixar a área que será editada, a ferramenta recomenda que essa área não seja muito grande. Em seguida, pode-se começar a editar o mapa, inserindo pontos, ruas, edificações, etc. No caso do protótipo desenvolvido, foi necessário criar uma via exclusiva para pedestres e inseri-la dentro do prédio. Em todos os corredores ou locais onde se pode trafegar pessoas dentro do prédio, foi inserido uma via para possibilitar, posteriormente, a criação da rota e guiar o usuário até o seu destino, conforme Figura 4 (b).

3.3.5 Implementação do código Java

Uma aplicação Android é muito gráfica, e como tal, a sua programação tem um grande foco nessa característica. Quando é criado um projeto Android, automaticamente é criado uma *Activity*. Uma *Activity* é uma componente da aplicação que fornece uma interface com a qual os utilizadores podem interagir, com um propósito, como enviar um e-mail ou tirar uma fotografia, escolher uma música ou interagir com um mapa.

Toda aplicação Android possui um arquivo de configurações chamado *AndroidManifest.xml*. Este arquivo armazena todas as configurações do aplicativo, como a versão mínima e máxima do Android que será suportada pelo App, dentro outras. Este arquivo é lido antes mesmo da aplicação ser visualizada pelo usuário, e contém todas as permissões que o aplicativo deverá ter para acessar componentes e recursos do dispositivo.

Segundo a documentação do Android, as permissões do sistema são divididas em vários níveis de proteção. Os dois níveis de proteção mais importantes que devem ser considerados são as permissões *normais* e *perigosas*: **Permissões normais** cobrem áreas onde o seu aplicativo precisa acessar dados ou recursos fora da *sandbox* do aplicativo, mas apresenta pouco risco à privacidade do usuário ou à operação de outros aplicativos; **Permissões perigosas** abrangem áreas onde o aplicativo precisa de dados ou recursos que envolvem informações pessoais do usuário ou que podem afetar os dados armazenados do usuário ou a operação de outros aplicativos.

No desenvolvimento do aplicativo, foram criados 4 *arrays* com todos os destinos possíveis cadastrados anteriormente. Cada *array* tem uma funcionalidade específica. O primeiro armazena os nomes de cada destino, o segundo as suas descrições, o terceiro armazena as latitudes de cada ponto e o quarto, a longitude de cada um. O primeiro *array* foi utilizado para criar a lista de destinos que é disponibilizada para o usuário no aplicativo. Quando o usuário escolhe um destino na lista, automaticamente são utilizados os dois últimos *arrays*, dependendo do index, para determinar as coordenadas (latitude e longitude) do ponto no mapa.

Para gerar o *QR Codes* que serão espalhados pelo local de estudo, foi utilizado um site específico para tal, que pode ser localizado pelo endereço www.qrcode-monkey.com. Em cada *QR code* criado foi inserido um texto com o index de cada ponto de interesse para, após ser escaneado, determinar qual a localização do usuário no prédio. Estes foram colados nas portas de todas as salas do nível escolhido, para permitir os usuários navegarem pelo prédio até o seu destino.

4. Estudo de Caso

4.1 Espaços de Estudo

As áreas de estudo são dois locais de amplo espaço, localizados na cidade de Fortaleza no Ceará, conforme podemos observar na figura 5.

Primeiramente foram feitas visitas técnicas para se conhecer a infraestrutura dos edifícios. Em seguida foi feito um mapeamento visual dos ambientes. Nesse mapeamento foram listados todos os Pontos de Interesse (destinos ou POI), como salas, banheiros e possíveis entradas e saídas, bem como escadas e elevadores.

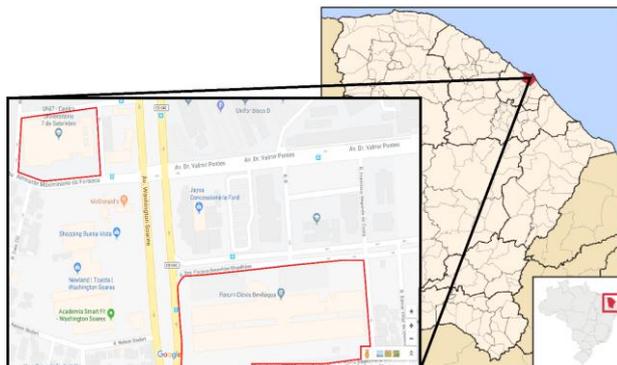


Figura 5 – Mapa de localização dos espaços de estudo.

Normalmente os visitantes destes locais, não têm familiaridade com o uso da estrutura e sofrem com a falta de compreensão do ambiente e com problemas de orientação, que em muitos casos ocasionam o estabelecimento de rotas mentais erradas e gastos maiores de tempo para se determinar a localização dos Pontos de Interesse. Isso justifica a escolha desses locais como espaços de estudo e teste do aplicativo de localização interna.

4.3 Testes, Validações e Resultados Obtidos

A primeira versão MVP do aplicativo foi desenvolvido utilizando a UNI7 como planta para se realizar a prova de conceito. O aplicativo, depois de finalizado, foi apresentado no festival de Software, onde foi demonstrado o uso real do aplicativo para participantes, que aprovaram a ideia, apresentando *feedback* positivo para o aplicativo.

Na Figura 6 é possível observar algumas interfaces da aplicação durante o processo de navegação, com um exemplo do seu uso real. As figuras estão em ordem de uso do aplicativo. Primeiro o usuário escolhe um destino na lista. Em seguida o destino é mostrado no mapa com um ponto vermelho. O usuário procura um *QR Code* mais próximo e o captura utilizando a câmera do seu celular. O aplicativo automaticamente apresenta para o usuário o ponto de destino no mapa, a rota mais curta, que é calculada pelo próprio Mapbox, e a de distância em metros até aquele destino.

Após esses passos o usuário pode se dirigir até o seu destino, capturando os *QR Codes* encontrados pelo caminho para atualizar a sua localização no mapa. Ao fim do percurso é mostrado uma mensagem de sucesso, informando que o usuário chegou ao seu destino. Se por acaso o usuário escolher uma rota diferente, após capturar um *QR Code*, a sua rota é recalculada automaticamente e exibida no mapa.

Antes do início dos testes, foi explicado ao usuário os princípios básicos da aplicação (leitura dos QR Codes e escolha do destino). O utilizador podia encontrar QR

Codes nas portas de praticamente todas as salas daquele andar. Durante os testes foi possível perceber que muitas pessoas se perdem com frequência no local, pois o mesmo possui muitas salas, o que pode tornar a procura muito trabalhosa e cansativa. A grande maioria dos visitantes, quando chegam, só portam consigo um documento com o nome da sala que devem ir, mas não recebem uma orientação clara de como chegar nesse local.



Figura 6 – Telas do aplicativo.

Vários visitantes testaram o aplicativo, e todos comentaram que o aplicativo é muito útil e viável para auxiliar na acessibilidade de pessoas, e ainda perguntaram se o mesmo já está disponível nas lojas para download. Pôde-se observar também nos testes que praticamente nenhum visitante conhecia esse tipo de tecnologia ou tinha visto um QR Code antes, o que deixa a dúvida se a aplicação seria de fato aceita no mercado pela população.

Em entrevista, o secretário executivo do Fórum Clóvis Beviláqua aprovou a ideia, acrescentando que seria de grande utilidade para a organização ter um aplicativo como esse. O mesmo ainda disse que outras soluções já foram propostas, mas este protótipo foi o que mais se aproximou da solução pretendida pela organização.

5. Conclusão e Trabalhos Futuros

Pelo presente trabalho mostramos que criar uma ferramenta offline de navegação indoor, sem o uso do GPS, não é uma tarefa de difícil execução, porém devemos ressaltar que foram encontradas dificuldades na adaptação das plataformas Mapbox e OSM, para serem utilizadas em ambientes *indoor*; A obtenção e adaptação das plantas arquitetônicas dos espaços de estudo para serem utilizadas no protótipo, uma vez que foi necessário a conversão das plantas para o modelo de planta baixa para facilitar a sua visualização por parte dos usuários; E obter as permissões para utilizar e espalhar os QR Codes nos espaços de estudo, onde teve-se que entrar em contato com a diretoria de cada estabelecimento para pedir tal permissão.

Percebemos que não existem aplicações comerciais voltadas para esse público, fazendo que muitas pessoas se percam e acabem perdendo tempo tentando encontrar destinos por meio de rotas mentais e letreiros que podem estar desatualizados.

Os testes realizados com o protótipo obtiveram resultados satisfatórios e esperamos que incentivem outros trabalhos e aplicações, uma vez que o protótipo desenvolvido resolveu satisfatoriamente a problemática proposta, porém ainda resta muitas funcionalidades que podem ser exploradas em versões futuras, tais como: disponibilizar as instruções de navegação para o usuário, onde será mostrada na tela

para onde o usuário deve seguir para chegar ao seu destino, semelhante a um software GPS; disponibilizar a união entre navegação outdoor e indoor; substituição do método de aquisição da localização do usuário (Beacons Bluetooth ou WIFI), onde a principal vantagem será o rastreamento contínuo da posição atual do usuário, que não necessitaria capturar outros QR Codes para atualizar a sua localização; inserção da realidade aumentada e/ou imagens em 360°; inclusão de recursos de acessibilidade para pessoas com deficiência visual e/ou física, como navegação guiada por voz; marketing geolocalizado, escoamento de pessoas para saídas mais próximas em casos de emergência, guardar e localizar veículos em estacionamentos, etc.

6. Referências Bibliográficas

- Android. (2018) <https://developer.android.com/studio/?hl=pt-br>.
- Amazon. (2018) <https://aws.amazon.com/pt/what-is-aws>.
- JSON. (2018) <http://json.org/index.html>.
- Kim, N.H. Lim, H. Jo e H. J. Lee. (2010) “Online Banking Authentication System using Mobile-OTP with QR-code,” em: 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT).
- Küpper, A. (2005) “Location-Based Services: Fundamentals and Operation. John Wiley and Sons”.
- Lei Fang, P. J. A., Montestruque, L., McMickell, M. B., Lemmon, M., Sun, Y., Fang, H., Koutroulis, I., Haenggi, M., Xie, M., and Xie, X. (2005). Design of a wireless assisted pedestrian dead reckoning system - the navmote experience.
- Nossum, A. S. (2013). “Developing a Framework for Describing and Comparing Indoor Maps”, *The Cartographic Journal*, 50(3):218–224.
- OGC, Li, K.-J., Lee, J., Kolbe, T. H., Zlatanova, S., Morley, J., Nagel, C., and Becker, T. (2014). “Open Geospatial Consortium Indoor GML. Technical report”.
- Pratama, A. R., Widyawan, W., and Hidayat, R. (2012). “Smartphone-based pedestrian dead reckoning as an indoor positioning system”.
- PEREIRA, Celso Almeida. (2016) “Sistema De Navegação Para O Interior De Edifícios”, em: Dissertação de Mestrado em Engenharia Informática.
- SAROT, R. V. (2015) “Avaliação de mapas indoor para dispositivos móveis para auxílio à tarefa de orientação”, em: Dissertação de Mestrado PPGCG, 135p, Curitiba.
- W3. (2018) <http://www.w3.org/XML/>.
- Zlatanova, S., Sithole, G., Nakagawa, M., and Zhu, Q. (2013) “Problems in indoor mapping and modelling”, em: In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, volume 40.
- Vazquez, C., SIMÕES, G. (2016) “Engenharia de Requisitos: Software Orientado ao Negócio” em: [S.l.]: Brasport.
- FARIAS, P. P. S., DELAZARI, L. S. (2017) “Cálculo De Rotas Com O Algoritmo Do Caminho Mais Curto” Em: Ambientes Indoor.
- SIMÕES, D. M. (2015) “Navegação indoor baseada na rede WiFi como suporte a serviços baseados na localização: Estudo de caso no campus da UL”.

Protótipo de Monitoramento Remoto da Saúde do Idoso

Cícero Joasyo Mateus de Moura¹, Otávio Calaçã Xavier²

¹Departamento IV – Instituto Federal de Goiás (IFG)
Goiânia – GO – Brasil

²Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

cicerojmm@gmail.com, otavio@inf.ufg.br

Abstract. *The aging population in the world has grown considerably, and the forecast for the year 2050 is around 22% of the world's population. People over 65 years of age tend to face some health-related problems such as falls, respiratory, cardiovascular, and other diseases. In this context, solutions that offer improvements in the quality of life are relevant. This project aims to build a health monitoring system for the elderly, in a way that is economically feasible and technologically possible. The system includes monitoring heart rate, body temperature and detection of falls. For family members or caregivers, a mobile application is offered to monitor data and receive alerts on monitoring situations.*

Resumo. *A população idosa no mundo vem crescendo consideravelmente, e a previsão para o ano de 2050 é que seja em torno de 22% da população mundial. Pessoas acima dos 65 anos de idade, tendem a enfrentar alguns problemas relacionados à saúde, como quedas, doenças respiratórias, cardiovasculares e outras. Neste quadro são relevantes soluções que venham oferecer melhorias na qualidade de vida. Este projeto traz o objetivo de construir um sistema de monitoramento da saúde do idoso, de forma que seja economicamente viável e tecnologicamente possível. O sistema contempla monitoramento da frequência cardíaca, temperatura corporal e detecção de quedas. Para os familiares ou cuidadores é oferecido um aplicativo móvel para acompanhamento dos dados e recebimento de alertas sobre situações do monitoramento.*

1. Introdução

A população idosa do Brasil tem aumentado a cada ano consideravelmente, e assim o aumento da qualidade de vida para esse grupo da sociedade, é alvo de diversos esforços. O objetivo é buscar soluções que possam trazer praticidade, independência, prevenção, auxílio e conforto para a vida das pessoas acima dos 65 anos de idade. No ano de 2010, foi constatado que são cerca de 20 milhões de pessoas que estão dentro da população idosa no Brasil, o que corresponde a cerca de 7,4% da população total [BRASIL 2012]. Segundo dados do IBGE, a expectativa média de vida do brasileiro subiu em 2012, chegando a 74,5 anos [IBGE 2012], o que tende aumentar ainda mais essa parcela da população.

Estudos apontam quais são as principais causas de atendimento e internações de idosos realizados pelos pontos de saúde [Loyola Filho et al. 2004], e dentre eles estão problemas respiratórios, cardiovasculares, ingerir remédios errados e quedas sofridas. Ainda

é importante destacar que no Brasil cerca de 70% dos idosos possuem pelo menos uma patologia crônica, ou seja, um problema de saúde que necessita de constante acompanhamento e monitoramento [Bernardo 2010].

Segundo estudo realizado no ano de 2010, as três maiores frequências de atendimentos hospitalares entre os idosos são: problemas no aparelho circulatório, que afeta 28% dos idosos brasileiros, quedas sofridas que somam 24% ao ano e doenças no aparelho respiratório que afetam cerca de 17% da população idosa [Batista de Góis and Peixoto Veras 2010].

Nesse aspecto, este trabalho apresenta uma proposta que possa reunir as principais necessidades relacionadas à saúde do idoso, provendo um sistema de monitoramento remoto para pessoas acima dos 65 anos, utilizando conceitos como a Internet das Coisas para monitorar batimentos cardíacos, temperatura corporal e detecção de quedas. Juntamente com algoritmos para analisar os dados, oferecendo uma resposta rápida e de forma automática a possíveis alterações na saúde do idoso. Assim emitir alertas para a família ou cuidadores que estejam acompanhando o monitoramento.

Este artigo está organizado da seguinte forma: na seção 2 é apresentado os trabalhos relacionados com esta pesquisa. Na seção 3 encontra-se a fundamentação teórica. Na seção 4 são apresentados os materiais e métodos para a construção do projeto. Os resultados são discutidos na seção 5, seguido da conclusão e das considerações finais na seção 6.

2. Trabalhos relacionados

Diversos esforços em estudos têm sido realizados para criar soluções tecnológicas que possam oferecer suporte aos idosos. Ao realizar o monitoramento de sua saúde, o idoso pode ter maior independência e conforto, além de possibilitar o controle e administração das situações por parte dos familiares e cuidadores.

Em relação ao monitoramento de possíveis quedas que o idoso possa sofrer, o artigo *Fall detection algorithms for real-world falls harvested from lumbar sensors in the elderly population: a machine learning approach* [Bourke et al. 2016] aborda o conceito de tecnologia assistiva, onde sensores identificam a queda do idoso, assim o sistema envia mensagens para parentes ou médicos, para que o auxílio seja feito rapidamente. A identificação da queda é feita através de sensores e algoritmos computacionais de Inteligência Artificial.

Outra solução é descrita no artigo *IoT system for monitoring vital signs of elderly population*, que faz a utilização de sistema embarcado capaz de monitorar os batimentos cardíacos para a prevenção de arritmia cardíaca e também módulos de monitoramento da respiração e da temperatura do corpo, utilizando algoritmos computacionais para identificar padrões e possíveis alterações na saúde da pessoa em monitoramento [Banuleasa et al. 2016].

Uma outra proposta feita no trabalho *We-care: An IoT-based health care system for elderly people* é o desenvolvimento de uma pulseira baseada no conceito de Internet das Coisas, que realiza o monitoramento da frequência cardíaca, temperatura, saturação de oxigênio e sinais eletrocardiográficos, usando informações em tempo real através de sensores internos e ainda notificar os membros da família ou cuidadores sobre mudanças

inesperadas [Pinto et al. 2017].

Levando em consideração os estudos e soluções apresentadas, podemos perceber que se trata de um assunto com importância, relacionando os idosos, familiares, médicos e o sistema de saúde.

3. Fundamentação Teórica

Nesta seção serão abordados conceitos que envolvem a construção deste trabalho, como frequência cardíaca (3.1), temperatura corporal (3.2), tecnologia assistiva (3.3), computação vestível (3.4), internet das coisas (3.5) e outros.

3.1. Frequência Cardíaca

A frequência cardíaca é medida com batimentos por minutos (BPM) e define a quantidade de vezes que o coração pulsa durante um período de tempo. Existem valores definidos em média para a frequência cardíaca, que variam por idade, situações temporárias e condições de saúde. A frequência cardíaca quando baixa, pode indicar problemas metabólicos, doenças cardíaca-vascular, exaustão e outras situações. Já a frequência cardíaca alta pode diagnosticar pressão alta, insuficiência cardíaca, infarto e outras situações de riscos [Viega 2017].

Normalmente um adulto saudável tem a medida entre 72 a 78 BPM como normal e pode variar de acordo com o sexo, sendo que mulheres tendem a ter frequência cardíaca maior. Através de estudo é indicado que idosos apresentam de 6 a 10 BPM menores que um adulto abaixo dos 65 anos [Fechine and Trompieri 2015]. Outra forma de realizar o cálculo da frequência cardíaca máxima (FCM) é subtrair a idade do paciente de 220, assim uma pessoa de 65 anos pode ter FCM de 155 BPM [Fechine and Trompieri 2015].

3.2. Temperatura Corporal

A temperatura corporal é uma combinação de calor gerado e perdido pelo corpo humano. O órgão responsável pela manutenção da temperatura corporal é o Hipotálamo e tem o papel de mantê-la mais ou menos com valor constante [Vinícius 2017].

Em geral, a temperatura normal média de uma pessoa adulta situa-se entre 36,7 e 37 graus *celsius* (°C), podendo variar cerca de 0,6 °C a mais, dependendo do lugar no corpo em que é medida, sendo que abaixo de 34 °C é muito baixa e acima de 37 muito alta.

Pelo valor da temperatura corporal é possível identificar algumas situações de riscos como a hipotermia, que se refere à perda de temperatura corporal, chegando a valores muito baixos. A hipertermia, que é elevação da temperatura corporal que pode caracterizar febre e situações de riscos à saúde [Vinícius 2017].

3.3. Tecnologia Assistiva

O termo *Assistive Technology*, traduzido para o português como Tecnologia Assistiva (TA), foi criado oficialmente em 1988, fazendo parte da legislação dos Estados Unidos, garantindo os direitos das pessoas com limitações físicas, especificando que devem ter acessos aos mesmos serviços e recursos como toda a população [Bersch 2008].

Com o passar dos anos, a TA ganhou maior destaque e um amplo arsenal de aplicações, expandindo as fronteiras e assim podendo ser utilizada para descrever o uso

da tecnologia como ferramenta de auxílio para pessoas com deficiências, idosos, crianças e qualquer um que tenha alguma limitação física. Soluções relacionadas ao uso da tecnologia que venham ajudar a melhorar vidas e auxiliar a execução de alguma atividade do dia-a-dia, podem ser classificadas como TA.

3.4. Computação Vestível

Computação vestível, ou no inglês *wearables*, são sistemas inteligentes que podem ser vestidos por uma pessoa. Estes dispositivos podem estar conectados a Internet ou a outros dispositivos em rede [Sugathan et al. 2013].

O termo surgiu em 1998, no artigo escrito por Sterve Mann [Mann 1998], que define como sendo um computador que está alocado no espaço pessoal do usuário, onde há controle do usuário, porém existe uma constância em sua operação e interação com usuário e o meio.

Dispositivos vestíveis estão cada vez mais presentes no cotidiano das pessoas, e podem ser encontrados através de relógios, roupas, pulseiras, óculos e acessórios em geral. Como o dispositivo proposto neste trabalho, que é vestível e possui um sistema de computação executando tarefas que interagem com o usuário.

3.5. Internet das Coisas

A Internet das Coisas, ou em inglês, *Internet of Things* (IoT), é um termo que foi criado em 1999 por Kevin Ashton, co-fundador do *Auto-ID Center* do *Massachusetts Institute of Technology* (MIT).

Esse termo foi criado pois o seu autor previa que todos os objetos físicos estariam conectados a Internet, com a capacidade de capturar informações por meio de sensores que detectam o ambiente em que estão inseridos, podendo assim identificar e compreender o mundo independentemente das pessoas [Gubbi et al. 2013].

Este trabalho tem como um dos objetivos a criação de um dispositivo para a Internet das Coisas, onde se possa realizar o monitoramento remoto da saúde de pacientes idosos utilizando sensores e conexão com a Internet.

3.6. Microcontrolador

Microcontroladores são sistemas computacionais inteiros contidos em apenas um único chip de circuito integrado [Akyildiz and Wang 2005]. Trata-se de um *hardware* específico para executar funções dedicadas através de uma programação feita em uma linguagem suportada pelo circuito.

As grandes vantagens dos microcontroladores são o seu preço acessível, pouco consumo de energia e adequação para sistemas de tempo real, pois executam tarefas de forma dedicada. Somando essas características e seus componentes agregados, é possível o desenvolvimento de diversas aplicações, em várias áreas diferentes e com muitas funcionalidades.

4. Materiais e métodos

O dispositivo de monitoramento é conectado a rede Wi-Fi local e para isso é utilizado o padrão 802.11 b/g/n¹. Os dados são salvos em plataforma de armazenamento em nuvem, e para serem consultados, o *smartphone* deve estar conectado a Internet. O acompanhamento do paciente pode ocorrer de forma remota e em qualquer lugar com acesso a Internet. O protocolo de comunicação dos dados é o HTTP² e seus métodos de transmissão POST e GET são utilizados, possibilitando salvar e recuperar informações.

Conforme apresentado na Figura 1 o dispositivo deve ser vestido pelo paciente em monitoramento. Todos os dados coletados pelos sensores serão encaminhados para a nuvem, onde serão tratados, analisados e enviados para armazenamento também em nuvem, no banco de dados em tempo real. Assim os dados são acessados pelo aplicativo móvel, podendo consultar histórico e acompanhar o monitoramento.

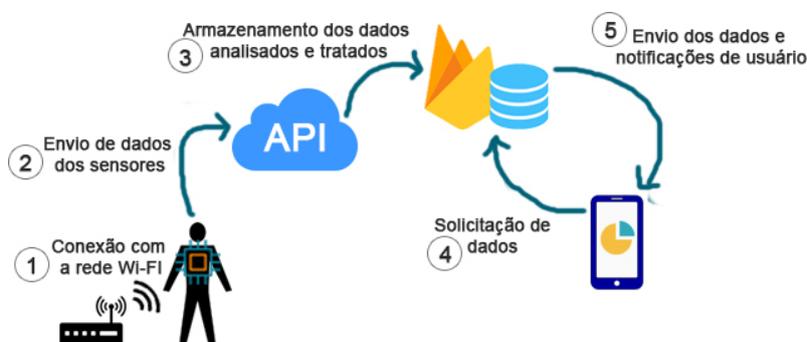


Figura 1. Figura ilustrativa da arquitetura do sistema de monitoramento proposto

O dispositivo funciona de forma reativa, e emite alertas para o usuário sobre estados da saúde do paciente. Em caso de temperatura acima ou abaixo do normal, anormalidades no batimento cardíaco ou até mesmo quando detectado uma queda do idoso.

Todos os materiais utilizados e estudados no desenvolvimento do projeto serão descritos individualmente abaixo, evidenciando suas funcionalidades e características.

- NodeMCU: é um módulo de prototipação para a Internet das Coisas, que possui o microcontrolador ESP8266 em sua arquitetura e conta com uma plataforma de desenvolvimento nas linguagens C/C++ ou em Lua [Oliveira 2017]. O módulo utiliza a versão ESP-12 do microcontrolador ESP8266, possui conversor micro USB para serial, um regulador de tensão interno para adequar a corrente elétrica de entrada no dispositivo, arquitetura de 32 bits com núcleo de processamento com 80 MHz que pode chegar até 160 MHz e memória disponível para os dados do programa de usuário tem cerca de 50 kB. O módulo trabalha com uma tensão de 3,3 volts, conforme padrão do microcontrolador. Ainda conta com antena embutida em seu circuito para conexão com a rede Wi-Fi. O seu modo de operação em rede pode ser como *Station* (STA), *Access Point* (AP) ou com os dois modos combinados.

¹Família de especificações desenvolvidas pelo IEEE (Institute of Electrical and Electronics Engineers) para a tecnologia Wireless LAN (WLAN/Rede Sem-Fio).

²*Hypertext Transfer Protocol*, em português Protocolo de Transferência de Hipertexto, é um protocolo de comunicação na camada de aplicação segundo o Modelo OSI utilizado para transmissão de dados na rede de computadores.

- *Pulse Sensor Amped*: é um sensor de frequência cardíaca, que realiza a coleta de dados do BPM de uma pessoa. O sensor utiliza a técnica de fotopletiografia, uma maneira não invasiva de medir a pulsação do coração e em alguns casos a oxigenação do sangue. Essa técnica faz a leitura por meios óticos, através de uma luz infravermelha, que é emitida e refletida de volta, computando assim o sinal ple-tismográfico, que indica o pulso e oxigenação do sangue [Leite and Lima 2017].
- MPU6050: é um módulo para Internet das Coisas que possui acelerômetro, gi-rosκόpio e um sensor de temperatura no mesmo chip. O acelerômetro faz a detecção da inclinação, movimentação e orientação no ambiente. O giroscópio identifica a direção do movimento. O sensor é capaz de fazer a leitura tridimensi-onal (eixo x, y e z) tanto para o acelerômetro quanto para o giroscópio. O sensor de temperatura integrado, possibilita a coleta de dados sobre a temperatura am-biente, a princípio em *fahrenheit* (°F) que conseqüentemente pode ser convertida em graus *celsius* (°C) [Arduino 2016].

Os recursos computacionais utilizados para a construção deste trabalho são des-critos individualmente abaixo:

- Plataforma *Heroku*: é uma plataforma de serviço em nuvem, que oferece suporte a várias linguagens de programação. O Heroku automatiza serviços de criação das máquinas virtuais para armazenar sites e aplicações Web. O serviço utiliza o conceito de *Dyno*, onde cada *Dyno* é uma pequena máquina virtual, com 4 núcleos de processamento e 512 *megabytes* de memória RAM. Utiliza-se o sistema opera-cional Debian, que é uma distribuição Linux gratuita [Salesforce 2017].
- Plataforma *Firebase*: é uma plataforma com serviços oferecidos em nuvem, que possui diversas ferramentas úteis e eficientes para desenvolvimento de aplicações Web e móveis. O Firebase atualmente conta com o suporte a aplicações desenvol-vidas para iOS, Web e Android nas linguagens C++, Java, Javascript, Objective-C e Swift [Inc 2017].
- *Ionic Framework*: é um *framework* para desenvolvimento de aplicativos móveis na forma híbrida [Ionic 2017]. O *framework* é composto pelo Cordova [Apache 2017], que possibilita a instalação de *plugins*, onde recursos nativos dos dispositivos são disponibilizados para acesso e o Angular [Google 2017] que é ou-tro *framework*, que utiliza a linguagem de programação JS (JavaScript), e fornece recursos para desenvolvimento Web, com HTML (*HyperText Markup Language*) e CSS (*Cascading Style Sheets*).
- *Flask Microframework*: é um *microframework* que possibilita o desenvolvi-mento de aplicações Web utilizando a linguagem Python. É um *microframework*, pois não define uma estrutura rígida de projeto a ser seguida, preserva a liber-dade de decisão das ferramentas a serem utilizadas por parte do desenvolvedor [Ronacher 2017].

@misc{carduinobri2c, title=I2C: Protocolo de Comunicação, author=Valdinei Rodrigues dos Reis, year=2014, howpublished="http://www.arduino.br/arduino/i2c-protocolo-de-comunicacao/", note = "[Acesso em 07 de fevereiro de 2018]

A modelagem do protótipo realizado com a ajuda do *Fritizing*³ pode ser conferida na Figura 2. No centro do sistema encontra-se o NodeMCU, fazendo a interface de conexão com o sensor de frequência cardíaca, acelerômetro, giroscópio e o sensor de temperatura. O NodeMCU ainda é responsável por realizar conexão com a rede Wi-Fi e fazer comunicação com a Internet. O sistema é alimentado com uma bateria recarregável de 9 volts.

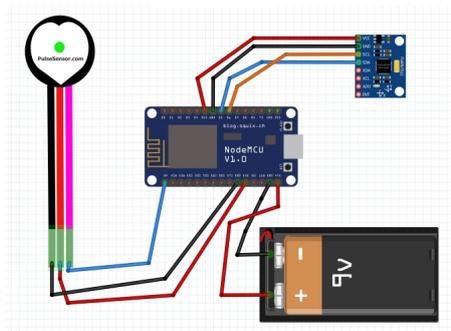


Figura 2. Figura ilustrativa com o protótipo do sistema embarcado

Para transformação do sistema embarcado em computação vestível todos os componentes foram fixados em uma capa protetora. Os fios foram organizados e a costura na roupa ficou mais simples, podendo ocorrer de formas diferentes e também facilmente ser removido e costurado em outras roupas. A Figura 3 mostra a esquerda os componentes fixados na capa protetora e a direita como foi costurado em uma roupa para sua utilização.

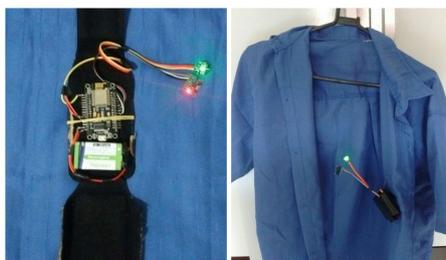


Figura 3. Protótipo final do sistema costurado na parte interna de uma camisa

Para obter corretamente o BPM e a temperatura corporal, é necessário que o MPU6050 e *Pulse Sensor* estejam em contato diretamente com a pele, na posição do coração, fixados com algum tipo de fita adesiva. A forma como o dispositivo está anexado à roupa facilita a mesma ser vestida e os componentes serem fixados à pele do idoso.

O algoritmo do sistema embarcado é construído na linguagem C/C++, de forma estruturada, onde cada função deve executar apenas uma tarefa, facilitando alterações que possam ser necessárias e o desenvolvimento de novas funcionalidades.

Na plataforma *Firebase*, são utilizados os recursos do banco de dados em tempo real, a autenticação que oferece criação de usuário com senha para se conectar aos

³Software com código aberto, onde o objetivo é ajudar com os esquemas da montagem do hardware e suas conexões físicas.

serviços da plataforma e as notificações, que fornecem uma interface para *smartphones*, possibilitando receber alertas através do aplicativo.

A Web API⁴ faz o meio entre o dispositivo de monitoramento, que é o sistema embarcado e a plataforma Firebase, desenvolvida utilizando o com a linguagem de programação Python na versão 3.5.2 e o *microframework* Flask na versão 0.12.2. O código da Web API é armazenado em nuvem com o intuito de disponibilizar o acesso de forma online, excluindo a necessidade de haver um servidor local para se comunicar com o dispositivo de monitoramento, simplificando a infraestrutura do sistema.

Além de salvar os dados recebidos, é necessário realizar o processo de análise dos dados, com o objetivo de verificar situações com possíveis quedas, frequência cardíaca e temperatura acima ou abaixo do normal. A Tabela 1 exemplifica como é feito a análise por intervalos dos dados.

Tabela 1. intervalos de comparação para definir situações da saúde

Variável analisada	Intervalo de Comparação	Situação da Saúde
BPM	Maior ou igual a 160	Frequência cardíaca acima do normal
	Menor ou igual a 40	Frequência cardíaca abaixo do normal
	Igual a 0	Frequência cardíaca não detectada
Temperatura	Maior ou igual a 37,4	Temperatura Corporal acima do normal
	Menor ou igual a 34	Temperatura Corporal abaixo do normal

Além das comparações mencionadas que são feitas por intervalos, é realizado o cálculo da mediana dos últimos 60 dados anteriores relacionados ao BPM, e feito a comparação com cada valor capturado da frequência cardíaca. Os últimos 60 BPM são relacionados a cerca de 3 minutos de monitoramento. pois os dados são gravados no banco em média a cada 3 segundos, ou seja, são cerca de 20 dados gravados por minuto, considerando o atraso que pode ocorrer entre a captura dos dados pelo dispositivo embarcado, o envio para Internet e o armazenamento no banco de dados.

5. Resultados

O objetivo desta seção é apresentar os resultados obtidos com o sistema finalizado e através dos estudos de casos em ambiente real, analisar o desempenho do sistema proposto neste trabalho.

O aplicativo possui quatro telas, sendo a de cadastro, que é apresentada apenas na primeira vez em que é usado, com objetivo de coletar dados do paciente. A tela de *dashboard* como mostrado na Figura 4, apresenta os dados relacionados ao monitoramento em tempo real da saúde do idoso. A de histórico conforme a Figura 5, apresenta detalhes do monitoramento, dividindo os dados, que são agrupados por datas e apresentados por horários. E a última tela do aplicativo, é a de perfil do usuário, que apresenta os dados do paciente que foram cadastrados anteriormente.

Para coletar dados e apresentar os resultados, foram realizados dois estudos de casos em ambiente real, com o objetivo de apresentar a viabilidade do sistema desenvolvido.

⁴Traduzido como Interface de Programação de Aplicativos, representa um conjunto de rotinas, funções e padrões de programação que podem ser acessados por outro *software* ou plataforma Web.

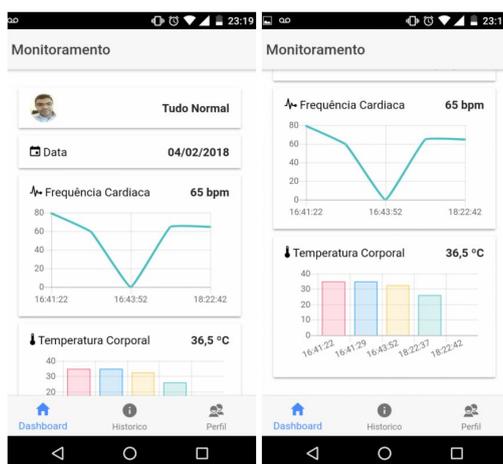


Figura 4. Página inicial do aplicativo mostrando os dados do monitoramento

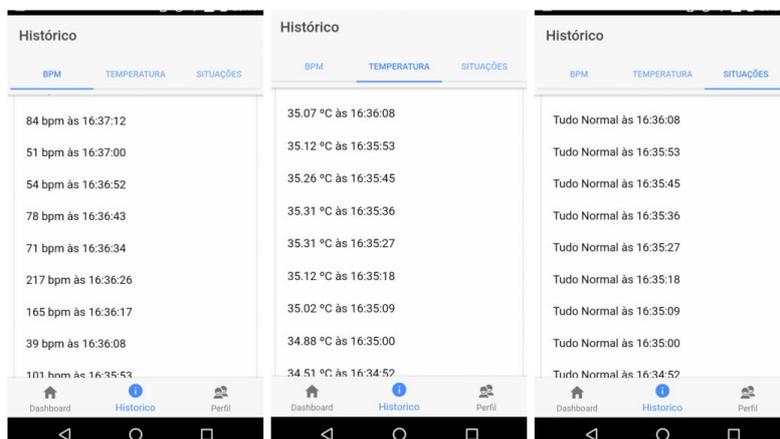


Figura 5. Página de histórico mostrando as três opções de visualização dos dados

O primeiro estudo de caso (EC1), foi realizado por uma pessoa do sexo masculino, com 22 anos de idade, 1,83 metros de altura e peso de 87 quilogramas, com objetivo de validar o funcionamento do sistema, corrigir erros e realizar testes para a implementação do algoritmo de detecção de quedas. O segundo estudo de caso (EC2), realizado por cerca de 1 hora, com uma pessoa do sexo masculino, 77 anos de idade, 1,60 metros de altura e peso de 68 quilogramas, com o objetivo de realizar a validação do sistema em ambiente real, sendo utilizado por uma pessoa idosa. As informações de altura, peso e idade são relevantes, pois a frequência cardíaca é diferente conforme a idade e o acelerômetro pode ser afetado pelo peso corporal no momento de detectar quedas.

O EC1 teve duração aproximada de três horas e trinta minutos, onde o dispositivo de monitoramento foi utilizado por este período contendo alguns minutos de intervalos para ajustes. Ao todo foram armazenadas 759 linhas de dados relacionados à saúde do paciente.

O gráfico, mostrado na Figura 6, realizado com os dados, auxiliou na construção do algoritmo que detecta quedas. O gráfico de linha mostra os dados do acelerômetro nos

eixos X (acIX), Y (acIY) e Z (acIZ). Para coleta destes dados, foram simuladas diversas situações de queda, buscando encontrar padrões de aceleração nos eixos do acelerômetro.

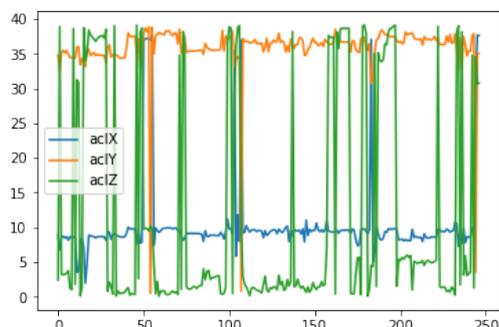


Figura 6. Gráfico do acelerômetro com os dados coletados no estudo de caso 1

O EC2 teve 154 linhas de dados coletados, em um período de quase uma hora de realização. O teste foi executado em um idoso, com o objetivo de validar se o protótipo exerce suas funções corretamente em ambiente real.

O gráfico da Figura 7 mostra as oscilações que ocorreram nos dados da frequência cardíaca. As situações que apresentam BPM abaixo do normal, foram simuladas, afastando o sensor do corpo da pessoa em monitoramento.

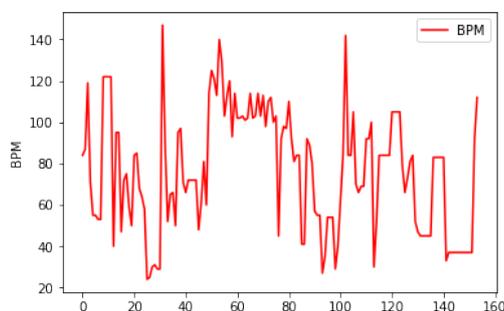


Figura 7. Gráfico do BPM do estudo de caso 2

No caso da temperatura corporal, também houve algumas oscilações e dados que podem ser considerados *outliers*, porém na maior parte do tempo a temperatura ficou entre 34 e 36 °C, o que pode ser considerado normal para este estudo de caso. A Figura 8 mostra com detalhes como ficou o gráfico da temperatura corporal com os dados coletados no EC2.

Com relação as situações de saúde, foi possível observar que foram informadas corretamente. Em momentos que a frequência cardíaca ficou abaixo dos 40 BPM, houve o aviso de "BPM Muito Baixo", em relação a temperatura abaixo de 34 °C, foi indicado como "Temperatura Muito Baixa", e quando não houve nenhum destes casos indicado como "Situação Regular". A Figura 9 mostra como ficou a tela de modo parcial das situações na página de histórico do aplicativo móvel no momento da realização do EC2.

A Figura 10 mostra como são visualizadas as notificações, caso haja alguma situação de risco detectada.

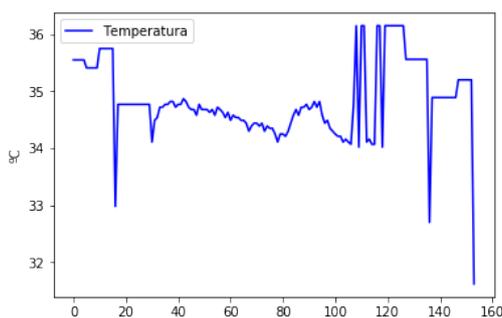


Figura 8. Gráfico da temperatura corporal do estudo de caso 2



Figura 9. Página de histórico mostrando as situações do estudo de caso 2

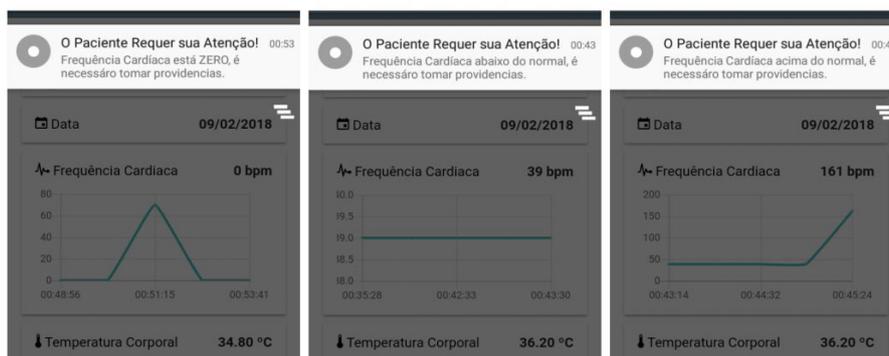


Figura 10. Notificações relacionadas ao batimento por minuto

A última notificação implementada é relacionada a mediana do BPM, podendo ser acima ou abaixo, dependendo da situação. Mesmo que o cálculo seja realizado pela mediana dos valores, é apresentado para o usuário como média, para facilitar a interpretação da informação. A Figura 11 mostra como são apresentadas essas notificações para o usuário.



Figura 11. Notificações relacionadas a média do batimento por minuto

6. Considerações finais

O projeto desenvolvido neste trabalho teve como objetivo principal construir um sistema de monitoramento à saúde do idoso. Podendo ser realizado de forma remota por familiares ou cuidadores, que possibilita acompanhar os dados por um *smartphone*, oferecendo monitoramento da frequência cardíaca, temperatura corporal e capaz de detectar quedas. O sistema possui ainda como funcionalidade, uma análise dos dados coletados e notificações de alertas em caso de alguma alteração que ofereça risco à saúde do idoso. Com o objetivo de contribuir com a qualidade de vida dos idosos, aumentando a sua independência e aos familiares acompanhamento da saúde de forma remota.

O sistema possui custo consideravelmente acessível na montagem do dispositivo embarcado, com o valor de US\$ 32,61 dólares americanos (levando em consideração a cotação de 07/2018), sendo que todos os componentes foram adquiridos no Brasil. Possui características de um sistema em tempo real, com a utilização de tecnologias como banco de dados em tempo real, que possibilita realizar o monitoramento com pouco atraso, em média de três segundos a cada coleta de dados. A não necessidade de um servidor local para a análise e armazenamento dos dados e a rápida resposta à alterações das situações de saúde, que são quase de modo instantâneo.

O equipamento apresentou precisão satisfatória nos seus objetivos de monitorar frequência cardíaca e temperatura corporal, apresentando algumas inconsistências na coleta dos dados em alguns momentos, porém com efetividade na maior parte do período testado nos estudos de casos. Em relação ao algoritmo para detectar quedas, apresentou algumas falhas, gerando falsos positivos e negativos, porém com acertos consideráveis. A análise dos dados pela Web API foi satisfatória, gerando notificações corretamente para todos os casos. O aplicativo móvel apresentou desempenho satisfatório, mesmo recebendo uma quantidade grande de dados por minuto, possibilita o acompanhamento personalizado e o recebimento de notificações quando necessário. O sistema ainda é facilmente expansível, desenvolvido de forma modular, fácil de agregar novos sensores e novas funcionalidades que contribuam para a sua evolução. Todas as tecnologias utilizadas são gratuitas e a maioria com código aberto.

O código completo desenvolvido neste trabalho, incluindo o sistema embarcado, Web API e aplicativo móvel, pode ser encontrado no GitHub [de Moura 2018], plataforma de repositórios para códigos fontes gratuita.

Como trabalhos futuros, deseja-se agregar novos sensores ao sistema, como o sensor de pressão arterial e oxigenação do sangue, ainda podendo agregar novos módulos, como 3G/GPRS para possibilitar o monitoramento fora de ambiente doméstico, sem rede Wi-Fi e GPS para identificar a localização do idoso. É possível melhorar as análises dos dados coletados, aplicando técnicas que possam prever situações de riscos e gerar avisos de forma preditiva com a utilização de Inteligência Artificial. Conseguir detectar movimentos bruscos do idoso, reduzindo assim os ruídos na leitura dos sensores relacionado ao contato com o corpo. Ainda deseja-se implementar maior segurança na transmissão dos dados entre dispositivo embarcado e servidor em nuvem, usando o protocolo HTTPS. O estudo dos sensores, módulos e técnicas deve ser considerado, agregando ainda o consumo de energia do sistema, que deve aumentar quando adicionado novas funcionalidades.

Referências

- Akyildiz, I. F. and Wang, X. (2005). A survey on wireless mesh networks. *IEEE Communications magazine*, 43(9):S23–S30.
- Apache, T. S. F. (2017). Cordova documentation. <http://cordova.apache.org/docs/en/latest/>. [Acesso em 30 de dezembro de 2017].
- Arduino (2016). MPU-6050 Accelerometer + Gyro. <https://playground.arduino.cc/Main/MPU-6050>. [Acesso em 11 de novembro de 2017].
- Banuleasa, S., Munteanu, R., Rusu, A., and Tonț, G. (2016). Iot system for monitoring vital signs of elderly population. In *Electrical and Power Engineering (EPE), 2016 International Conference and Exposition on*, pages 059–064. IEEE.
- Batista de Góis, A. L. and Peixoto Veras, R. (2010). Informações sobre a morbidade hospitalar em idosos nas internações do sistema único de saúde do brasil. *Ciência & Saúde Coletiva*, 15(6).
- Bernardo, A. M. (2010). Proposta de sistema embarcado para auxílio e monitoramento do idoso.
- Bersch, R. (2008). Introdução à tecnologia assistiva. *Porto Alegre: CEDI*.
- Bourke, A. K., Klenk, J., Schwickert, L., Aminian, K., Ihlen, E. A., Mellone, S., Helbstadt, J. L., Chiari, L., and Becker, C. (2016). Fall detection algorithms for real-world falls harvested from lumbar sensors in the elderly population: a machine learning approach. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 3712–3715. IEEE.
- BRASIL, P. D. R. (2012). Dados sobre envelhecimento no brasil. <http://www.sdh.gov.br/assuntos/pessoa-idosa/dados-estatisticos/DadosobreoenvelhecimentonoBrasil.pdf>. [Acesso em 07 de fevereiro de 2018].
- de Moura, C. J. M. (2018). Mirsi - repositório no github. <https://github.com/cicerojmm/mirsi/>.
- Fechine, B. R. A. and Trompieri, N. (2015). O processo de envelhecimento: as principais alterações que acontecem com o idoso com o passar dos anos. *InterSciencePlace*, 1(20).

- Google (2017). Angular get started. <https://angular.io/>. [Acesso em 30 de dezembro de 2017].
- Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. (2013). Internet of things (iot): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7):1645–1660.
- IBGE (2012). Tábua Completa de Mortalidade para o Brasil de 2012. www.ibge.gov.br/home/estatistica/populacao/tabuadevida/2012/. [Acesso em 24 de maio de 2017].
- Inc, G. (2017). Firebase by Platform. <https://firebase.google.com/docs>. [Acesso em 14 de novembro de 2017].
- Ionic (2017). Ionic framework: Docs. <https://ionicframework.com/docs/intro/installation/>. [Acesso em 30 de dezembro de 2017].
- Leite, E. R. and Lima, R. A. (2017). Sistema não invasivo de monitorização da pressão arterial. *Anais da Mostra de Extensão, Inovação e Pesquisa*, 4.
- Loyola Filho, A. I. d., Leite Matos, D., Giatti, L., Afradique, M. E., Viana Peixoto, S., and Lima-Costa, M. F. (2004). Causas de internações hospitalares entre idosos brasileiros no âmbito do sistema único de saúde. *Epidemiologia e serviços de saúde*, 13(4):229–238.
- Mann, S. (1998). Definition of wearable computer. *International Conference on Wearable Computing ICWC-98, Fairfax VA, May 1998*.
- Oliveira, S. (2017). Internet das coisas com esp8266, arduino e raspberry pi. *São Paulo: Novatec*.
- Pinto, S., Cabral, J., and Gomes, T. (2017). We-care: An iot-based health care system for elderly people. In *Industrial Technology (ICIT), 2017 IEEE International Conference on*, pages 1378–1383. IEEE.
- Ronacher, A. (2017). Welcome to flask. <http://flask.pocoo.org/docs/0.12/>. [Acesso em 30 de dezembro de 2017].
- Salesforce, C. (2017). Heroku Dev Center. <https://devcenter.heroku.com/>. [Acesso em 14 de novembro de 2017].
- Sugathan, A., Roy, G. G., Kirthivijay, G., and Thomson, J. (2013). Application of arduino based platform for wearable health monitoring system. In *Condition Assessment Techniques in Electrical Systems (CATCON), 2013 IEEE 1st International Conference on*, pages 1–5. IEEE.
- Viega, S. (2017). Lista de batimentos cardíacos normais por idade. <https://saude.umcomo.com.br/artigo/lista-de-batimentos-cardiacos-normais-por-idade-28323.html>. [Acesso em 05 de fevereiro de 2018].
- Vinícius, M. (2017). Temperatura corporal. <http://www.enfermagemesquemematizada.com.br/temperatura-corporal/>. [Acesso em 12 de fevereiro de 2018].

Experiências usando PERT-CPM para desenvolvimento ágil de software

Thiago Jabur Bittar¹, Patrícia Duarte Silva¹, Larissa P. C. dos Santos¹, Luanna Lopes Lobato¹

¹Departamento de Computação – Universidade Federal de Catalão (UFCat)
Av. Dr. Lamartine Pinto de Avelar, 1120 - CEP 75704-020 – Catalão – GO – Brazil

{thiagojabur, luannalobato}@ufg.br

{patricia, larissap}@cc.grad.ufg.br

Abstract. *Identifying activities, establishing precedence, deadlines, and those responsible are initial tasks for good planning of software development. For this, there are important tools to help managers, such as the application of Critical Path Method (CPM) methods and the development and monitoring of the Program Evaluation and Review Technique (PERT) network diagram. In the combination of these methods we have the PERT-CPM, which, in this work, was applied on a case study about the production of a Web and mobile educational game using agile methods. As a result, with PERT-CPM it was possible to identify experiences and lessons for other projects.*

Resumo. *Identificar as atividades, estabelecer as precedências, os prazos de execução e os responsáveis são tarefas iniciais para o bom planejamento do desenvolvimento de software. Para tanto existem instrumentos importantes para auxiliar os gestores, como a aplicação dos métodos Critical Path Method (CPM) e a elaboração e acompanhamento do diagrama de redes Program Evaluation and Review Technique (PERT). Na combinação desses métodos tem-se o PERT-CPM, que, neste trabalho, foi aplicado sobre um estudo de caso sobre produção de um software do tipo jogo educativo Web e mobile usando métodos ágeis. Como resultados, com o PERT-CPM foi possível identificar experiências e lições para outros projetos.*

1. Introdução

Um produto de qualidade no cenário competitivo atual do sistema industrial mundial é de significativa importância para uma instituição. Manter o nível de qualidade de um produto está diretamente ligado com o controle da produção do mesmo, da sua concepção à entrega e uso pelo cliente. O controle sobre o desenvolvimento está inserido no complexo arcabouço de custos, prazos, posicionamento de mercado e níveis de qualidade desejados. Este controle permite a viabilidade, acompanhamento gerencial, produtividade no desenvolvimento, operação e manutenção de produtos [Vargas, 2009].

Nesse sentido, a partir do momento em que o desenvolvimento de um produto é planejado, com a identificação das atividades de produção e o tempo que cada uma pode levar para ser cumprida, é possível determinar um prazo ideal para a entrega do mesmo.

Mas não basta simplesmente estabelecer datas, é necessário estar atento a imprevistos, analisar diariamente a evolução das atividades e reorganizar as mesmas e/ou a equipe, se necessário. Por exemplo, ao identificar o atraso de uma atividade que afeta outra, pode haver a alocação de mais membros na equipe para a conclusão da atividade em questão. Dessa maneira, a qualidade do produto não precisará ser diminuída para acelerar o processo de entrega das atividades pendentes [Martins e Laugeni, 2005].

Na indústria de software a necessidade da qualidade é a mesma. Um software de qualidade é aquele que satisfaz os requisitos solicitados pelo cliente, é de fácil manutenção, tem boa performance, é confiável e fácil e acessível de se usar. Manter a qualidade com um preço final justo e respeitando prazos de entrega é essencial. Então, para reduzir o custo de venda de um software é feita a distribuição do custo entre vários clientes, até mesmo em âmbito global. Ou seja, o software deve ser planejado para atender o máximo possível de clientes, sendo assim, desenhado uma única vez e distribuído diversas.

Desta maneira o desenvolvimento de software deixou de ter características artesanais e começou a assemelhar-se com produções industriais. As empresas de desenvolvimento de software passaram a focar seus esforços na produção de componentes, objetivando o reuso, com entregas rápidas e com qualidade.

Assim, satisfazer o cliente cumprindo o prazo estabelecido para entrega com todos os atributos prometidos para o produto deve ser o foco de todas instituições. O planejamento e o gerenciamento são as ferramentas necessárias para o alcance e manutenção da qualidade mantendo os prazos previstos.

Na falta de planejamento e gerenciamento dos projetos, os mesmos são cancelados, não terminam no prazo determinado, são encerrados prematuramente ficando sem qualidade, entre outras consequências desastrosas. A quebra dos prazos de entrega de softwares foi uma das características mais identificadas durante a Crise do Software [Pressman e Maxim, 2016], revelando a suma importância de um gerenciamento produtivo eficiente durante o desenvolvimento de softwares.

Em termos de gestão, princípios de controle de atividades foram aplicados, nesta pesquisa, em um estudo de caso referente ao desenvolvimento de um jogo educacional para as plataformas *Web* e *mobile*. As atividades para o desenvolvimento do produto foram estabelecidas de acordo com regras contemporâneas da Engenharia de Software (ES) e de metodologia ágil (incluindo, por exemplo, gerenciamento constante de riscos), juntamente com a determinação do seu prazo de conclusão e suas atividades precedentes. Para a gestão destas atividades foi aplicado o método PERT-CPM para identificar os prazos e o caminho crítico para o desenvolvimento planejado do projeto. Tal método não é novo, sendo proposto inicialmente por Wiest e Levy (1969) e mais tarde com um tutorial prático de Swanson e Woolsey (1974), porém a sua utilização nos moldes aqui mencionados, em abordagem prática voltada para produção de produtos de software com metodologia ágil (geralmente utilizada em pequenas empresas e *startups*) tem aspectos de inovação, podendo os preceitos aqui descritos auxiliarem outras equipes.

Este artigo está organizado da seguinte maneira: na Seção 2 tem-se a apresentação dos conceitos envolvidos; já na Seção 3 os trabalhos relacionados são relatados, com a identificação de diferenciais desta pesquisa; na Seção 4 tem-se as fases de ES com métodos ágeis detalhadas e incluídas na gestão, algo que difere esta pesquisa das demais verificadas na Seção 3 e que possivelmente auxiliará outras equipes. Por fim, na Seção 5 é apresentado o estudo de caso realizado, seus resultados e análise e na Seção 6 tem-se as considerações finais deste trabalho.

2. Conceituação

2.1. Desenvolvimento ágil

A primeira e principal proposição para o desenvolvimento ágil de software foi realizada no período de 11 a 13 de fevereiro de 2001 no chamado “Manifesto Ágil” por um grupo de 17 especialistas em desenvolvimento de software que estavam insatisfeitos com insucessos dos métodos tradicionais. Eles também formaram na ocasião a Aliança Ágil (Beck *et al.*, 2001). Os valores principais defendidos por esses especialistas são:

- Indivíduos e interações mais que processos e ferramentas;
- Software em funcionamento mais que documentação abrangente;
- Colaboração com o cliente mais que negociação de contratos;
- Responder a mudanças mais que seguir um plano.

Nas metodologias de desenvolvimento ágil o software é produzido de modo incremental, fazendo com que alterações sejam absorvidas ao longo das sucessivas versões, sempre com partes funcionais [Teles, 2004] [Teles, 2005] [Cockburn, 2002] [Sbrocco, 2012].

Foram estabelecidos 12 princípios para se alcançar a agilidade, que são:

- **Valor:** satisfaça o cliente com entrega adiantada e contínua;
- **Flexibilidade:** aceite os pedidos de alterações, pois os processos ágeis devem aproveitar essas mudanças como vantagem competitiva;
- **Frequência:** entregue software em funcionamento frequentemente, em intervalos curtos;
- **União:** a equipe da área comercial e os desenvolvedores devem trabalhar em conjunto;
- **Motivação:** use técnicas e ferramentas para ter a equipe sempre motivada;
- **Comunicação:** tenha conversas abertas para comunicação mais efetiva e eficiente;
- **Funcionalidade:** use como medida de progresso o software em funcionamento;
- **Sustentabilidade:** nunca esgote todos os seus recursos, processos ágeis devem promover o desenvolvimento sustentável;
- **Revisão:** valorize a excelência técnica;
- **Simplicidade:** maximize o trabalho realizando apenas o essencial;

- **Organização:** tenha as equipes auto organizadas, com divisão de tarefas e responsabilidades feitas com auxílio de cada integrante do time;
- **Autoavaliação:** a equipe deve avaliar seu trabalho em intervalos regulares.

A partir desses valores foram desenvolvidas várias metodologias ágeis, sendo que entre as mais populares estão a eXtreme Programming (XP) e a SCRUM, com foco significativo nas atividades práticas de desenvolvimento e na gestão dos projetos [Fernandes e Almeida, 2010].

2.2. Método do Caminho Crítico e o diagrama de redes PERT

O Método do Caminho Crítico (em inglês *Critical Path Method*, CPM) e o diagrama de redes PERT (do inglês: *Program Evaluation and Review Technique*), são técnicas de planejamento baseadas em modelos de rede de atividades [Wiest e Levy, 1969] [Swanson e Woolsey, 1974].

Esses métodos (que geralmente são aplicados em conjunto) têm sido utilizados em programas de gestão nas mais diversas áreas, tais como indústria, agricultura, defesa nacional e pesquisa científica [Swanson e Woolsey, 1974]. A motivação para seu uso advém do fato de que na maioria dos projetos do cotidiano humano tem-se um rol complexo de atividades inter-relacionais com tempos definidos, umas que podem ser executadas em paralelo e outras que devem aguardar o desenvolvimento e conclusão de outras.

O PERT-CPM, então, fornece uma significativa base teórica gráfica e matemática para os gestores de projetos para auxílio ferramental no planejamento: i) organizando e explicitando o progresso das atividades em seus caminhos e ii) lidando com o impacto de fatores incertos e no acompanhamento de custos e atrasos [Swanson e Woolsey, 1974] [Christoph, 2001].

No CPM, o caminho é a ordem em que as tarefas são realizadas de acordo com a sequência a ser seguida. O chamado Caminho Crítico é, então, a sequência que leva mais tempo para ser finalizada, indicando o tempo máximo total que um projeto levará. É importante ressaltar que um projeto pode ter vários caminhos críticos, com um empate na soma de tempos totais.

3. Trabalhos Relacionados

Para a busca de trabalhos relacionados foram utilizadas inicialmente três bibliotecas digitais científicas mundiais, a ACM DL¹, a IEEE Xplore² e a Springer Link³, com a meta específica de encontrar publicações com aderência total ao foco desta pesquisa, envolvendo ES, PERT-CPM e desenvolvimento ágil. No entanto tais bases não retornaram resultados substanciais considerando a *string* definida de busca em todos os metadados que endereçam essa meta: `pert AND cpm AND software AND agile`. O uso

¹ <https://dl.acm.org>

² <https://ieeexplore.ieee.org>

³ <https://link.springer.com>

de uma variação dessa *string* excluindo-se o termo “cpm” se mostrou ineficaz por conta desse item ser utilizado para outras definições.

Uma possível explicação para o resultado encontrado nestas bases é que o método buscado, PERT-CPM, é genérico, servindo para várias áreas do conhecimento e não ter uma maior exploração específica em desenvolvimento ágil após seu lançamento por Swanson e Woolsey (1974) na área da programação matemática no veículo ACM Special Interest Group on Mathematical Programming, SIGMAP. Na verdade, é possível perceber que o método nasceu de um rigoroso formalismo matemático, o que pode, numa primeira e rasa análise, distanciar seu uso nos métodos ágeis, que procuram desburocratizar ao máximo o desenvolvimento. Análise essa que é discutida nas próximas seções deste artigo.

Como modo de ampliar a busca e enriquecer esta investigação foi aberta, então, a possibilidade de uso da pesquisa do Google considerando publicações científicas e também incluída a análise do uso do método fora do desenvolvimento de software. O resultado pode ser visto na discussão do restante desta seção.

Göksu e Catovic (2012) relatam que devido aos efeitos crescentes da globalização em vários ambientes de negócios, espera-se que a indústria de manufatura seja eficiente e eficaz, com entregas de qualidade no prazo correto. De acordo com isso, no planejamento, programação e controle de um projeto, que é uma combinação de várias atividades, são testadas técnicas de gerenciamento de projetos PERT-CPM. Para esse teste empírico, a pergunta de pesquisa que é feita foi: “Como a implementação do PERT-CPM influencia na eficácia e a eficiência de uma empresa de móveis chamada Dallas”.

A resposta para essa pergunta é relevante para indicar a importância desses métodos na redução do tempo e dos custos de conclusão do projeto. Os dados retirados da empresa de móveis foram combinados com revisões de literatura. Esse estudo verificou como determinar as atividades envolvidas nos processos de fabricação da empresa selecionada e como demonstrar possíveis benefícios e inconvenientes que esses métodos de planejamento podem criar na organização. Implicações dessa pesquisa são a avaliação do tempo de conclusão do projeto e o controle dos recursos, para que o projeto fosse concluído dentro do tempo e custo planejados usando os métodos mencionados. No final do estudo, tem-se relatos para ajudar outros indivíduos, bem como as empresas, a entender melhor o conceito do método PERT-CPM na redução do tempo e dos custos de conclusão do projeto.

Tal trabalho é análogo ao realizado nesta pesquisa, no entanto o foco aqui está na indústria de software ágil, que apresenta complexidades próprias do domínio, como uma severa e constante análise de riscos, feita em paralelo com as demais. Outro diferencial desta pesquisa apresentada aqui é que são apresentadas as atividades e todo arcabouço gráfico feito, pois Göksu e Catovic (2012) focam na parte inicial teórica, não mostrando visualmente as atividades e gráficos relacionados.

Lu *et al.* (2011), por sua vez, apresentam a combinação das técnicas PERT e CPM para o planejamento do projeto “Pré-emprego” para graduandos universitários. O projeto “Pré-emprego” consiste em combinar atividades desde a elaboração e distribuição de currículos até a preparação para as entrevistas. O foco do trabalho foi,

então, o de resolver questões comuns a todos projetos como: “Quando o projeto inteiro pode ser concluído?”, “Qual é o melhor método para concluir o projeto com custo mínimo?”, “Os recursos são suficientes para concluir o projeto no prazo?”, entre outras questões de mesma importância.

O trabalho mais alinhado ao tema desta pesquisa foi o de Ashkenazi *et al.* (2010), que realizaram um estudo de caso em uma empresa de desenvolvimento de software utilizando o método ágil SCRUM e a técnica COCOMO II para estimar o esforço de atividades. Apesar do foco da pesquisa em questão estar no uso do SCRUM e não existir muitos detalhes na publicação (é um resumo estendido de apenas 4 páginas sem a imagem da rede utilizada) a conclusão é de que o PERT-CPM trouxe benefícios no *design* de cada fase do projeto, permitindo a detecção do(s) caminho(s) crítico(s) e chamando a atenção para o progresso das atividades ali presentes. No entanto os autores indicam que o PERT-CPM não é estritamente necessário no SCRUM, pois são utilizados projetos bem particionados em “semi independentes” subprojetos com pequenas equipes de desenvolvedores com grande flexibilidade de execução de atividades.

Ashkenazi *et al.* (2010), terminam relatando que os líderes de projeto devem saber sim fundamentalmente o caminho crítico, usando ou não um método formal. Ou seja, PERT-CPM pode ser um bom instrumento para o acompanhamento de projeto por estes líderes, especialmente quando se tem complexos caminhos de atividades. Atrasos são péssimos em todos os projetos, incluindo os ágeis e todos os métodos devem ser utilizados para ajustar os desenvolvimentos.

4. Fases da Engenharia de Software com Métodos Ágeis

Na ES, com uso de métodos ágeis, o processo de desenvolvimento de sistema computacional é constituído de um conjunto de atividades com entregas de partes do software em períodos curtos de tempo, chamados genericamente de iterações. Essas atividades são agrupadas em fases, como: levantamento de requisitos, análise de riscos, definição sobre a arquitetura do projeto, elaboração de testes e implementação do código. Em cada fase são definidas, além das suas atividades, as funções e responsabilidades de cada membro da equipe, e como produto resultante, os artefatos que são incrementalmente melhorados.

As fases descritas nas próximas subseções são um apanhado dos conceitos de metodologias ágeis que foram usados neste projeto de desenvolvimento e que deram azo para a criação das tarefas gerenciadas com os métodos ágeis. Tais fases são similares às dos métodos tradicionais, amplamente descritas na literatura; a mudança está no uso de princípios da metodologia ágil.

4.1. Levantamento de requisitos

Nesta fase o problema a ser solucionado pelo software deve ser compreendido, fornecendo aos desenvolvedores e clientes a mesma visão do que deve ser construído. É feito, então, um levantamento e priorizadas as necessidades dos futuros usuários do software. Essas necessidades definidas são denominadas como requisitos. O levantamento e a análise de requisitos irão definir o que o sistema deve fazer, antes de definir como o sistema irá fazer [Pressman e Maxim, 2016].

Os requisitos de software podem ser divididos em dois tipos básicos: os funcionais e os não funcionais. Os primeiros são aqueles que especificam uma função que o sistema ou componente deve ser capaz de realizar. Esses são requisitos de software que definem o comportamento do sistema, ou seja, o processo ou transformação que componentes de software ou hardware efetuam sobre as entradas para gerar as saídas. Tais requisitos capturam as funcionalidades sob o ponto de vista do usuário. Já os não funcionais não estão diretamente relacionados à funcionalidade de um sistema, definindo características do software como método de desenvolvimento, tempo, espaço, sistema operacional, ambiente, usabilidade, ergonomia e acessibilidade.

4.2. Análise de riscos

A análise de riscos consiste na identificação de possíveis eventos que possam afetar o projeto. Esses eventos são denotados como riscos. Os riscos identificados são avaliados e atividades são definidas para reduzir a possibilidade de ocorrência ou o impacto dos riscos mais prejudiciais.

O gerenciamento e monitoramento dos riscos devem ocorrer durante todo o processo de desenvolvimento do projeto com o objetivo de mitigar os riscos encontrados, trazendo o menor impacto negativo ao projeto.

Em uma metodologia ágil a identificação de riscos contínua e a entrega de partes funcionais em pequenos ciclos com grande interação com o cliente auxilia na mitigação e na resolução dos problemas. A tendência é que o risco seja rapidamente controlado numa metodologia ágil. No caso investigado aqui, tem-se a possibilidade de rápidas mudanças no cronograma com ajustes sendo feitos na rede de atividades do projeto.

4.3. Definição sobre arquitetura do projeto

A definição da arquitetura do projeto considera como o sistema funcionará e estará organizado internamente, para que os requisitos do cliente possam ser atendidos. Os artefatos produzidos por essa fase consistem na descrição computacional de como o código-executável deve ser construído.

A arquitetura descreve tanto a estrutura lógica do funcionamento do software quanto a arquitetura física de componentes físicos que pertencem ao software. Na arquitetura lógica é descrito claramente o funcionamento lógico do software em termos de funções, variáveis, componentes e classes. Já na arquitetura física é descrito o conjunto de arquivos fontes, arquivos de dados, bibliotecas, executáveis e outros.

4.4. Elaboração e execução dos testes

A fase de elaboração e execução dos testes busca identificar as situações de uso do software e o comportamento desejável que o software deve entregar. O objetivo dos testes é validar o produto, testando cada funcionalidade de cada módulo levando em consideração as especificações feitas na fase de projeto da arquitetura. Os testes elaborados são executados conforme os componentes do sistema são implementados, produzindo um relatório com as informações relevantes sobre os erros encontrados.

4.5. Codificação

Nesta fase, o sistema é codificado a partir da materialização dos artefatos descritos na fase de definição da arquitetura do projeto. O projeto passa a ser executado computacionalmente em linguagem de programação atendendo os requisitos levantados.

Vale ressaltar que em equipes inexperientes em relação à linguagem utilizada é necessário incluir uma etapa anterior à implementação para haver o treinamento da equipe.

4.5. Manutenção

Mesmo no melhor cenário de identificação de requisitos e desenvolvimento o software precisa software modificações evolutivas e corretivas depois de ser entregue. A manutenção deve ser uma operação viável em que há a aplicação de cada uma das fases descritas anteriormente aos módulos existentes em vez de criar um novo software.

Na utilização de metodologias ágeis é preciso que as equipes pensem na estruturação arquitetural do código fonte do projeto como um todo e utilizem boas práticas, garantindo uma futura manutenção rápida para viabilidade da evolução do software. Não basta então, entregar partes funcionais do software funcionando, elas têm que estarem calcadas numa estrutura sólida e bem projetada.

5. Estudo de Caso

Segundo Yin (2010, p. 19), “O estudo de caso é apenas uma das muitas maneiras de se fazer pesquisa em ciências sociais. Experimentos, levantamentos, pesquisas históricas e análise de informações em arquivos são alguns exemplos de outras maneiras de se realizar pesquisa. Cada estratégia apresenta vantagens e desvantagens próprias, dependendo basicamente de três condições: a) o tipo de questão da pesquisa; b) o controle que o pesquisador possui sobre os eventos comportamentais efetivos; c) o foco em fenômenos históricos, em oposição a fenômenos contemporâneos”.

Em geral, os estudos de caso representam a estratégia preferida quando se colocam questões do tipo “*como*” e “*por que*”, quando o pesquisador tem pouco controle sobre os eventos e quando o foco se encontra em fenômenos contemporâneos inseridos em algum contexto da vida real. Pode-se complementar esses estudos de casos com dois outros tipos - estudos “*exploratórios*” e “*descritivos*”.

Para a investigação proposta neste trabalho foi utilizada a metodologia de Yin (2010) em conjunto com as diretrizes de Runeson e Höst (2009), que são específicas para ES.

Nesse cenário, foi utilizado o seguinte estudo de caso com contexto fictício para testar a eficiência da gestão de atividades usando o PERT-CPM: “Você acabou de entrar em uma equipe de desenvolvimento de software de uma pequena empresa que solicitou o desenvolvimento de um jogo, desde a sua criação até a implementação e testes. Essa empresa não se preocupa com ES, seus gestores alegam que não tem pessoal e recursos para realizarem as atividades de ES. A empresa não faz documentação. O jogo deve ser simples e educativo (abordando conceitos básicos de matemática, biologia ou geografia) e deve ser feito para as plataformas *Web* e *mobile*”.

A partir deste contexto, para o desenvolvimento deste jogo foram definidas por dois pesquisadores especialistas em ES as atividades principais, o tempo de execução estimado e a precedência destas atividades conforme apresentado na Tabela 1. Apesar do contexto ser fictício a implementação do jogo e a interação entre membros da equipe e uso de metodologia ágil foi real, com software sendo entregue ao final de uma disciplina “Engenharia de Software II” do curso de Ciência da Computação da Universidade Federal de Catalão (UFCat).

Para efeitos operacionais e para simplificar o estudo de caso foi considerado que após a conclusão de uma atividade a outra subsequente não paralela só começa no dia seguinte.

Tabela 1. Alocação das atividades (Fonte: Dados da pesquisa).

Atividade	Precedência	Duração em dias
Início	-	0
A - Definição do tema	Início	3
B - Pesquisar sobre jogos similares	A	2
C - Pesquisar dados sobre o público alvo	A	3
D - Levantamento de requisitos e <i>game design</i>	B, C	3
E - Análise de riscos	Início	2
F - Definição do cronograma	D	2
G - Definição sobre arquitetura do projeto	D	2
H - Escrever documentação	A	21
I - Elaboração dos Layouts	D	2
J - Elaboração dos níveis do jogo	I	3
K - Implementação do código	J	10
L - Elaboração dos testes	K	3
M - Execução dos testes	L	3
N - Escrita do artigo	M	17
O - Escrita do ofício	N	5
P - Apresentação/Entrega	O	1
Fim	O,P,H,F,G,E	-

Com os dados da Tabela 1 é possível identificar que existem 6 caminhos, que são rotas seguindo os arcos a partir do nó “Início” até o nó “Fim”. O comprimento de um caminho é a soma das durações das atividades sobre o caminho. Esses caminhos são apresentados na Tabela 2.

Tabela 2. Caminhos e seus comprimentos (Fonte: Dados da pesquisa).

Caminho	Comprimento
Início - A - C - D - I - J - K - L - M - N - O - P - Fim	53
Início - E - Fim	2
Início - A - H - Fim	24
Início - A - C - D - G - Fim	11
Início - A - B - D - I - J - K - L - M - N - O - P - Fim	52
Início - A - B - D - G - Fim	10
Início - A - B - D - F - Fim	10
Início - A - B - D - I - J - K - L - M - N - O - Fim	51
Início - A - C - D - I - J - K - L - M - N - O - Fim	52

Com a somatória dos tempos das atividades foi possível identificar claramente o caminho crítico, que é aquele de maior comprimento. Em nosso caso trata-se do primeiro caminho da Tabela 2, com 53 dias totais para completar o projeto.

Na Figura 1, apresentada a seguir, tem-se o primeiro esboço do diagrama de rede sobre as atividades levantadas, feito com a técnica de lápis e papel. A opção pelo uso dessa técnica foi o de ter maior liberdade de trabalho e para teste dos conhecimentos dos conceitos estudados em sua totalidade. Outro ponto importante e que fica como lição aprendida é de não ter sido encontrado pelos autores desta pesquisa um software simples e gratuito para geração deste gráfico.

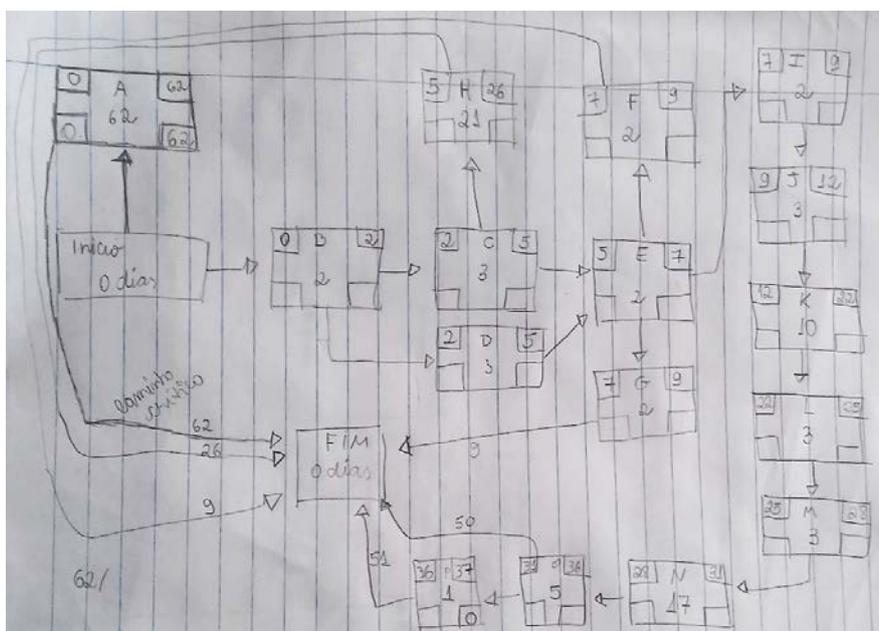


Figura 1. Primeira versão do Gráfico PERT-CPM feito em lápis e papel (Fonte: Autoria própria).

Na Figura 1, cada caixa representa uma atividade e as setas representam a precedência das atividades. A seta é direcionada para a(s) atividade(s) seguinte(s). A leitura do diagrama começa pela atividade “Início”, que começa com 0 dias. Em seguida, a atividade “Início” irá apontar para as atividades “A” e “B”, que ocorrerão em paralelo após o início do projeto. A atividade “B” aponta para a atividade “C” que só poderá ser iniciada assim que a atividade “B” for concluída.

Para facilitar a identificação das atividades e suas informações referentes, foi definida a estrutura da representação das caixas. Em todas as caixas, no meio é colocada uma letra maiúscula do alfabeto que representa uma atividade. E logo abaixo da atividade, é alocado o tempo de execução da atividade.

Na Figura 2, apresentada a seguir, é descrito o que cada componente da caixa representa.



Figura 2. Descrição do que cada componente da caixa (Fonte: Autoria própria).

Na Figura 3, apresentada a seguir, tem-se a finalização do diagrama de rede, com o uso de ferramenta computacional de editoração gráfica.

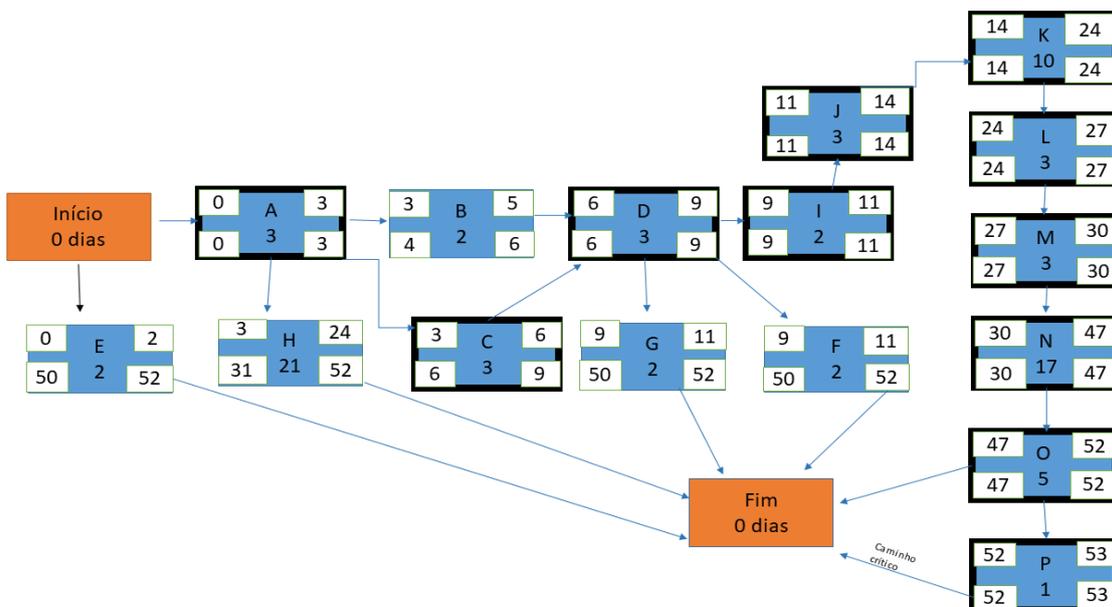


Figura 3. Gráfico CPM final de atividades do desenvolvimento do produto (Fonte: Dados da pesquisa)

Na Figura 3 o gráfico foi refeito e recalculado após um estudo mais aprofundado sobre a aplicação do método PERT-CPM. Este gráfico possui caixas, descritas anteriormente, que representam cada atividade planejada do projeto; as setas, que apontam a sequência das atividades; a representação do caminho crítico, o caminho com as atividades mais longas; e as informações de início e término mais cedo e mais tarde de cada atividade.

Para construir o gráfico da Figura 3 primeiramente foi definido o “Início” com 0 dias para execução. Em seguida, todas as precedências foram ajustadas conforme a Tabela 1, até chegar na atividade “Fim” que também contém 0 dias como tempo de execução. Posteriormente, foi definido o caminho crítico (destacado pela cor preto no gráfico de rede CPM da Figura 3) com 53 dias. É importante ressaltar que essa informação define o tempo total que o projeto leva para terminar.

Após identificar o caminho crítico, foram calculados o início mais cedo, término mais cedo, início mais tarde e término mais tarde de todas as atividades do caminho, exceto das atividades “Início” e “Fim” (que são marcos gerenciais). Estes cálculos foram feitos do seguinte modo:

- **Início mais cedo (IMC):** para calcular o início mais cedo da primeira atividade, sempre é definido o valor da quantidade de dias de execução da atividade “Início”. No restante das atividades, é somado o início mais cedo e a quantidade de dias de execução da atividade anterior.
- **Término mais cedo (TMC):** no término mais cedo da primeira atividade, sempre é colocada a quantidade de dias da primeira atividade. Para as demais atividades, é somado o término mais cedo da anterior com a quantidade de dias da atual.
- **Início mais tarde (IMT):** é feito calculando o término mais tarde da atividade atual menos a quantidade de dias da atividade atual.
- **Término mais tarde (TMT):** é calculado da atividade final para a inicial (no inverso da ordem natural de execução). Para as atividades que estão ao lado do fim, é utilizado o valor do dia do início mais tarde da última atividade do caminho crítico.

Depois de identificar o caminho crítico do projeto, é importante identificar também as folgas das atividades. Existem três tipos de folga, a saber:

- **Folga total:** quantidade de tempo que uma atividade pode atrasar sem atrasar a data do projeto.
- **Folga livre:** quantidade de tempo que uma atividade pode atrasar sem atrasar as atividades sucessoras.
- **Folga do projeto:** quantidade de tempo que o projeto pode atrasar sem que atrase a data de entrega imposta pelo cliente.

Com o gráfico concluído é possível identificar com mais clareza as folgas de cada atividade. Na Figura 3 é possível verificar que a atividade H tem **28 dias de folga livre**. Ela pode ser iniciada a partir do terceiro dia ou no máximo no trigésimo primeiro dia de execução do projeto, mas deve ser concluída até no quinquagésimo segundo dia.

6. Conclusão

Entregar um produto de qualidade no prazo é o objeto de desejo da maioria das instituições de bens e serviços, característica essa que agrega maturidade produtiva a essas organizações. O alcance dessa tão sonhada qualidade pode ser facilitado por processos de gerenciamento e planejamento ágil de projetos. Controlar o desenvolvimento de um produto se resume em monitorar custos, prazos e níveis de qualidade desejáveis. O método de controle correto irá permitir a produtividade no desenvolvimento, operação e manutenção de produtos.

A produção de software dos dias atuais apresenta características de processos industriais, portanto os métodos de controle de produção utilizados nas indústrias de grande porte também passam a ser aplicáveis com benefícios no desenvolvimento deste produto, até mesmo no desenvolvimento ágil. Isso pôde ser verificado com sucesso no ensaio do estudo de caso apresentado neste artigo.

O controle de desenvolvimento envolveu a identificação de atividades elementares, suas precedências e seus prazos de execução, o que não burocratizou o trabalho de planejamento e que frequentemente já é feito pelos líderes de projeto de modo não sistematizado. O gerenciamento destas atividades possibilitou determinar o prazo ideal de entrega do produto e as atividades que devem ser tratadas com mais atenção. Com base nisso, foi aplicado o método PERT-CPM envolvendo o planejamento de um software de jogo educacional para as plataformas *Web* e *mobile*.

A partir da construção do gráfico PERT-CPM foi possível determinar com custo de trabalho baixo a melhor estratégia de gerenciamento das atividades. Ao encontrar o caminho crítico e tendo as atividades pré-estabelecidas foi possível determinar os responsáveis e impor datas de entrega com acompanhamento contínuo nas iterações. As atividades identificadas com folga podem ser adiadas e colocadas em iterações finais, priorizada a entrega no prazo de todos os requisitos do produto solicitado.

Não planejar bem as atividades ou contar apenas com o feeling dos líderes agilistas pode dificultar a identificação de gargalos de prazos em atividades e afetar gravemente nas entregas do software. Além disso, atividades executadas ao acaso podem causar atraso em cascata, situação essa em que as atividades mais onerosas são deixadas para o final, estourando o prazo de entrega e afetando a qualidade.

Por fim, verificou-se que o número de horas para se especificar as atividades, responsáveis e para desenvolver o diagrama PERT-CPM é pequeno se comparado ao tempo total estimado para o projeto, sendo uma atividade benéfica mesmo em empresas pequenas que utilizam desenvolvimento ágil.

Referências

- Almeida M. e Fernandes J. M. Classification and Comparison of Agile Methods. 2010 Seventh International Conference on the Quality of Information and Communications Technology (QUATIC), Porto, Portugal, 2010, pp. 391-396.
- Ashkenazi L., Shmilovici A., Ben-Gal I., Agile Software Development: A Case Study in a Software Company, Proc. of the 16th Industrial Eng. & Management Conference, Tel Aviv, March 23-24, 2010.

- Beck, K.; Beedle, M.; Van Bennekum, A.; Cockburn, A.; Cunningham, W.; Fowler, M.; Grenning, J.; Highsmith, J.; Hunt, A.; Jeffries, R.; Kern, J.; Marick, B.; Martin, R. C.; Mellor, S.; Schwaber, K.; Sutherland, J.; Thomas, D. Manifesto para Desenvolvimento Ágil de Software. 2001. Disponível em: <<http://www.agilemanifesto.org/iso/ptbr/>>. Acessado em 14 de fevereiro de 2018.
- Christoph. S.T.: Job shop scheduling with alternative process plans. 2001. Int. J. Production Economics 74; Páginas 125-134.
- Cockburn, A. Agile Software Development. ISBN 0201699699, 978-0201699692. 304 p. Addison-Wesley, 2002.
- Göksu, A.; Catovic, S. (2012). Implementation Of Critical Path Method And Project Evaluation And Review Technique. 3rd International Symposium on Sustainable Development. 31th May-1st June 2012. Sarajevo, Bosnia and Herzegovina.
- Lu X., Guan S., Tian R., Zhang W. Using PERT/CPM Technology for the Development of College Graduates Seeking Employment in Project Planning. In: Liu C., Chang J., Yang A. (eds) Information Computing and Applications. ICICA 2011. Communications in Computer and Information Science, vol. 244. Springer, Berlin, Heidelberg. 2011.
- Martins, P. G.; Laugeni, F. P. Administração da Produção. São Paulo: Saraiva, 5ª Ed., 2005.
- Pressman, R. S.; Maxim, B. R. Engenharia de Software. Uma Abordagem Profissional. 2016. 968 páginas. 8a. Edição. Editora AMGH. ISBN: 8580555337.
- Runeson, P., Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. Empirical software engineering, 14(2), 131.
- Sbrocco, J. H. T. C.; Macedo, P. C. de. Metodologias Ágeis: Engenharia de Software sob medida. 1a ed. São Paulo: Érica, 2012. ISBN 978-85-365-0979-2. 254 páginas.
- Swanson, H. S.; Woolsey, R. E. D. 1974. A PERT-CPM tutorial. SIGMAP Bull. 16 (Abril de 1974), 54-62. DOI=<http://dx.doi.org/10.1145/1217031.1217043>.
- Teles, V. M. Extreme Programming. São Paulo: Novatec Editora, 2014. 2ª. Ed. ISBN: 978-85-7522-400-7. 328 páginas.
- Teles, V. M. Um estudo de caso da adoção das práticas e valores do Extreme Programming. 181 p. Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, UFRJ, 2005.
- Vargas, R. Gerenciamento de projetos – Estabelecendo diferenciais competitivos. 7ª Ed. Rio de Janeiro: Brasport, 2009. 276 páginas.
- Wiest, J. D. e Levy; F.K. A Management Guide to PERT/CPM. Prentice-Hall, Englewood Cliffs, New Jersey, 1969.
- Yin, R. Estudo de caso: planejamento e métodos. Porto Alegre: Bookman, 2ª Ed., 2001. 205 páginas. ISBN 8573078529.

Parâmetros de Avaliação Multidimensionais: Aspiração para Modelos de Maturidade de Gestão do Conhecimento

Douglas de F. Alves¹, Eliomar A. de Lima¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

dougfalves@gmail.com, eliomar@inf.ufg.br

Abstract. *The reality in the organizational context is perceived by its complexity and manage it has been a challenge to your leaders. The dynamic inherent to the organizations difficulties the adoption of the Knowledge Management (KM) practices, because the instability, uncertainties and turbulences imply in constants changes of the business. The objective of this study is to define evaluation parameters that can contribute to the maturity specification models of KM, in way to assure the relational and conceptual precepts inherent to organizational systems. This is a narrative review that culminated with a discover of multidimensional nature parameters to evaluation of the maturity level of KM in complex scenarios.*

Resumo. *A realidade no contexto organizacional deve ser percebida pela sua complexidade e gerenciá-la tem sido um desafio para os seus líderes. A dinâmica inerente à atuação das organizações dificulta a adoção das práticas de Gestão do Conhecimento (GC), pois a instabilidade, incertezas e turbulências implicam em mudanças constantes no negócio. O objetivo deste estudo é definir parâmetros de avaliação que possam contribuir para a especificação de modelos de maturidade de GC, de modo a assegurar os preceitos relacionais e contextuais inerentes aos sistemas organizacionais. Trata-se de um trabalho de revisão narrativa que culminou com a descoberta de parâmetros de natureza multidimensional para avaliação do nível de maturidade de GC em cenários complexos.*

1. Introdução

As organizações têm buscado formas de entender e gerenciar a complexidade. Diversos estudos teóricos e práticos foram aplicados com o objetivo de diagnosticar e parametrizar o ambiente organizacional, porém, compreender esta realidade ainda é um desafio e objeto de estudo para diversos especialistas e pesquisadores, que buscam novas metodologias para entender a dinâmica e a quantidade de informação presente no ambiente organizacional [Naves et al. 2000].

Antes de tentar diagnosticar qualquer um dos aspectos das organizações, é necessário compreender como as partes desta se comunicam. Desde o século XVII as organizações são reconhecidas como estruturas compartimentadas e com pouco relacionamento entre as partes. Para superar as dificuldades de acoplamento e coesão, uma composição mais fluida e guiada por relacionamentos interpessoais e trabalho em equipe vem se estabelecendo [Lodi 1984, Rummler and Brache 1994, Wheatley 1996].

Para Drucker [1997], o modelo de organização tradicional, sólido e permanente é substituído por relações de alianças, parcerias, participações e acordos que possuem

caráter temporário e são guiados por políticas e estratégias comuns em que participam atores da própria organização e stakeholders, caracterizando a necessidade de uma visão compartilhada. Organizações guiadas por esta visão, organizações que aprendem [Senge 2013], possuem a sistematização do conhecimento como parte do seu processo de negócio, ou seja, a criação e o compartilhamento do conhecimento são essenciais para impulsionamento dos objetivos e metas organizacionais [Goldoni 2007].

Diante do exposto, nota-se a necessidade do desenvolvimento de práticas, normas, políticas e processos que possibilitem o gerenciamento do conhecimento no ambiente organizacional e cabe à própria organização definir quais modelos se encaixam melhor à sua realidade. Posto isto, em que medida diagnosticar o nível da maturidade das organizações no que é tocante às práticas de gestão do conhecimento? E como avaliar se o modelo aplicado é adequado no contexto (segmento) em questão?

Em uma pesquisa realizada por Lima et al. [2015], foi proposto um construto multidimensional para entendimento da complexidade sistêmica das organizações. Este estudo guiou a realização de uma revisão sistemática da bibliografia no universo de modelos de maturidade em GC. Na proposta do artigo, apresentamos um quadro relacional entre os principais modelos de maturidade existentes e as dimensões complexas que são contempladas por estes. A revisão bibliográfica é apresentada na seção 2. A metodologia de pesquisa está na seção 3 e na seção 4 é formalizada a proposta citada.

Em síntese, pretende-se relacionar os avanços dos estudos no campo de modelos de maturidade em GC, obtidos através da revisão bibliográfica, à necessidade de compreensão da realidade complexa das organizações.

2. Revisão da Literatura

O meio corporativo dispõe de fontes inesgotáveis de conhecimento, quer seja associado a processos, normas, políticas, sistemas de informação e rotinas administrativas, quer seja vinculado aos colaboradores e grupos na organização. Heerdt [2002] defende que é preciso sistematizar as ações de GC, o que ainda não é realidade em grande parte das organizações, para se obter saltos qualitativos a partir das descobertas e dos achados que o conhecimento capturado revela.

Ainda que o conhecimento seja um ativo importante e essencial no ambiente organizacional, captar, armazenar e transformar grandes quantidades de informação em conhecimento é um desafio para as corporações na atualidade. Neste contexto, o que percebemos é a criação e desenvolvimento de instrumentos, mecanismos e a formação de táticas para gerir o conhecimento nas organizações [Castro 2005].

As dificuldades encontradas em aplicar a GC (GC) dentro das organizações devem-se principalmente ao fato do conhecimento ser um ativo intangível, ou seja, está presente em todas as partes das organizações, mas não é percebido de forma concreta [Del-Rey-Camorro et al. 2003]. Dessa forma, analisando o conhecimento sob a ótica do capital intelectual e dos ativos intangíveis, chega-se a estudos e pesquisas que atribuem à melhoria de serviços prestados e aumento de ganhos financeiros das empresas a aplicação do conhecimento captado, o que vai além da sua retenção, por meio do compartilhamento deste de forma estratégica e aderente aos objetivos organizacionais [Rauniar et al. 2007].

Na perspectiva das organizações que aprendem, aplicar GC nas é mais do que identificar os ativos de conhecimento existentes. A gestão estratégica do conhecimento, resultante da integração de dimensões estratégicas e práticas gerenciais, depende da construção de uma estrutura organizacional bem definida, englobando processos, pessoas e tecnologias simultaneamente [Senge 2013]. Nesse sentido, Trigo et al. [2006] asseveram que a GC deve ser vista como uma área interdisciplinar, em que são agregadas noções de sistemas de informação, gestão e ciências sociais.

Na literatura, existem diversas abordagens voltadas para avaliar a GC dentro das organizações. Muitas destas abordagens são definidas como modelos de maturidade, que são formas de avaliar organizações em determinado aspecto. Para Lin [2007], modelos de maturidade são estruturas desenvolvidas para avaliar a situação da organização em determinada área, como por exemplo, gestão da qualidade, gestão de recursos humanos, inovação e GC, relacionando a área em questão com determinado objetivo, ou conjunto de objetivos organizacionais, geralmente, identificando possíveis melhorias.

Na próxima seção serão estudados alguns modelos de maturidade de GC. A maioria desses modelos é estrutura por meio de estágios e recomendações de fatores-chave, buscando a formação de uma matriz de referência [Smith et al. 1985].

2.1. Modelos de Maturidade em GC

Para Van-de-Vem e Poole [1995], em geral, os modelos de maturidade em gestão aplicados às organizações servem para a detecção de mudanças no ambiente corporativo e baseiam-se no ciclo de vida organizacional. Tais mudanças são padronizadas e podem ser categorizadas em estágios. Para cada estágio existem fatores-chave que incluem pessoas, procedimentos, estratégias, entre outros.

A avaliação da GC é feita através do estabelecimento de indicadores e níveis de maturidade em GC. Ao aplicar um modelo de diagnóstico em uma realidade limitada é possível estabelecer o nível de maturidade da organização em GC naquele momento, desta forma, fatores críticos e deficiências podem ser detectadas e melhoradas [Gonçalo, Junges and Borges 2010, Kuriakose et al. 2010]. Sinteticamente, a organização, para ser capaz de gerenciar seus ativos de conhecimento, deve conhecer sua capacidade atual [Natale et al. 2016].

O trabalho realizado por Kraemer et al. [2017] apresenta uma revisão sistemática ampla dos modelos de maturidade em GC, trazendo os principais pontos e comparativos dos modelos estudados. Nessa revisão, os modelos de maturidade são separados em quatro categorias: proposição de um novo modelo teórico, relacionamento de maturidade em GC com outras variáveis, modelos de mudança de níveis e modelos aplicados em diferentes empresas.

O modelo de maturidade em GC definido em Mehta et al. [2007] foi criado inicialmente para atender as necessidades da organização Infosys. O modelo é dividido em cinco níveis. O primeiro nível é o nível *default*, em que a GC não é percebida na organização ou os membros não reconhecem a necessidade de gerir conhecimento. O segundo nível é o reativo, onde se inicia o compartilhamento do conhecimento de processos e práticas da organização em momentos oportunos. O terceiro nível é o consciente, onde uma abordagem mais ampla de GC é iniciada na organização, incluindo processos, pessoas e tecnologias utilizadas. O quarto nível é o de

convencimento, em que a GC passa a integrar a cultura organizacional e torna-se sustentável. O quinto e último nível é o de compartilhamento, considerado o estágio máximo por este modelo, em que a disseminação e o aproveitamento efetivo do conhecimento estão institucionalizados.

O modelo de cinco níveis, na revisão de Kraemer et al. [2017], foi categorizado como um modelo de mudança de níveis. Em sua aplicação na empresa Infosys foi questionado o investimento financeiro necessário para atingir o último nível (compartilhamento), conseqüentemente, as métricas foram alteradas para avaliação dos fluxos de conhecimento, tornando-as mais tangíveis e possíveis para a organização.

Por sua vez, Kruger e Johnson [2010] avaliaram o crescimento dos níveis de GC em diferentes setores de empresas africanas. Seu estudo apresentou segmentos em que a aplicação de GC mostrou crescimento elevado: construção civil, materiais de construção e mineração. Setor de bens de consumo, serviços públicos, bancos e seguros, transportes e governo tiveram crescimento moderado.

Lin et al. [2012] fizeram uma sistematização das barreiras existentes em cada nível de maturidade, definindo quatro tipos: existência, as barreiras não mudam com o avanço da GC; queda, as barreiras diminuem à medida que se aplica GC; aumento, as barreiras crescem à medida que se aplica GC; aleatória, barreiras diferentes em cada nível de GC.

Através da análise de diversos modelos de maturidade de mudança de níveis, Jiuling et al. [2012] identificaram quatro áreas que chamou de “áreas-chave”, que são mais utilizadas quando se fala sobre GC. As áreas são: pessoas, cultura, tecnologia e processos.

Paulzen et al. [2002] apresentam o modelo de maturidade para melhoria de qualidade em GC que categoriza a maturidade em cinco níveis: início, em que não existe nenhuma prática de GC na organização; consciência, em que as primeiras práticas de GC começam a aparecer nos processos e rotinas organizacionais; estabelecido, em que os papéis das pessoas, processos de GC, tecnologias, ativos informacionais importantes já estão bem definidos na organização; gerido, em que os indicadores de GC passam a ser coletados e analisados para verificação dos investimentos; otimização, a partir dos indicadores, traçam-se estratégias de melhoria.

Lee e Kim [2001] também desenvolveram um modelo em estágios de GC semelhante ao modelo utilizado na Infosys. Possui quatro estágios: o estágio de iniciação, em que a organização prepara as primeiras iniciativas para realização de práticas de GC; o estágio de propagação, etapa em que a organização levanta e investe na infraestrutura necessária para que a GC seja realidade na organização; o estágio da integração, em que se verificam os resultados do investimento em GC para a organização, seja em produtos ou serviços prestados; o estágio de rede de trabalho, em que são incorporados parceiros externos à organização para contribuição nos ativos de conhecimento.

Seguindo a mesma lógica dos modelos em níveis, Kulkarni et al. [2004] propõem os seguintes estágios para a GC: difícil, possível, encorajada, praticada, gerida e melhoria contínua. A organização deve avançar por todos esses estágios, até conseguir monitorar os investimentos em GC e os resultados para o negócio.

O modelo de maturidade em GC utilizado na empresa Siemens, conhecido como *Knowledge Management Maturity Model-KMMM*, possui cinco níveis de maturidade estabelecidos, associados a cada uma das seguintes áreas: estratégia e objetivos do conhecimento, ambiente interno e externo (parcerias), pessoas e responsabilidades, cultura organizacional, liderança, suporte, tecnologia, processos e estrutura do conhecimento. O início é o momento em que não há iniciativas de GC e o conhecimento não é reconhecido. A repetição é o momento em que a organização reconhece a importância da GC e planeja-se um projeto piloto, ou seja, os processos de GC são estabelecidos e aplicados a uma realidade limitada. Gerido é o estágio em que as atividades de GC estão bem definidas e aceitas na organização e é possível fazer a coleta de indicadores. O último estágio é o de otimização, em que é possível reformular estratégias de captação e utilização de conhecimento através dos indicadores obtidos [Ehms and Langen 2002].

O modelo proposto por Teah et al. [2007] é constituído de 4 níveis, semelhante ao apresentado por Lee e Kim [2001], porém, neste, em cada nível relacionam-se pessoas, processos e tecnologias necessárias como principais parâmetros elencados.

No modelo de Robinson et al. [2006] há cinco estágios, a saber: *start-up*, em que os processos de GC começam a ser elaborados e cria-se a consciência da importância da GC para o negócio; *descolagem*, define-se os objetivos da GC na organização; *expansão*, é quando os objetivos de GC alinham-se com as necessidades do negócio, conseqüentemente, a GC ganha visibilidade na organização; *progressivo*, ocorre avaliação da GC na organização; *sustentabilidade*, momento em que a GC é uma realidade na organização. Na revisão sistemática de Kraemer et al. [2017], esse estudo é categorizado como proposição de um novo modelo.

Já o modelo de Kathibian et al. [2010] está subdividido em cinco níveis: inicial, gerido, definido, gerido quantitativamente e otimizado. Consideram-se fatores como estratégia, liderança, cultura organizacional, estrutura organizacional, recursos humanos, processos e tecnologias.

O padrão proposto por Kruger e Snyman [2007] possui sete níveis que descrevem etapas progressivas: na primeira etapa a tecnologia deve ser estabelecida como capacitadora da GC; na segunda, devem ser estabelecidos os princípios norteadores de GC na organização; em seguida, deve-se formalizar e documentar a política de GC da organização; na quarta etapa, devem ser construídas as estratégias de captação do conhecimento; conseqüentemente, deve ser estruturada a estratégia de gestão deste conhecimento; na sexta etapa o conhecimento torna-se onipresente e gerenciado em toda organização; por fim, chega-se o nível considerado como “futuro” das organizações, em que o conhecimento é utilizado como habilitador para melhoria das práticas em toda a corporação. O estudo de Kruger e Snyman [2007] é definido por Kraemer et al. [2017] como a proposição de um novo modelo.

A proposta de Lin [2007] resume a organização da maturidade em GC através de três estágios: iniciação, desenvolvimento e maturidade. Na iniciação, reconhece-se a importância da GC na organização. No desenvolvimento, estruturam-se os processos e práticas de GC, definem-se as políticas e investe-se na infraestrutura necessária para aplicação. Na maturidade deve ser formada uma rede de colaboração que inclua o ambiente interno e externo da organização.

Já em Oliveira et al. [2011], a partir de uma revisão da literatura é sintetizado em um quadro comparativo os modelos de maturidade estudados. São mostrados, por modelo de maturidade em GC, quais fatores são vinculados durante a avaliação dos níveis: clientes, fornecedores, parceiros, competidores, legislação, cultura, estrutura, suporte da alta administração, alinhamento com os objetivos de negócio, objetivos da GC, benefícios, conhecimento tácito, conhecimento explícito, conhecimento crítico, tecnologia, líder de GC, sistema de recompensa, tempo, treinamento, fases do processo, comunicação e orçamento.

Os estudos que relacionaram a maturidade em GC com outras variáveis também trouxeram resultados interessantes. Kruger e Johnson [2010] realizaram experimentos com empresas africanas e Hartono et al. [2016] com empresas da indonésia, ambos relacionando o nível de maturidade com o desempenho organizacional, foram encontrados resultados positivos para as organizações analisadas. Maturidade e crescimento foram relacionados nos estudos de Salojarvi et al. [2005] através da avaliação da maturidade em pequenas e médias empresas (PMEs) eslovacas e os resultados também foram satisfatórios. Pérez [2013] relacionou maturidade e inovação e encontrou métricas positivas no que diz respeito a processos organizacionais e marketing [Kraemer et al. 2017].

A avaliação do nível de maturidade em empresas de diferentes características ajuda a construir uma visão genérica e viabiliza a percepção de tendências do mundo organizacional no universo da GC. Ao analisar a GC em grandes empresas na Eslováquia percebeu-se contrastes entre o entendimento de GC e sua implementação [Sujanova et al. 2012]. Posteriormente, o estudo foi refeito e notou-se o avanço da aderência entre estratégias do negócio e práticas de GC [Sujanova et al. 2015].

O trabalho de Kruger e Johnson [2010] concentrou-se na definição de princípios norteadores da GC para cada setor produtivo. A pesquisa foi realizada em setores industriais da África. Oliveira et al. [2014] realizaram um estudo em PMEs brasileiras e percebeu que a maioria não possui maturidade em tecnologia e as práticas de GC ainda são muito incipientes.

Na última década, foram apresentados alguns novos modelos de maturidade em GC que afirmam suprir as deficiências dos modelos anteriores, que tinham muito foco em tecnologia. Porém, percebe-se que grande parte destes “novos” modelos ainda se baseia em práticas do CMM, que foi criado para um universo tecnológico. Hsieh et al. [2009] apresentam o KNM (Knowledge Navigator Model), baseado no CMM. O modelo possui itens e processos bem detalhados e separa a GC em três dimensões básicas: cultura, tecnologia e processos.

O estudo de Pee e Kankanhali [2009] traz outra vertente para os modelos de GC, afirmando que a GC em organizações deve ser realizada a partir de uma abordagem holística. O modelo apresenta a GC em três dimensões: pessoas, processos e tecnologia, apresentando maior flexibilidade que anteriores para ser aplicável em partes da organização ou na organização inteira. O foco é na criação de unidades intensivas de captação e aproveitamento do conhecimento.

Por sua vez, Oliveira e Pedron [2014] relacionam a GC com três variáveis: capacidade de captação de conhecimento, inovação e desempenho organizacional. Eles afirmam que devem ser levados em consideração para aplicar GC o ambiente interno e externo, os processos e o conteúdo (conhecimento).

Já Serna [2012] apresenta a seguinte abordagem: o conhecimento depende do seu contexto, pois é construído a partir da interação entre as pessoas e não pode ser armazenado, pois depende de cada situação em que este é gerado. A partir dessa constatação, o modelo é construído por meio de uma abordagem cognitiva, pois o foco é o aprendizado organizacional. A partir dos estudos de Serna [2012], Galati [2015] propõe um modelo de três dimensões para a aprendizagem organizacional: fontes de conhecimento (internas e externas), processos de GC e estratégia de GC. A dimensão de processos de GC possui três etapas: descongelando, movendo-se e institucionalizando, identificando o processo de amadurecimento das práticas de GC na organização. A dimensão estratégica de GC possui quatro estágios: sem estratégia, estratégia de codificação (o conhecimento é explicitado), estratégia de simbiose (gerenciamento do conhecimento explícito e tácito) e aprendizagem organizacional (a GC chega ao nível máximo e o aprendizado é uma realidade contínua na organização).

3. Metodologia

Este trabalho se desenvolveu com base na revisão narrativa. Esta técnica constitui, basicamente, análise da literatura publicada em diversas fontes (livros, artigos, revistas e outros) baseando-se na visão dos autores. O papel principal deste tipo de revisão é a educação continuada do leitor, possibilitando com que este adquira e atualize conhecimento sobre determinada temática em um curto espaço de tempo [Rother 2007].

Partindo de um mapeamento amplo da literatura, os esforços foram concentrados nos achados da pesquisa bibliográfica sobre GC, sendo selecionados modelos de maturidade pertinentes ao objetivo desta pesquisa. Em seguida, foram expostas percepções dos estudiosos, assim como pontos de intersecção e diferenças entre os padrões.

A pesquisa bibliográfica teve por finalidade levantar as referências sobre modelos de maturidade em GC mais recorrentes. Contudo, não houve um critério detalhado e específico para a seleção da fonte material, pois o tema investigado parte de uma premissa holística.

Por meio da revisão narrativa permitiu-se estabelecer relações com produções anteriores, identificar temáticas recorrentes, apontar novas perspectivas, mapear uma área de conhecimento e apropriar-se de orientações de práticas de GC para a definição dos parâmetros de formação de um sistema de avaliação mais amplo.

4. Parâmetros de Avaliação Multidimensionais

Para gerir o conhecimento corporativo é preciso lidar com a complexidade organizacional. Resolver a complexidade de um sistema de atividade humana, como é o caso das organizações, faz-se necessário construir uma visão macro, que envolva as diversas perspectivas do ambiente. Essas perspectivas também podem ser enxergadas como dimensões de avaliação [Checkland 1993, Morins 2011]. Assimilar a realidade contextual do negócio e os fatores condicionantes do ambiente organizacional consiste em construir uma visão de múltiplas dimensões [Gharajedaghi 2011].

O pensamento complexo afirma que a realidade não deve ser fragmentada, pois não seriam levados em consideração relacionamentos importantes, portanto, deve-se compreender a multidimensionalidade dos processos [Moraes and Valente 2008].

As organizações, principalmente aquelas orientadas ao conhecimento e à aprendizagem colaborativa, estão inseridas em mercados inovadores e dinâmicos. Como foi relatado, as práticas de GC são importantes para a vantagem competitiva das empresas e estão intimamente ligadas com a elaboração das estratégias do negócio. Desta forma, não há como realizar GC sem compreender a realidade complexa do mundo corporativo, levando em consideração o contexto multidimensional de cada organização [Johnson et al. 2007].

O estudo de Lima et al. [2015] propõe um construto para entendimento da complexidade sistêmica das organizações. Esse construto propõe uma visão de múltiplas dimensões que buscam sintetizar o contexto organizacional, promovendo uma visão holística. As dimensões definidas são: perspectiva interna e externa, domínio de arquitetura e tomada de decisão, escopo e tempo e domínio sistêmico – Figura 1.

A perspectiva interna diz respeito às funções, processos e estrutura da organização com foco no ambiente interno. A externa refere-se à stakeholders, contexto em que a organização se encontra e modelo de negócio. O domínio de arquitetura inclui a arquitetura técnica, de negócio, social e de informação. A tomada de decisão inclui os processos de compreensão, decisão e monitoramento do negócio. O escopo apresenta os cursos de ação estratégico e tático. O parâmetro temporal sistematiza a análise dos objetivos da organização para curto, médio e longo prazo. O domínio sistêmico define a necessidade da existência de uma visão holística, incluindo o modelo do mundo real, que apresenta parâmetros de prospecção e percepção da realidade da organização e o modelo de sistemas, onde são definidos os sistemas e subsistemas que compõem o negócio, com o objetivo de mapear o contexto organizacional [Lima et al. 2015].

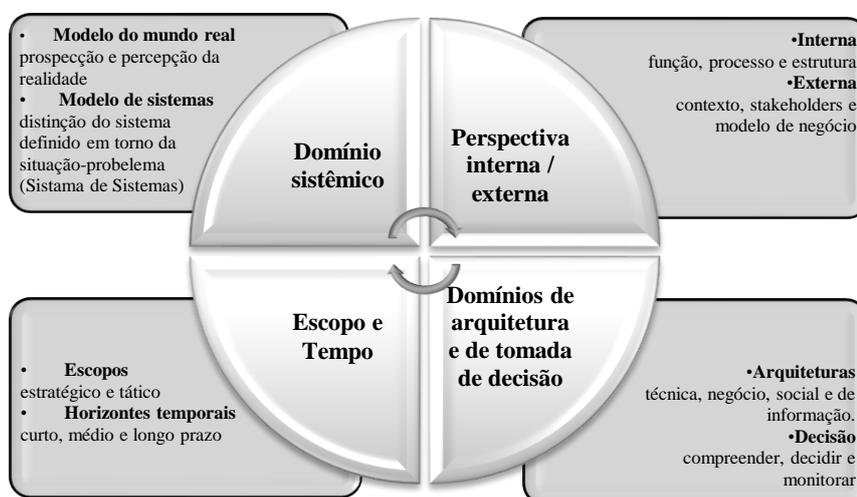


Figura 1. Construto de Visão Multidimensional [Lima et al. 2015]

Em alusão à visão da empresa que aprende de Senge [2013], da visão holística de GC de Kankanhali [2009] e da visão multidimensional de Lima et al. [2015], para avaliar a maturidade em GC nas organizações, os modelos devem levar em consideração a complexidade sistêmica, ou seja, o conhecimento pode ser captado e aproveitado à medida que as práticas de GC levam em consideração a realidade dinâmica e multidimensional das organizações. Com base na pesquisa bibliográfica foi construído o Quadro 1, relacionando quais dimensões são contempladas pelos modelos de GC pesquisados. A marcação “C” significa que a dimensão é contemplada pelo

modelo. A marcação “NC” significa que não contempla ou que não foram encontradas evidências suficientes. Por fim, a sigla “CP” indica que a dimensão é contemplada parcialmente.

Quadro 1. Modelos de Maturidade versus Dimensões de Complexidade Sistêmica Organizacional

Modelos de Maturidade	Domínio Sistêmico		Perspectiva Interna e Externa		Domínios de Arquitetura e Tomada de Decisão		Escopo e Tempo	
	Mundo Real	Sistemas	Interna	Externa	Arquitetura	Tomada de Decisão	Escopos	Horizontes Temporais
Mehta et al. [2007]	NC	NC	C	NC	CP	CP	CP	CP
Kruger e Johnson [2009]	CP	NC	C	NC	CP	NC	NC	CP
Lin et al. [2012]	NC	NC	C	NC	NC	NC	NC	CP
Jiuling et al. [2012]	NC	CP	C	NC	CP	NC	NC	NC
Paulzen et al. [2002]	CP	NC	C	NC	CP	CP	C	CP
Lee e Kim [2001]	CP	NC	C	C	CP	CP	C	CP
Kulkarni e Freeze [2004]	CP	NC	C	NC	CP	CP	NC	CP
Ehms e Langen [2002]	CP	CP	C	C	C	C	C	C
Teah et al. [2007]	NC	NC	C	NC	CP	NC	NC	CP
Robinson et al. [2006]	CP	NC	C	NC	CP	C	C	CP
Kathibian et al. [2010]	CP	CP	C	NC	CP	C	C	CP
Kruger e Snyman [2007]	NC	NC	C	NC	CP	C	C	C
Lin [2007]	CP	NC	C	C	CP	CP	CP	CP
Kruger e Johnson [2010]	CP	CP	C	NC	CP	CP	CP	CP
Hsieh et al. [2009]	NC	NC	C	NC	CP	CP	CP	CP
Pee e Kankanhali [2009]	C	C	C	NC	CP	CP	CP	CP
Oliveira e Pedron [2014]	NC	NC	C	C	CP	CP	CP	CP
Serna [2012] e Galati [2015]	C	CP	C	C	CP	C	C	C

Fonte: Autoria própria [2017]

Por meio da análise da pesquisa bibliográfica, destacando-se o quadro relacional apresentado por Oliveira et al. [2011], percebe-se que indiretamente, os modelos de maturidade levam em consideração diversos fatores da complexidade sistêmica proposta no construto de Lima et al. [2015], porém não de forma sistematizada, ou seja, não são

claramente evidenciadas as dimensões organizacionais. Para englobar todos estes fatores em um modelo, é essencial a compreensão de como eles são categorizados. O quadro 1 repercute os avanços dos modelos de maturidade na compreensão do contexto complexo das organizações e é possível verificar quais as deficiências e necessidades de melhoria de cada um dos padrões.

Em síntese, pode-se promover a análise interpretativa de modelos de maturidade e ferramentas de avaliação de GC, considerando os níveis de satisfação em termos das dimensões da complexidade sistêmica organizacional. Ainda que nenhum modelo avaliado tenha contemplado integralmente as dimensões avaliadas, o que importa nesse momento é observar os fatores organizacionais existentes na perspectiva sistêmica, fatores esses que devem ser considerados ao se aplicar GC nas práticas estratégicas, gerenciais e operacionais das organizações.

5. Conclusão

A partir dos achados da pesquisa bibliográfica, ao relacionar os principais modelos de maturidade com as dimensões do construto de visão multidimensional da complexidade sistêmica, um novo alento pode ser alcançado, tendo em vista os preceitos relacionais, contextuais, funcionais, informacionais e processuais presentes no âmbito organizacional.

A revisão narrativa promovida nesta pesquisa permitiu revelar aspectos recorrentes nos modelos de avaliação de maturidade de GC. Destes, a visão holística da organização predomina quando se busca uma avaliação mais criteriosa da realidade observada. A visão holística vem sendo incorporada progressivamente como critério de avaliação dos modelos de maturidade mais recentes, ou seja, mesmo que não seja de forma sistematizada, os modelos têm abandonado a visão cartesiana de avaliar as organizações, substituindo-a pela visão sistêmica.

A perspectiva interna do Quadro 1 é contemplada em todos os modelos, porém, em relação à perspectiva externa, existem iniciativas tímidas de alguns deles. Considerando a organização como uma realidade complexa e guiada por relacionamentos, o fluxo de informação e conhecimento inevitavelmente atingirá o ambiente externo. Nesse contexto, é essencial que tal perspectiva seja considerada.

Em relação aos horizontes temporais, nota-se que os modelos consideram o avanço da organização através dos níveis de maturidade (que levará determinado tempo), porém, não foi apresentada em nenhum deles a necessidade enfática do alinhamento estratégico para o estabelecimento de metas de curto, médio e longo prazo.

Como trabalho futuro, sugere-se que o modelo que deu origem às dimensões da complexidade sistêmica, adotada neste artigo, seja considerado para a elaboração de uma ferramenta contextual de avaliação de maturidade em GC no ambiente organizacional.

Do mesmo modo, a estruturação dos sistemas de informação em geral envolve a percepção da realidade complexa e multidimensional das organizações para que estas atinjam de forma satisfatória o objetivo de configurar-se como agente facilitador e integrador das diversas dimensões que culminam com a maior propensão em produzir, compartilhar, sistematizar e disseminar o conhecimento. Promover discussão envolvendo o modelo utilizado como parâmetro para a avaliação de maturidade em

cenários complexos e a estruturação da análise dos modelos de maturidade em GC, contribuiriam para entendimento da realidade complexa das organizações, podendo embasar a sistematização da informação no âmbito corporativo.

Referências Bibliográficas

- Castro, G. D. (2005). Gestão do conhecimento em bibliotecas universitárias: um instrumento de diagnóstico.
- Checkland, P. (1981). Systems thinking, systems practice.
- Chen, F., e Burstein, F. (2006). A dynamic model of knowledge management for higher education development. In Information Technology Based Higher Education and Training, 2006. ITHET'06. 7th International Conference on (pp. 173-180). IEEE.
- Cruz, C. A. (2007). Gestão Estratégica do Conhecimento: estudo exploratório em empresas instaladas nos parques tecnológicos do estado de São Paulo (Doctoral dissertation, Universidade de São Paulo).
- Del-rey-camorro, F. M., Roy, R.; Wegen, B. V. e Steele, A. (2003), A framework to create key performance indicators for knowledge management solutions. Journal os Knowledge Management, vol. 7(2), pp. 46-62.
- Davenport, T. e Prusak L. (1998), Conhecimento Empresarial: Como as Organizações Gerenciam o seu Capital Intelectual. Rio de Janeiro: Campus, 6ª ed.
- Drucker, P. F. (1997). Rumo à nova organização. HESSELBEIN, Frances, 15-19.
- Ehms, K. e Langen, M. (2002), Holistic development of knowledge management with KMMM. Siemens AG, Alemanha.
- Galati, F. (2015). At what level is your organization managing knowledge? Measuring Business Excellence, 19(2), 57-70.
- Gharajedaghi, J. (2011). Systems thinking: Managing chaos and complexity: A platform for designing business architecture. Elsevier.
- Goldoni, V. (2007). Indicadores para avaliação da gestão do conhecimento: o caso de empresas de desenvolvimento de software (Master's thesis, Pontificia Universidade Católica do Rio Grande do Sul).
- Gonçalo, C. R., Junges, F. M. e Borges, M. L. (2010). Avaliação da gestão do conhecimento: modelos de mensuração. Encontro Nacional de Engenharia de Produção, São Carlos. Anais São Carlos: ABEPRO. p.15.
- Hartono, B., Ramadhani, V. S., Ratnadilla, Y., Indarti, N. e Chai, K. H. (2015). Knowledge management maturity and organizational performance in project-based organizations: Initial evidence from Indonesia. In Industrial Engineering and Engineering Management (IEEM), 2015 IEEE International Conference on (pp. 656-660). IEEE.
- Heerdt, A. P. S. (2002). Competências essenciais dos coordenadores de curso em uma instituição de ensino superior.
- Hooff, B. V. D. e Huysman, M. (2009), Managing knowledge sharing: Emergent and engineering approaches. Information & Management, vol. 46(1), pp. 1-8.

- Hsieh, P. J., Lin, B. e Lin, C. (2009). The construction and application of knowledge navigator model (KNM™): An evaluation of knowledge management maturity. *Expert Systems with Applications*, 36(2), 4087-4100.
- Jasimuddin, S. M. (2007), Exploring knowledge transfer mechanisms: the case of a UK-based group within a high-tech global corporation. *International Journal of Information Management*, vol.27(4), pp. 294-300.
- Johnson, G., Scholes, K. e Whittington, R. (2009). *Explorando a estratégia corporativa: texto e casos*. Bookman Editora.
- Jiuling, X., Jiankang, W. e Hongjiang, Y. (2012, October). Study on maturity level transition mechanism of knowledge management. In *Information Management, Innovation Management and Industrial Engineering (ICIII), 2012 International Conference on* (Vol. 1, pp. 325-328). IEEE.
- Jones, P. M. (2001). Collaborative knowledge management, social networks, and organizational learning. *Systems, Social and Internationalization Design Aspects of Human-Computer Interaction*, 2, 306-309.
- Kraemer, R., de Sá Freire, P., de Souza, J. A. e Dandolini, G. A. (2017). Maturidade de Gestão do Conhecimento: Uma Revisão Sistemática da Literatura para Apoiar o Desenvolvimento de Novos Modelos de Avaliação. *Perspectivas em Gestão & Conhecimento*, 7(1), 66-79.
- Kruger, C. J. e Johnson, R. D. (2010). Principles in knowledge management maturity: A South African perspective. *Journal of Knowledge Management*, 14(4), 540-556.
- Kruger, C. J. e Snyman, M. M. M. (2007), Guidelines for assessing the knowledge management maturity of organizations. *South African Journal of Information Management*, vol. 9(3), pp. 1--11.
- Kulkarni, U.; Ravidran, S. e Freeze, R. (2007), A knowledge management success model: Theoretical development and empirical validation. *Journal of Management Information Systems*, vol. 23(3), pp. 309-347.
- Kuriakose, K. K., Raj, B., Satya Murty, S. A. V. e Swaminathan, P. (2010). Knowledge management maturity models—a morphological analysis. *Journal of Knowledge Management Practice*, 11(3), 1-10.
- Lee, J. H. e Kim, Y. G. (2001), A stage model of organizational knowledge management: a latent content analysis. *Expert Systems with Applications*, vol. 20(4), pp. 299-311.
- Lima, E. A. (2007). *Plataformas de DTV: Impactos da Convergência Digital*. Monografia de Especialização em Telecomunicações, Escola de Engenharia, UFG, Goiânia-GO.
- Lima, E. A. D. (2009). *Factores habilitadores de gestão do conhecimento: estudo de caso no Sector de Educação Superior brasileiro* (Doctoral dissertation).
- Lima, E. A., Molinaro, L. F., Garrossini, D. F. e Soares, C. M. (2015). *Complexidade Sistêmica: Uma Proposta de Construto de Visão Multidimensional*.
- Lin, H.F. (2007), A stage model of knowledge management: an empirical investigation of process and effectiveness. *Journal of Information Science*, vol. 33(6), pp. 643-659.

- Lin, C., Wu, J. C. e Yen, D. C. (2012). Exploring barriers to knowledge flow at different knowledge management maturity stages. *Information & Management*, 49(1), 10-23.
- Lodi, J. B. (1984). *História da Administração*. 8ª. São Paulo: Pioneira.
- Luchesi, E. S. F. (2012). *Gestão do conhecimento nas organizações*. Companhia de Engenharia de Tráfego de São Paulo (CET): Notas Técnicas (NT), 221.
- Mehta, N., Oswald, S. e Mehta, A. (2007). Infosys Technologies: improving organizational knowledge flows. *Journal of Information Technology*, 22(4), 456-464.
- Moraes, M. C. e Valente, J. A. (2008). *Como pesquisar em educação a partir da complexidade e da transdisciplinaridade*. São Paulo: Paulus.
- Morin, E. (2011), *Introdução Ao Pensamento Complexo*, Sulina, Porto Alegre, RS, 4ª Ed.
- Natale, C. H. C., Neves, J. T. R. e Carvalho, R. B. (2016). Maturidade em gestão do conhecimento: análise das percepções dos gestores de uma grande empresa de construção civil. *Informação & Informação*, 21(1), 375-406.
- Naves, F. L., Mafra, L. A. S., Gomes, M. A. O. e Amâncio, R. (2000). Diagnóstico organizacional participativo: potenciais e limites na análise de organizações. *Organizações & Sociedade*, 7(19), 53-66.
- Oliveira, M., Pedron, C., Romão, M. e Becker, G. (2011). Proposta de um modelo de maturidade para Gestão do Conhecimento: KM3. *Revista Portuguesa e Brasileira de Gestão*, 10(4), 11-25.
- Oliveira, M. e Pedron, C. D. (2014, September). Maturity Model for Knowledge Management and Strategic Benefits. In *European Conference on Knowledge Management (Vol. 2, p. 748)*. Academic Conferences International Limited.
- Oliveira, M., Pedron, C. D., Nodari, F. e Ribeiro, R. (2014). Knowledge management in small and micro enterprises: applying a maturity model. In *European Conference on Knowledge Management (Vol. 2, p. 757)*. Academic Conferences International Limited.
- Paulzen, O., Doumi, M., Perc, P. e Cereijo-Roibas, A. (2002), A maturity model for quality improvement in knowledge management. *ACIS Proceedings*.
- Pee, L. G. e Kankanhalli, A. (2009). A model of organisational knowledge management maturity based on people, process, and technology. *Journal of Information & Knowledge Management*, 8(02), 79-99.
- Pérez, J. E. A. (2012). Associations between the maturity of knowledge management and innovative performance: organization and people, and interpretation. *Revista Lasallista de Investigación*, 9(1), 86-95.
- Perrotti, E. (2004), *Estrutura Organizacional e Gestão do Conhecimento*. Tese de Mestrado em Administração, USP, São Paulo-SP.
- Rauniar, R., Rawski, G. e Meacham, J. (2007). Collective Ambition, Creative Chaos, Information Redundancy, and Shared Knowledge in Integrated Product Development—Case Study. *Journal of Information & Knowledge Management*, 6(02), 79-89.

- Rother, E. T. (2007). Revisão sistemática X revisão narrativa. *Acta paulista de enfermagem*, 20(2), v-vi.
- Robinson, H. S., Anumba, C. J., Carrillo, P. M. e Al-Ghassani, A. M. (2006), STEPS: a knowledge management maturity roadmap for corporate sustainability. *Business Process Management Journal*, vol. 12(6), pp. 793-808.
- Rummler, G. A., & Brache, A. P. (1994). *Melhores desempenhos das empresas*.
- Salojärvi, S., Furu, P. e Sveiby, K. E. (2005). Knowledge management and growth in Finnish SMEs. *Journal of knowledge management*, 9(2), 103-122.
- Senge, P. M. (2009). *A quinta disciplina: arte e prática da organização que aprende*. Best Seller.
- Serna, E. (2012). Maturity model of Knowledge Management in the interpretivist perspective. *International Journal of Information Management*, 32(4), 365-371.
- Smith, K. G., Mitchell, T. R. e Summer, C. E. (1985), Top level management priorities in different stages of the organizational life cycle. *Academy of Management Journal*, vol. 28(4), pp. 799-820.
- Sujanova, J., Gabris, P., Licko, M., Pavlenda, P. e Stasiak-Betlejewska, R. (2012). Aspects of knowledge management in Slovak industrial enterprises. In *Proceedings of the 13th European Conference on Knowledge Management (Vol. 60, pp. 1135-1144)*.
- Sujanova, J., Caganova, D. e Soos, L. (2015). Innovation, knowledge and multicultural management influence on intellectual capital in industrial enterprises. *Turkish Online Journal of Educational Technology*, v. 2015, p. 289-300.
- Terra, J. C. C. (2005). *Gestão do conhecimento: o grande desafio empresarial*.
- Teah, H. Y., Pee, L. G. e Kankanhalli, A. (2006). Development and application of a general knowledge management maturity model. *Tenth Pacific Asia Conference on Information Systems*.
- Trigo, M. R., Vasconcelos, J. B., Camelo, C. e Quoniam, L. (2007). *A organização em comunidades de prática com o objetivo de facilitar a implementação do planejamento estratégico numa Instituição do Ensino Superior*.
- Van-de-vem, A. H. e Poole, M. S. (1995). Explaining development and change in organizations. *Academy of Management Review*, vol. 20(3), pp. 510-540.
- Wheatley, M. J. (1996). *Liderança e a nova ciência*. Editora Cultrix.

Caracterização de Mensagens Em Língua Portuguesa Com Traços de Racismo No Twitter

Rodolfo Costa Cezar da Silva¹, Deborah Silva Alves Fernandes¹,
Marcio Giovane Cunha Fernandes²

¹Instituto de Informática – Universidade Federal De Goiás(UFG)
CEP 74690-900 - Goiânia - GO - Brazil

²CCET – Universidade Estadual de Goiás(UEG)
Durham, U.K.

{rodolfocosta.ufg, debmarfer}@gmail.com, marcio.giovane@ueg.br

Abstract. *This article provides the characterization of textual data with traces of racism in Portuguese obtained from Twitter. It discusses racism in ethical and legal terms for conceptual identification. It explains Sentiment Analysis and some approaches and possible levels of analysis. At the end, the facts perceived during the data collection process are discussed.*

Resumo. *Este artigo provê a caracterização de dados textuais com traços de racismo em língua portuguesa obtidos na RSO Twitter. Discute racismo em termos éticos e legais para a identificação conceitual. Explica Análise de Sentimento e algumas abordagens e os níveis possíveis de análise. Ao final, discute-se os fatos percebidos durante o processo de coleta de dados.*

1. Introdução

As Redes Sociais Online (RSO) conduziram o Processamento de Linguagem Natural (PLN) para a Análise de Sentimentos (AS), cujo objetivo principal é automatizar a busca de opinião (o boca-boca na internet) sobre algum tema no grande volume de dados não estruturados disponíveis eletronicamente na Internet.

Opinião não é um fato objetivo, mas um conceito formado por diversas experiências vivenciadas por um indivíduo ou por um grupo. A natureza dinâmica do conceito é percebida em cada relação estabelecida entre os entes sociais ou a partir de um fato novo observado. Tal comportamento dinâmico é potencializado quando o universo de dados e relações aumenta vertiginosamente, o que é observado com a absorção das RSO no cotidiano mundial.

Não obstante, tal potencialidade ressalta aspectos positivos e negativos do cognitivo humano. Esse artigo perpassa pelo interesse nos meios de automatização do reconhecimento daqueles negativos relacionados aos discursos de ódio e preconceito. A Análise de Sentimentos é um desses meios.

Análise de Sentimentos, também conhecida por Mineração de Opinião, é a área de estudo que analisa opiniões, sentimentos, avaliações, apreços, atitudes e emoções das pessoas sobre entidades tais como produtos, serviços, organizações, indivíduos, problemas, eventos, tópicos e seus atributos [Liu 2012].

Essa técnica tem sido amplamente utilizada em uma variedade de aplicações, em diversos nichos de pesquisa. Cada aplicação utiliza as técnicas para fazer previsões, auxiliar na tomada de decisões, classificar sentimento público, entre outras finalidades.

Em 2008 [Pang and Lee 2006] descreveu técnicas e abordagens que visam auxiliar sistemas de informação orientados a opinião. O foco da pesquisa é introduzir métodos que tratam os desafios apresentados pelas aplicações que envolvem análise de sentimento e mineração de opinião tais como subjetividade e ambiguidade da linguagem natural, e trazer aplicações práticas para esse métodos.

Desta feita, este artigo objetiva o desenvolvimento em primeira instância de elementos viabilizadores à caracterização de traços de racismo em mensagens textuais em língua portuguesa publicadas na RSO Twitter. Na seção 2 são tratados racismo e injúria racial, além de exemplos de racismo na internet, especificamente em RSO. A seção 3 aborda Análise de Sentimento. A seção 4 trata da coleta e caracterização de dados. As próximas etapas do trabalho são relatadas na seção 5.

2. Racismo

2.1. Definições Gerais

Apesar da população brasileira ser formada por aproximadamente 50.47% de pessoas pretas e pardas, segundo o Censo Demográfico do IBGE de 2010¹, o racismo ainda é uma questão recorrente no Brasil.

Embora seja um problema frequente, sua definição não é tão trivial, tendo em vista que ideias racistas podem ser expressas e percebidas de várias maneiras. Segundo [Martins 2014], "racismo é o conjunto de teorias e crenças que estabelecem uma hierarquia entre as raças e etnias. É uma doutrina ou sistema político fundado sobre o direito de uma raça (considerada pura ou superior) de dominar as outras. Por fim, é um preconceito extremado contra indivíduos pertencentes a uma raça ou etnia diferente, considerada inferior".

Tratando-se de leis brasileiras, o crime de injúria racial está associado ao uso de palavras depreciativas referentes à raça ou cor com a intenção de ofender a honra da vítima, já o crime de racismo, previsto na Lei nº 7.716/1989, implica em conduta discriminatória dirigida a um determinado grupo ou coletividade e, geralmente, refere-se a crimes mais amplos.

2.2. Racismo na Internet

O racismo sempre existiu, porém com o desenvolvimento da internet e das redes sociais, as pessoas passaram a ter uma plataforma para expressar e propagar suas opiniões, crenças e sentimentos com maior facilidade e visibilidade para outros usuários. Alguns casos podem ser citados:

- Em 2005, o estudante de Letras na Universidade de Brasília (UnB), Marcelo Mello discutia o sistema de cotas para negros na sua universidade pelo Orkut (rede social descontinuada em 2014). Durante essa discussão, Marcelo se referiu aos negros e afrodescendentes como "burros", "urubus", "macacos subdesenvolvidos", entre outras ofensas [ISTOÉ Independente 2015].

¹Disponível em : <https://www.ibge.gov.br/>

- A jornalista Maria Júlia Coutinho, a Maju do "Jornal Nacional", foi vítima de comentários preconceituosos na página oficial do programa no Facebook (Figura 1), em julho de 2016, logo quando se destacou pela sua cobertura da previsão do tempo.
- No dia 30 de junho de 2018, um *youtuber* conhecido (Figura 2) fez comentários racistas durante a Copa do Mundo FIFA de 2018, sobre o jogador da seleção francesa, Kylian Mbappé, associando o jogador com suas possíveis habilidades de realizar arrastões na praia.

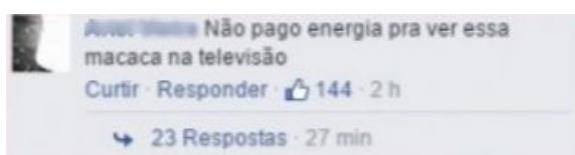


Figura 1. Comentário racista na página do Facebook de Maju



Figura 2. Youtuber faz comentário racista sobre jogador francês no Twitter.

Devido a subjetividade e ambiguidade, identificar traços de racismo automaticamente em textos, é uma tarefa bastante desafiadora.

2.3. Trabalhos relacionados a Racismo nas Redes Sociais

Um sistema de classificação automática de textos com traços racistas é proposto em [Greevy and Smeaton 2004]. Treinaram uma *Support Vector Machine*(SVM) com os padrões encontrados a partir de *Bag-of-Words*(BOW) e bigramas.

O corpus de 3 milhões de palavras foi dividido em *datasets* de diversos tamanhos que continham o mesmo número de documentos racistas e não-racistas. Utilizando um conjunto de treinamento de 2 mil documentos e outro de teste com 410 documentos, concluíram que a técnica de SVM combinada com BOW obteve melhor resultado comparada a técnica de SVM combinada com bigramas cujas taxas de precisão foram de 87.33% e 84.77%, respectivamente.

Modelos para classificação de discurso de ódio da internet são apresentados por [Waseem and Hovy 2016], especificamente no Twitter, para uma gama de características como etnia, deficiência e orientação sexual. Para tal, utilizaram técnicas para extração de características dos textos, relação sintática e gramatical entre palavras. O artigo explora a influência de características diferentes na tarefa de classificação, e para isso testa todas as combinações de características para verificar qual leva a um melhor resultado. A combinação de bigramas até 4-gramas combinada com o gênero obteve os melhores resultados, com 73.66% de acurácia.

Pesquisadores propuseram em [José De Alencar 2012] mapear e mensurar a ocorrência de *cyberbullying* contra professores no Twitter através de técnicas de aprendizado de máquina. Os dados foram coletados durante uma semana através da API do Twitter, foram também pré-processados e classificados em três categorias: positivo, negativo, e neutro. Para classificação utilizaram um classificador bayesiano (*Naive-Bayes*) que obteve uma acurácia de 87.1%.

3. Análise de Sentimentos

Análise de sentimentos, também chamada de mineração de opinião, é uma área que estuda as opiniões, sentimentos, avaliações, atitudes e emoções sobre entidades, tais como produtos, serviços, organizações, indivíduos, eventos, tópicos e seus atributos [Hirst and Liu 2012].

A análise de sentimentos trabalha em diferentes níveis de granularidade. Em geral, essa análise pode ser feita em três níveis : (a) *Nível de documento*, em que o trabalho consiste em classificar se um documento como um todo expressa um sentimento positivo ou negativo. Este nível de análise assume que cada documento expressa opiniões sobre uma única entidade (b) *Nível de sentença* consiste em analisar a polaridade do sentimento de uma apenas sentença. Assume-se que em um documento pode conter várias sentenças que podem possuir um sentimento individual. Cabe ressaltar que, em geral, postagens e comentários em mídias sociais seguem um padrão de sentenças curtas [Benevenuto 2015] (c) *Nível de entidade e aspecto*, em que a granularidade é menor, tenta definir um sentimento sobre uma entidade, analisa diretamente a opinião em si, e é baseado na ideia em que uma opinião consiste em um *sentimento*, e um *alvo*. O objetivo desse tipo de análise é descobrir as entidades e as respectivas opiniões sobre elas separadamente.

3.1. Abordagens

3.1.1. Aprendizado de Máquina

O aprendizado de máquina pode ser supervisionado, semi-supervisionado, ou não supervisionado. O primeiro consiste em fornecer um conjunto de dados de treino previamente coletados e anotados (com alguma polaridade de sentimentos) à um classificador que pode então, classificar novos dados após esse treinamento. Um dos métodos que podem ser utilizados é o *Support Vector Machines* (SVM) que classifica textos em alguma polaridade dado certo treinamento. Seu princípio é determinar separadores lineares no espaço de busca, chamados de *hiperplanos*. Através dos dados de treinamento, a SVM cria um hiperplano que divide as classes.

Outro método classificador é a Regressão Logística, um modelo probabilístico que atribui pesos (B) para *features* (X), dado um conjunto de treino conhecido (y). Esse encontra o conjunto de pesos que maximiza a probabilidade de $P(X|B, y)$. Para o aprendizado de máquina supervisionado também pode ser adotado um classificador *Naive-Bayes*, que é uma especificação da Regressão Logística que usa o *Teorema de Bayes* para determinar o conjunto de pesos. Um dos componentes necessários para os classificadores supervisionados são conjuntos de dados rotulados para o treinamento. Esses dados geralmente são manualmente anotados por seres humanos, ou por algum tipo de serviço de colaboração coletiva (*crowdsourcing*, do inglês) como o *Figure-Eight*².

Na abordagem não-supervisionada, também conhecida como aprendizado por observação e descoberta, não há o uso de dados de treinamento, nesse tipo de abordagem o classificador deve encontrar padrões automaticamente sem conhecimento prévio dos dados. Esse tipo de método geralmente usa abordagens léxicas para classificação que utilizam dicionários léxicos de sentimentos. Tais dicionários associam uma palavra com

²<https://www.figure-eight.com/>

um significado quantitativo, que varia de $[-1, 1]$ de acordo com sua polaridade, ou qualitativo, que associam uma palavra a uma certa polaridade (positivo/negativo, feliz/triste).

É possível utilizar algoritmos de aprendizado para descobrir padrões nos dados a partir de alguma caracterização de regularidade, sendo esses padrões denominados *clusters* [Decker and Focardi 1995]. Exemplos que estão contidos em um mesmo *cluster* são mais similares do que exemplos contidos em clusters diferentes. O processo de formação de *clusters* é conhecido por *clustering*, e pode ser feitas usando várias técnicas que incluem : Cluster K-Médio, Cluster Hierárquico e Mapas Auto-Organizadores.

Existe também a abordagem semi-supervisionada, que parte do pressuposto que o conjunto dos dados disponíveis para treinamento são formados por uma parte rotulada, em menor número, e outra não rotulada, em maior número [Vicente et al.]. A ideia é utilizar os exemplos rotulados para obter informações sobre o problema e utilizá-las para guiar o processo de aprendizado a partir dos exemplos não rotulados [Bruce 2001].

4. Caracterização dos Dados

4.1. Proposta de Arquitetura

A Figura 3 ilustra o modelo da arquitetura proposta para o problema. O processo se inicia através de um *crawler*, que é um software desenvolvido para extrair dados vistos como importantes de *tweets* do servidor do Twitter em tempo real. A medida em que essas informações são capturadas, elas são armazenadas em um banco de dados junto com outras informações sobre o *tweet*. Após formar um *corpus* considerável, com cerca de 28.000 tweets, decidiu-se realizar uma análise inicial dos dados a fim de extrair e evidenciar as características do dados, e também de identificar a viabilidade e relevância da pesquisa. Em trabalhos futuros, pretende-se construir um classificador automático de mensagens com traços de racismo através de um modelo supervisionado de aprendizado de máquina. Para tal, é necessário realizar um pré-processamento dos dados com o intuito de prover uma formatação, como a seleção de atributos, tratamento de valores ausentes(*missing values*). Esse pré-processamento é uma parte crucial na tarefa de classificação, pois sem essa técnica, a acurácia do classificador é comprometida.

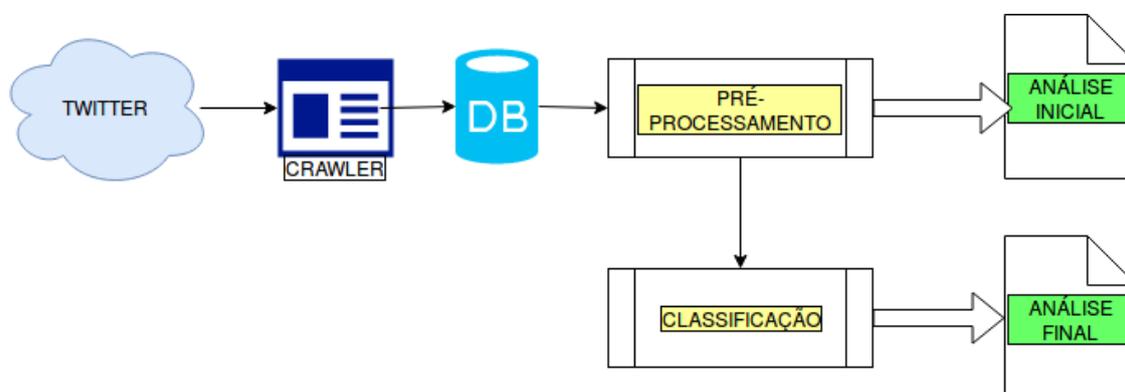


Figura 3. Arquitetura proposta para análise de traços de racismo em tweets.

4.2. Coleta

Criado em 2006, Twitter³ é uma rede social popular com um serviço de *microblogging* com cerca de 330 milhões de usuários (27.7 milhões no Brasil, que é o segundo país com mais usuários ativos). Cada usuário possui um conjunto de inscritos chamados de *seguidores*, e pode publicar atualizações periódicas, chamadas de *tweets*, que consistem em pequenas mensagens de até 280 caracteres.[Aslam 2018].

A coleta de dados foi realizada entre os dias 04/06/2018 e 19/07/2018 no Twitter. A escolha dessa rede social ao invés de outras disponíveis, se deu pelo fato de ser uma rede social popular no Brasil, e da plataforma oferecer uma *Application Programming Interface*(API) que permite, de forma fácil, coletar dados em tempo real.

Os *tweets* coletados continham termos⁴ que foram definidos previamente através de leituras e investigação de textos e casos sobre racismo na internet, redes sociais, jornais, palavras usadas em sistemas de buscas online, e também por meio de um questionário próprio(desenvolvido e disponibilizado no Google Forms) que ficou disponível durante uma semana no Twitter, a rede social alvo da pesquisa, e nos permitiu defrontar com outros termos antes desconhecidos. Dentre as perguntas do questionário, destacam-se: "Você já sofreu por e/ou presenciou algum tipo de comentário que continha traços racistas em seu conteúdo?" e "O que caracterizava aquele texto como racista? Algum(ns) termo(s) em específico?".

Um dos obstáculos foi o fato da API coletar *tweets* cujo o nome de usuário, e não o texto do *tweets* em si, continham os termos de busca, *i.e.*, um *tweet* enviado por um usuário chamado "spider gorila"era coletado pelo *crawler*, mesmo que o corpo da mensagem não apresentasse nenhum comentário com traços racistas. Outra dificuldade encontrada foi a lapidação do conjunto de termos que seriam buscados. Nas coletas de teste, os termos "nega" e "neguinha" estavam incluídos nos termos de busca, mas após análise percebeu-se que estes termos só se tratavam de palavras de cunho racista em contextos específicos. O primeiro termo, na maioria da vezes, se referia a uma conjugação do verbo "negar", e o segundo retornava várias mensagens de caráter apreciativo e afetuoso, e não discriminativo, portanto eles foram retirados do conjunto final de termos de busca.

4.3. Usuários mais ativos

O gráfico da Figura 4 apresenta a frequência em que certos usuários publicaram mensagens na coleta. Após análise do gráfico, decidiu-se investigar os três usuários que publicaram mais no conjunto de mensagens coletadas. Segue abaixo, uma tabela de características dos usuários e exemplos das mensagens publicadas por eles.

Após investigação, notou-se que o usuário 1, quem mais publicou mensagens durante o período desse estudo, mostrou-se engajado em debater, comentar e expressar suas opiniões na rede social sobre notícias que envolvem racismo, para imputar e culpabilizar outros por suas atitudes. Descobriu-se que o motivo das publicações do usuário 2 serem tão recorrentes no conjunto de *tweets* coletados é que, em todas suas mensagens coletadas, o

³<https://twitter.com/>

⁴Termos de busca: "senzala", "gorila", "cabelo de bombril", "cabelo de esfregão", "nariz de nego", "nariz de nega", "tinha que ser preto", "tinha que ser preta", "preto da senzala", "preta da senzala", "preto da macumba", "nega macaca", "preto macaco", "preta macaca", "preta nojenta", "preto nojento", "criola", "crioula", "crioulo".

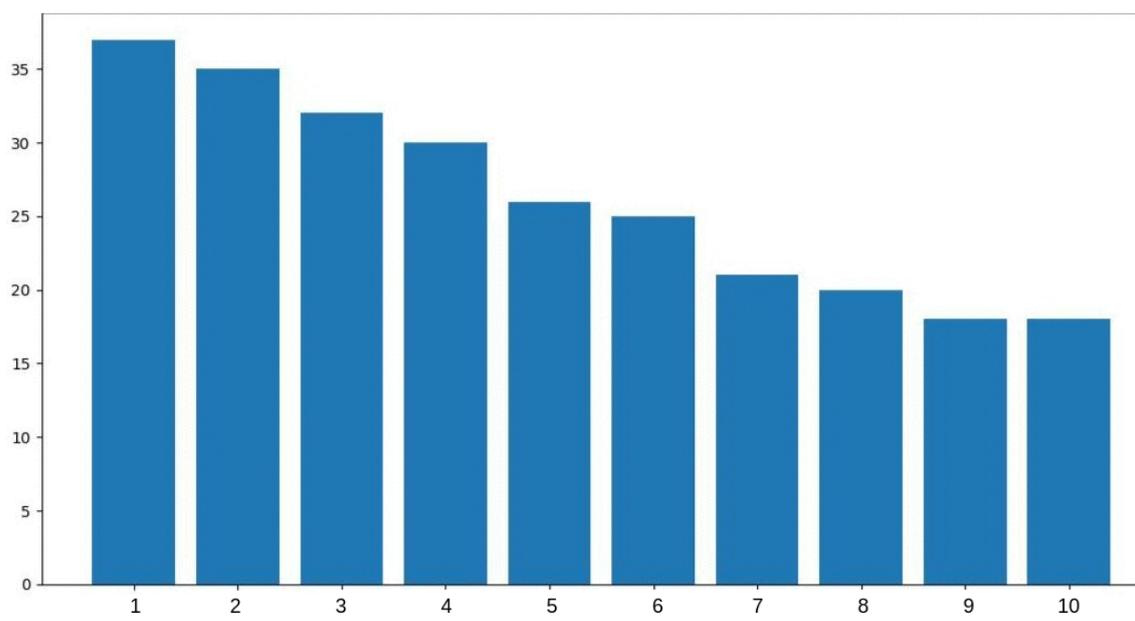


Figura 4. Quantidade de tweets por usuário.

Identificação do usuário	Localização	Usuário do Twitter desde	data	Exemplo de Tweet
usuário1	Rio de Janeiro - RJ	março de 2011	18/06/2018	"Só vota em Ciro Gomes , um racista safado , negros que não tem vergonha na cara e gostam de serem tratados como animais por um Coronel(...)"
usuário2	Rio de Janeiro - RJ	março de 2017	07/06/2018	"como o MC Gorila fala numa de suas grandes obras : "Olha a Evolução" https://t.co/1QxKgrulLc "
usuário3	-local inválido-	janeiro de 2017	08/06/2018	"O gorila do respeito é como o centeio do passado"

Figura 5. Tabela de características dos usuários.

usuário faz menções a um cantor de funk da cidade do Rio de Janeiro, cujo nome artístico é "Mc Gorila", e por isso a mensagem era capturada pelo *crawler*. Após averiguação do perfil do usuário 3, concluiu-se que se tratava de um *Twitter Bot*⁵. Notamos que a periodização de publicação do usuário 3 era exatamente de 10 minutos, e o número de *tweets* (52.800) era muito exorbitante tendo em vista que o usuário3 participa do Twitter a pouco mais de um ano, o que permitiu comprovar que se tratava de um software, de acordo com os critérios definidos por [Chu et al. 2012].

Como verificado, a detecção automática de mensagens com traços de racismo em redes sociais é uma tarefa árdua, por conta da ambiguidade e subjetividade que existe no Processamento de Linguagem Natural.

⁵Twitter Bot é um tipo de software que controla uma conta do Twitter através da API do Twitter. O software realiza automaticamente a maioria das funções que um usuário humano pode fazer.

4.4. Tweets x Termo

Analisando os dados da coleta é possível descobrir padrões nas mensagens postadas pelos usuários. Baseado na frequência de *tweets* que contém certo termo, é possível entender o que os usuários mais escrevem quando estão incitando racismo na internet. Na imagem a seguir é mostrado o número de *tweets* que contém cada palavra dos termos de busca definidos para esse artigo.

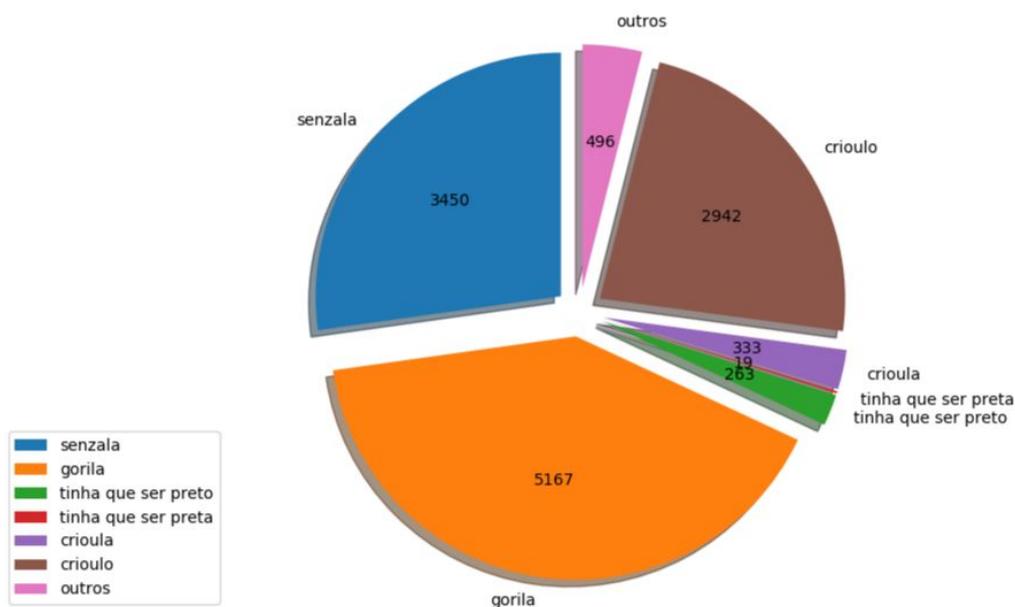


Figura 6. Quantidade de *tweets* por termo de busca

Na Figura6 é possível notar que os termos mais utilizados por usuários com tendências racistas na amostra coletada são "gorila" e "senzala". Apesar de nem sempre serem usados de forma preconceituosa, a análise dos dados permitiu constatar que a maioria dos usos dessa palavra é para denegrir e depreciar pessoas negras.

4.5. Tweets x Data

Durante a coleta dos dados, observou-se que em determinados dias o número de mensagens coletadas era muito maior do que em outros. Após análise percebeu-se que o número de *tweets* aumentava abundantemente quando algum episódio de racismo acontecia e era publicado na internet. Os usuários compartilhavam, comentavam e debatiam sobre a notícia, o que muitas vezes ocasionava em mais comentários racistas, e isso fazia com que o número de *tweets* naquele dia aumentasse. A seguir (Figura7), é apresentado uma relação entre a quantidade de *tweets* capturados, e suas respectivas datas de publicação.

No gráfico (Figura7) nota-se picos de publicação nos dias 18 e 21 de junho, e 1, 3 e 6 de julho, e é possível relacionar esse valores com alguns fatos que ocorreram no dia.

No dia 18 de junho de 2018, o candidato a presidência da república, Ciro Gomes, foi acusado de racismo após chamar Fernando Holiday, um político brasileiro negro, de "capitãozinho do mato", que era um termo usado na época da escravatura para descrever o serviçal de uma fazenda que era responsável por capturar escravos fugitivos.

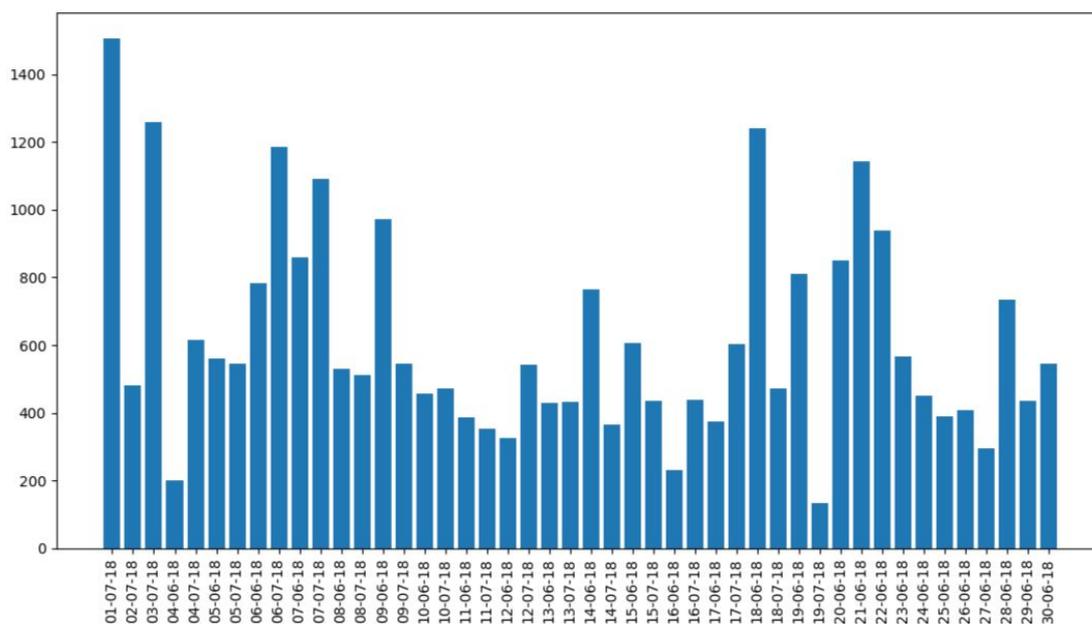


Figura 7. Quantidade de tweets por dia

No dia 21/06/2018 um movimento negro organizou postagens no Facebook a fim de mostrar depoimentos de racismo cotidiano, institucional, violência policial e assédio na cidade de Fortaleza através da hashtag #NaFortalezaRacista. Apesar do veículo escolhido pelos idealizadores do movimento ser o Facebook, muito era comentado e compartilhado no Twitter.

O *youtuber* Cocielo publicou em sua conta de Twitter, no dia 01/07/2018, um comentário racista durante a Copa do Mundo FIFA de 2018. O influenciador associou a velocidade do atacante francês, Kylian Mbappé, a uma habilidade necessária para "fazer arrastão", e gerou uma repercussão enorme nas redes sociais. Após isso, os usuários encontraram várias publicações anteriores em que Cocielo comentava sobre negros, mulheres e minorias. Outro fator que elevou o número de postagens nesse dia também é relacionado a Copa do Mundo : a presidente da Comissão Parlamentar para Família, Mulheres e Crianças do Congresso russo, Tamara Pletnyova, orientou que mulheres evitassem relações sexuais com estrangeiros "de outras raças".

No dia 03/07/2018, Itaú e Submarino, que mantinham vínculo publicitário com o *youtuber* Cocielo, emitiram notas de repúdio em suas páginas de redes sociais informando que eram veemente contra qualquer tipo de manifestação de racismo e preconceito e que haviam encerrado a parceria com o influenciador.

No dia 06/07/2018, após a eliminação do Brasil na Copa do Mundo FIFA de 2018, com a derrota de 2 a 1 para Bélgica, que incluiu um gol contra do volante brasileiro Fernandinho, o jogador foi alvo de racismo na internet. O atleta foi vítima de ofensas racistas através de comentários na sua conta do Twitter e no Instagram.

5. Próximas etapas do trabalho

Em trabalhos futuros pretende-se aplicar o pré-processamento dos dados, filtrando os *tweets* que foram coletados devido ao nome de usuário, e outras técnicas para normalizar os

dados para o uso no classificador, assim como organizar a aplicação conjunta de técnicas de Análise de Sentimento e Aprendizado de Máquina para classificar mensagens com traços racistas em Redes Sociais Online em língua portuguesa.

Referências

- Aslam, S. (2018). *Twitter by the Numbers: Stats, Demographics & Fun Facts*.
- Benevenuto, F. (2015). *Métodos para Análise de Sentimentos em mídias sociais*.
- Bruce, R. (2001). *A Bayesian Approach to Semi-Supervised Learning*.
- Chu, Z., Gianvecchio, S., Wang, H., and Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing*.
- Decker, K. M. and Focardi, S. (1995). *Technology Overview: A Report on Data Mining*.
- Greevy, E. and Smeaton, A. F. (2004). Classifying racist texts using a support vector machine. In *Proceedings of the 27th annual international conference on Research and development in information retrieval - SIGIR '04*.
- Hirst, G. and Liu, B. (2012). *SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES Sentiment Analysis and Opinion Mining Sentiment Analysis and Opinion Mining*.
- ISTOÉ Independente (2015). *O criminoso da internet*.
- José De Alencar, R. (2012). ESTUDO DA OCORRÊNCIA DE CYBERBULLYING CONTRA PROFESSORES NA REDE SOCIAL TWITTER POR MEIO DE UM ALGORITMO DE CLASSIFICAÇÃO BAYESIANO. pages 5–1.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*.
- Martins, I. C. (2014). *O racismo nas redes sociais : O mundo virtual é feito por pessoas de carne e osso!*
- Pang, B. and Lee, L. (2006). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Informatio*Pang, B., & Le*Foundations and Trends® in Information Retrieval, 1(2), 91–231., 1(2):91–231.*
- Vicente, B., De Lima, A., Machado, V. P., De Melo, R., and Veras, S. *Abordagem Semi-supervisionada para Rotula ao de Dados*.
- Waseem, Z. and Hovy, D. (2016). *Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter.* pages 88–93.

Extração de Regras de Associação de Dados Criminais no Município de Goiânia

Sílvio Passos Severino¹, Nádia Félix F. da Silva¹

¹Instituto de Informática – Universidade Federal de Goiás – Goiânia, GO – Brasil

silviopass@gmail.com, nadia@inf.ufg.br

Abstract. *The present work aims to extract associations in data of criminal occurrences in the city of Goiânia, identifying the neighborhoods where there is a greater concentration of crimes, considering the spatial and temporal distribution, as well as the socioeconomic profile of the place of occurrence. The developed process uses all phases of knowledge discovery - KDD (Knowledge Discovery in Databases), including the selection of attributes, cleaning, standardization, pre-processing and data transformation. The data set used in this study refers to the record of criminal occurrences provided by the Information Analysis Management of the Public Security Secretariat of the State of Goiás. The result obtained through the use of association rules gathers important information about criminal occurrences, such as the identification of the most frequent crimes in a specific place, in a time period, with a specific victim profile. The results are relevant to help decision making and also to inform and alert the population about the places with the most criminal occurrences at certain times in the municipality of Goiânia.*

Resumo. *O presente trabalho tem como objetivo extrair associações em dados de ocorrências criminais no município de Goiânia, identificando os bairros onde há maior concentração de crimes considerando a distribuição espacial e temporal, bem como o perfil socioeconômico do local da ocorrência. O processo desenvolvido utiliza-se de todas as fases da descoberta de conhecimento – KDD (Knowledge Discovery in Databases) compreendendo a seleção de atributos, limpeza, padronização, pré-processamento e transformação dos dados. O conjunto de dados utilizado no presente estudo refere-se ao registro de ocorrências criminais fornecido pela Gerência de Análise de Informações da Secretaria de Segurança Pública do Estado de Goiás. O resultado obtido através do uso de regras de associação reúne informações importantes sobre as ocorrências criminais como, por exemplo, a identificação dos crimes mais frequentes em um determinado local, em um período de horário, com um perfil de vítima específico. Os resultados são relevantes para o auxílio à tomada de decisão, bem como informar e alertar a população sobre os locais com mais ocorrências criminais em determinados horários no município de Goiânia.*

1. Introdução

Não são poucas as notícias e imagens, expondo o sério problema da violência no Brasil, especialmente nas capitais. A alta criminalidade no país tem gerado custos altíssimos. De acordo com o Anuário de Segurança Pública, só em 2016, o governo brasileiro gastou

na área de segurança pública mais de 67 bilhões de reais. No estado de Goiás a situação não é diferente, em 2016 foram gastos mais de 3 bilhões de reais em segurança pública [Fórum 2017].

Devido ao elevado número de crimes, em 2015 Goiânia chegou a ocupar o 13º lugar no ranking das capitais mais violentas do Brasil [Fórum 2017]. Buscando reduzir os índices de crimes, o governo de Goiás investiu no ano 2015, somente em inteligência de informações, o valor de 1,5 milhões reais. Um dos investimentos foi a implantação da Plataforma de Sistemas Integrados (PSI). Lançada em abril de 2016, a PSI é composta pelos programas de Registro de Atendimento Integrado (RAI), Sistema Geográfico de Informação (GisGestão), Mapeamento de Operações Policiais Integradas (MOPI), Mapeamento de Ações Sociais Integradas (MASI) e o Aplicativo de Integração entre Polícia e Cidadão (I9X). Segundo a Secretaria, o RAI foi desenvolvido para que as instituições que compõem o Centro Integrado de Inteligência, Comando e Controle (CIICC) possam ter base de informações sobre os registros das ocorrências policiais integradas e em tempo real.

É importante ressaltar que em análise de dados criminais, fatores relacionados ao espaço geográfico do crime devem ser considerados como, por exemplo, demografia, perfil socioeconômico, entre outros. Abordando sobre o relacionamento entre crime e níveis socioeconômicos [Villarreal and Silva 2006] foi encontrada uma associação entre o nível social e as taxas de criminalidade no Brasil. De um modo geral a violência atinge, sem distinção, todos os segmentos sociais. No entanto, é relevante considerar fatores socioeconômicos para identificar padrões de crimes no município de Goiânia e o conhecimento adquirido a partir de bases de dados criminais unificadas a informações socioeconômicas pode auxiliar na análise sobre os crimes em determinados bairros.

2. Trabalhos Relacionados

Na literatura existem diversos estudos sobre análise de dados criminais, bem como técnicas com a finalidade de auxiliar os gestores da área na tomada de decisão ou para outra finalidade, a seguir são apresentados alguns estudos sobre o tema:

- [Mande et al. 2012] apresenta uma metodologia que usa Modelo Gaussiano de Mistura Generalizada para mapear informações especificadas pela testemunha ocular utilizando o agrupamento k-means com base no tipo de crime.
- No trabalho [Nath 2006] é utilizado o agrupamento K-means para encontrar alguns padrões criminais, implementando uma estrutura que funciona com a informação geoespacial, o tipo de crime, o perfil da vítima e o perfil do criminoso.
- [Keyvanpour et al. 2011] trata da relação dos crimes entre si e os criminosos através de uma abordagem sistemática do uso de Redes Neurais Auto-Organizáveis - SOM e Multicamadas Perceptron - MLP para agrupar e classificar dados do crime.
- [Wang et al. 2013] utiliza o algoritmo representado como GDPatterns para encontrar a informação espacial e o relacionamento dos crimes com outros fatores, apresentando uma ferramenta de otimização de hotspots que considera outros crimes ocorridos nas proximidades.

- Já [Vijayakumar et al. 2014] apresenta os resultados da predição e de padrões criminais focando em como incorporar o fator tempo na análise de dados utilizando o modelo STEM “Modelo de espaço-tempo-evento”.
- O trabalho proposto por [Almanie et al. 2015] apresenta dois modelos classificadores e regras de associação para analisar dados criminais considerando a localização, tempo e tipo de crime.

3. Fundamentação Teórica

A tecnologia da informação possui vários recursos que pode auxiliar a análise de dados e a compreender padrões e tendências dos crimes ocorridos a partir de um conjunto de dados. Dentre estes se destaca o processo de Descoberta de Conhecimento em Bases de Dados (KDD, do inglês *Knowledge Discovery in Databases*). Uma definição clássica para o KDD, segundo [Fayyad et al. 1996], *é o processo não trivial de identificar informações válidas, novas, potencialmente úteis e padrões compreensíveis nos dados*, e contém uma série de etapas a saber: seleção, pré-processamento, transformação, mineração de dados e interpretação/avaliação dos dados.

3.1. Mineração de Dados

A mineração de dados é uma etapa do KDD, compreende os principais algoritmos que permitem obter conhecimento em grandes conjuntos de dados [Zaki and Jr 2014] e tem como finalidade encontrar anomalias, padrões e correlações para depois os apresentarem na forma de representação e visualização [Silva 2004]. Em mineração de dados existem diferentes técnicas e tarefas para diferentes propósitos, cada uma com vantagens e desvantagens. A escolha da técnica está relacionada com o tipo de conhecimento que se deseja extrair ou com o tipo de dado no qual ela será aplicada.

A Classificação é a tarefa de mineração de dados que associa ou classifica objetos a determinadas classes buscando prever um rótulo de classe para um determinado objeto não rotulado [Zaki and Jr 2014].

A Clusterização busca reunir instâncias com características comuns em grupos, que posteriormente podem ser classificados. Exemplos de tarefas de agrupamento são: identificar grupos de clientes para direcionamento de campanhas, identificar fraude ou até mesmo classificar instâncias, quando não houver classe conhecida. [Fayyad et al. 1996] define Clusterização como uma tarefa descritiva comum em que se procura identificar um conjunto finito de categorias ou *clusters* para descrever os dados.

A tarefa de Associação busca encontrar relacionamentos significativos entre os itens de dados armazenados através de regras de associação, identificando associações entre atributos e apresentando padrões frequentes em um conjunto de dados [Kantardzic 2011]. Neste trabalho a técnica utilizada foi a Associação.

3.2. Regras de Associação

Uma Regra de Associação dá-se por meio da utilização de métodos probabilísticos sobre a ocorrência simultânea de determinados eventos em uma base de dados [Zaki and Jr 2014]. Por exemplo, para a regra R assumem-se as seguintes variáveis como binárias R : $(SE A=1 \wedge B=1 \wedge C=1)$ com probabilidade P , Assim, tem-se:

$$P = P(C = I \mid A = I, B = I) \quad (1)$$

Uma regra de associação é composta de dois conjuntos de itens, um antecedente e um conseqüente e são representadas na forma: Antecedente (A) \rightarrow Conseqüente (B), interpretada da seguinte forma: Se A então B, e ambos formam um conjunto de itens (*itemsets*) [Zaki and Jr 2014]. Para um item ser considerado frequente deve satisfazer alguma condição previamente definida, para tanto é necessário definir medidas, entre as mais usadas estão o suporte e a confiança.

O suporte representa o número de transações incluindo todos os itens na posição de antecedente e conseqüente da regra. Assim, para uma a regra de associação $\{A\} \rightarrow \{B\}$ o suporte mede o número total de registros de transação que contêm os conjuntos de itens A e B. Neste caso o suporte da regra $\{A\} \rightarrow \{B\}$, em que A e B são conjuntos de itens é dado pela seguinte expressão:

$$\text{Suporte}(A \rightarrow B) = (\text{Frequência de } A \text{ e } B) / (\text{Total de } T) \quad (2)$$

A confiança de uma regra é a probabilidade condicional que uma transação contém A, dado que contém B é representada pela seguinte expressão:

$$\text{Confiança}(A \rightarrow B) = \text{suporte}(A \cup B) / \text{suporte}(A) \quad (3)$$

Existem muitos algoritmos que utilizam regras de associação e entre os mais populares estão o Apriori e o FPGrowth [Amaral 2016].

3.2.1. Algoritmo FPGrowth

O algoritmo FPGrowth leva uma abordagem diferente do Apriori que usa o paradigma “gerar e testar”, a abordagem do FPGrowth consiste no desenvolvimento de uma estratégia baseada na técnica de dividir para conquistar, na qual o problema é fracionado em subproblemas [Han et al. 2000]. O FPGrowth codifica o conjunto de dados através de uma estrutura compacta chamada Frequent Pattern tree (FP-tree) e extrai os conjuntos de itens frequentes diretamente desta estrutura. Isso possibilita um melhor desempenho na geração das regras de associação, pois o número de varreduras na base de dados é menor [Tan et al. 2005]. Neste presente trabalho o algoritmo utilizado para descoberta das regras de associação é o FPGrowth.

4. Experimentos

Nesta seção são descritos o conjunto de dados e a metodologia do processo utilizado.

4.1. Conjunto de Dados de Ocorrências Criminais

Fornecido pela Gerência do Observatório de Segurança Pública do Estado de Goiás – GEOSP, o conjunto de dados compreende o registro de ocorrências no Estado de Goiás obtidas através da integração da Polícia Militar, Polícia Civil, Corpo de Bombeiros Militar e da Superintendência de Polícia Técnico-Científica (SPTC). O conjunto é composto por 16 atributos, contém 1.785.488 instâncias e os principais atributos fornecem a natureza do crime, data, hora, localização geográfica e informações do comunicante e/ou

envolvido como, por exemplo, sexo e data de nascimento da vítima. O período referente às ocorrências registradas foi entre os dias 1º de abril de 2016 a 31 de março de 2017, totalizando doze meses.

4.2. Conjunto de Dados Socioeconômicos

O conjunto de dados contém a renda domiciliar dos bairros do município de Goiânia e foram extraídos do Censo Demográfico 2010 - Resultados Agregados por Setor Censitário, os dados compreendem as características dos domicílios particulares e das pessoas que foram investigadas para a totalidade da população e são denominados por convenção resultados do universo [IBGE 2017].

4.3. Pré-processamento de Dados

A seguir são descritos os passos da etapa de pré-processamento dos dados, os quais foram realizados através das ferramentas IBM SPSS Statistics [IBM 2011] e Orange Canvas [Demšar et al. 2018]

O conjunto de dados criminais foi disponibilizado em 12 arquivos no formato CSV. Sendo assim, o primeiro passo foi fazer a integração dos arquivos para facilitar a manipulação dos dados.

Após a integração dos arquivos, verificou-se a existência de valores ausentes nos atributos 'profissão', 'escolaridade', 'orientação_sexual', 'cor_raça'. Como os índices eram altos, todos estes atributos foram eliminados da base.

Em seguida foram eliminadas as instâncias que não eram do município de Goiânia, em sequência foram eliminadas todas as instâncias que não se caracterizavam como uma ocorrência criminal. Depois de observado que uma ocorrência gerava várias instâncias, ou seja, uma instância para cada envolvido na ocorrência, considerando a abordagem do trabalho que foi com foco na qualificação da vítima foram eliminadas da base todas as instâncias em que o atributo 'qualificação' era diferente de 'vítima' ou 'vítima comunicante'. O passo seguinte foi padronizar o atributo 'natureza', pois em algumas ocorrências havia mais de uma 'natureza'. Os dados foram reestruturados novamente, criando uma instância para cada crime.

Com o objetivo de delimitar o escopo do estudo e levando em consideração os objetivos específicos do trabalho, os crimes foram delimitados nas seguintes categorias: crimes contra a dignidade sexual, crimes contra a pessoa, crimes contra o patrimônio e crimes contra o estatuto da criança e do adolescente. Assim, foi necessário identificar as instâncias que não se enquadravam na categoria desses crimes e excluí-las da base de dados, classificando-as segundo a Tabela de Naturezas - RAI¹

Terminado o processo de limpeza dos dados, o próximo passo foi obter a idade da vítima, através da subtração do atributo 'data_ocorrencia' e 'data_nascimento'.

Com o intuito de gerar análises por dia do mês, o atributo 'data' foi transformado em três atributos, 'data_ano', 'data_mes' e 'data_dia'.

Ao final deste processo o conjunto de dados contava com 10 atributos: data_dia, data_mes, dia_semana, hora, bairro, natureza, logradouro, qualificacao, sexo e idade.

¹<http://observatorio.ssp.go.gov.br/>

Foi observada a necessidade de reduzir a diversidade dos valores de alguns atributos. Assim, foi aplicada a transformação de dados aos atributos, mapeando seus valores para que se enquadrem em grupos menores. O objetivo foi obter padrões mais frequentes e aumentar a precisão do modelo. Para o atributo ‘natureza’, foi minimizada a lista de crimes criando uma categoria denominada ‘outros crimes’, agrupando os crimes com frequência menor que 1%. Ao final, a lista de natureza de crimes ficou com 20 categorias. Para o atributo ‘hora’ do crime, foram mapeados os valores em intervalos de 4 horas.

Finalmente, utilizando o atributo ‘bairro’ presente nos dois conjuntos (criminal e socioeconômico) as bases foram unificadas, gerando assim um único conjunto de dados. A figura 1 apresenta o conjunto de dados após o pré-processamento.

	Atributo	Tipo	Categorias
1	id	numérico	...
2	data_dia	numérico	...
3	data_mes	numérico	...
4	dia_semana	nominal	Domingo; Segunda-feira; Terça-feira; Quarta-feira; Quinta-feira, Sexta-feira; Sábado
5	turno	nominal	T1: 01:00 as 04:59; T2: 05:00 as 08:59; T3: 09:00 as 12:59; T4: 13:00 as 16:59; T5: 17:00 as 20:59; T6: 21:00 as 00:59
6	bairro	nominal	(64 categorias)
8	crime	nominal	(20 categorias)
9	sexo	nominal	Masculino; Feminino
10	faixa_etaria	nominal	Criança (<11); Adolescente (11-17); 2 Jovem (18-30); 3 Adulto (31-59); 4 Idoso (>59)
11	renda_domicilio	numérico	...

Figura 1. Conjunto de dados após pré-processamento

4.4. Obtenção de Regras de Associação

Para encontrar relações entre atributos — por exemplo, data, local, gênero — o foco foi aplicar algoritmos de obtenção de regras de associação e avaliar se tais regras obtidas podem auxiliar na tomada de decisões e em estratégias de políticas de segurança pública. Desta forma, a abordagem proposta está focada nos três principais elementos sobre os crimes, que são a natureza do crime, o tempo de ocorrência e a localização do crime. Na tentativa de extrair os possíveis padrões frequentes interessantes baseados nas variáveis do crime, foi aplicada regras de associação.

O Algoritmo FPGrowth é um dos algoritmos mais usados para encontrar padrões frequentes de mineração (3.2.1) e sua estratégia é baseada na técnica de dividir para conquistar, obtendo um melhor desempenho na geração das regras de associação, pois o número de varreduras na base de dados é menor que a do algoritmo APriori. Para FPGrowth encontrar padrões frequentes são necessários dois parâmetros de entrada, o suporte mínimo e a confiança. A implementação deste modelo foi realizada com uma ferramenta de código aberto, Orange [Demšar et al. 2018].

O objetivo inicial foi encontrar todos os possíveis padrões frequentes, independentemente da natureza do crime, criando uma lista de todos os locais com maior ocorrência criminal, juntamente com o dia e o horário mais frequente relacionado.

O algoritmo foi implementado utilizando somente os atributos relacionados à localização e tempo de ocorrência. Além disso, para obter padrões mais frequentes, foi

aplicada a mineração baseada em restrição, restringindo o processo de extração com essa fórmula de três conjuntos de itens específicos (bairro → dia → turno).

O objetivo foi encontrar aproximadamente 50 padrões frequentes. Assim, foram realizados vários experimentos usando diferentes valores mínimos de suporte e de confiança. Finalmente, define-se o valor de suporte mínimo em 0.01 e confiança 0.04, que representou 13.216 instâncias, 12% da base de dados. Com estas definições foram identificados 62 padrões, contemplando 14 bairros.

O segundo objetivo foi identificar que natureza de crime pode estar associada a um local específico, dentro de um tempo particular, e um perfil específico de vítima. Assim, foram realizados outros experimentos, inserindo os atributos anteriormente desconsiderados.

5. Resultados

Nesta seção é apresentado o resumo dos principais resultados obtidos com a aplicação das Regras de Associação utilizando o Algoritmo FPGrowth.

5.1. Análise espacial e temporal

Visando encontrar os pontos criminais espaciais e temporais foram extraídos todos os padrões interessantes com base nos limites predefinidos. A figura 2 mostra os resultados da aplicação das regras de associação, revelando 62 padrões, que contemplaram 13.216 instâncias. É importante mencionar que neste primeiro momento o objetivo foi encontrar associações entre o bairro, dia da semana e horário da ocorrência. Com esses conjuntos de itens frequentes são revelados os locais de crimes juntamente com seu dia e horário de ocorrência associados.

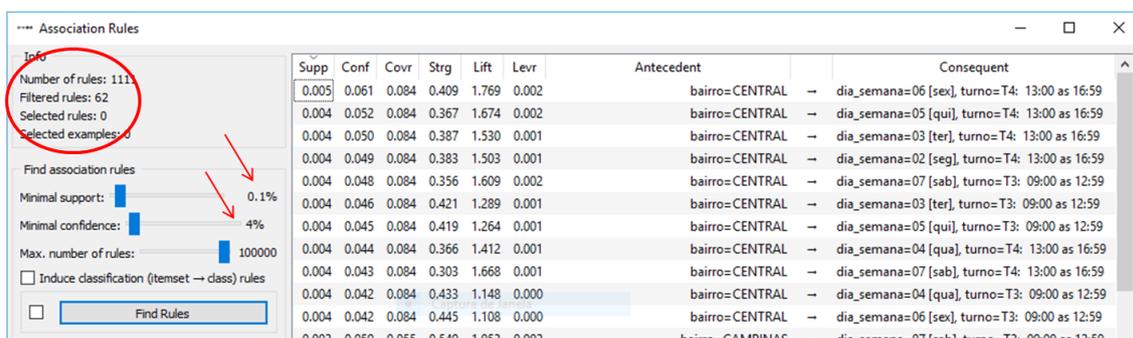


Figura 2. Padrões encontrados

Na sequência são apresentadas as tabelas com os resultados extraídos, onde são listadas a regra (associação de itens), o suporte e a confiança de cada associação. No rodapé de cada tabela são apresentados os parâmetros definidos para extração das referidas regras.

A tabela 1 apresenta os padrões frequentes, suporte mínimo e confiança das regras extraídas entre a associação dos atributos — bairro, dia da semana e turno — indicando que 14 bairros são locais que têm padrões frequentes de crimes. O bairro Central e Campinas são os que têm maior número de padrões. Além disso, nota-se que sexta-feira é o dia da semana com o maior número de padrões. Observa-se também que a maioria

A seguir, são apresentados os resultados obtidos com a indução para associar as características da ocorrência à natureza do crime. Na identificação dos padrões foi feita a filtragem das regras informado o número mínimo de itens do antecedente (lado esquerdo da regra). Sendo assim, inicialmente foi deixado como 1 (um). Com essa configuração é mostrado na tabela 2 que o padrão com o maior suporte, ou seja, ocorre com maior frequência é SE (SEXO = FEMININO ENTÃO CRIME = ROUBO A TRANSEUNTE). Percebe-se que a associação entre ‘FEMININO’ e ‘ROUBO A TRANSEUNTE’ é maior que entre ‘MASCULINO’ e ‘ROUBO A TRANSEUNTE’ (linha 3), indicando que o crime de roubo a transeunte é mais frequente com a vítima do sexo feminino.

Regra	Antecedent	Consequent	Supp	Conf
1	FEMININO	→ crime=ROUBO A TRANSEUNTE	0.106	0.215
2	Jovem (18-30)	→ crime=ROUBO A TRANSEUNTE	0.098	0.275
3	MASCULINO	→ crime=ROUBO A TRANSEUNTE	0.097	0.191
4	Adulto (31-59)	→ crime=ROUBO A TRANSEUNTE	0.081	0.160
5	17h00min às 20h59min	→ crime=ROUBO A TRANSEUNTE	0.056	0.241
6	MASCULINO, Jovem (18-30)	→ crime=ROUBO A TRANSEUNTE	0.051	0.279
7	FEMININO	→ crime=AMEACA	0.050	0.101
8	FEMININO, Jovem (18-30)	→ crime=ROUBO A TRANSEUNTE	0.047	0.271
9	FEMININO, Adulto (31-59)	→ crime=ROUBO A TRANSEUNTE	0.047	0.187
10	MASCULINO	→ crime=OUTROS FURTOS	0.046	0.090
11	05h00min às 08h59min	→ crime=ROUBO A TRANSEUNTE	0.045	0.344
12	Adulto (31-59)	→ crime=OUTROS FURTOS	0.045	0.088
13	Adulto (31-59)	→ crime=AMEACA	0.043	0.085
14	MASCULINO	→ crime=FURTO DE DOCUMENTOS	0.042	0.084
15	MASCULINO	→ crime=ROUBO DE VEICULO	0.041	0.081
16	Adulto (31-59)	→ crime=ROUBO DE VEICULO	0.040	0.080

Settings (Supp mín: 4.000; Conf: 1.000; Antecedent min items: 1; Antecedent max items: 99)

Tabela 2. Regras geradas com a indução da classe (1 item no antecedente)

Na sequência, foram filtradas as regras com no mínimo “dois itens no antecedente”, ou seja, serão apresentadas somente associações com no mínimo de dois atributos (tabela 3). Os padrões com maiores suportes são: ‘MASCULINO’, ‘JOVEM’ → ‘ROUBO A TRANSEUNTE’; ‘FEMININO’, ‘jovem’ → ‘ROUBO A TRANSEUNTE’; ‘FEMININO’, ‘ADULTO’ → ‘ROUBO A TRANSEUNTE’, indicando que a associação entre vítima ”jovem” do sexo masculino e roubo a transeunte é maior que, vítima ”jovem” do sexo feminino.

Regra	Antecedent	Consequent	Supp	Conf
1	MASCULINO, Jovem (18-30)	→ crime=ROUBO A TRANSEUNTE	0.051	0.279
2	FEMININO, Jovem (18-30)	→ crime=ROUBO A TRANSEUNTE	0.047	0.271
3	FEMININO, Adulto (31-59)	→ crime=ROUBO A TRANSEUNTE	0.047	0.187
4	MASCULINO, Adulto (31-59)	→ crime=ROUBO A TRANSEUNTE	0.034	0.133
5	17h00min às 20h59min, FEMININO	→ crime=ROUBO A TRANSEUNTE	0.030	0.253
6	17h00min às 20h59min, Jovem (18-30)	→ crime=ROUBO A TRANSEUNTE	0.028	0.307
7	FEMININO, Adulto (31-59)	→ crime=AMEACA	0.028	0.110
8	05h00min às 08h59min, FEMININO	→ crime=ROUBO A TRANSEUNTE	0.026	0.407
9	17h00min às 20h59min, MASCULINO	→ crime=ROUBO A TRANSEUNTE	0.026	0.228
10	MASCULINO, Adulto (31-59)	→ crime=ROUBO DE VEICULO	0.025	0.100
11	MASCULINO, Adulto (31-59)	→ crime=OUTROS FURTOS	0.025	0.097
12	21h00min às 00h59min, MASCULINO	→ crime=ROUBO A TRANSEUNTE	0.024	0.268
13	21h00min às 00h59min, Jovem (18-30)	→ crime=ROUBO A TRANSEUNTE	0.023	0.329
14	17h00min às 20h59min, Adulto (31-59)	→ crime=ROUBO A TRANSEUNTE	0.022	0.189
15	05h00min às 08h59min, Adulto (31-59)	→ crime=ROUBO A TRANSEUNTE	0.021	0.298

Settings (Supp mín: 1.000; Conf: 4.000; Antecedent min items: 2; Antecedent max items: 99)

Tabela 3. Regras geradas com a indução da classe (2 itens no antecedente)

Deste modo, o número mínimo de itens do antecedente foi alterado progressivamente, até a quantidade de cinco (máximo de atributos considerados na análise, exceto a classe). Com cinco atributos (mais a classe), o padrão com maior suporte é bairro ‘CENTRAL’, ‘sexta-feira’, ‘13h00min às 16h59min’, ‘FEM’, ‘Adulto (31-59)’ → crime=‘FURTO A USUÁRIOS DE TRANSPORTE COLETIVO’, essa associação equi-

vale a 67 ocorrências criminais da base de dados, representando o suporte de 0.001 (tabela 4).

Regra	Antecedent	Consequent	Supp	Conf
1	CENTRAL, sexta-feira, 13h00min às 16h59min, FEM, Adulto (31-59)	→ crime=FURTO A USUARIOS DE TRANSP	0.001	0.435
2	CENTRAL, quinta-feira, 13h00min às 16h59min, FEM, Adulto (31-59)	→ crime=FURTO A USUARIOS DE TRANSP	0.001	0.504
3	CENTRAL, segunda-feira, 13h00min às 16h59min, FEM, Adulto (31-59)	→ crime=FURTO A USUARIOS DE TRANSP	0.001	0.420
4	CENTRAL, terça-feira, 13h00min às 16h59min, FEM, Adulto (31-59)	→ crime=FURTO A USUARIOS DE TRANSP	0.001	0.413
5	CENTRAL, sexta-feira, 13h00min às 16h59min, FEM, 4 Idoso (60..)	→ crime=FURTO A USUARIOS DE TRANSP	0.001	0.505

Settings (Supp min: 0.005; Conf: 1.000; Antecedent min items: 5; Antecedent max items: 99)

Tabela 4. Regras geradas com a indução da classe (5 itens no antecedente)

5.3. Análise integrada: ocorrências e perfil socioeconômico do local

A geração de regras de associação envolvendo a informação socioeconômica foi limitada aos seguintes atributos: dia da semana, horário, renda do bairro (renda média por domicílio em salários mínimos). As linhas 1 a 3 da tabela 5 revela a associação entre o crime ‘roubo a transeunte’, o horário ‘17h00min às 20h59min’ e os bairros com renda domiciliar ‘acima de 3,5 salários mínimos’ é mais frequente. Porém, observa-se que na linha 4, o mesmo crime (roubo a transeunte) é mais frequente nos bairros com renda ‘abaixo de 3,5’ no horário de ‘05h00min às 08h59min’. Deste modo, nota-se que é possível existir a relação entre um tipo específico de crime, um perfil específico de bairro, em um espaço de tempo específico.

Regra	Antecedent	Consequent	Supp	Conf
1	17h00min às 20h59min, renda domicilio 5.5 - 8.5	→ crime=ROUBO A TRANSEUNTE	0.013	0.238
2	17h00min às 20h59min, renda domicilio ≥ 8.5	→ crime=ROUBO A TRANSEUNTE	0.012	0.198
3	17h00min às 20h59min, renda domicilio 3.5 - 4.5	→ crime=ROUBO A TRANSEUNTE	0.012	0.269
4	05h00min às 08h59min, renda domicilio < 3.5	→ crime=ROUBO A TRANSEUNTE	0.011	0.490
5	13h00min às 16h59min, renda domicilio ≥ 8.5	→ crime=FURTO A USUARIOS DE TRAN	0.011	0.169
6	05h00min às 08h59min, renda domicilio 3.5 - 4.5	→ crime=ROUBO A TRANSEUNTE	0.011	0.392
7	09h00min às 12h59min, renda domicilio ≥ 8.5	→ crime=ESTELIONATO	0.011	0.147
8	05h00min às 08h59min, renda domicilio 5.5 - 8.5	→ crime=ROUBO A TRANSEUNTE	0.010	0.319
9	09h00min às 12h59min, renda domicilio ≥ 8.5	→ crime=OUTROS FURTOS	0.010	0.135
10	17h00min às 20h59min, renda domicilio 4.5 - 5.5	→ crime=ROUBO A TRANSEUNTE	0.010	0.244
11	21h00min às 00h59min, renda domicilio ≥ 8.5	→ crime=ROUBO A TRANSEUNTE	0.010	0.237
12	17h00min às 20h59min, renda domicilio < 3.5	→ crime=ROUBO A TRANSEUNTE	0.010	0.279

Settings (Supp min: 0.010; Conf: 1.000; Antecedent min items: 3; Antecedent max items: 99)

Tabela 5. Regras geradas com a inclusão da renda do bairro da ocorrência

Regra	Antecedent	Consequent	Supp	Conf
1	domingo, 17h00min às 20h59min, renda domicilio 5.5 - 8.5	→ crime=AMEACA	0.011	1.000
2	terça-feira, 09h00min às 12h59min, renda domicilio ≥ 8.5	→ crime=AMEACA	0.011	1.000
3	domingo, 17h00min às 20h59min, renda domicilio < 3.5	→ crime=AMEACA	0.010	1.000
4	segunda-feira, 17h00min às 20h59min, renda domicilio 5.5 - 8.5	→ crime=AMEACA	0.010	1.000
5	domingo, 17h00min às 20h59min, renda domicilio 3.5 - 4.5	→ crime=AMEACA	0.010	1.000
6	segunda-feira, 09h00min às 12h59min, renda domicilio 5.5 - 8.5	→ crime=AMEACA	0.010	1.000
7	terça-feira, 09h00min às 12h59min, renda domicilio 5.5 - 8.5	→ crime=AMEACA	0.010	1.000
8	quarta-feira, 09h00min às 12h59min, renda domicilio ≥ 8.5	→ crime=AMEACA	0.010	1.000
9	quinta-feira, 09h00min às 12h59min, renda domicilio ≥ 8.5	→ crime=AMEACA	0.010	1.000

Settings (Supp min: 0.010; Conf: 1.000; Antecedent min items: 3; Antecedent max items: 99)

Tabela 6. Regras geradas somente para o crime ameaça

Em um segundo momento foi testado a extração de regras de associação isolando dois crimes e tratando-os de formas distintas. Primeiro foi testado o crime de ‘ameaça’. Desta forma, a base de dados só conteve as instâncias deste crime. Igualmente, o mesmo procedimento foi feito para o crime ‘roubo a residência’. A tabela 6 revela que o crime ‘ameaça’ é mais frequente em bairros com domicílios de renda média acima de 3,5 salários mínimos. Uma associação interessante na tabela 7 é que o ‘roubo em residência’

nos bairros (com renda inferior a 3,5 salários mínimos) tem um mais frequência a partir das 21h00min, enquanto que nos bairros com renda superiores o padrão mais frequente é no período da tarde ou pela manhã.

Regra	Antecedent	Consequent	Supp	Conf
1	quarta-feira,21h00min às 00h59min, renda_domicilio < 3.5	→ crime=ROUBO EM RESIDENCIA	0.020	1.000
2	quinta-feira,05h00min às 08h59min, renda_domicilio 5.5 - 8.5	→ crime=ROUBO EM RESIDENCIA	0.019	1.000
3	sábado,21h00min às 00h59min, renda_domicilio < 3.5	→ crime=ROUBO EM RESIDENCIA	0.017	1.000
4	sexta-feira,21h00min às 00h59min, renda_domicilio 3.5 - 4.5	→ crime=ROUBO EM RESIDENCIA	0.016	1.000
5	quarta-feira,17h00min às 20h59min, renda_domicilio 5.5 - 8.5	→ crime=ROUBO EM RESIDENCIA	0.016	1.000
6	segunda-feira,17h00min às 20h59min, renda_domicilio < 3.5	→ crime=ROUBO EM RESIDENCIA	0.015	1.000
7	quinta-feira,17h00min às 20h59min, renda_domicilio 3.5 - 4.5	→ crime=ROUBO EM RESIDENCIA	0.015	1.000
8	quarta-feira,21h00min às 00h59min, renda_domicilio 3.5 - 4.5	→ crime=ROUBO EM RESIDENCIA	0.013	1.000
9	terça-feira,17h00min às 20h59min, renda_domicilio 5.5 - 8.5	→ crime=ROUBO EM RESIDENCIA	0.013	1.000
10	sexta-feira,17h00min às 20h59min, renda_domicilio < 3.5	→ crime=ROUBO EM RESIDENCIA	0.013	1.000
11	sexta-feira,21h00min às 00h59min, renda_domicilio < 3.5	→ crime=ROUBO EM RESIDENCIA	0.013	1.000
12	terça-feira,21h00min às 00h59min, renda_domicilio 3.5 - 4.5	→ crime=ROUBO EM RESIDENCIA	0.013	1.000
13	segunda-feira,21h00min às 00h59min, renda_domicilio < 3.5	→ crime=ROUBO EM RESIDENCIA	0.013	1.000

Settings (Supp mín: 0.010; Conf: 1.000; Antecedent min items: 3; Antecedent max items: 99)

Tabela 7. Regras geradas somente para o crime de roubo a residência

6. Conclusão e Trabalhos Futuros

O trabalho apresentou a construção de um modelo completo para descoberta do conhecimento e utilizou-se de todas as fases da descoberta (da seleção à interpretação). Como fonte de estudo fez uso de uma base criminal do município de Goiânia, visando extrair padrões criminais no que diz respeito à vítima, espaço local e temporal do crime.

Através do estudo foi possível identificar padrões de ocorrências criminais, como por exemplo, os locais onde há mais ocorrências em determinados horários, relacionados a um perfil específico de vítima. Portanto, conclui-se que a pesquisa foi satisfatória alcançando os objetivos propostos.

Em relação à análise dos dados socioeconômicos, de um modo geral não foi um fator identificado como “forte” influência para o “alto” índice de crimes. No entanto, fica a ressalva: a análise de cada natureza criminal deverá ser analisada de forma independente e outros elementos característicos do local do crime deverão ser inseridos.

Finalmente, diante do estudo realizado acerca de um tema tão relevante, que de certa forma todos estão inseridos, espera-se que o conteúdo aqui apresentado, de certa forma possa contribuir na descoberta de conhecimento sobre o referido tema e na decisão de adotar a utilização de técnicas de Mineração de Dados como ferramenta de apoio a tomada de decisão.

Visando a continuidade deste trabalho, no intuito de alcançar novas descobertas que possam contribuir um pouco mais sobre o tema, ficam como sugestão para trabalhos futuros as seguintes linhas de pesquisa: aplicação de outros modelos e técnicas não inseridas neste estudo, como por exemplo, técnicas de visualização de dados, modelos de classificação, entre outros; abordagem de outros aspectos não tratados neste estudo, como por exemplo, associação entre o perfil da vítima e criminoso; expansão da análise considerando o âmbito estadual.

Referências

Almanie, T., Mirza, R., and Lor, E. (2015). Crime prediction based on crime types and using spatial and temporal criminal hotspots. *CoRR*, abs/1508.02050.

- Amaral, F. (2016). *Introdução à Ciência de Dados: mineração de dados e big data*. ALTA BOOKS.
- Demšar, J., Curk, T., Erjavec, A., Črt Gorup, Hočvar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., and Zupan, B. (2018). Orange data mining. version 3.13.0. <https://orange.biolab.si/>.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *Ai Magazine*, 17:37–54.
- Fórum, B. d. P. S. (2017). Anuário brasileiro de segurança pública. <http://www.forumseguranca.org.br/11o-anuario-brasileiro-de-seguranca-publica/>. acesso em: 30 abril 2018.
- Han, Pei, and Yin (2000). Mining frequent patterns without candidate generation. *SIG-MODREC: ACM SIGMOD Record*, 29.
- IBGE, I. B. d. G. e. E. (2017). Estimativa da população. <https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9103-estimativas-de-populacao.html?t=resultados>. acesso em: 02 fev 2018.
- IBM (2011). Ibm spss statistic. version 20.0.0. <https://www.ibm.com/products/spss-statistics>.
- Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms*. Wiley.
- Keyvanpour, M. R., Javideh, M., and Ebrahimi, M. R. (2011). Detecting and investigating crime by means of data mining: a general crime matching framework. volume 3, pages 872–880. Elsevier.
- Mande, U., Srinivas, Y., and Murthy, J. V. R. (2012). Feature specific criminal mapping using data mining techniques and generalized gaussian mixture model. *International Journal of Computer Science and Communication Networks*.
- Nath, S. V. (2006). Crime pattern detection using data mining. pages 41–44. IEEE Computer Society.
- Silva, M. P. d. S. (2004). Mineração de dados - conceitos, aplicações e experimentos com weka.
- Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining*. Addison-Wesley.
- Vijayakumar, M., Balamurugan, P., and Alhadidi, B. (2014). Crime classification algorithm for mining crime hot spot and cold spot.
- Villarreal, A. and Silva, B. F. (2006). Social cohesion, criminal victimization and perceived risk of crime in brazilian neighborhoods. *Social Forces*, 84(3):1725–1753.
- Wang, D., 0003, W. D., Lo, H. Z., Stepinski, T. F., Salaza, J., and Morabito, M. (2013). Crime hotspot mapping using the crime related factors - a spatial data mining approach. *Appl. Intell*, 39(4):772–781.
- Zaki, M. J. and Jr, W. M. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press, New York, NY, USA.

Utilização de Sistemas Multiagentes em Rede de Energia Elétrica para Identificação de Irregularidades no Consumo

Gislainy Crisostomo Velasco¹, Luana Silva Carvalho¹, Vinícius Sebba Patto¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 — 74001-970 — Goiânia — GO — Brasil

{gislainyvelasco, luanacarvalho, viniiciussebba}@inf.ufg.br

Abstract. *This article aims to use Multiagent Systems, through the GAMA platform, to simulate an electric power distribution network and its regular and irregular consumers. To identify the irregularities is proposed a computing analyzing agent that, from the simulated data, suggests the possible irregular consumers. Through the simulator and the analyzing agent, it is possible to do complementary studies about the efficiency and effectiveness of the methods used to detect non-technical losses (PNT).*

Resumo. *Este artigo tem como objetivo utilizar Sistemas Multiagentes, por meio da plataforma GAMA, para simular uma rede de distribuição de energia elétrica e seus consumidores regulares e irregulares. Para identificar as irregularidades foi proposto um agente computacional analisador que, a partir dos dados simulados, sugere os possíveis consumidores irregulares. Com o simulador e o agente analisador, é possível fazer estudos complementares acerca da eficiência e da eficácia dos métodos utilizados para detecção de perdas não-técnicas (PNT).*

Palavras-Chave: Sistemas Multiagentes, Simulação, Rede de Distribuição de Energia Elétrica, Fraude.

1. Introdução

Em uma rede de distribuição de energia elétrica existem casos de irregularidades no consumo. Elas são realizadas por seus consumidores, residenciais e comerciais, e são caracterizadas como perdas não-técnicas (PNT). No Brasil, 313 milhões de megawatts são consumidos por ano. Desse total, 5% é utilizado por ligações clandestinas, de acordo com observações feitas desde o início das privatizações do setor energético nos anos 90. Segundo a Aneel, esses furtos provocam aumento das tarifas para todos os consumidores, para compensar o que é desviado pelos infratores; além disso, dificultam o fornecimento de serviço de qualidade e a cobrança adequada do serviço prestado [ANEEL 2018a].

As concessionárias possuem dificuldades em identificar consumidores irregulares, pois as ligações clandestinas são feitas de diversas sortes e demandam considerável tempo de investigação [Penin 2008]. Para identificar consumidores irregulares, as concessionárias utilizam técnicas variadas que indicam possíveis infratores, *e.g.* média de consumo por categoria; monitoramento por períodos regulares; alterações nas faixas de consumo e outras. Porém, essas técnicas necessitam em alguns casos de inspeções presenciais nos locais indicados.

Quando isso acontece, as concessionárias enviam equipes de campo a determinado local. Segundo Morais et al., essas equipes são em média compostas por dois eletricitistas e que têm como objetivo identificar irregularidades nos sistemas de medição de consumidores alimentados em baixa, média e alta tensão em toda a área de concessão [Morais et al. 2011]. Esse tipo de procedimento demanda tempo e recursos consideráveis; todavia, nem sempre é o mais eficaz, porque as indicações nem sempre apontam para consumidores irregulares.

Por consequência das irregularidades, as tarifas regulatórias para os consumidores finais são impactadas. A redução das perdas não-técnicas pelas distribuidoras geram benefícios que vão além da redução desse item na tarifa, tais como a incorporação desses consumidores no rateio de todos os custos operacionais, a redução do consumo inconsciente ou perdulário e melhorias na qualidade do fornecimento [ANEEL 2018b].

Motivado pela necessidade de reduzir custos e analisar técnicas adotadas pelas companhias de distribuição de energia elétrica para identificar consumidores irregulares, esse trabalho tem como objetivo apresentar uma proposta de simulador, que se equivale a um "laboratório virtual", e verificar se ele poderia servir de insumo para análise e verificação dos métodos usados pelas companhias de distribuição de energia elétrica para identificar PNT.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta os conceitos fundamentais que envolvem as perdas de energia e os métodos existentes para a sua identificação, bem como conceitos de Sistemas Multiagentes; na Seção 3 são apresentados a modelagem da rede proposta e o simulador multiagente; na Seção 4, são descritos os aspectos relevantes do analisador de dados dos consumos e identificação dos consumidores irregulares; na Seção 5, são descritos os experimentos e resultados alcançados; e, por fim, na Seção 6 são feitas as considerações finais e as possibilidades de trabalhos futuros.

2. Fundamentação Teórica

Neste trabalho, as perdas de energia elétrica foram tratadas com Sistemas Multiagentes. Esses dois assuntos são detalhados nas Sub-seções 2.1 e 2.2.

2.1. Perdas de energia elétrica

O sistema elétrico é composto por geração, transmissão e distribuição, conforme ilustra a Figura 1. As perdas se referem à energia elétrica gerada que passa pelas linhas de transmissão e redes da distribuição [ANEEL 2018b]. As perdas globais são constituídas pela diferença entre a energia gerada e/ou comprada por uma empresa de energia elétrica e a energia efetivamente faturada por ela (Equação (2.1)) [Penin 2008].

$$\text{Perdas Totais} = \text{Energia Gerada ou Comprada} - \text{Energia Faturada} \quad (2.1)$$

Tais perdas podem ser divididas em perdas técnicas (PT) e perdas não-técnicas (PNT) ou comerciais:

- **Perdas Técnicas:** energias perdidas no transporte, na transformação e nos equipamentos de medição, portanto inerentes ao processo. As PT compõem o custo operacional esperado pela concessionária e são expressas em megawatt-hora (MW/h)

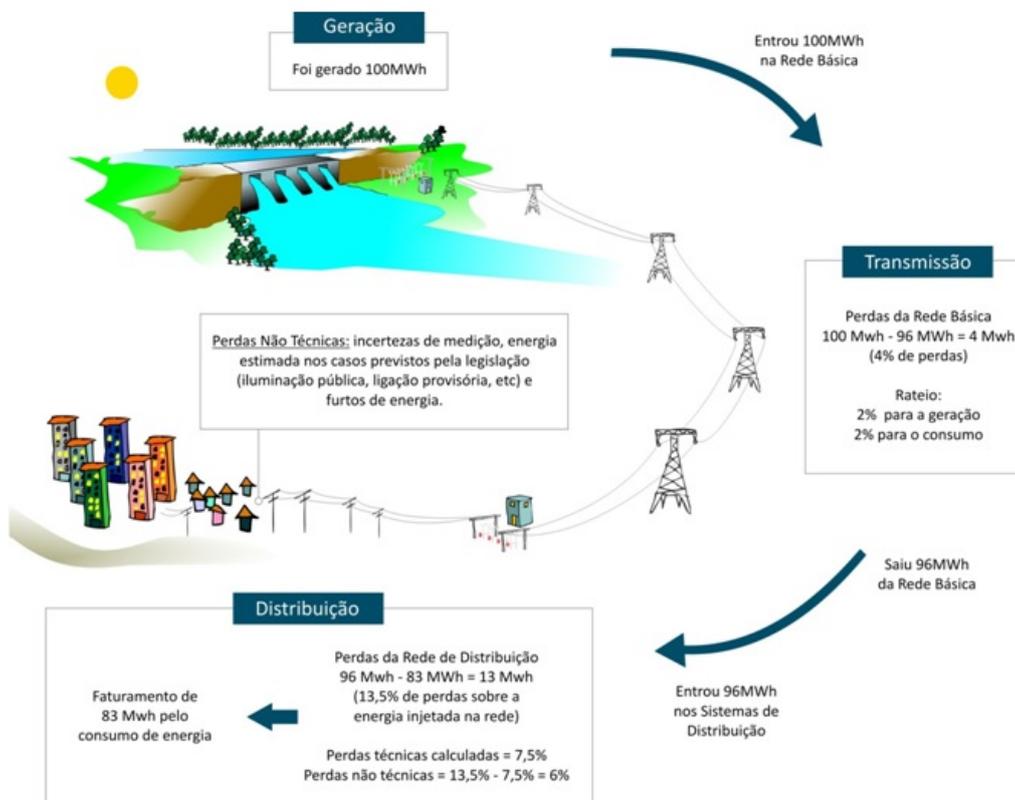


Figura 1. Estrutura de um Sistema Elétrico [ANEEL 2018b].

[Penin 2008]. Essas perdas são apuradas mensalmente pela Câmara de Comercialização de Energia Elétrica (CCEE) e o seu custo é definido anualmente nos processos tarifários, sendo rateado pela metade para geração e para os consumidores [ANEEL 2018b]. São calculadas conforme as regras definidas no Módulo 7 do PRODIST¹.

- **Perdas Não-Técnicas ou Perdas Comerciais:** São as perdas associadas à comercialização da energia até o usuário final, dada em MW/h. [Penin 2008] define a estimativa desse tipo de perda pela Equação (2.2). Os valores regulatórios são calculados pela ANEEL conforme as regras definidas no Submódulo 2.6 do PRO-RET², por meio de uma metodologia de comparação de desempenho das distribuidoras, observando critérios de eficiência e as características socioeconômicas das áreas de concessão [ANEEL 2018b].

$$\text{Perdas Não-Técnicas} = \text{Perdas Totais} - \text{Perdas Técnicas} \quad (2.2)$$

As causas mais comuns das PNT são:

- (A) Inexistência de medidores.

¹<http://www.aneel.gov.br/modulo-7>

²http://www2.aneel.gov.br/cedoc/aren2015660_Proret_Submod_2_6_V3.pdf

- (B) Falha ou falta de aferição dos medidores.
- (C) Erros na leitura dos medidores ou no faturamento das unidades consumidoras.
- (D) Furto de energia.
- (E) Fraude no fornecimento ou no sistema de medição de energia.

O foco deste trabalho está nos itens D e E. Com isso, há a necessidade de caracterizar o que são furtos e fraudes nessa área. A fraude é o ato consciente de uma pessoa para eliminar ou reduzir o montante cobrado da energia consumida no período apurado. O furto, popularmente conhecido como “gato”, é dado quando a unidade consumidora se liga diretamente à rede da distribuidora caracterizada por ligações irregulares, clandestinas [Penin 2008].

2.2. Sistemas Multiagentes

Um Sistema Multiagente (SMA) é um sistema que compreende um conjunto de agentes, software e/ou hardware distribuídos, com metas específicas que interagem e negociam uns com os outros no mesmo ambiente para resolver um problema global [Wooldridge and Jennings 1995] [Silva 2009]. Em [Russell et al. 1995], os autores definem um agente como sendo uma entidade autônoma que percebe seu ambiente por meio de sensores e age sobre o mesmo através de atuadores. Dessa forma, a operação de agentes em um SMA consiste em três etapas: percepção, tomada de decisão e ação [SAMPAIO 2017]. Na Figura 2, é apresentada a operação de um SMA e interações dos agentes.

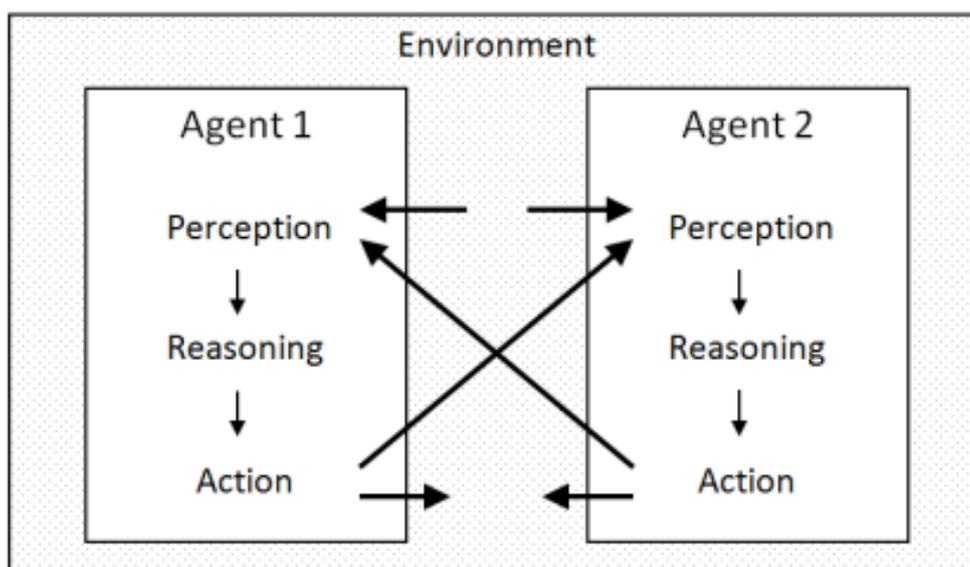


Figura 2. Operação de um SMA e interações dos agentes [Roche et al. 2010]

Neste trabalho, foi utilizado a plataforma de simulação GAMA (*Gis & Agent-Based Modeling Architecture*) para implementar o sistema multiagente para simulação de uma rede elétrica com geração de dados de consumo e para análise. O GAMA é uma plataforma permite que a modelagem baseada em agentes com a capacidade de

representar vários níveis de interação ao longo de uma representação detalhada (complexa). Em particular, integra ferramentas poderosas provenientes de Sistemas de Informações Geográficas (GIS) e *Data Mining*, facilitando os esforços de modelagem e análise [Taillandier et al. 2010].

2.3. Trabalhos de detecção de irregularidades na rede de distribuição de energia

Nesta subseção, são apresentados estudos no contexto de perdas não-técnicas, sendo que parte deles usa sistemas inteligentes e métodos estatísticos para classificar, identificar e caracterizar os consumidores com instalações elétricas irregulares [Ramos 2014]. Existem trabalhos que abordam essa temática que utilizam de sistemas computacionais, *e.g* inteligência artificial, como Rede Neural Artificial (RNA), e também, estudos voltados para a engenharia elétrica. Dentre esse trabalhos, tem-se projetos P&D regulados pela ANEEL; teses e dissertações; e artigos publicados em periódicos e conferências. A seguir, são apresentados alguns trabalhos relevantes que utilizam técnicas computacionais.

Em 2000, o projeto de P&D da Companhia de Eletricidade do Estado do Rio de Janeiro (CERJ), em conjunto com o Centro de Pesquisas de Energia Elétrica (CEPEL), desenvolveu um sistema com RNA treinadas com padrões de fraudes existentes na concessionária que fornece ferramentas que auxiliam na análise do consumo, gerenciamento das inspeções, e acompanhamento dos fraudadores [ANEEL 2006].

Outro projeto de P&D que também utiliza de RNA é o desenvolvido pela Companhia de Energética de Pernambuco (CELPE) e pesquisadores da Universidade Federal Fluminense (UFF) e Universidade Federal do Rio de Janeiro (UFRJ), realizado entre os ciclos de 2005/2006 e 2007/2008. O trabalho consiste na proposta de um método que, segundo os pesquisadores, indica os transformadores que precisam ser inspecionados, o que reduz o custo elevado de verificar todos ou de se instalar medidores comparativos ao longo de todo o sistema [ANEEL 2011].

Faria, implementou computacionalmente um sistema inteligente híbrido intercomunicativo específico que se baseia no emprego de diferentes técnicas oriundas da área de sistemas inteligentes tais como RNA e lógica fuzzy em módulos independentes e que se comunicam entre si. O sistema é baseado em três pilares: extração automática de conhecimento a partir da base de dados da concessionária, incorporação na metodologia do conhecimento e experiência de especialistas e, em último, consultas na base de dados por características específicas de cada cliente [Faria 2012].

Eller et al. propõem um método para auxiliar no gerenciamento de perdas comerciais de energia elétrica através da mineração de dados, destacando-se o uso das RNA para descobrir comportamentos suspeitos de consumidores de energia, indicando a possibilidade de fraude ou não. Nesse sentido, tem-se o trabalho realizado por [Ortega 2008] que desenvolveu uma metodologia que identifica, com maior precisão, o perfil do cliente irregular (comprovada fraude no medidor, furto por ligação clandestina ou irregularidade técnica). O sistema inteligente resultante, denominado SIIPERCOM, baseia-se em RNA para a filtragem agrupando clientes com comportamentos semelhantes e classificação dos clientes de cada grupo em normais ou irregulares [Eller et al. 2003].

No trabalho de Monedero et al., também usaram técnicas de mineração de dados, propondo a utilização de RNA, juntamente com uma análise estatística de fraudes no consumo elétrico [Monedero et al. 2006]. Em um trabalho recente, Monedero et al.

também propuseram para a detecção de fraudes uma análise através da utilização do coeficiente de Pearson em conjunto com o uso de redes bayesianas e árvores de decisões [Monedero et al. 2012].

Vale ressaltar que é difícil comparar os métodos desenvolvidos por outros autores com o utilizado neste trabalho com relação a identificação das perdas não-técnicas, visto que a maioria deles trabalha diretamente com um histórico de curvas de carga ou, indiretamente, com características extraídas dessas curvas. Ao contrário, neste trabalho, no qual as curvas de carga não foram disponibilizadas por nenhuma concessionária e os dados gerados foram por meio da simulação. Além disso, as características usadas neste trabalho nem sempre são iguais às características usadas por outros autores. Outro fato importante é a utilização de técnicas de SMA ao invés das outras áreas da IA que foram usadas pelos autores citados.

3. Modelagem da rede de energia elétrica

Esta seção apresenta a modelagem da rede de energia elétrica. Em seguida, a criação do SMA por meio do *framework* GAMA.

3.1. Rede elétrica

Uma rede de energia elétrica possui uma quantidade de variáveis que a torna complexa. Nesta rede, os consumidores estão ligados à determinada sub-rede ou transformador de energia elétrica. Para construção do simulador SMA e análise dos dados gerados, foi modelada uma rede simplificada, face à maioria das redes de energia elétrica encontradas nos grandes centros urbanos; porém, foram considerados aspectos suficientes para simular os dados necessários para a aplicação de técnicas usadas para detecção de PNT.

A topologia modelada é composta por uma única rede na qual trafega toda a corrente elétrica e é distribuída por seus consumidores sem restrição de consumo. Os consumidores são de dois tipos: comerciais e residenciais. O tipo residencial, tem baixas faixas de consumo e possuem variações nas categorias de consumidores, *e.g.* uma residência que possui somente um único morador tende a apresentar consumo diferente de uma residência com mais que 3 moradores. Nesse sentido, os comerciais tem as variações que dependem do porte da empresa (quantidade de funcionários) e da atuação (indústria geralmente consome mais que comércio).

Para representar essas faixas de consumo foram criadas, com auxílio de especialista do setor elétrico, categorias de faixas de consumos para os dois tipos, conforme dados da Tabela 1. Observa-se que os valores das faixas não se tratam de valores reais, somente estipulações para que a simulação fique mais próxima da realidade dos perfis de consumidores do estado de Goiás.

Cada consumidor possui propriedades que o identifica na rede, como:

- Unidade consumidora: identificação única na rede.
- Tipo de consumidor: residencial ou comercial.
- Consumo registrado por período: geralmente medido em KW/h, tratando-se da base para cálculo do valor cobrado no período - é o valor registrado pelo medidor.
- Faixa de consumo: trata-se da média de consumo da unidade consumidora. Foram utilizadas as variações da Tabela 1.

Tabela 1. Faixa de consumo dos consumidores de acordo com suas categorias

Categorias	Faixa de Consumo	
	Residenciais	Comerciais
CAT0	1, 100	1000, 1500
CAT1	101, 200	1501, 2000
CAT2	201, 300	2001, 3000
CAT3	301, 400	3001, 4000
CAT4	401, 500	4000, 5500
CAT5	501, 1000	5501, 8000

Para atender a necessidade de se saber quais são os consumidores irregulares (saber se o agente computacional teve sucesso) foi adicionado mais duas propriedades:

- É irregular: é uma propriedade que possui dois valores possíveis: verdadeiro, caso o consumidor seja irregular e falso, caso contrário.
- Consumo real por período: se o consumidor é irregular, logo o seu consumo registrado por período não corresponde ao valor real consumido, para fins de análise foi adicionado essa propriedade para conseguir identificar posteriormente esses consumidores. Caso o consumidor não seja irregular, o valor do período corresponde ao mesmo do registrado.

Com isso, a rede possui as propriedades:

- Quantidade de consumidores: valor numérico correspondente a soma total de cada tipo.
- Consumo registrado por período: somatória de todos os consumos registrados pelo consumidor no período.
- Consumo real por período: somatória de todos os consumos reais pelo consumidor no período.

3.2. Simulador SMA

A plataforma GAMA foi a utilizada no desenvolvimento do SMA. Para isso, foi necessário a modelagem de um agente genérico para a simulação. O agente possui todas as propriedades apresentada no consumidor. O meio que os agentes interagem é a rede de distribuição de energia elétrica que também possui as propriedades apresentadas na seção anterior.

O foco do simulador é permitir que as distribuidoras consigam reproduzir cenários de irregularidades próximas ao mundo real. Por isso, existem uma gama de possibilidades de combinações possíveis de informações que podem facilitar nas análises posteriores. Com isso, o simulador permite definir algumas propriedades, como o número de consumidores na rede, por padrão foi utilizado mil consumidores. Para cada ciclo é utilizado o intervalo de um mês por ser o convencional e usado por grande parte das concessionárias como o período faturado, o histórico de consumo é feito por um período de doze meses (um ano). Vale ressaltar que se for empregado o método de análise por histórico, períodos curtos podem não ser adequados para detecção de irregularidades.

Com a definição do meio estabelecida, são instanciados os agentes consumidores. Cada agente possui a probabilidade X de ser residencial e Y de ser irregular, considerando X e Y como 70% e 10%, respectivamente. Esses valores podem ser modificados dependendo do cenário. Além disso, é realizado um sorteio para classificar o agente a uma faixa de consumo, como mostrado na Tabela 1.

Para cada ciclo de interação os agentes realizam os seus comportamentos, neste caso, o consumo de energia. Como o intervalo definido é mensal, o consumo do ciclo é dado por meio de um sorteio entre as faixas de consumo, ou seja, se o consumidor é residencial e pertence à categoria CAT2 o sorteio estará entre 101 à 200 KW/h. O valor do consumo fará parte do consumo registrado no período, caso a propriedade irregular seja verdadeira, o consumo real é dado pela Equação (3.1). O Percentual do qual a Equação (3.1) se refere é explicado na Seção 4.

A cada ciclo de interação, o agente consumidor interage com a rede (meio) que possui uma quantidade de energia que é disponibilizada para consumo, para efeitos de cálculos das perdas (técnicas e não-técnicas), são registrados na rede global a energia registrada e a real consumida.

$$\text{Consumo Real} = \text{Consumo Registrado} * \text{Percentual} \quad (3.1)$$

Por fim, após todos os ciclos, as informações geradas são armazenadas em arquivos CSV (*Comma-Separated Values*); um arquivo contém todas as informações geradas pelo consumidor, suas propriedades, bem como a energia registrada e consumida por período; o outro arquivo, contém somente a energia registrada, pois este é utilizado para detecção de irregularidades.

4. Analisador de irregularidades

A segunda etapa deste trabalho consiste na proposta de um agente computacional analisador para detecção de irregularidades na rede. O analisador é um agente que verifica os dados gerados pela simulação a fim de indicar os possíveis consumidores que podem estar fraudando o sistema. Foi utilizado o método de detecção por desvio padrão por meio de percentuais abaixo e acima da média pela faixa de consumo.

As informações necessárias para realizar a análise estão contidas no arquivo CSV que possui somente as informações de consumos simulados. Em cada registro é identificado o tipo e a categoria dos quais o consumidor pertence; assim, é possível obter a média de consumo por categoria. Após a instanciação, o agente analisador faz a leitura do arquivo e inicia as verificações dos consumos mensais simulados com o objetivo de verificar se algum consumo (Consumo Acima ou Consumo Abaixo - ver Eq. 4.1 e Eq. 4.2) ficou acima ou abaixo de um limiar estipulado para a média de determinada categoria (Média Consumo Acima ou Média Consumo Abaixo - ver Eq. 4.3 e Eq. 4.4). Na Seção 5, foram estipulados 6 percentuais (para cima e para baixo): 0%, 2%, 6%, 14%, 30% e 62%.

$$\text{Consumo Acima} = \text{Consumo Mensal} + \text{Percentual Acima} \quad (4.1)$$

$$\text{Consumo Abaixo} = \text{Consumo Mensal} - \text{Percentual Abaixo} \quad (4.2)$$

$$\text{Média Consumo Acima} = \text{Média Consumo Por Categ.} + \text{Percentual Acima} \quad (4.3)$$

$$\text{Média Consumo Abaixo} = \text{Média Consumo Por Categ.} - \text{Percentual Abaixo} \quad (4.4)$$

Caso o Consumo Acima ou Consumo Abaixo seja superior ou inferior à Média Consumo Acima ou Média Consumo Abaixo, respectivamente, o agente sinaliza o usuário da rede como possível consumidor irregular e essa informação é armazenada em um arquivo CSV com todos os potenciais consumidores fraudulentos.

5. Experimentos e resultados

Durante os experimentos, foram identificados possíveis cenários que impactam nas detecções dos consumidores irregulares por meio do analisador proposto. No mundo real, existem várias formas de realizar furtos e fraudes nas redes de energia. Como não foi obtido dados reais das distribuidoras para serem analisados, foram utilizadas durante os testes algumas combinações do simulador para geração de dados para verificar se o modelo proposto no analisador é eficiente e eficaz.

Nas simulações, foram considerados dois tipos de consumidores irregulares: os que furtam e permanecem na sua faixa de consumo e os que furtam e saem da sua faixa de consumo. O primeiro caso acontece com consumidores que porventura realizam furtos e, às vezes, realizam de maneira mais sofisticadas, geralmente acompanhada de pessoas que possuem práticas nesse ramo e sabem como ludibriar a concessionária sem serem identificados. O segundo acontece com mais frequência e sem auxílio de métodos sofisticados, os consumidores em muitos casos querem realizar o furto para diminuir o montante pago, os motivos para essa prática são inúmeras, *e.g* condição financeira, inadimplência, histórico-cultural e outros fatores [Penin 2008].

Com isso, foram realizados os experimentos para esses dois tipos de consumidores irregulares. As variações foram: quantidade de consumidores totais, taxa de consumidores residenciais e comerciais, taxa de consumidores regulares e irregulares e mudanças nas faixas de consumo nos casos de irregularidade, seguindo os padrões identificados anteriormente. Os consumidores residenciais foram representados por círculos amarelos e comerciais por brancos, para os consumidores irregulares as bordas ficaram na cor vermelha e regulares na cor azul, conforme ilustra a Figura 3.

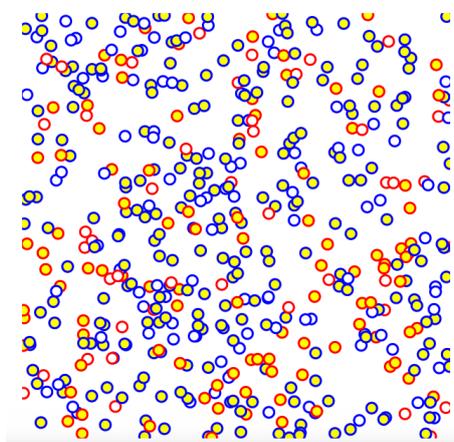


Figura 3. Representação de 500 consumidores, sendo 70% residenciais e 30% irregulares no simulador SMA proposto.

Foram realizadas um total de 12 simulações; nas 6 primeiras, os consumidores irregulares tiveram a soma do consumo com furto dentro da sua faixa usual de consumo; nas outras 6 simulações, esses valores ficaram fora da faixa usual de consumo. No analisador foram utilizados os seguintes percentuais para cálculos dos Consumo Acima e Consumo Abaixo: 0%, 2%, 6%, 14%, 30% e 62% (detalhes na Seção 4). As Tabelas 2 e 3, mostram a média de indicações para as simulações, a taxa de acertos comparada a quantidade dos consumidores irregulares na rede e a taxa de erros das indicações.

A taxa de acerto (Eq. 5.1) é obtida pela quantidade de acertos de consumidores irregulares indicados pelo agente analisador dividido pela quantidade de consumidores irregulares gerados na simulação dividida. Já a taxa de erro (Eq. 5.2), é obtida subtraindo de 100% o resultado da divisão da quantidade de indicações corretas de consumidores irregulares pela quantidade total de consumidores indicados pelo agente analisador.

$$Taxa\ Acerto = \frac{Número\ Acertos}{Total\ Irregulares} \quad (5.1)$$

$$Taxa\ Erro = 100 - \left(\frac{Número\ Acertos}{Total\ Indicações} \right) \quad (5.2)$$

Tabela 2. Resultados dos experimentos com valores de consumo dentro da faixa de consumo

Percentual	Dentro da Faixa de Consumo		
	Média de Indicações (500 -- 1000)	Taxa de acertos na rede (500 -- 1000)	Taxa de erros pelo total de indicações (500 -- 1000)
0%	500 -- 1000	100% -- 100%	23,14% -- 27,6%
2%	449 -- 902	89,28% -- 88,3%	27,79% -- 27,6%
6%	339 -- 694	69,31% -- 66,85%	28,27% -- 28,37%
14%	138 -- 300	32,58% -- 30,1%	30,18% -- 28,50%
30%	16 -- 34	2,94% -- 3,15	26,93% -- 25,11%
62%	1 -- 1	0% -- 0%	100% -- 100%

Tabela 3. Resultados dos experimentos com valores de consumo fora da faixa de consumo

Percentual	Fora da Faixa de Consumo		
	Média de Indicações (500 -- 1000)	Taxa de acertos na rede (500 -- 1000)	Taxa de erros pelo total de indicações (500 -- 1000)
0%	500 -- 1000	100% -- 100%	30% -- 30,1
2%	455 -- 917	96,11% -- 94,7%	29,31% -- 31,14%
6%	364 -- 725	87,21% -- 83,47%	33,84% -- 34,33%
14%	192 -- 397	68,17% -- 66,42%	47,03% -- 48,56%
30%	61 -- 120	35,69% -- 31,33%	75,84% -- 74,64%
62%	1 -- 1	4,41% -- 3,23%	88,89% -- 83,34%

6. Considerações finais

O simulador apresentado neste artigo tem como propósito simular vários cenários de consumos em redes de distribuição de energia elétrica com consumidores regulares, irregulares, residenciais e comerciais. Apesar de ser uma proposta simplificada, conseguiu abstrair os aspectos necessários para gerar os dados de estudos acerca de métodos de identificação de PNT. A realização dos experimentos mostrou que um dos métodos mais usados (o método de detecção por desvio padrão por meio de percentuais abaixo e acima da média pela faixa de consumo) possui uma pequena eficiência e eficácia. Por consequência, métodos precisam ser repensados e aperfeiçoados. Como existem várias formas de irregularidades, a identificação dos consumidores irregulares se torna complexa e depende de como esses consumidores cometem fraudes/furtos. Observa-se que se utilizar do método de detecção por meio do desvio padrão há uma taxa considerável de acertos; porém, ainda resta uma taxa alta de erros. Isso indica que, se for enviado uma equipe de técnicos ao local, tem-se um uso pouco eficiente e de baixa eficácia dos recursos financeiros e humanos.

A partir desse trabalho, destaca-se possibilidades de estudos complementares, por exemplo, de utilizar multiagentes para analisar dados simulados para verificar a eficiência e eficácia de outros métodos de detecção de fraude. Há também a possibilidade de aumentar a complexidade da rede de distribuição de energia elétrica para que a simulação fique ainda mais parecida com ambientes encontrados nos grandes centros urbanos. Dessa forma, há de se destacar que o estudo, abre uma gama de possibilidades de estudos complexos e de grande importância tanto para as companhias de distribuição de energia elétrica quanto para a academia.

Referências

- ANEEL (2006). Revista Pesquisa e Desenvolvimento da ANEEL P&D. 1.
- ANEEL (2011). Revista Pesquisa e Desenvolvimento da ANEEL P&D. 4.
- ANEEL (2018a). Medição, faturamento e combate a perdas comerciais. http://www2.aneel.gov.br/visualizar_texto.cfm?idtxt=1623/. Online; acessado 24 Junho 2018.
- ANEEL (2018b). Perdas de energia. http://www.aneel.gov.br/conteudos-relacionados-distribuicao/-/asset_publisher/hrGWK1e8xMyp/. Online; acessado 18 Julho 2018.
- Eller, N. A. et al. (2003). Arquitetura de informação para o gerenciamento de perdas comerciais de energia elétrica.
- Faria, L. T. d. (2012). Sistema inteligente híbrido intercomunicativo para detecção de perdas comerciais.
- Monedero, Í., Biscarri, F., León, C., Biscarri, J., and Millán, R. (2006). Midas: Detection of non-technical losses in electrical consumption using neural networks and statistical techniques. In *International Conference on Computational Science and Its Applications*, pages 725–734. Springer.
- Monedero, I., Biscarri, F., León, C., Guerrero, J. I., Biscarri, J., and Millán, R. (2012). Detection of frauds and other non-technical losses in a power utility using pearson

- coefficient, bayesian networks and decision trees. *International Journal of Electrical Power & Energy Systems*, 34(1):90–98.
- Morais, F. J. d. O. et al. (2011). Equipamento eletrônico de baixo consumo com comunicação sem fio para auxílio nas inspeções de detecção de roubo de energia elétrica.
- Ortega, G. V. C. (2008). Redes neurais na identificação de perdas comerciais do setor elétrico. *Rio de Janeiro*.
- Penin, C. A. d. S. (2008). *Combate, prevenção e otimização das perdas comerciais de energia elétrica*. PhD thesis, Universidade de São Paulo.
- Ramos, C. C. O. (2014). *Caracterização de perdas comerciais em sistemas de energia através de técnicas inteligentes*. PhD thesis, Universidade de São Paulo.
- Roche, R., Blunier, B., Miraoui, A., Hilaire, V., and Koukam, A. (2010). Multi-agent systems for grid energy management: A short review. In *IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society*, pages 3341–3346. IEEE.
- Russell, S., Norvig, P., and Intelligence, A. (1995). A modern approach. *Artificial Intelligence*. Prentice-Hall, Englewood Cliffs, 25(27):79–80.
- SAMPAIO, R. F. (2017). Sistema de automação distribuído uma abordagem baseada em multiagente aplicada a sistemas de distribuição de energia elétrica em média tensão. *Tese (Doutorado em Engenharia Elétrica) Centro de Tecnologia*, 25(27):225.
- Silva, B. C. d. (2009). *Utilização de um sistema multi-agentes em redes de comunicação para a proteção digital de distância adaptativa*. PhD thesis, Universidade de São Paulo.
- Taillandier, P., Vo, D.-A., Amouroux, E., and Drogoul, A. (2010). Gama: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 242–258. Springer.
- Wooldridge, M. and Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The knowledge engineering review*, 10(2):115–152.

Governança de TI no Setor Público: Um Estudo Sobre o Impacto na Efetividade Organizacional

Luciano da Costa Duarte¹, Eliomar A. de Lima¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Câmpus Samambaia, CEP 74690-900 – Goiânia, GO – Brasil

lucianoduarte2@hotmail.com, eliomar@inf.ufg.br

Abstract. *This article aims to promote an organizational impact analysis from the description of the IT governance architecture within the Sistema de Informática do Serviço Público (SISP) in Brazil. Based on the assumption that organizations with well defined IT governance processes are more effective when compared to others that do not have them, it should be noted that in the federal public sector, despite the planning, control and supervision actions emanating from the competent Organs, there is still a predominant framework of incomprehensions and low level of maturity in terms of governance and IT management processes in the bodies and entities that make up the SISP. Considering this scenario, an exploratory research was undertaken in order to broaden the horizons of study on the state of things that characterize the main contexts that are contained in the IT governance system of SISP and its understood entities, and then to evaluate the level of organizational effectiveness, guided by the meta-model of subsystems of directing, development and delivery before the levels of performance organization, processes and implementation. The analysis revealed the current model of IT governance within the scope of the SISP, as well as the level of satisfaction achieved in the most diverse dimensions evaluated.*

Key-words: *IT governance, organizational effective, IT subsystems.*

Resumo. *Este artigo visa promover uma análise de impacto organizacional proveniente da descrição da arquitetura de governança de TI no âmbito do Sistema de Informática do Serviço Público (SISP). Partindo do pressuposto de que organizações com processos de governança de TI bem definidos obtêm maior efetividade quando comparadas com outras que não os possui, é preciso observar que no setor público federal, apesar das ações de planejamento e de controle e fiscalização, emanadas dos Órgãos competentes, ainda persiste um quadro predominante de incompreensões e de baixo nível de maturidade em termos de processos de governança e de gestão de TI nos órgãos e entidades que compõem o SISP. Diante desse cenário, empreendeu-se uma pesquisa de caráter exploratório, ampliando os horizontes de estudo sobre o estado de coisas que caracterizam os principais contextos que se encerram no sistema de governança de TI do SISP e seus órgãos compreendidos, para em seguida avaliar o nível de efetividade organizacional, pautada pelo modelo de subsistemas de direcionamento, desenvolvimento e entrega ante aos níveis de desempenho organização, processos e implementação. As análises permitiram revelar o modelo atual da governança de TI no âmbito do SISP, bem como o nível de satisfatoriedade alcançado nas mais diversas dimensões avaliadas.*

Palavras-chave: *governança de TI, efetividade organizacional, subsistemas de TI.*

1. Introdução

Vive-se em uma sociedade cada vez mais dependente de Tecnologia da Informação (TI), dada a ultraconectividade presente nas mais diversas instâncias e manifestações organizacionais. No ambiente corporativo, a dependência da TI não é diferente, cada vez mais a complexidade e os desafios que se encerram nas organizações públicas e privadas aumentam. As soluções computacionais auxiliam o negócio a contrapor essa situação.

De acordo com uma pesquisa realizada pelo Information Technology Governance Institute (ITGI), com quase 800 entrevistados, de 21 países, entre eles executivos de TI e de negócios, 94% disseram que a utilização da TI para a estratégia de negócio é considerada essencial [Isaca 2011], o que sugere que a TI efetiva pode agregar valor aos serviços e produtos de uma organização.

Somente o fato de a organização ter um departamento ou função de TI, não é suficiente para que esta agregue valor efetivo ao negócio. A Governança de Tecnologia da Informação (GovTI) atua como um direcionador da TI, ela espelha conceitos da governança corporativa, ao mesmo tempo em que se concentra no uso racional e efetivo da TI para alcançar as metas da organização. Os estudos de Weill e Ross [2006] revelam que empresas onde a GovTI é efetiva acabam tendo um retorno financeiro de até 20% mais alto.

Haja vista a grande importância da GovTI, o setor público federal brasileiro vem intensificando ações de planejamento e de fiscalização e controle por meio de seus órgãos competentes, de modo a buscar um aprimoramento contínuo dos processos de GovTI e de gestão de TI. A partir de 2007 um modelo de avaliação passou a ser adotado para tratar o assunto de forma mais ampla e profunda, utilizando-se de instrumentos de avaliação bianual e inspeções regulares promovidos pelo Tribunal de Contas da União (Tcu).

Em grande medida, a TI no setor público sugere um desalinhamento das estratégias de TI com as estratégias corporativas, desencadeando gastos desnecessários e baixo nível de agregação de valor da TI às organizações públicas, levando ao seguinte questionamento: em que medida o modelo de governança de TI no setor público está adequado para promover a TI de forma sustentada? Nesse sentido, o objetivo deste estudo é mapear os elementos essenciais para estruturação de um modelo de governança de TI no setor público, pautado pelo princípio da sustentação.

Para tanto, busca-se identificar os fatores candidatos a compor o sistema de governança de TI, descrever o Sistema de Administração dos Recursos de Tecnologia da Informação da administração direta, autárquica e fundacional do Poder Executivo Federal (SISP), levantar os instrumentos de diagnóstico do modelo atual de governança de TI no setor público, para ao final avaliar o estado atual de desenvolvimento do SISP em termos de GovTI.

2. Governança de TI como um fator habilitador para a efetividade organizacional

A definição de Governança de TI (GovTI) é bastante controversa, não obstante haver um consenso quanto à importância de sua aplicabilidade no contexto organizacional. Em termos práticos, a GovTI se traduz em um conjunto de políticas, processos, papéis e responsabilidades associados a estruturas e pessoas da organização, de modo a estabelecer claramente o processo de tomada de decisões e as diretrizes para o gerenciamento e uso da TI, de forma alinhada à visão, missão e metas estratégicas da organização [Tcu 2017].

De acordo com o Itgi [2007], a GovTI apresenta cinco áreas focais: alinhamento estratégico, entrega de valor, gestão de riscos, gestão de recursos e mensuração de desempenho.

Para a GovTI ser implantada no setor público, para além dessas áreas, algumas diretrizes devem ser observadas. Segundo Rodrigues et al. [2010] faltam iniciativas estruturadas por meio da administração pública brasileira para implantação de GovTI e que não basta somente haver instruções normativas se não há instrução para sua implantação. As diretrizes recomendadas são [*ibid.*]: criação do segmento de GovTI na estrutura dos órgãos, implantação de mecanismo de relacionamento, evolução da maturidade dos principais processos de TI, introdução de procedimentos e práticas já legitimadas na APF.

No âmbito da Administração Pública Federal (APF) foi constatado por meio de levantamento feito pelo TCU em 2014 que 78% dos órgãos e entidades possuem comitês de TI, um percentual razoavelmente bom se estivessem efetivamente realizando suas funções precípuas [Cunha and Neto 2014]. Para superar esse quadro divergente, os autores sugerem fatores críticos de sucesso para se alcançar maior efetividade dos comitês de TI.

Desde 1986, a efetividade organizacional é um termo que não apresenta um consenso sobre sua definição e conceito [Cameron 1986], há também uma dificuldade na diferenciação da efetividade organizacional aplicada no setor público e no privado [Parhizgari and Gilbert 2004]. Por vezes, a efetividade organizacional é vista como um indicador que mensura o distanciamento ou aproximação do que foi planejado e executado. Outra definição de efetividade é a convergência dos conceitos de eficiência e eficácia; eficiência está relacionada com produzir o máximo com o mínimo de recurso e a eficácia se relaciona com a capacidade de alcançar o que foi planejado [Cavaliere and Macedo 2004].

Do mesmo modo que falta um consenso do conceito da efetividade organizacional, o mesmo se aplica para a efetividade da TI. Para Silva e Moraes [2011], a efetividade da TI pela perspectiva de negócio é voltada aos vínculos entre os resultados da TI e o sucesso organizacional, buscando compreender o contexto da organização e sua efetividade numa visão que engloba o ambiente externo, as características setoriais e a estrutura funcional.

Do ponto de vista da GovTI, a efetividade se caracteriza pelo cumprimento de quatro objetivos levantados por Weill e Ross [2006]: custo, crescimento, utilização de ativos e flexibilidade de negócio. Adicionalmente, Bowen et al. [2007] sugerem que se deve cumprir requisitos legais e regulamentares previstos para o negócio.

Logo, o cumprimento desses objetivos, aliada aos requisitos propostos, torna a GovTI efetiva porque passa a atender alguns fatores críticos para a efetividade organizacional, a saber: transparência, participação ativa, redesenho frequente, governança, simplicidade, processo de tratamento de exceções, governança desenhada para múltiplos níveis organizacionais e alinhamento nos sistemas de incentivo e recompensa [Weill and Ross 2006]. Contudo, uma GovTI efetiva deve envolver também um conjunto de processos com o objetivo de realização das expectativas em relação aos fatores críticos de sucesso para a organização [Isaca 2011].

3. Metamodelo de Subsistemas e Níveis de desempenho de TI

A TI é um dos sistemas de uma organização, sendo compreendida por três subsistemas: direcionamento, desenvolvimento, e entrega. Esses subsistemas interagem entre si, possibilitando o alinhamento da TI com a organização, como mostra a Figura 1 [Betz 2011].

O subsistema direcionamento está envolvido com a estratégia geral da área de TI e de sua arquitetura, com funções de: planejamento e controle de alto nível, monitoramento das demandas e priorização de serviços ou produtos a serem desenvolvidos ou entregues [Molinari and Ramos 2013]. Já o subsistema desenvolvimento é responsável pelo empreendimento de ações e projetos autorizados pelo subsistema de direcionamento. Possui

quatro processos a fim de mitigar os riscos inerentes, são eles: gerenciamento de projetos, concepção e construção de soluções, garantia de qualidade e gerenciamento de configurações [Betz 2011].

A entrega, por seu turno, é o subsistema que contempla a implantação dos projetos que foram autorizados pelo subsistema de direcionamento e concebido pelo subsistema de desenvolvimento. Sua principal funcionalidade é a operação e a manutenção de serviços [Molinaro and Ramos 2013].

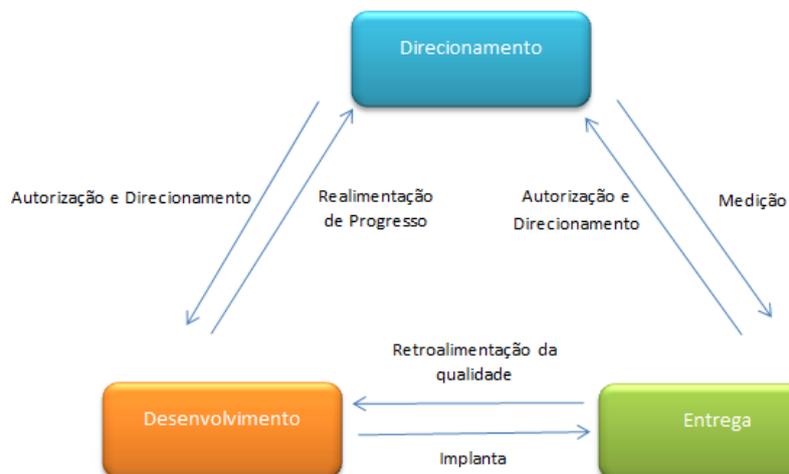


Figura 1. Interações dos subsistemas de TI [BETZ 2011]

O Quadro 1 apresenta uma heurística para avaliação das funcionalidades e as preocupações de cada área [Molinaro and Ramos 2013].

Quadro 1. Heurística para avaliação dos subsistemas de TI

Subsistema	Questões da Heurística
Direcionamento	Quais são os investimentos futuros mais promissores no portfólio de TI?
	Quais são os investimentos correntes aceitáveis?
	Quais são os investimentos correntes questionáveis?
	Quais são os investimentos correntes não aceitáveis?
	Quais são os custos totais de aquisição e operação?
Desenvolvimento	Quais são os serviços ou sistemas que necessitam melhorar?
	Que sistemas usam um determinado elemento de dados?
	Qual o nível corrente de mudanças nos sistemas?
Entrega	Como a organização está consumindo os recursos de TI?
	Qual o status operacional e as tendências dos sistemas?
	Como métricas de incidentes e problemas se relacionam com as atividades de mudança?

Fonte: Betz [2011]

Dentre as várias formas de classificação de um sistema organizacional, Rummler e Branche [1994] propõem a compreensão da organização em três níveis de desempenho, a saber: Organização, Processos e Trabalho/Executor. Quando estão voltados para mesma direção impactam significativamente a eficiência e eficácia organizacional.

No nível da Organização são definidas as características gerais do negócio, o que permite as outras áreas funcionais uma orientação ao alcance dos objetivos. São consideradas

e avaliadas as estratégias, a governança, a estrutura, as arquiteturas corporativas e o emprego dos recursos [Rummler and Branche 1994 apud Molinaro and Ramos 2013].

No nível de Processos é possível ver o fluxo de trabalho e o esforço geral empreendido. A principal técnica para utilizada nesse modelo é o mapa de processos onde são definidas as atividades que produzem valor para o negócio, destacando suas entradas e saídas, o fluxo de atividades, as pessoas, as funções e as regras de negócio [Damelio 1996].

No último nível Trabalho/Executor são verificados os cargos e as pessoas, com o propósito de obter mão de obra qualificada para que o ambiente organizacional apoie a realização dos objetivos do trabalho [Rummler and Branche 1994 apud Molinaro and Ramos 2013].

O BPTrends – organização reconhecida como uma importante fonte de informações sobre melhores práticas de gestão de processos de negócio, traz uma nova terminologia quanto ao nível de Trabalho/Executor proposto por Rummler e Branche [1994]. Nessa contribuição com a inclusão da Tecnologia da Informação, passou a ser denominado nível de implementação, o qual será aplicado nesse trabalho por ser mais adequado quando se aborda organizações de TI [Bptrends 2017].

Para cada subsistema de TI pode-se aplicar os níveis de desempenho Organização, Processos e Implementação. O objetivo dessa aplicação é proporcionar o alinhamento da TI com ela mesma e também com o negócio, fazendo com que os subsistemas de TI sejam avaliados na perspectiva de cada nível de desempenho [Molinaro and Ramos 2013].

Em síntese, um exemplo de arquitetura organizacional que contempla os subsistemas de TI está ilustrado na Figura 2. Caberá ao sistema de GovTI atuar sobre cada um dos compartimentos do triângulo para se buscar a efetividade organizacional, segundo os parâmetros de alinhamento de TI e negócio [*ibid.*].

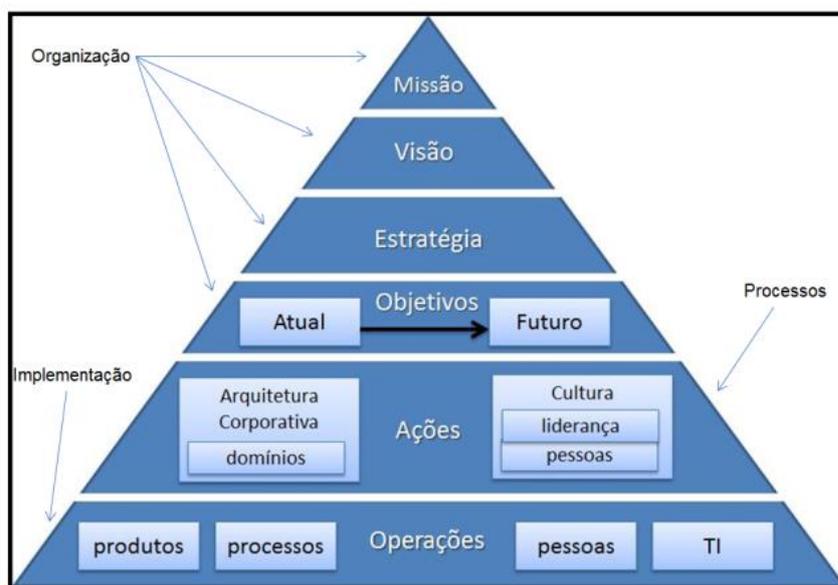


Figura 2. Elementos descritores dos subsistemas de TI [adaptado de Betz 2011]

4. Método

O presente trabalho caracteriza-se como sendo exploratório, tendo em vista que busca uma maior compreensão da Governança de TI no setor público federal brasileiro, com vistas à avaliação do nível de efetividade organizacional. Especificamente, o interesse dessa pesquisa

recai nos órgãos que compõem o SISP, por se tratar de uma instância de representação por excelência do Governo Federal.

A estrutura do SISP comporta diversas categorias de Órgãos da APF, totalizando 224 unidades. O SISP organiza a operação, controle, supervisão e coordenação dos recursos de informação de toda a APF. Por esta razão, escolheu-se o SISP enquanto unidade de análise, devido à sua grande representatividade para o setor público federal.

Para atingir o objetivo desta pesquisa, qual seja o mapeamento dos elementos essenciais para estruturação de um modelo de governança de TI no setor público, duas etapas foram empreendidas: 1) estudo exploratório, no qual se efetuou pesquisa bibliográfica e documental, além da coleta de dados em fontes primárias e secundárias no âmbito do SISP e, 2) análise arquitetural, na qual foram produzidos os seguintes artefatos: mapa de situação de GovTI do SISP, mapa de contexto de governança de TI e controle dos subsistemas de TI.

5. Resultados e Discussão

Para mapear os elementos essenciais do modelo de GovTI do SISP artefatos informacionais e documentais foram levantados no segundo semestre de 2017, incluindo leis, instruções normativas, acórdãos, normas, guias de boas práticas, entre outros. Devido à alta fragmentação desses artefatos no âmbito dos órgãos emanadores de dispositivos institucionalizantes, normativos e legais no âmbito do SISP, um grande esforço foi empreendido no sentido de levantá-los, conforme lista a seguir:

- Tribunal de Contas da União (TCU), por intermédio da Secretaria de Fiscalização de Tecnologia da Informação (SeFTI), atuando como um órgão regulador e fiscalizador;
- Ministério do Planejamento, Desenvolvimento e Gestão (MPDG), por intermédio da Secretaria de Tecnologia da Informação e Comunicação (SETIC), tendo a incumbência de propor políticas, planejar, coordenar, supervisionar e orientar normativamente as atividades de governança e gestão dos recursos de TI.
- Gabinete de Segurança Institucional (GSI), por intermédio do Departamento de Segurança da Informação e Comunicações (DSIC), atuando nas atividades relacionadas ao credenciamento de segurança e ao tratamento de informação sigilosa.
- Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC), com competência para planejar, coordenar, supervisionar e controlar as atividades de ciências, tecnologia, inovações e comunicações.
- Sistema de Administração dos Recursos de Tecnologia da Informação (SISP), com o objetivo de organizar a operação, controle, supervisão e coordenação dos recursos de tecnologia da informação da APF.
- Demais entes legislativos, executivos e jurídicos com competência para atuar na TI.

Após análises preliminares dos artefatos coletados foi possível obter uma visão das principais partes interessadas que compõem o Sistema de Governança de TI (SGTI) no âmbito do SISP, assim sintetizada na Figura 3.

Para que o SGTI pudesse ser descrito, o primeiro quadrante do modelo proposto por Lima [2015] foi adotado para descrever os contextos de análise que se encerram no SISP. O propósito ao descrever os blocos de construção do SGTI é propiciar a elucidação da consciência situacional [Lima 2015], contendo os seguintes artefatos:

- Mapa de Situação de Governança de TI no Setor Público
- Mapa de Contexto de Governança de TI no Setor Público
- Controle dos Subsistemas de TI

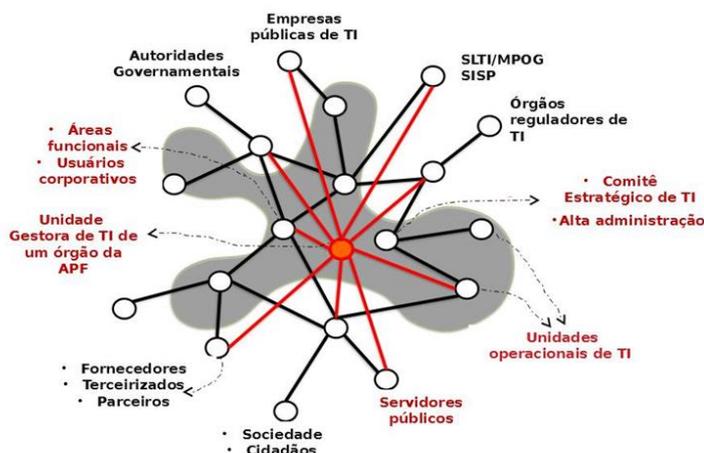


Figura 3. Organização do Sistema de GovTI do SISP [Autoria própria 2018]

5.1 Mapa de Situação e Contexto de GovTI

O setor público apresenta um ambiente organizacional bastante complexo, pois os processos não são bem definidos, atuando com diferentes frameworks, com prestação de contas para diversos stakeholders, como: ministros, parlamento, cidadãos, cliente e o público em geral, nenhum com qualquer direito de proprietário [August, 2001], isso reflete negativamente no modelo de GovTI, já que os elementos que a definem não estão desfragmentados.

Dada à complexidade optou-se por utilizar o pensamento sistêmico crítico com base na metodologia Heurística Sistêmica Crítica, proposto por Ulrich [2002], para que se possa entender melhor a GovTI no setor público. O pensamento sistêmico crítico pode ser utilizado como método de análise de ambientes e relações que envolvem atores que estão em posições de desigualdade diante de uma dada situação. O teor do pensamento sistêmico leva em consideração questões que envolvem diversos fatores ou variáveis a partir de padrões organizados de interações. Isto significa dizer que concepções teóricas e princípios que procuram explicar entidades, fenômenos e situações, cujo entendimento não pode ser compreendido em sua totalidade pelo pensamento analítico.

A Heurística Sistêmica Crítica contempla tanto os atores que são envolvidos, os que têm maior poder de decisão, assim como os atores que são afetados, os que suportam as decisões tomadas pelos envolvidos. A compreensão de ambiente complexo se dá quando são entendidas as relações entre atores envolvidos e afetados [Ulrich, 2002]. Nesse sentido, o mapa de situação de GovTI é proposto para descrever os atores envolvidos e afetados, o tipo de participação que eles apresentam no âmbito do SISP e os principais dispositivos legais normativos emanados – vide Quadro 2. Adicionalmente, o mapa de contexto de GovTI possibilita avaliar as principais influências emanadas dos atores, segundo a força indutora presente em cada perspectiva contextual – ambiental, social, legal, tecnológico, política e institucional, conforme descrito no Quadro 3.

Quadro 2. Mapa de Situação de GovTI no SISP

Ator [Papel]	Tipo de Participação	Principais Artefatos Produzidos
--------------	----------------------	---------------------------------

TCU/SEFTI [envolvido]	Órgão Regulador / Fiscalizador	Acórdão 1.603/2008-TCU-Plenário, Acórdão 2.471/2008, Acórdão 2.308/2010-TCU-Plenário, Índice de Governança de TI (iGovTI) 2010, Acórdão 1233/2012, Acórdão 2585/2012, Acórdão 3051/2014, Nota Técnica 7/2014
Ministério do Planejamento [envolvido]	Órgão Regulador	Portaria nº 68, de 7 de Março de 2016, Portaria nº 290, de 29 de setembro 2016, Estratégia de Governança Digital 2016 – EGD, Portaria nº 19, de 29 de maio de 2017
SISP [envolvido/ afetado]	Órgão Gerenciador	Portaria MP 208/2006, Metodologia de Gerenciamento de Portfólio de Projetos do SISP 2011, Guia de Comitê de TI do SISP 2013, Guia de boas práticas em contratação de soluções de TI 2014, Guia de PDTIC do SISP 2016, Guia de Governança de TIC do SISP 2017
SETIC [envolvido]	Órgão Regulador	Instrução normativa nº 4, de 11 de setembro de 2014, Plano Diretor de Tecnologia da Informação (PDTI), Estratégia de Governança Digital 2016
Governo Federal [envolvido]	Órgão regulador	Artigo 165 da Constituição Federal 1988, Decreto 3.751/2001, Decreto nº 3.751, de 15 de fevereiro de 2001, Decreto nº 5.151, de 22 de julho de 2004, Decreto nº 5.482, de 30 de junho de 2005, Decreto 5.707/2006 Decreto de 15 de setembro de 2011, Lei nº 12.527, de 18 de novembro de 2011, Decreto nº 8.638 de 15 de Janeiro de 2016, Lei nº 4.320/64, Projeto de Lei Orçamentária Anual (LOA)
GSI/DSIC [envolvido]	Órgão regulador	Decreto no 3.505, de 13 de junho de 2000, NC à IN Nº 01 GSI/PR/2008, Guia básico de orientações ao gestor em segurança da informação e comunicações 2015, Estratégia de SIC da Administração Pública Federal

Fonte: Autoria própria [2018]

Quadro 3. Mapa de Contexto de GovTI no SISP

Contexto	Atores/ Artefatos	Impacto
		Influência do contexto sobre o sistema de TI?
Ambien- tal Força: TI Verde	SETIC - Instrução normativa nº 4, de 11 de setembro de 2014	Impacta a escolha e definição dos recursos tecnológicos a empregar. Exerce baixo impacto , havendo apenas recomendações, sem caráter punitivo ou coercitivo Efeito positivo: - <i>Redução de gases, baixo consumo de energia, reutilização de materiais.</i>
	SISP - Guia de boas práticas em contratação de soluções de TI	Ameaças: - <i>Aumento do aquecimento global, falta de matéria-prima, alto custo de energia, insustentabilidade ambiental.</i> Implicações: - <i>Incorporar requisitos ambientais na formação das políticas e planos da TI</i> - <i>Necessidade de adequação de sistemas e tecnologias da informação aos preceitos ambientais</i>
Social Força: <i>Transpa- rência (e- Gov)</i>	APF - Lei nº 12.527, de 18 de novembro de 2011 Decreto nº 5.482, de 30 de junho de 2005 - Decreto de 15 de setembro de 2011 - Lei nº 4.320/64	Prestações de contas das ações empreendidas pela TI. Desenvolvimento e disponibilização de serviços eletrônicos. Exerce alto impacto Efeito positivo: - <i>Melhoria da percepção das ações da TI</i> <i>Aumento da credibilidade da TI</i>
	SETIC Estratégia de Governança Digital –	Oportunidades: - <i>Tecnologias emergentes para tratar grandes volumes de dados.</i> <i>Facilidade de acesso dos serviços públicos.</i> <i>Sociedade requer transparência, prestação de contas, cada vez mais.</i> Ameaça: <i>Falta de mão obra capacitada.</i> <i>Falta de recurso financeiro para contratações/manutenção de</i>

	EGD	<i>equipamentos de TI.</i>
Legal Força: <i>Regulamentação da TI</i>	Governo Federal Artigo 165, Constituição Federal de 1988 Decreto 3.751/2001 Decreto nº 5.482/2005 Decreto 5.707/2006 Lei nº 12.527/ 2011 Decreto nº 8.638/2016 Lei nº 4.320/64 LOA	Regras e normas que a TI deve obedecer e cumprir. Exerce alto impacto Efeito positivo: - <i>Aumento da credibilidade da TI.</i> Implicação: - <i>Obriga o sistema de TI se adequar as Leis, Decretos e Normas.</i> - <i>Medidas de cobrança nem sempre tem um resultado positivo, se o próprio governo não prover meios para que o sistema de TI cumpra os requisitos legais.</i> Oportunidade: - <i>Aprimorar os serviços da TI</i> - <i>Falta de competências da TI para atender as novas demandas.</i>
Tecnológico Força: <i>Tecnologias emergentes, inovadoras, habilitadoras</i>	SETIC - EGD GSI/ DSIC - Estratégia de SIC da Administração Pública Federal 2015 – 2018 Guia básico de orientações ao gestor em segurança da informação e comunicações 2015	Desenvolvimento e disponibilização de serviços eletrônicos. Implementação da Segurança de Informação. Exerce médio impacto Efeito Positivo: - <i>Melhoria nos serviços oferecidos pela TI</i> - <i>Maior segurança dos serviços.</i> Oportunidades: - <i>Grande disponibilidade de novas tecnologias.</i> - <i>Custo menor.</i> - <i>Fragilidade na segurança de serviços da TI.</i> Ameaças: - <i>Custo alto de novas tecnologias.</i> - <i>Mão de obra desqualificada.</i>
Política Força: <i>Diretrizes gerais para a TI</i>	MCTIC - Estratégia Brasileira para a Transformação Digital SETIC - EGD GSI/ DSIC - Decreto 3.505/2000	Políticas públicas que viabiliza o sistema de TI. Exerce alto impacto Positivo: - <i>Meio de fomentação do sistema de TI</i> Oportunidade: - <i>Crescimento de novas tecnologias</i>
Institucional Força: <i>Transparência</i>	Governo Federal - Instrução normativa nº 4/ 2014 - Lei nº 4.320/64 (LOA)	Instruções quanto ao financeiro. Exerce alto impacto. Efeito Positivo: - <i>Aumento da credibilidade da TI</i> - <i>Transparência dos gastos da TI</i>

Fonte: Autoria própria [2018]

5.2 Subsistemas de TI no Setor Público

Como vimos, o sistema de TI é composto de 3 (três) subsistemas, Direcionamento, Desenvolvimento e Entrega. Para cada subsistema de TI pode-se aplicar os níveis de desempenho Organização, Processos e Implementação. A interação desses subsistemas e os níveis de desempenho possibilita o alinhamento da TI com ela mesma e com a organização.

Desse modo optou-se por utilizar esse modelo, pois abrange toda organização de TI. Molinaro e Ramos [2013] descrevem os objetivos de controle de modo a verificar se o modelo atual da TI abrange ou não os subsistemas de TI. A partir dos artefatos que foram levantados,

aplicamos os objetivos e controles para verificar se o modelo atual da governança de TI presente no setor público contempla os subsistemas de TI e os seus níveis de desempenho, presentes nos quadros 4, 5 e 6.

Quadro 4. Direcionamento – Objetivos de Controle

Objetivos de Controles	Controles
Financeiro	<p>O sistema de TI possui um processo formal de gerenciamento financeiro? <i>Sim.</i> Lei no 4.320, de 17 de março de 1964, institui Normas Gerais de Direito Financeiro para elaboração e controle dos orçamentos e balanços da União, dos Estados, dos Municípios e do DF. Lei complementar nº 101, de 4 de maio de 2000 – Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. LDO – Lei de Diretrizes Orçamentárias: compreende as metas e prioridades da APF. PPA – instrumento previsto no art. 165/CF destinado a organizar e viabilizar a ação pública, com vistas a cumprir os fundamentos e os objetivos da República.</p> <p>O sistema de TI possui um sistema automatizado para gerenciamento financeiro? <i>Sim.</i> SIAFI - Sistema Integrado de Administração Financeira do Governo Federal.</p>
Priorização	<p>O sistema de TI possui um processo formal de priorização dos investimentos de TI? <i>Sim.</i> Acórdão 1603/2008 – TCU - Aborda questões de priorização dos investimentos. Acórdão 1233/2012 – TCU - Aborda questões de priorização dos investimentos. Estratégia de Governança Digital 2016 – Definem as estratégias, conceitos, desafios, oportunidades, objetivos estratégicos, as metas que norteará o sistema e TI. Instrução normativa nº 4, de 11 de setembro de 2014 – Estabelece a obrigatoriedade do órgão em aprovar periodicamente o plano de metas e ações no PDTI. Guia de Governança de TI do SISP – “Prática 02 – Especifica os direitos decisórios sobre TIC”</p> <p>O sistema de TI possui um sistema automatizado para priorização dos investimentos de TI? <i>Não.</i> Fica a cargo de cada órgão da APF.</p>
Orçamentário	<p>O sistema de TI possui um processo formal para preparação e acompanhamento orçamentário? <i>Sim.</i> Lei no 4.320, de 17 de março de 1964, estatui Normas Gerais de Direito Financeiro para elaboração e controle dos orçamentos e balanços da União.</p> <p>O sistema de TI possui um processo formal para preparação e acompanhamento das contratações? <i>Sim.</i> Acórdão 2.471/2008- Plenário – Trata de pontos a observar nas contratações de soluções de TI. Instrução normativa nº 4, de 11 de setembro de 2014 – Dispõe sobre o processo de contratação de Soluções de Tecnologia da Informação pelos órgãos integrantes do SISP. Guia de boas práticas em contratação de soluções de TI 2014 elaborado pelo SISP.</p> <p>O sistema de TI possui um sistema automatizado de acompanhamento da execução orçamentária? <i>Sim.</i> SIAFI – Sistema Integrado de Administração Financeira do Governo Federal.</p>
Custos	<p>O sistema de TI possui um processo formal de acompanhamento de custos? <i>Sim.</i> Art. 50, § 3º da Lei Complementar nº 101/2000 – Lei de Responsabilidade Fiscal (LRF).</p> <p>O sistema de TI possui um sistema automatizado para acompanhamento de custos? <i>Sim.</i> SIC – Sistema de Informações de Custos do Governo Federal – SIC. Observatório da Despesa Pública do Ministério da Transparência e Controladoria-Geral da União.</p>
Benefícios	<p>O sistema de TI possui um processo formal de análise dos benefícios dos investimentos? <i>Sim.</i> Acórdão 1603/2008 – TCU - Aborda questões de análise dos benefícios dos investimentos. Acórdão 1233/2012 – TCU - Aborda questões de análise dos benefícios dos investimentos. Dispõe de mecanismos: Guia de Governança de TI cita na “Prática 05 – Portfólio de TIC - Capacidade de mensuração dos benefícios dos investimentos realizados em TIC.”</p> <p>O sistema de TI possui um sistema automatizado de acompanhamento dos benefícios dos investimentos? <i>Sim.</i> SIC – Sistema de Informações de Custos do Governo Federal.</p>
Segurança	<p>O sistema de TI possui um processo formal de segurança da informação? <i>Sim.</i> Lei nº 12.527, de 18 de novembro de 2011 – regula o acesso a informações. Instrução Normativa GSI Nº 1, de 13 de junho de 2008 – disciplina a Gestão de Segurança da Informação e Comunicações na APF, direta e indireta, e dá outras providências. Instrução Normativa GSI Nº 2, de 5 de fevereiro de 2013 – dispõe sobre o Credenciamento de segurança para o tratamento de informação classificada, em qualquer grau de sigilo, no âmbito do Poder Executivo Federal.</p>

	Instrução Normativa GSI Nº 3, de 6 de março de 2013 – dispõe sobre os parâmetros e padrões mínimos dos recursos criptográficos baseados em algoritmos de Estado.
--	--

Fonte: Autoria própria [2018]

Quadro 5. Desenvolvimento – Objetivos de Controle

Objetivos de Controles	Controles
Alteração de escopo	O sistema de TI possui um processo formal para autorizar alterações de escopo no projeto? Sim, presente no artefato Metodologia de Gerenciamento de Projetos do SISP.
	O sistema de TI possui um sistema automatizado para acompanhar as alterações de escopo do projeto? Fica a cargo do órgão da APF definir a metodologia e as ferramentas de Gestão de Projeto, incluindo a metodologia de desenvolvimento de software.
	O sistema de TI possui um processo formal para rastrear os relacionamentos das atividades a serem realizadas? <i>Sim</i> , presente no artefato “Metodologia de Gerenciamento de Projetos do SISP”.
	O sistema de TI possui um sistema automatizado para rastrear os relacionamentos das atividades a serem realizadas? Fica a cargo do órgão da APF definir a metodologia e as ferramentas de GP, incluindo a metodologia de desenvolvimento de software.
	O sistema de TI possui um processo formal para avaliar o impacto da alteração de escopo proposta? <i>Sim</i> , presente no artefato “Metodologia de Gerenciamento de Portfólio de Projetos do SISP”.
	O sistema de TI possui um sistema automatizado para mensurar o impacto da alteração de escopo proposta? Fica a cargo do órgão da APF definir a metodologia e as ferramentas de GP, incluindo a metodologia de desenvolvimento de software.
Erro de quantificação	O sistema de TI possui um processo formal para estimar o trabalho a ser realizado? Sim, presente no artefato Metodologia de Gerenciamento de Projetos do SISP.
	O sistema de TI possui um sistema automatizado para gerenciamento de estimativa? Fica a cargo do órgão da APF definir a metodologia e as ferramentas de GP, incluindo a metodologia de desenvolvimento de software.
Erro de produtividade	O sistema de TI possui um processo formal para analisar a produtividade do trabalho a ser realizado? <i>Não</i> .
	O sistema de TI possui um sistema automatizado para gerenciar a produtividade? <i>Fica a cargo do órgão da APF definir um sistema automatizado para gerenciar a produtividade.</i>

Fonte: Autoria própria [2018]

Quadro 6. Entrega – Objetivos de Controle

Objetivos de Controles	Controles
Acesso	O sistema de TI possui um processo formal para requisitar, estabelecer conceder contas de usuários? <i>Sim</i> . Presente no Guia Básico de Orientações ao Gestor em Segurança da Informação e Comunicações elaborado pelo Departamento de Segurança da Informação e Comunicações.
	O sistema de TI possui um sistema automatizado para mapear contas de usuários e acessos de usuários não autorizados? <i>Fica a cargo do órgão da APF definir a metodologia de gestão de acessos de usuários.</i>
	O sistema de TI possuem papéis com responsabilidades bem definidas? <i>Sim</i> . Lei nº 11.907, de 2 de fevereiro de 2009, que define papeis e responsabilidades de uma maneira geral do profissional de Tecnologia da Informação.
Mudança	O sistema de TI registra e avalia as suas taxas de mudanças realizadas com sucesso? <i>Fica a cargo do órgão da APF registrar e avaliar as taxas de mudanças realizadas com sucesso.</i>
	O sistema de TI registra e avalia periodicamente o numero de mudanças autorizadas e implementadas? <i>Sim</i> . IN/STI/MPDG nº 4/2014, previsto no processo de gestão e fiscalização processos/contratos.

	<p>Existem consequências definidas para mudanças intencionais não autorizadas? <i>Sim.</i> Acórdão 2.471/2008- Plenário – Trata de pontos a observar nas contratações de soluções de TI.</p> <p>IN/STI/MPDG n° 4/2014 – Dispõe sobre o processo de contratação de Soluções de Tecnologia da Informação pelos órgãos do SISP.</p>
	<p>O sistema de TI utiliza a experiência de taxas de mudança com sucesso para prevenir proativamente potenciais riscos de mudança? <i>Não.</i></p>
Configuração	<p>O sistema de TI possui um processo formal para gerenciamento de configuração de TI? Acórdão 2613/2011 – Trata de Gestão de Serviços e contempla a gestão de configuração.</p>
	<p>O sistema de TI possui um sistema automatizado para gerenciamento de configuração? <i>Fica a cargo do órgão da APF.</i></p>
	<p>O sistema de TI está preparado para fornecer aos profissionais de TI que tenham acesso privilegiado às informações corretas e precisas das configurações atuais da infraestrutura de TI, incluindo as suas especificações físicas e funcionais? <i>Não.</i></p>

Fonte: Autoria própria [2018]

5.3 Discussão

O objetivo principal que direciona esse trabalho é mapear os elementos essenciais para estruturação de um modelo de governança de TI no setor público, pautado pelo princípio da sustentação.

Não se pretendia fazer um levantamento que fosse focado somente em um órgão da Administração Pública Federal, pelo fato da pesquisa querer retratar o modelo de governança de TI abrangendo a APF como um todo, por isso o universo de pesquisa foi os órgãos pertencentes ao SISP.

A partir da aplicação dos mapas de situação e de contexto de GovTI no âmbito do SISP, foi possível organizar todos os documentos levantados, possibilitando uma melhor compreensão do ambiente organizacional inerente ao setor público que se encerra no modelo atual de GovTI.

Com a aplicação dos objetivos de controles proposto por Molinaro e Ramos [2013], permitiu a verificação do sistema de TI do setor público, se estão contemplando os três subsistemas de TI definidos por Betz [2011], Direcionamento, Desenvolvimento e Entrega, e os níveis de desempenho definidos por Rummler e Branche [1994] Organização, Processo e Implementação, nas suas integridades.

Com essa aplicação permitiu diagnosticar que o setor público está mais presente nos subsistemas de direcionamento e no nível de desempenho da Organização, significa que está mais preocupado em definir as características gerais do negócio. Com isso, cada Órgão da APF executa as atividades pertencentes aos subsistemas de desenvolvimento e entrega de e os níveis de desempenho de processo e implementação de modo assimétrico, desencadeando uma grande variabilidade em termos de arranjo arquitetural da TI.

De um modo geral, o setor público federal dispõe de uma estrutura de governança de TI definida, ainda que não contemple integralmente os subsistemas de TI. Mas, a partir dos métodos de análise empregados nesta pesquisa, foi possível detectar a insatisfatoriedade dos elementos essenciais para estruturação de um modelo de governança de TI adequado. Permitiu evidenciar que os esforços nem sempre estão alinhados com o órgão central do SISP, a exemplo da documentação que define os modelos de GovTI de cada órgão. Por isso, estes acabam executando suas ações e aplicando os instrumentos e mecanismos de GovTI de modo discricionária, devido à falta de orientação adequada em termos de direcionamento, desenvolvimento e entrega.

Finalmente, ao resgatar a pergunta de pesquisa que deu origem a este estudo, verifica-se que o modelo atual de governança de TI não contempla satisfatoriamente os três

subsistemas de TI e seus níveis de desempenho, impossibilitando promover a TI de forma sustentada.

6. Conclusão

Este estudo amplia a noção de governança para a perspectiva dos subsistemas de TI com o propósito de avaliar o modelo de governança de TI no âmbito do SISP. Além disso, permitiu identificar óbices e lacunas que comprometem o nível de satisfatoriedade em termos de desempenho, especialmente os níveis de processo e implementação.

Não há um marco regulatório de GovTI devidamente estruturado no âmbito do SISP, uma vez que os dispositivos legais e normativos estão fragmentados e não refletem o sistema de TI na sua integralidade. Conquanto que os métodos de mapeamento aplicados permitiu descrever o ambiente que compreende o sistema de GovTI, revelando um quadro repleto de inconsistências.

A configuração atual do modelo de GovTI do SISP é pautada pelos preceitos de planejamento, fiscalização e controle, contudo não contempla os requisitos e os mecanismos que orientam devidamente os subsistemas de direcionamento, desenvolvimento e entrega e os níveis de desempenho corporativo, de processos e de implementação.

Em síntese, ao compreender o sistema de governança de TI na APF, em particular, e por extensão ao setor público como um todo, é esperado que os subsistemas e os níveis avaliados possam ser tomados como referência para assegurar a efetividade organizacional.

Referências

- AUGUST, I. *Governance in the Public Sector: A Governing Body Perspective*, 2001.
- BETZ, Charles T. *Architecture and patterns for IT service management, resource planning, and governance: Making shoes for the cobbler's children*. Elsevier, 2011.
- BOWEN, P.; CHEUNG, M.; ROHDE, F. (2007). Enhancing IT governance practices: A model and case study of an organization's efforts. *International Journal of Accounting Information Systems*, v. 8, n. 3, p. 191-221, set.
- BPTRENDS. (2017). *BPM Analysis, Opinion and Insight*. Disponível em: <<https://www.bptrends.com/>>. Acesso em: 10 out 2017.
- CAMERON, Kim S. (1986). Effectiveness as paradox: Consensus and conflict in conceptions of organizational effectiveness. *Management science*, 32.5: 539-553.
- CAVALIERI, A.; MACEDO, S. (2004). *Avaliando o desempenho da universidade*. Editora PUC-Rio.
- CUNHA, M.; NETO, J. (2014). *Comitês de governança de tecnologia da informação na administração pública federal brasileira: fatores críticos de sucesso*.
- DAMELIO, R. (1996). *The basics of process mapping*. New York: Productivity Press.
- ISACA (2011). *Global Status Report on the Governance of Enterprise IT (GEIT)*.
- ITGI. (2007). *Cobit 4.1: framework, control objectives, management guidelines, maturity models*. 213p.
- MOLINARO, L.; RAMOS, K. (2013). *Gestão de Tecnologia da Informação, Governança de TI: Arquitetura e Alinhamento Entre Sistemas de Informação e o Negócio*. LTC, 2013.

- PARHIZGARI, A.; GILBERT, G. (2004). Measures of organizational effectiveness: private and public sector performance. *Omega: The International Journal of Management Science*, Vol. 32, No. 3, p. 221-229.
- RODRIGUES, J. G. L.; et al. (2010). Diretrizes para implantação da governança de TI no setor público brasileiro à luz da teoria institucional.
- RUMMLER, G.; BRACHE, A. (1994). *Melhores desempenhos das empresas*. Makron Books.
- SILVA, B. A. M., MORAES, G. (2011). Influência dos direcionadores do uso da TI na Governança de TI. *Revista Brasileira de Gestão e Negócios*, São Paulo, v13, n.38, p 41-60.
- TCU. Governança de TI: entendendo a governança de TI. Tribunal de Contas da União. Governança de TI. Disponível em: <<http://portal.tcu.gov.br/comunidades/governanca-de-ti/entendendo-a-overnanca-de-ti/>> Acesso em: 29 set. 2017.
- ULRICH, W. (2002). Critical Systems Heuristics. In: Daellenbach, H.G. *The Informed Student Guide to Management Science*. London: Thomson Learning.
- WEILL, P.; ROSS, J. (2006). *Governança de TI: Tecnologia da Informação*. Makron Books.

Avaliação de Metodologia para Implementação do BPM

Carlos Hernane de Oliveira¹, Nivaldo Pereira de Morais Júnior², Eliomar Araújo de Lima³

^{1,2,3}Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74.690-900 – Goiânia – GO – Brazil
{carloshernane,nivaldo,eliomar}@inf.ufg.br

Abstract. *This research intends to evaluate the level of learning achieved with the application of methodology based on the discipline of Business Process Management (BPM). The study was developed through an exploratory analysis of a set of companies, located in the State of Goiás, that underwent a process of intervention at the level of consulting or training to improve business processes. In order to do so, an evaluation tool was structured based on factors considered critical for success in the implementation of business processes in the corporate scope. The results allowed to reveal an acceptance framework favorable to the assimilation of the applied methodology, but with a strong tendency of unsatisfactory in terms of support and accommodation of the planned practices and actions. Keywords: Business process management. Processes improvement. Critical success factors.*

Resumo. *Este trabalho pretende avaliar o nível de aprendizado alcançado com a aplicação de metodologia baseada na disciplina de Gestão de Processos de negócios (BPM). O estudo se desenvolveu por meio de análise exploratória sobre um conjunto de empresas, instaladas no Estado de Goiás, que passaram por um processo de intervenção em nível de consultoria ou capacitação para melhoria de processos de negócio. Para tanto, estruturou-se um instrumento de avaliação a partir de fatores apontados como críticos para o sucesso na implementação de processos de negócio no âmbito corporativo. Os resultados permitiram revelar um quadro de aceitação favorável à assimilação da metodologia aplicada, mas com forte tendência de insatisfação em termos de sustentação e acomodação das práticas e ações planejadas. Palavras-chave: Gestão de processos de negócio. Melhoria de processos. Fatores críticos de sucesso.*

1. Introdução

Vive-se uma explosão tecnológica sem precedentes na história da humanidade. Uma das questões de ordem no âmbito corporativo é a efetividade organizacional para geração de valor. Para obtê-la é necessário haver um posicionamento pautado pela capacidade de assimilar mudanças e inovar em seus processos, serviços e negócios. [Osterwalder and Pigneur 2010] sugere que para inovar é necessário confiança, diversidade, tempo disponível, troca de ideias e certo grau de informalidade, ou seja, não é tão simples assim, as variáveis são muitas. Essa é uma época onde as grandes inovações não serão em produtos, porém na geração de valor.

Paradoxalmente, ainda há uma forte tendência de as empresas adotarem o modelo de gestão orientado à abordagem funcional que, de certo modo, contraria a ideia de inovação e promoção de mudanças organizacionais. Nesse modelo, a gestão é baseada nas

divisões departamentais em que cada função está associada um cargo que busca potencializar seu próprio ganho em detrimento de uma preocupação com a organização como um todo. Já no modelo de gestão centrado em processos, a mudança decorre da adoção de um posicionamento que privilegia a visão com foco no cliente, contando com a integração tecnológica e a revisão de competências. Embora a estruturação funcional continue válida, a especialização por processos leva à produtividade, a geração de valor passa a ser gerenciada horizontalmente em uma visão notadamente interfuncional ponta a ponta [CBOK 2013].

A qualidade competitiva de uma empresa vem sendo determinada pela sua capacidade de aprender rapidamente, de gerar conhecimento aplicado à inovação de processos e de gerenciar suas competências para desenvolver melhores produtos, serviços e negócio [Antonello et al. 2005].

A busca constante pelo alcance dos objetivos estratégicos tem gerado nas organizações uma preocupação em melhorar continuamente seus processos de negócio. Esse contexto gera a necessidade de respostas rápidas aos mercados onde atua e para que se possa desenvolver essa capacidade é esperado que existam práticas de gestão baseada em processos de negócio.

Este trabalho se insere nesse contexto para tratar da melhoria de resultados organizacionais propiciados pelo aprimoramento de seus processos de negócio. O desafio é buscar uma maior compreensão dos resultados obtidos com o uso da disciplina Gestão de Processos de Negócio (do inglês, *Business Process Management* - BPM). As empresas precisam se modernizar para melhorar os resultados e sobreviver. Existem disciplinas e métodos que possibilitam desenvolvimento de processos, que servem como apoio para a mudança e aprimoramento, entretanto, o uso desses recursos nem sempre é simples de se aplicar, e em uma parte considerável das iniciativas de mudanças, os resultados esperados não são alcançados [Osterwalder and Pigneur 2010].

O objetivo aqui é expor os conceitos acerca da disciplina de BPM e avaliar o nível de aprendizado alcançado por meio da aplicação prática de uma metodologia baseada em BPM. Antes porém, será promovida uma breve caracterização do modelo de gestão baseado em processos de negócio.

Para tanto, em busca dessa avaliação metodológica de BPM, um estudo empírico é empreendido junto a empresas que passaram por um processo de intervenção para melhoria de processos de negócio.

2. Fundamentos de Processos de Negócio

Para compreender a disciplina de gestão de processos de negócio é necessário compreender o significado de processo de negócio. No contexto de BPM, um processo de negócio é um trabalho que entrega valor aos clientes, ou apoia, ou gerencia outros processos. A noção de trabalho ponta a ponta interfuncional é chave, pois envolve todo o trabalho, cruzando os limites funcionais necessários para entregar valor para clientes [CBOK 2013].

Processo é uma cooperação de atividades e recursos distintos voltados à realização de um objetivo global, orientado para o cliente final, que é comum ao processo ou ao produto/serviço [Zarifian 1997]. Outra forma de pensar a respeito de processos é apresentada pela equipe do Centro de Ciências da Coordenação do MIT, que considera os

processos como sequências semi-repetitivas de eventos que, geralmente, estão distribuídas de forma ampla no tempo e espaço, possuindo fronteiras não muito bem definidas [Feldman and Pentland 2003].

Assim, um processo dispõe de entradas, saídas, duração, espaço, ordenação, objetivos e valores que, interligados logicamente, irão resultar em uma estrutura para fornecer produtos ou serviços ao cliente, que se pode denominar de "entrega de valor ao cliente". Sua compreensão é importante, afinal uma organização é tão efetiva quanto os seus processos, já que eles são responsáveis pelo que será ofertado ao cliente [JOHANSSON et al. 1995].

Ao analisar um processo, a equipe de projeto deve partir sempre da perspectiva do cliente (interno ou externo), de forma a atender às suas necessidades e preferências, ou seja, quase sempre, o processo começa e termina no cliente.

Portanto, o BPM é uma abordagem voltada à gestão por processos contemplando as atividades de modelar, documentar, simular, automatizar, medir, monitorar e melhorar processos de negócio, automatizados ou não, para alcançar resultados consistentes e alinhados com os objetivos estratégicos da organização [CBOK 2009].

O CBOK define BPM [CBOK 2013] como:

Gerenciamento de Processos de Negócio (BPM - Business Process Management) representa uma nova forma de visualizar as operações de negócio que vai além das estruturas funcionais tradicionais. Essa visão compreende todo o trabalho executado para entregar o produto ou serviço do processo, independente de quais áreas funcionais ou localização estejam envolvidas. Começa em um nível mais alto do que o nível que realmente executa o trabalho e, então, subdivide-se em subprocessos que devem ser realizados por uma ou mais atividades. Nessa visão, o gerenciamento pode envolver indicadores de desempenho de processos de significância para cada um dos componentes do processo, permitindo às partes do negócio uma base de comparação em termos de tempo, custo, capacidade e qualidade com outros provedores internos e externos.

2.1. Organização Funcional versus Organização por processos

O modelo funcional está baseado na departamentalização, onde cada departamento possui uma função. Esses departamentos são individualizados. Mais de um século se passou desde a revolução industrial e o Taylorismo continua a ser o grande modelo de gerenciamento organizacional: áreas funcionais especializadas no assunto, promovendo uma cadeia de relações clientes-fornecedores internos à organização, de forma que o trabalho das unidades funcionais resulte em um produto demandado pelo cliente final.

A visão tradicional é voltada a núcleos de especialidade dentro da organização, cujos esforços somados resultarão em clientes tendidos.[Scucuglia 2011] assevera, no entanto, que desde meados da década de 90 este paradigma gerencial começou a ser questionado. Isto porque ele carrega consigo alguns problemas que poderiam ser resolvidos de maneira mais aperfeiçoada se a ótica de gestão fosse diferente. Não são incomuns conflitos internos entre os departamentos, já que cada área possui sua própria percepção da continuidade do processo.

A Gestão por Processos (BPM) consiste essencialmente em uma quebra deste paradigma funcional, propondo uma visão interfuncional de como os processos ponta a ponta poderiam ser melhor geridos visando eliminar os efeitos ruins. Trata-se fundamentalmente de uma mudança filosófica da maneira de gerenciamento da organização. É preciso dentro desse novo modelo de trabalho reavaliar os velhos processos e adaptá-los, para entregar o mesmo produto ou serviço, de maneira mais eficiente, aumentando a entrega de valor de cada atividade envolvida no processo.

A análise e a melhoria de processo de negócio é uma técnica de benefícios duplos, ela permite que se obtenha resultados a curto e em longo prazo, identificando oportunidades para melhorias importantes e ao mesmo tempo serve de base para outras técnicas contábeis que podem manter a empresa no caminho de aperfeiçoamento permanente [Ostrega et al. 1993].

Empresas que ainda utilizam sistemas tradicionais de organização e gestão por funções perdem em agilidade e competitividade quando comparadas com empresas que são gerenciadas por processos, modelo este, caracterizado fundamentalmente por valorizar a estrutura horizontal, ser interdepartamental e basear-se na divisão de responsabilidades, comunicação e transferência de informações [Gonçalves 2000]. A figura 1 mostra as diferenças entre estas duas visões.

Características	Visão Funcional	Visão Por Processos
Distribuição dos funcionários	Isolados juntos aos seus pares em departamentos	Times de processos multitarefas
Tomada de decisão	Rígida supervisão hierárquica	Autonomia para tomada de decisão
Avaliação de desempenho	Desempenho medido por indicadores funcionais	Desempenho medido pelos resultados do processo de negócio
Estrutura Hierárquica	Rígida supervisão hierárquica	Colaboração e envolvimento
Capacitação dos funcionários	Especialização voltada para assuntos do departamento	Capacitação em múltiplas competências
Estrutura organizacional	Departamentalizada e desagregada	Transversalidade e sinergia
Valores organizacionais	Competição e falta de confiança entre os funcionários	Transparência e pró-atividade por parte e entre os funcionários
Indicadores de desempenho	Medem o desempenho dos departamentos	Medem o desempenho dos processos
Tipo do trabalho	Repetitivo e burocrático	Variado e complexo
Estrutura do trabalho	Constante e moroso	Sistêmico e evolutivo
Relacionamento externo	Quase inexistente. O que interessa é o ambiente interno	Incentivo à colaboração com entes externos através dos processos
Utilização da área de Tecnologia da Informação e Comunicação	Sistemas desintegrados com visão funcional	Sistemas Integrados através dos processos

Figura 1. Gestão por processos versus funcional. Fonte: Adaptado de De Sordi (2012)

Na visão de [Capote 2011a] a transformação da gestão corporativa tradicional para a gestão por processos se dá a partir do momento em que a empresa que adota BPM, gerencia seus processos com o intuito de buscar o melhor resultado geral na realização de suas atividades, independentemente de onde suas atividades são realizadas, afinal não importa o setor, mas o resultado do processo. O foco na entrega deve estar no cliente. Os processos devem contribuir na entrega de valor a esses mesmos clientes.

2.2. Propostas de Aplicação de BPM em Diferentes Contextos

Em seu estudo [Mariano 2012], assevera que o gerenciamento de processos é importante para aprimorar o desempenho de processos. Apresenta um relato de uma aplicação da metodologia BPM (Business Process Management) adaptada para o contexto de uma organização pública. A partir da análise do processo corrente foram propostas melhorias a fim de otimizar recursos, melhorar a confiabilidade e aumentar a satisfação dos cidadãos.

No trabalho de [Torres 2015] é apresentada uma proposta de modelo estendido para operacionalização do Business Process Management (BPM), observando fatores culturais e interpessoais. Para atingir o objetivo, o autor promoveu uma revisão da literatura explorando a relação entre os pressupostos culturais básicos das organizações e as novas tecnologias de gestão, para em seguida levantar os Fatores Críticos de Sucesso (FCS) por meio de um estudo de caso. Finalmente, propôs um modelo estendido para operacionalização do BPM baseado nos novos FCS identificados.

Para [Pizza 2012] o foco está na utilização da Metodologia BPM (Business Process Management) e de suas ferramentas no âmbito corporativo. Isso permite que, através de uma execução e de um controle mais eficazes, processos possam ser melhorados em qualquer uma destas áreas, a saber: setor industrial, governamental, comercial, entre outros. O autor sugere que BPM é estatisticamente um investimento seguro e de rápido retorno.

Já no trabalho de [for double-blind review] avalia-se o uso do BPM para melhoria dos processos na construção civil. Ressalta-se a importância dos processos de negócio para as organizações e apresenta a disciplina BPM com suas principais características. Discute-se conceitos e métodos que são fundamentais para aplicação do BPM, tais como: gestão de indicadores, redução de desperdícios com a produção e construção enxuta e notação de modelagem. Os resultados obtidos são confrontados com os fatores críticos de sucessos para iniciativas de BPM e apontam para a viabilidade do uso do método proposto, possibilitando constatar a aplicação da gestão de processos de negócio em um ambiente real de negócios da construção civil.

3. Metodologia

A natureza da pesquisa aplicada neste trabalho é exploratória, em que se caracteriza quando há pouco conhecimento sobre a temática a ser abordada [Gil 2008]. É por meio deste estudo que se pretende conhecer com maior profundidade as unidades de análise de interesse para possibilitar a avaliação do nível de aprendizado na apropriação de melhorias de processos de negócio.

Uma característica da pesquisa exploratória é esclarecer as questões superficialmente abordadas, definindo um aprofundamento dos conceitos preliminares

[de Andrade 2000], ressalta algumas finalidades primordiais, como: proporcionar maiores informações sobre o assunto; facilitar a delimitação do tema da pesquisa; orientar a fixação dos objetivos e formulação das hipóteses; ou descobrir um novo tipo de enfoque sobre o assunto.

As empresas correspondentes às unidades de análise deste estudo implementaram parcial ou integralmente a metodologia desenvolvida por [Morais Junior 2014], doravante denominada MI-BPM, que se baseou não só na disciplina BPM, como também nas abordagens Balanced ScoreCard e Lean Construction. A metodologia está organizada em um ciclo de aplicação contendo fases e etapas. Cada fase compreende um conjunto de etapas e estas correspondem a grupos de atividades ou atividades. Segundo [Morais Junior 2014],

As fases da metodologia em questão são direcionadas por um ganho central que é definido junto com os responsáveis pela organização e pelo processo na primeira etapa do trabalho - Entendimento do Negócio. Este ganho, invariavelmente, está relacionado com um problema central que se espera resolver ou minimizar com a melhoria dos processos. A definição do ganho parte da definição de uma hipótese que envolve a melhoria do processo para a solução do problema. Durante todas as fases e etapas do ciclo do método é este ganho que direciona as ações e esforços, orienta a análise e as melhorias, delimita o escopo para desenho dos métodos e ampara a construção e execução do plano de ação de transformação.

A metodologia está compreendida em 5 fases - Tabela 1. Na primeira fase busca-se o entendimento do negócio. Há a preparação de todos os instrumentos que serão utilizados em todas as fases da intervenção. Define-se o ciclo específico do projeto delimitando os processos que serão trabalhados. Também são definidos os ganhos esperados para o processo a partir da definição de um problema. O entendimento da estratégia da organização é um requisito básico, e ao final, é preciso haver uma compreensão do negócio da organização como um todo, de como que os processos ou grupo de processos estejam adequados ao contexto da organização e sua relação com os demais processos.

Na segunda fase ocorre a modelagem de processos correntes (AS IS). Neste momento são levantados os processos em seu estado atual utilizando a notação BPMN – Notação para gerenciamento de processos de negócio. O foco desta fase é reproduzir os processos em diagramas descritivos que expressam seu estado atual e possibilita a análise de oportunidade de melhorias.

Já na terceira fase é promovida a análise do processo. Este é o momento para analisar todas as informações levantadas a fim de identificar, além de oportunidades de melhorias, desconexões do processo que afetam seu desempenho e os objetivos do negócio. Na sequência, quarta fase, empreende-se o projeto de processos almejados (TO-BE), onde serão redesenhados todos os processos previamente mapeados, implementando melhorias levantadas pela fase anterior. Por fim, a fase de transformação. Aqui, desenvolvem-se os planos específicos direcionados pelas análises e ideias de melhorias, que permitam o acompanhamento e controle de ações para promover as mudanças definidas, transformando o processo AS-IS em TO-BE.

Tabela 1. Fases do ciclo de implementação da metodologia MI-BPM

Fases	Atividades
1ª FASE - Entendimento do Negócio	Definição e preparação de Formulários, Consolidação de Informações Estratégicas, Reunião de Alinhamento, Definição do Plano de Gerenc. do Projeto, Treinamento do Projeto.
2ª FASE - Modelagem de Processos (AS IS)	Realizar Entrevistas, Modelagem Inicial, Modelagem Validação, Modelagem Validação AS-IS Final, Descrição de Atividades.
3ª FASE - Análise do Processo	Treinamento Painel de Gestão de Mudança, Reunião de Painel de Gestão de Mudança, Check-list análise de Processo, Análise de Riscos e Mudanças, Revisão Ferramentas de Gestão, Estudo das Melhores Práticas, Medição de Desempenho Atual, Indicadores de Desempenho do Processo.
4ª FASE - Projeto de Processos (TO-BE)	Redesenho dos Processos, Validação dos novos Processos, Projeção de Cenário.
5ª FASE - Transformação	Plano de Ação Treinamento TO BE Relatório Final Acompanhamento

4. Resultados e Discussão

A população considerada para esta pesquisa é o conjunto de organizações que passaram por uma intervenção parcial ou integral, por meio de atividades de consultoria e/ou capacitação, pautada pela metodologia MI-BPM de [for double-blind review]. As intervenções ocorreram no período de Maio de 2011 a Maio de 2017, totalizando 289 empresas. Destas, apenas 37 aceitaram participar da pesquisa. Trata-se de empresas de diferentes portes, atuando em segmentos e regiões distintas, com forte concentração na cidade de Goiânia.

Para cada unidade de análise foi escolhido um ou mais pontos focais - indivíduos que participaram ou detinham conhecimento privilegiado sobre a implementação da metodologia MI-BPM.

O propósito inicial é capturar as impressões de cada organização no que tange à implementação das melhorias nos processos de negócio. Para tanto, um questionário foi aplicado junto aos pontos focais de cada organização ao longo do período de maio a setembro de 2017. O questionário foi estruturado a partir de fatores apontados como críticos para o sucesso na implementação de processos de negócio, baseando-se nas obras de [Capote 2011b], [Baldam et al. 2007], e [Al-Mashari and Zairi 1999]. Isto posto, estes são os fatores críticos (FCS) de sucesso considerados para a avaliação:

- FCS para a Organização:
 - Transformação de processos requer definições claras de responsabilidade e propriedade.
 - Acompanhamento de desempenho permite respostas ágeis.
 - Medições de desempenho contribuem para controle de custos, qualidade e melhoria contínua.

- Monitoramento melhora a conformidade.
 - Visibilidade, entendimento e prontidão para mudança aumentam a agilidade.
 - Acesso a informações úteis simplifica a transformação de processos.
 - Avaliação de custos de processos facilita controle e redução de custos.
 - Melhor consistência e adequação da capacidade de negócio.
 - Operações de negócio são bem compreendidas e o conhecimento é gerenciado.
 - Alinhamento estratégico de TI, uso efetivo de ferramentas de software.
- FCS para os Clientes:
 - Transformação dos processos impacta positivamente os clientes.
 - Colaboradores atendem melhor às expectativas de partes interessadas.
 - Compromissos com clientes são bem mais controlados.
 - FCS para a Gestão:
 - Confirmação de que as atividades realizadas em um processo agregam valor.
 - Otimização do desempenho ao longo do processo.
 - Melhoria de planejamento e projeções.
 - Superação de obstáculos de fronteiras funcionais.
 - Facilitação de benchmarking interno e externo de operações.
 - Organização de níveis de alerta em caso de incidente e análise de impactos.
 - FCS para os Atores de processos:
 - Maior segurança e ciência sobre seus papéis e responsabilidades.
 - Maior compreensão do todo.
 - Clareza de requisitos do ambiente de trabalho.
 - Maior contribuição para os resultados da organização e, por consequência, maior possibilidade de visibilidade e reconhecimento pelo trabalho que realiza.

4.1. Análise dos Resultados

Em meio aos esforços que são requeridos para promover o sucesso da gestão baseada em processos e aos fatores críticos de sucesso para dimensionar o nível de satisfatoriedade de sua implementação, a avaliação por meio do autopreenchimento dos questionários por parte dos pontos focais das organizações pesquisadas possibilitou encontrar achados e artefatos informacionais em função dos resultados das atividades de intervenção pós aplicação da metodologia MI-BPM. A figura 2 mostra o questionário apresentado para obtenção dos resultados, desenvolvido a partir dos Fatores Críticos e de Sucesso.

Em síntese, as respostas revelaram um quadro satisfatório quanto à assimilação da abordagem orientada à disciplina de BPM ante às ações incipientes de acomodação corporativa. Um pouco mais de 55% dos respondentes aplicou a metodologia MI-BPM por meio de atividades de consultoria externa e os demais passaram a conhecê-la por meio de capacitação focada na melhoria de processos com a MI-BPM.

Em geral, as respostas mostram um alto índice de satisfação parcial ou completa, pouco mais de 94% das empresas, com o projeto de melhoria de processos empreendido

Nome da empresa
Quantidade de funcionários da empresa
Foi realizado na sua empresa Consultoria ou Treinamento?
<input type="checkbox"/> Consultoria <input type="checkbox"/> Treinamento
1 Quando a consultoria / treinamento foi realizado, um projeto de melhoria com Desenhos, Indicadores e Plano de Ação, entre outros, foi entregue. A empresa ficou satisfeita com o que foi entregue?
<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Parcialmente
2 No projeto de melhoria foram definidas ações. Estas ações foram executadas?
<input type="checkbox"/> Sim, Todas ou Boa Parte <input type="checkbox"/> Não, Nenhuma ou Quase Nenhuma <input type="checkbox"/> Parcialmente
3 No projeto de melhoria, foram entregues indicadores para medir o desempenho do processo. Estes indicadores foram ou estão sendo medidos?
<input type="checkbox"/> Sim, todos ou Boa Parte <input type="checkbox"/> Não, Nenhum ou Quase Nenhum <input type="checkbox"/> Parcialmente.
4 Hoje, algum tempo após a execução das ações de melhoria do projeto, pode-se dizer que foram obtidos resultados positivos?
<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Parcialmente
5. Você conseguiu levar para as outras áreas/processo da empresa o que foi feito na área/processo atendida pela consultoria / treinamento?
<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Parcialmente
6. Hoje, algum tempo após a execução das ações de melhoria do processo, quais resultados foram obtidos? (Escolha quantos forem necessários)
<input type="checkbox"/> Passou a existir uma padronização das ações e processos; <input type="checkbox"/> Houve uma redução de desperdício de recursos; <input type="checkbox"/> Existe um controle da execução dos serviços em todas as fases; <input type="checkbox"/> Houve uma Redução de prazos; <input type="checkbox"/> Melhoria da produtividade e satisfação do pessoal; <input type="checkbox"/> Diminuição de mão-de-obra por serviço executado; <input type="checkbox"/> Melhoria da qualidade dos trabalhos entregues; <input type="checkbox"/> Maior integração entre as diversas áreas; <input type="checkbox"/> Identificação e uso de indicadores para a gestão; <input type="checkbox"/> Alinhamento dos processos com as diretrizes estratégicas. <input type="checkbox"/> Aumento da entrega de valor e satisfação dos Clientes.
7. Quais dos fatores abaixo você considera como causa para que os resultados não tenham sido satisfatórios ao ainda mais satisfatórios? (Escolha quantos forem necessários)
<input type="checkbox"/> Faltou envolvimento dos colaboradores que operam o processo; <input type="checkbox"/> Faltou envolvimento da liderança (Sócios, Diretores, Gerentes); <input type="checkbox"/> Os resultados, ganhos, a serem alcançados com a melhoria não estavam claros; <input type="checkbox"/> A razão, o motivo da mudança não estava claro; <input type="checkbox"/> Faltou acompanhamento da ações e indicadores; <input type="checkbox"/> As ações e indicadores não estavam claros; <input type="checkbox"/> Houve tentativa de executar todas as ações ao mesmo tempo, faltou tempo e recurso para todas; <input type="checkbox"/> As mudanças planejadas não eram prioridade, por isso não houve foco e tempo para elas.

Figura 2. Questionário aplicado para obtenção dos resultados, construído a partir dos Fatores Críticos de Sucesso.

durante a execução da consultoria para apoio do uso da metodologia. Esta indicação sugere uma assimilação consideravelmente alta da metodologia, haja vista a percepção de acessibilidade da metodologia e a melhoria de expectativas quanto aos serviços de negócio das empresas.

Quanto à busca pela efetividades das ações decorrentes da intervenção, as respostas mostram que, mesmo com um plano de ação definido, entre 55 e 60% dos casos as ações planejadas não são executadas ou são executadas parcialmente. Percentual similar evidencia o acompanhamento por meio de indicadores de desempenho, indicando um quadro insatisfatório em termos de sustentação e acomodação nas atividades e operações de negócio.

Alternativas	Quantidade de respostas	
Passou a existir uma Padronização dos processos;	17	11,89%
Houve uma redução de desperdício de recursos;	14	9,79%
Existe um controle da execução dos serviços em todas as fases;	13	9,09%
Houve uma Redução de prazos;	12	8,39%
Melhoria da produtividade e satisfação do pessoal;	14	9,79%
Diminuição de mão-de-obra por serviço executado;	6	4,20%
Melhoria da qualidade dos trabalhos entregues;	16	11,19%
Maior integração entre as diversas áreas;	11	7,69%
Identificação e uso de indicadores para a gestão;	16	11,19%
Alinhamento dos processos com as diretrizes estratégicas;	12	8,39%
Aumento da entrega de valor e satisfação dos Clientes.	12	8,39%
Total de respostas	143	

Figura 3. Resultados obtidos satisfatoriamente com a intervenção

A percepção de valor decorrente da assimilação da metodologia MI-BPM é altamente positiva, mais de 84% dos respondentes. Contudo, a maioria afirma que não é possível replicar as ações de melhoria em outros processos de negócio a partir de uma única aplicação da metodologia estruturada.

Tendo em vista o nível de satisfatoriedade em função dos resultados produzidos com a intervenção, as respostas sugerem que, dentre as melhorias que podem ser obtidas com a aplicação de BPM, abordadas pelo CBOOK e por obras como [Capote 2011b], [Baldam et al. 2007], e [Al-Mashari and Zairi 1999], todas elas foram percebidas pelas empresas respondentes, porém com baixo número de respostas favoráveis, conforme consta na Figura 3.

Do mesmo modo, as questões que remetem à principal causa de insucesso na implementação da metodologia MI-BPM foram apontadas - vide Figura 4. Ainda que os percentuais de resposta sejam relativamente baixos, a falta de envolvimento dos Líderes, a falta de envolvimento dos colaboradores que operam o processo e as mudanças planejadas com baixa prioridade se sobressaíram como fontes de insucesso.

Alternativas	Quantidade de respostas	
	Faltou envolvimento dos colaboradores que operam o processo	11
Faltou envolvimento da liderança (Sócios, Diretores, Gerentes)	17	25,76%
Os resultados, ganhos, a serem alcançados com a melhoria não estavam claros	4	6,06%
A razão, o motivo da mudança não estava claro	2	3,03%
Faltou acompanhamento das ações e indicadores	7	10,61%
As ações e indicadores não estavam claros	4	6,06%
Houve tentativa de executar todas as ações ao mesmo tempo, faltou tempo e recurso para todas	10	15,15%
As mudanças planejadas não eram prioridade, por isso não houve foco e tempo para elas	11	16,67%
Total de respostas	66	

Figura 4. Resultados insatisfatórios com a intervenção

4.2. Discussão

A avaliação baseada nos FCS permite encontrar subsídios que sugerem um alto nível de assimilação da metodologia MI-BPM, porém, as ações de sustentação e acomodação são pouco evidentes. Uma explicação plausível é o fato de que, em geral, o processo de intervenção nas empresas pesquisadas se deu por meio de uma única sessão transversal, impedindo que haja um devido acompanhamento e ajustes na implementação.

Os resultados demonstram que é preciso planejar e envolver os atores chave para alcançar resultados melhores. Uma alternativa é adotar uma plataforma de engajamento que potencialize o diálogo dos atores e donos de cada processo, que facilite o acesso aos recursos, que avalie os riscos e que deixe transparente os ganhos de cada parte interessada.

Os resultados também sugerem que por meio da melhoria e controle dos processos de negócio é possível haver um melhor alinhamento ao planejamento estratégico da organização. Outrossim, ainda que os procedimentos, técnicas e ferramentas de melhorias aplicadas em um processo de negócio específico possam ser assimiladas satisfatoriamente, nem sempre haverá um entendimento suficiente para transladar os conhecimentos para outros processos de negócio, haja vista a necessidade de maturação que deverá estar presente na estratégia de acomodação das boas práticas para implementação dos processos de negócio.

Por fim, observou-se que, para se obter uma melhoria dos processos mais significativa, é preciso dar uma maior atenção ao controle e monitoramento dos indicadores e das ações de mudança entregues durante a aplicação de uma abordagem BPM.

5. Conclusão

O estudo empreendido sugere que a aplicação de BPM para melhoramento dos processos nas organizações foi assimilada parcialmente pelas organizações pesquisadas, cuja metodologia MI-BPM foi bem avaliada a partir das intervenções promovidas, seja por meio da capacitação promovida nas organizações, seja por meio de ações de consultoria. Alguns benefícios ficaram notabilizados como o aumento da padronização, maior agilidade

e melhor controle dos processos de negócio. O foco, portanto, recai sobre a melhoria da entrega de valor ao cliente.

O trabalho não tem a pretensão de sugerir possíveis relações de causa e efeito, nem sequer de generalizar resultados quanto à efetividade da aplicação da metodologia MI-BPM. Contudo, o nível de aprendizado em função dos fatores críticos de sucesso poderá contribuir para a formulação de hipóteses que poderá subsidiar novos estudos com maior grau de assertividade, bem como balizar o refinamento da própria metodologia empregada.

Não obstante os vieses das respostas dadas em termos de confiabilidade das respostas, da consistência interna com a realidade percebida e encontrada e da preocupação em validar rigorosamente o instrumento, o principal interesse da pesquisa foi alcançado, uma vez que foi possível identificar subsídios por meio da ampliação dos horizontes de percepção que permitam avaliar o grau de assimilação e de acomodação das ações de intervenção nas unidades de análise.

A principal dificuldade encontrada na aplicação do instrumento envolveu a grande resistência das empresas candidatas em participar da pesquisa, dos respondentes em repassar informações que remetem ao desempenho e aos resultados de negócio e a indisposição no preenchimento do questionário. Mesmo assim, o estudo permitiu avaliar com razoável acurácia os quesitos constantes no FCS.

Referências

- Al-Mashari, M. and Zairi, M. (1999). Bpr implementation process: an analysis of key success and failure factors. *Business process management journal*, 5(1):87–112.
- Antonello, C. S., Boff, L. H., et al. (2005). *Aprendizagem organizacional e competências. Porto Alegre.*
- Baldam, R., VALLE, R., PEREIRA, H., HILST, S., ABREU, M., and SOBRAL, V. (2007). *Gerenciamento de processos de negócios: Bpm–business process management. São Paulo: Érica.*
- Capote, G. (2011a). *Guia para formação de analistas de processos. Business Process Management. Rio de Janeiro: Bookess.*
- Capote, G. (2011b). *Guia para formação de analistas de processos. Business Process Management. Rio de Janeiro: Bookess.*
- CBOK, B. (2009). *Guide to the business process management common body of knowledge. Versão 2.0. 2009. Disponível em: www.abmp.org. Acesso em: 25 nov 2012.*
- CBOK, B. (2013). *Guia para o gerenciamento de processos de negócio corpo comum de conhecimento. Association of Business Process Management Professionals. ABPMP BPM CBOK, 3.*
- de Andrade, M. M. (2000). *Como Preparar Trabalhos Para Cursos de Pós-graduação: Noções Práticas*. Editora Atlas SA.
- Feldman, M. S. and Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative science quarterly*, 48(1):94–118.
- for double-blind review, R.

- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. 6. ed. Editora Atlas SA.
- Gonçalves, J. E. L. (2000). As empresas são grandes coleções de processos. *Revista de administração de empresas*, 40(1):6–9.
- JOHANSSON, H. J., Mchugh, P., Pedlebury, J., and Wheller III, W. (1995). Processos de negócios. *São Paulo: Pioneira*.
- Mariano, I. C. (2012). Melhoria de processos pelo bpm: Aplicação no setor público.
- Morais Junior, N. P. (2014). Melhoria do processo de gestão da construção cível com uso da disciplina bpm. *Dissertação de mestrado apresentada em TEDE - Sistemas de Publicação eletrônica de Teses e Dissertações, Pontifícia Universidade Católica de Goiás*.
- Osterwalder, A. and Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley & Sons.
- Ostrenge, J., Haleblian, J., Poulsen, B., Ferrell, B., Mueller, N., and Shastri, S. (1993). Vehicle design for a new tropical steroid, fluocinonide. *Journal of Investigative Dermatology*, 56(5):392–399.
- Pizza, W. R. (2012). A metodologia business process management (bpm) e sua importância para as organizações. *Monografia apresentada no curso de Tecnologia em processamento de dados na Faculdade de Tecnologia de São Paulo-FATEC SP*.
- Scucuglia, R. (2011). Mapeamento e gestão por processos: Bpm. *Gestão orientada a entrega por meio de objetos*. São Paulo: M. Books do Brasil.
- Torres, I. d. S. (2015). Aplicação da metodologia bpm em uma ifes: proposição de um modelo estendido.
- Zarifian, P. (1997). Organização e sistema de gestão: à procura de uma nova coerência. *Gestão & Produção*, 4(1):76–87.

Etnografia como base no desenvolvimento de um ERP voltado para o Hospital HAB.

Pétersson S. Jesus¹, Alessandro R. Silva¹

¹Instituto Federal de Goiás (IFG)
75131-457 – Anápolis – GO – Brazil

petersondejesus99@gmail.com, alessandro.rodriques@ifg.edu.br

Abstract. *This work consists of displaying the study, analysis and survey of requirements of administrative relations, interconnections with the production of information and systems operation in the Support Hospital of Brasília. And subsequent construction of an ERP capable of reformulating the communication process and optimizing the administrative activities, current reason of disorders and impediment in the agility of tasks. The process of gathering requirements was performed, through ethnography, technique used in Software Engineering for this purpose. The artifacts found and built: responsibilities of the sectors, diagrams and description of use cases, class diagrams, demonstrate how efficient the technique is and able to uncover problems and operations not perceptible to the requirements analyst only with quick techniques such as interview.*

Resumo. *Este trabalho consiste em exibir o estudo, a análise e o levantamento de requisitos das relações administrativas, interconexões com a produção da informação e operação de sistemas no Hospital de Apoio de Brasília. E posterior construção de um ERP capaz de reformular o processo de comunicação e otimizar as atividades administrativas, atual motivo de transtornos e impedimento na agilidade de tarefas. O processo de levantamento de requisitos foi executado, por intermédio da etnografia, técnica utilizada em Engenharia de Software para tal finalidade. Os artefatos encontrados e construídos: responsabilidades dos setores, diagramas e descrição de casos de uso, diagramas de classe, demonstram o quanto a técnica é eficiente e capaz de desocultar problemas e operações não perceptíveis ao analista de requisitos somente com técnicas rápidas como entrevista.*

1. Introdução

O Hospital de Apoio de Brasília (HAB) possui cuidados em doenças oncológicas e geriátricas, contando com vários departamentos para ajudar no gerenciamento das informações e cuidado dos pacientes do hospital. Cada departamento é responsável por certa tarefa formando o conjunto burocrático. Durante as atividades e processos dentro de cada departamento que são realizados por um, dois ou até mais funcionários, atualmente são realizadas, em sua maioria, por meio de planilhas eletrônicas ou manuscritas.

São encontradas dificuldades no decorrer desses processos. A comunicação entre os setores é uma das principais dentro do HAB, impossibilitando o desempenho e eficiência na realização do ofício dos funcionários envolvidos. Pelo motivo da distância

entre os setores, a locomoção entre eles é lenta e não há meios de comunicação em comum de todos. Para entregar um documento importante ou até mesmo enviar uma mensagem ao setor na região contrária do hospital, é necessário ir até o local para tal ação. Atividade que consome muito tempo, e que pode ser enviado simplesmente por uma plataforma online em alguns segundos. Isso, quando não se trata de uma informação que todos os setores devem receber, exemplo: informações do paciente que chegará ao hospital em certo dia, um aviso importante que necessita que todos recebam.

O tempo de execução de uma tarefa em alguns setores está atualmente muito lento, alguns setores já avançaram um pouco, porém é necessário avançarem mais se desejarem que a execução de suas atividades seja rápida. Atualmente as tarefas na administração que não são executadas nos sistemas “TrakCare” e “Sei!”, sistemas atuais utilizados no HAB para algumas funcionalidades, são realizadas em planilhas, manuscritas, ou não são nem realizadas. Algumas dessas atividades: Persistência dos registros de crianças e dados de seus exames em alguma plataforma onde pode ser acessados de forma rápida e eficiente; a busca de crianças que não compareceram para os exames; geração de relatórios e emissão de relatórios automáticos; solicitação de palestra para o NEPS; realização do gerenciamento dos alunos e informações sobre todas as palestras e cursos realizados no HAB; Requisitar alimentos da cozinha para os pacientes pelo Núcleo de Nutrição; criar chamados para o NTI por meio online sem a necessidade de se locomover até o local para o chamado; possibilitar o NTI gerar um relatório de suas atividades diárias automáticas. Desafios que o HAB enfrenta e que com a criação de um ERP simplificará todos.

A intranet ou ERP se encontra no topo dos veículos disponíveis para o funcionário da empresa como meio de comunicação no Brasil com 87,2%, seguido de jornal-mural com 83,5% e comunicação face-a-face com 50,6%, seguido de jornal-mural. [Marchiori 2010]. Portanto, chega-se a conclusão que a construção de um ERP é a solução ideal para o problema do hospital. ERP que necessitará de realizar atividades cotidianas dos funcionários e atividades específicas que são citadas na seção 10.2, construídas por módulos e sendo cada módulo, um núcleo, obtendo-se mais facilidades no exercício das atividades realizadas. E para o levantamento dos requisitos utilizou-se a etnografia.

A etnografia é um meio que observa os indivíduos em seu ambiente, com o intuito de compreender suas atividades do cotidiano, o modo de vida de certa sociedade, e nesse conceito de observar, tentar trazer para o nosso contexto e entender como aquela sociedade vive, nisso, entramos no conceito de interação dos indivíduos com o seu ambiente [Mattos and Castro 2011]. E na computação esse conceito é trazido para a interação homem-computador, definido como o modo que o usuário (indivíduo) interage com o computador (ambiente), porém não há uma definição estabelecida para esse termo muito utilizado não somente em nossa área [Rocha and Baranauskas 2003].

Um ERP (Enterprise Resource Planning ou Sistema de Gestão Empresarial) é um sistema construído por pacotes, que permite integrar e automatizar os processos de negócios, compartilhar informações com toda empresa, produzir e acessar informações em tempo real. [Deloitte Consulting 1999] E segundo Stamford (2000), um ERP possibilita a empresa observar e gerenciar um fluxo de dados contínuo e coeso com todas as informações em apenas uma base de dados. Permite visualizar as transações efetuadas pela empresa. Aperfeiçoa os processos de negócio e exhibe as informações em

tempo real.

Através da teoria dos Grupos Operativos e da Evolução dos Grupos, áreas fundamentadas na etnografia e Ethnomining (etnografia com mineração de dados). Foi analisado e apresentado que o vínculo entre analista de requisitos e usuários é a peça chave para compreensão do comportamento dos envolvidos. E constataram que este vínculo sempre ocorre, sendo de maneiras iguais ou distintas, porém existente. Por meio da observação (etnografia) nesse vínculo, é perceptivo em alguns casos que o analista de requisitos se titula um co-pensador da atividade, ou seja, apresenta a função de amadurecer o fluxo administrativo da empresa na ocasião do seu modelo administrativo não ser compatível com o sistema ERP. [Balthazar 2012]

A etnografia realizada no Hospital HAB teve como objetivo apresentar a especificação de objetivos e requisitos de um ERP, que se pretende usar no hospital para aprimorar as atividades realizadas nos núcleos administrativos, as interconexões com a produção de informação, operação de sistemas e eficácia do serviço público de saúde. Atualmente, propomos uma reformulação do processo de comunicação utilizado no hospital por meio da criação de uma Intranet, que abrangerá o registro e acompanhamento das atividades desenvolvidas com o intuito de sanar os problemas administrativos do hospital.

2. Metodologia

Os requisitos do software que será futuramente construído foram adquiridos por meio da etnografia realizada durante o período de quatro semanas dentro do hospital em cada departamento, observando todos os funcionários, nos seguintes departamentos: Diretoria Administrativa; Núcleo de Serviço Social; Laboratório de Biologia Molecular (BIOMOL); Laboratório de Citogenética; Núcleo de Qualidade e Segurança do Paciente (NQSP); Núcleo de Controle de Escalas (NCE); Núcleo de Controle de Infecções e Epidemiologia Hospitalar (NCIEH); Núcleo de Farmácia Hospitalar (NFH); Núcleo de Gerência Interna de Regulação (NGIR); Núcleo de Captação e Análise de Informações do SUS (NCAIS); Núcleo de Educação Permanente em Saúde (NEPS) / Núcleo de Ensino e Pesquisa (NEP); Núcleo de Nutrição; Núcleo de Planejamento Monitoramento e Avaliação (NPMA); Núcleo de Tecnologia da Informação (NTI); Núcleo De Atividades Gerais, Manutenção Predial e Transporte (NAGMPT); Lavanderia; Laboratório de Triagem.

A etnografia foi realizada durante o período de quatro semanas com os funcionários de cada setor citado acima, ao decorrer desse período foi analisada a maioria das atividades realizadas nesses setores. A análise dessas atividades e responsabilidades de cada setor foi concretizada por meio de perguntas e convivência por um tempo limitado com o responsável pela tarefa, para visualizar quais os processos realizados e quanto tempo demorava até se chegar ao fim da atividade. As perguntas foram apresentadas aos funcionários de forma informal, em momentos simples e que não era necessário nenhuma formalidade de nenhuma das partes. O que possibilitou a análise explorar a etnografia, deixando o funcionário livre, se dirigindo a resposta sem o receio das questões levantadas durante o diálogo. As perguntas efetuadas durante os diálogos de forma discreta foram as seguintes:

- Qual a função do setor.

- Quais atividades são realizadas no computador e no papel.
- Problemas no processo atual.
- Sugestões de melhoria.
- Suas sugestões para o processo.

Essa análise explorou a etnografia observando os indivíduos transmitirem informações sobre seu respectivo setor, suas atividades desenvolvidas e desfrutando do contato direto com os funcionários durante o processo de cada tarefa descrito nos resultados e da verificação se o que era dito pelo funcionário ocorria verdadeiramente em sua essência. O contato, a convivência com os funcionários dentro do hospital, durante o período do trabalho, fez com que todos envolvidos fossem afetados, funcionários quanto o etnógrafo, de forma direta ou indireta.

Como dito por Jeanne Favret-Saada em seu caderno de Campo, durante seu trabalho com os nativos não era concedido a ela as respostas desejadas, porque ela não participava dos rituais de feitiçaria, ela não queria ser “afetada” pelo fenômeno do rito. A partir do momento que ela participou das sessões de enfeitiçamento e desenfeitiçamento, os nativos se disponibilizaram a responder todas as questões levantadas por ela [Siqueira 2005]. Esse trecho reforçou que para essa etnografia ocorrer de maneira mais favorável, era necessário participar do contexto inserido dos nativos. Esse conceito foi utilizado na etnografia no HAB, durante todo o período da etnografia, o etnógrafo ficou inserido no contexto do hospital percebendo os mínimos detalhes que poderia incrementar no levantamento de requisitos.

3. Resultados

Nesse período foi realizado um raio-X do hospital HAB e como resultado se obteve artefatos para a construção do software para melhor qualidade e eficiência no processo de cada atividade realizada dentro do âmbito do hospital. Abaixo segue o exemplo de alguns dos artefatos adquiridos:

3.1. Responsabilidades de Cada Setor

Como artefato da etnografia realizada no Hospital HAB, cada departamento possui sua responsabilidade dentro do hospital. Abaixo estão citadas a de cada departamento em que foi realizada a etnografia.

3.1.1 Diretoria Administrativa

Setor responsável por comandar toda a parte administrativa do hospital, dentre as atividades realizadas são: manter o hospital com suas atividades realizadas corretamente e normalmente; verificar se as compras estão em dias; não deixarem faltar insumo e materiais no Hospital HAB. Necessita muito de seus braços direitos, os chefes de cada departamento. Cobrando aqueles que não estão realizando as atividades com êxito.

3.1.2. Laboratório de Citogenética

Nesse departamento são realizadas as seguintes atividades: Realização de exames cariótipos; o controle de estoque dos insumos utilizados no laboratório; comunicação com a farmácia para obter materiais para os exames; emissão de relatórios para os

médicos.

3.1.3. Laboratório de Biologia Molecular (BIOMOL)

Esse núcleo tem a responsabilidade de realizar exames biomoleculares que são encaminhados para eles através de médicos geneticistas; emitir laudos desses exames, encaminhando-os para os médicos responsáveis.

3.1.4. Laboratório de Triagem

Esse setor é responsável por recolher os dados para a realização dos exames do pezinho, realizar os exames e encaminhar para o Núcleo de Serviço Social para que aconteça o contato com a família se acontecer algo de diferente após o exame ou quando o resultado estiver pronto para marcação da próxima consulta se necessário. O laboratório de Triagem, departamento de recolhimento de sangue para o exame e o Núcleo de Serviço Social são totalmente interligados entre si, por meio da chamada “Busca Ativa”.

3.1.5. Lavanderia

Esse setor possui a responsabilidade de cuidar e preservar as roupas usadas pelos pacientes dentro do hospital, lavar, centrifugar, passar e armazená-las, cuidar para que os pacientes vistam uma roupa macia, com conforto e limpa. Economizar água e produtos é a maior responsabilidade e maior desafio deles, pois os produtos com o tempo gastam a roupa e cria furos nelas. E a água utilizada ali, não pode ser reutilizada por conter microrganismos de doenças hospitalares, uma situação difícil que o Hospital HAB presencia e tenta ao máximo cuidar.

3.1.6. Núcleo De Atividades Gerais, Manutenção Predial e Transporte (NAGMPT)

Esse núcleo é responsável por supervisionar, toda manutenção predial necessária ser feita no hospital, incluindo os materiais, empresa responsável, tudo é supervisionado e gerenciado por esse núcleo; receber os pedidos de cada departamento requisitando o serviço do núcleo em alguma parte do hospital. Todo o núcleo de transporte, os motoristas, os veículos, os horários de entrada e saída de cada veículo, escala de cada funcionário motorista são feitos pelo responsável pelo núcleo.

3.1.7. Núcleo de Captação e Análise de Informações do SUS (NCAIS)

O setor é responsável por subsidiar a alta gestão a produtividade hospitalar que são todos os atendimentos provenientes de internações hospitalares financiadas pelo SUS e a produtividade ambulatorial que são todos os atendimentos provenientes dos ambulatorios financiados pelo SUS do Hospital HAB, bem como análises básicas, ou seja, emitir relatórios de quantos pacientes foram atendidos por funcionário, de quantos funcionários possui no momento em cada departamento do hospital, os nomes deles pelo sistema “TrackCare”; passar esses dados para os sistemas do ministério da saúde, sendo quatro sistemas distintos. E com todos esses dados em planilhas e documentos, gerar um relatório trimestralmente com os dados e informações sobre as atividades

realizadas por funcionário e, por fim, enviar via e-mail para a SES-DF.

3.1.8. Núcleo de Controle de Escalas (NCE)

Esse núcleo é responsável por verifica se as cargas horarias estão corretas no “TrackCare” e no “ForPonto”; realiza as escalas e controla os abonos, férias e licenças prêmios para cada servidor, podendo liberar ou não as notificações de férias, abonos e licenças; controla o sistema de pontos dos funcionários, percebendo se algum funcionário esqueceu-se de bater ponto em um dia, se o funcionário não compareceu ao local de trabalho.

3.1.9. Núcleo de Controle de Infecções e Epidemiologia Hospitalar (NCIEH)

Esse núcleo tem a responsabilidade de analisar as epidemias no hospital, cuidar para que os pacientes não se contaminem por meio de bactérias hospitalares. Eles necessitam de indicadores, mas indicadores sobre as doenças ocorridas nos pacientes e qual sua origem (hospital ou comunitária), tendo controle de como as doenças estão ocorrendo dentro do Hospital HAB; prescrevem também os remédios que os pacientes com infecções tomarão; gerenciam o custo de álcool e sabonete utilizado pelos pacientes; recebem as notificações de infecções e doenças por um formulário preenchido pelo responsável do local ou do paciente, sendo 37 formulários, cada um diferente conforme as doenças; responsável também por cuidar da higienização das mãos de todos dentro do Hospital HAB, daqueles que entram e saem; emitir relatório de todas as atividades realizadas pelo NCIEH para a diretoria.

3.1.10. Núcleo de Educação Permanente em Saúde (NEPS) / Núcleo de Ensino e Pesquisa (NEP)

Esse departamento é responsável pela capacitação dos servidores desenvolvendo e ministrando cursos necessários para tal aplicação; gerenciar as informações de estágios realizados no Hospital HAB, necessitando conhecimento do local de cada estagiário dentro dos departamentos do hospital; divulgações de eventos dentro de todo o Distrito Federal; controlar as visitas técnicas realizadas ao Hospital HAB; emissão de certificados e declarações para os alunos capacitados e frequentes; verificação de planos de trabalho dos estagiários que chegam ao Hospital HAB.

3.1.11. Núcleo de Farmácia Hospitalar (NFH)

Esse núcleo é responsável por gerenciar o estoque de medicamentos do Hospital HAB; gerar relatórios da entrada e saída de cada medicamento; realizar pedidos pelo “SisMateriais” para SES-DF de medicamentos necessários para os pacientes do hospital; colocar data de validade em cada medicamento que chega novo na farmácia;

3.1.12. Núcleo de Gerência Interna de Regulação (NGIR)

Esse setor trabalha com tudo relacionado com o paciente, internação e ambulatório, prontuários. Esse setor só possui de todos os departamentos, as alas de internação submissas a ela. Responsável por avisar os pacientes que chegarão no dia seguinte, na

semana decorrida.

3.1.13. Núcleo de Nutrição

Esse núcleo é responsável por fazer o cardápio de alimentação de cada paciente separadamente; realizar mudanças se ocorrer algo diferente no estado do paciente; emitir relatórios sobre essas mudanças; inserir no “TrackCare” as informações de cada visita ao paciente diariamente, mostrando a melhora, estabilidade ou recaída do estado dele; se comunicarem com a cozinha para informá-los quais alimentos serão necessários para aquele dia; realizar relatórios de quantos pacientes estão sendo analisados e acompanhados por nutricionista e no geral; gerenciar o sistema “ProgDate”, sistema utilizado para algumas atividades e pela comunicação entre Núcleo de Nutrição e pela cozinha do hospital.

3.1.14. Núcleo de Planejamento, Monitoramento e Avaliação (NPMA)

Este núcleo tem por responsabilidade mapear os setores, como está o funcionamento atual de cada um, observando os pontos fracos, pontos fortes, oportunidades e ameaças; analisar o processo de cada um; redesenhar os processos de cada um; implantar os novos processos desenhados; monitorar os resultados dos novos processos.

3.1.15. Núcleo de Qualidade e Segurança do Paciente (NQSP)

Esse núcleo tem responsabilidade de manter a segurança do paciente e qualidade dos serviços prestados a eles. Fazendo com que os fatores de queda do paciente sejam menores a cada dia, verificando a causa da queda de cada paciente e diminuindo as possíveis causas que ocorrem à queda de cada paciente.

3.1.16. Núcleo de Serviço Social

Setor responsável pelo atendimento aos pais, familiares, comunidade externa; emissão de relatórios, pareceres; contato com redes de serviços externos do DF; trabalho de busca ativa, ou seja, a busca (localização e contato) de crianças que fizeram o teste do pezinho, e que precisam refazê-lo por algum motivo, porém perderam alguns dados que são necessários. Se não encontrar telefones, procuram o serviço de saúde para visitá-los e pedirem para virem ao Hospital HAB, se a família negar, é contatado o conselho tutelar.

3.1.17. Núcleo de Tecnologia da Informação (NTI)

Este núcleo é responsável por todo tipo de tecnologia do hospital, por controlar os acessos aos sistemas internos, quem pode acessar, quais atividades podem realizar nos sistemas; suporte aos usuários das máquinas e no funcionamento dos computadores; entregar relatórios mensais das atividades de suporte que ele realiza.

3.2. Funcionalidades Necessárias para o Software

Abaixo, segue a lista de funcionalidades necessárias do software que foram possíveis de

serem levantadas graças a etnografia:

- Administrador: Criar pacientes; Editar pacientes;
- Assistência Social: Gerar etiqueta de crianças; Notificar laboratório a entrada de exames; Visualizar notificação; Pesquisar criança; Ver criança; Emitir resultado de exames; Ver resultado de exames; Agendar recoleta; Encaminhar criança para busca ativa; Encerrar o caso e/ou óbito; Solicitar palestra; Criar palestra; Editar palestra;
- NEPS: Criar alunos; Editar alunos; Criar estagiários; Editar estagiários; Emitir certificados; Criar cursos; Editar cursos; Criar frequência; Realizar frequência; Ver inscrições de curso; Editar inscrição de curso; Gerar relatório de estagiários; Gerar relatório de cursos;
- Nutrição: Criar requisição de alimentos; Gerar relatório de requisições;
- Telecomunicações & Transporte: Visualizar chamado; Emitir resposta;

3.2.1. Descrição de Funcionalidades Específicas

Algumas funcionalidades citadas acima foram perceptíveis como um diferencial pela etnografia, são elas:

- [Assistência Social] Gerar etiqueta de crianças: Após o usuário inserir uma criança, o sistema gerará uma etiqueta única para esta criança que será útil para buscar a criança posteriormente.
- [Assistência Social] Emitir resultado de exames: Adiciona uma descrição breve com os resultados obtidos no exame realizado, mudando o estado do exame para finalizado.
- [Assistência Social] Agendar recoleta: Inclui a identificação e armazenamento de um agendamento para recolher o material do exame novamente, por motivo de a primeira extração não ocorrer corretamente.
- [Assistência Social] Encaminhar criança para Busca Ativa: Altera o estado do registro da criança para nível “busca ativa”, pois não se encontrou nenhum registro ou nenhum contato da família no sistema, não podendo informá-los de algo nos exames.
- [Assistência Social] Encerrar o caso e/ou óbito: Altera o estado do caso do paciente por não encontrar contato para informação sobre recoletas, retornos para exames, não encontrando contato da família para informações ou encerramento do caso por óbito do paciente.
- [NEPS] Emitir certificados: Ao encerramento do curso, o aluno que possuir matrícula com frequência maior ou igual a 80% (por cento) no curso, recebe o certificado de conclusão do mesmo. Certificado emitido em PDF para envio por e-mail para aluno.
- [NEPS] Gerar relatório de estagiários: Emite a relação de estagiários com suas informações para controle do departamento dos estagiários no Hospital HAB.
- [NEPS] Gerar relatório de cursos: Emite a relação de cursos que foram concluídos, em andamento e que irá ocorrer. Cada um com suas informações básicas (nome, sala, quantidade de alunos, horário e professor).
- [Nutrição] Criar requisição de alimentos: Inclui a identificação e armazenamento de um chamado para o departamento da cozinha com os alimentos necessários para certo paciente e não possui no estoque do Hospital HAB.
- [Nutrição] Gerar relatório de requisições: Emite a relação de requisições com suas

informações para controle do departamento da cozinha no Hospital HAB.

3.3. Diagramas de Caso de Uso

Abaixo são exibidos os Diagramas de Caso de Uso em forma de texto recorrentes da etnografia realizada:

- NEPS: Criar Estagiário; Editar Estagiário; Gerar relatório; Gerar relatório de estagiários (Extends: Gerar relatório); Gerar relatório com dados dos cursos e matriculados (Extends: Gerar relatório); Criar frequência; Criar curso; Editar curso; Realizar frequência de curso; Criar aluno; Editar aluno; Criar inscrição; Editar inscrição; Emitir certificado; Visualizar palestras;
- NCIEH: Criar documento; Criar infecção; Realizar diagnóstico; Gerar relatório de infecções;
- Assistência Social: Encaminhar criança para Busca ativa; Agendar coleta; Encerrar o caso; Inserir resultado de exame; Exibir criança; Pesquisar criança; Visualizar notificações; Notificar laboratório da entrada de exames; Solicitar palestra ao NEPS; Gerar relatórios de casos; Gerar relatório de casos novos (Extends: Gerar relatórios de casos); Gerar relatório de casos encerrados (Extends: Gerar relatórios de casos);
- Administrador: Criar funcionário; Editar funcionário;
- NCE: Criar férias de funcionário; Ver férias de funcionário; Criar licença prêmio de funcionário; Ver licenças atuais;
- Citogenética: Criar produto; Ver validade de produto; Editar produto; Ver quando estoque do produto encerra; Ver quais produtos está vencido; Fazer pedido de produto; Gerar relatórios de encaminhamento; Ver encaminhamentos; Preencher e gerar laudo; Gerar laudo de cariótipos (Extends: Preencher e gerar laudo); Gerar laudo de medula óssea (Extends: Preencher e gerar laudo);
- Médicos: Encaminhar requisição de exame cariótipo; Preencher hipótese diagnóstica;
- Biomol: Preencher dados do exame realizado; Preencher laudo; Gerar código de amostra; Gerar relatório de estatísticas de doenças; Enviar laudo para médico; Atestar resultado; Gerar laudo (Include: Atestar resultado);
- NAGMTP: Visualizar chamado; Emitir resposta;
- Nutrição: Criar requisição de alimentos; Gerar relatório de requisições; Criar atividade; Editar atividades; Registrar atividade diária; Gerar relatório de atividades mensais;

4. Considerações Finais

Conclui-se, demonstrando os resultados de uma etnografia com objetivo dito no início de receber informações para o desenvolvimento de um ERP para contribuir e melhorar atividades de alguns departamentos e a eficiência hídrica do Hospital HAB. Já foram alcançados como resultados a documentação do sistema, da especificação de requisitos do sistema e das informações necessárias para o desenvolvimento do sistema ERP. Todos os artefatos exemplificados desse documento são parte da documentação e especificação de requisitos. Espera-se que no decorrer e andamento desse projeto sejam adquiridos o sistema ERP concluído com alguns módulos citados para o Hospital HAB, que será muito útil para o hospital, pois necessitam de agilidade e eficiência na realização das atividades administrativas.

Os resultados obtidos nos mostraram que a etnografia como técnica para especificação de requisitos, é sim um meio pelo qual os resultados são obtidos de maneira simplificada e de grande completude. Pois, com as perguntas certas e a análise do comportamento de cada funcionalidade dentro do local que está realizando a etnografia, os requisitos para um software são muito bem obtidos de forma simples e mais completos, por haver uma forma mais relacional e humanizada entre os funcionários e o etnógrafo, transformando a pesquisa em seu âmbito geral eficaz e sem necessidade de relevância. O que em outras formas de levantamento de requisitos definidas por Sommerville (2007) em seus livros, não é tratado o conceito de cliente-pesquisador de maneira igualitária, mas sim, o Pesquisador diferente, superior ao Cliente. Não menosprezando o cliente, mas deixando desejar na relação com o Cliente por completo.

Em cada núcleo foram discutidos os pontos da pesquisa, levantando as dificuldades e desafios que possuem por meio da etnografia e pesquisa. Termo utilizado tanto na antropologia quanto na Ciência da Computação. Ao final desse trabalho etnográfico, foram perceptíveis alguns desafios que devem ser levados à solução, sendo eles: A dificuldade de comunicação de departamentos internos, atividades com tempo de realização muito longo que poderiam ser realizadas com maior eficiência com o uso adequado de alguma tecnologia atual. Desafios apresentados que podem ser melhorados com o uso da Intranet que será construída.

5. Referência

- Wazlawick, R. (2003) Engenharia de Software: Conceitos e Práticas. Elsevier.
- Sommerville, I. (2007) Engenharia de Software. Pearson Addison-wesley.
- Pressman, R. S. (2006) Engenharia de Software. Mcgraw-hill Higher Education.
- Rocha, H. V. and Baranauskas, M. C. C. (2003) Design E Avaliação De Interfaces Humano-Computador.
- Mattos, CLG., and Castro, PA., orgs. (2011) Etnografia e educação: conceitos e usos [online]. Campina Grande: EDUEPB.
- Consulting, D. (1999) ERP's Second Wave – Maximizing the Value of ERP-Enabled Processes. New York: Deloitte Consulting.
- Balthazar, G. R. (2012) Análise de interação social de Analistas de Requisitos e Usuários na etapa de levantamento de requisitos. Dissertação (Mestrado) - Curso de Engenharia da Computação, Instituto de Pesquisas Tecnológicas do Estado de São Paulo, São Paulo, 2012. http://cassiopea.ipt.br/teses/2012_EC_Glauber_Rocha.pdf, Setembro.
- Marchiori, M. (2010) Os desafios da comunicação interna nas organizações. In *Conexão-Comunicação e cultura, volume 9(17)*.
- Siqueira, P., (2005) “Ser afetado”, de Jeanne Favret-Saada. In *Cadernos de Campo (São Paulo, 1991)*, 13(13), pages 155-161.
- Stamford, P. P. (2000). ERP: prepare-se para esta mudança. Recife: Ed. da UFPE.

Aplicabilidade da Norma ABNT NBR ISO/IEC 27002 em uma Empresa de Médio Porte

Yan S. Rocha¹, Eliomar A. de Lima²

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Câmpus Samambaia, CEP 74690-900 – Goiânia, GO – Brasil

²Instituto de Informática – Universidade Federal de Goiás (UFG)
Câmpus Samambaia, CEP 74690-900 – Goiânia, GO – Brasil

{yanrocha,eliomar}@inf.ufg.br

Abstract. *In the face of constant changes and cyber threats, information and information assets become a fundamental and extremely valuable asset for many companies, determining their competitiveness. In order to ensure that the information is not with unauthorized persons, nor corrupted or even inaccessible, it is necessary to invest in confidentiality, integrity, and availability, which are the pillars of information security. In this sense, this paper approaches this theme through a case study involving the evaluation of information security in a medium-sized company under the protection of ISO / IEC 27002: 2013. The study revealed a picture full of nonconformities that allowed to raise hypotheses for future studies.*

Resumo. *Diante de constantes mudanças e ameaças cibernéticas, a informação e os ativos informacionais tornam-se um bem fundamental e extremamente valioso para muitas empresas, determinando a competitividade entre elas. Visando assegurar que as informações não estejam com pessoas desautorizadas, não sejam corrompidas ou mesmo inacessíveis, faz-se necessário investir em confidencialidade, integridade e disponibilidade, que são os pilares da segurança da informação. Nesse sentido, este trabalho aborda essa temática por meio de um estudo de caso envolvendo a avaliação de segurança da informação em uma empresa de médio porte sob o amparo da Norma ABNT NBR ISO/IEC 27002:2013. O estudo revelou um quadro repleto de não conformidades que permitiram levantar hipóteses para estudos futuros.*

1. Introdução

A importância dada à informação no século 21 produz implicações e desencadeia uma série de novas situações, representando grande poder a quem detém informações privilegiadas, críticas e/ou sensíveis a determinado contexto e ao mesmo tempo consegue subsidiar a tomada de decisão e a resolução de problemas corporativos.

O mundo dos negócios, imbricado no superlativismo cibernético, está sob ameaça de fenômenos e circunstâncias produzidas em meio a um cenário de grande volatilidade, incertezas, complexidades e ambiguidades. A informação passa a ser um ativo cobiçado devido ao seu grau de revelação do estado de coisas que caracteriza determinado objeto, entidade ou sistema, necessitando que as organizações se mantenham atualizadas para garantir a correta tomada de decisões estratégicas. Segundo Barbosa [1997 p.52] o processo de busca e utilização de informações externas subsidiam decisões estratégicas.

Os avanços tecnológicos, por seu turno, são necessários para que haja maior eficácia na captação, processamento, armazenamento e disponibilização das informações pertinentes. Sem a utilização da tecnologia e seus recursos, é extremamente difícil uma empresa continuar atuando satisfatoriamente no nicho que atua. É nesse contexto que é desejável o uso e aplicação de sistemas de informação para gerenciar toda a demanda de informação gerada. Segundo Turban, Mclean e Wetherbe [2004], sistemas de informação são responsáveis por coletar, armazenar, recuperar e disseminar informações pra fins específicos. Para Laudon [1999], as informações coletadas fazem parte da análise do processo decisório em organizações.

Nos últimos anos, com o advento e popularização da Internet, o acesso a informação fica cada vez mais fácil, tornando a competitividade entre as empresas cada vez maior. Do mesmo modo, potencializam-se os riscos relacionados à segurança da informação, tornando a necessidade de proteção das informações uma preocupação recorrente.

Considerada chave e parte da estratégia da organização, a segurança da informação se torna cada vez mais necessária para a proteção de informações e ativos informacionais. Para Beal [2005 p.71 *apud* Santos and Silva 2012] a segurança da informação é “o processo de proteger a informação das ameaças para garantir a sua integridade, disponibilidade e confidencialidade”.

Para que a segurança da informação seja garantida, inúmeras abordagens e práticas são recomendadas na literatura técnica-especializada, incluindo precauções que devem ser tomadas. Análises de riscos, por exemplo, são ações necessárias para mitigar situações que podem ser prejudiciais a organização, identificando assim a possibilidade de ameaças. Quando mapeadas as ameaças, um apontamento de soluções cabíveis é realizado visando minimizar o risco. De acordo com Furnell e Thomson [2009], os usuários são identificados como grande ameaça na implantação de práticas e procedimentos de segurança da informação.

Para Oliveira [2001 p.43 *apud* Spanceski 2004],

Nenhuma área da informática é tão apreciada como a segurança da informação, todo processo de segurança inicia e tem seu término em um ser humano. Segurança não é uma questão técnica, mas uma questão gerencial e humana. Não adianta adquirir uma série de dispositivos de hardware e software sem treinar e conscientizar o nível gerencial da empresa e todos os seus funcionários. Mas, uma ameaça também pode ser considerada todo e qualquer indício de acontecimentos desfavoráveis, sobre um ativo ou pessoa.

As organizações, em geral, planejam ou criam normas de segurança da informação sem considerar as pessoas, o que dificulta a elaboração e a regulamentação, podendo levar a um resultado ineficiente ou inexecutável [LORENS 2007].

Em meio aos desafios e incompreensões existentes para a adoção de medidas e práticas de segurança da informação no meio corporativo, o objetivo deste trabalho é propor hipóteses a partir da avaliação de segurança da informação em uma organização de médio porte que atua no segmento de prestação de serviços em tecnologia da informação, à luz dos requisitos e controles preconizados na Norma NBR ISO/IEC 27002:2013 [ABNT ISO 27002 2013], a qual prevê boas práticas de controles de segurança da informação.

Os conceitos e requisitos especificados na norma ABNT ISO 27002 [2013] são basilares para avaliação pretendida, haja vista a possibilidade de implantação de um sistema de gestão de segurança da informação no contexto organizacional, bem como identificação e mitigação de ameaças. Para tanto, o estudo inicia-se com a identificação das principais

necessidades da organização, buscando reconhecer riscos de segurança da informação, tendo em vista os ativos de maior relevância, face aos objetivos estratégicos estabelecidos. Em seguida, uma análise ampla dos processos e práticas que remetem à segurança da informação é empreendida, em conformidade com os objetivos de controle da norma ABNT ISO 27002 [2013]. Por fim, o resultado alcançado produz subsídios para a definição de hipóteses reveladas a partir de uma visão geral acerca dos pontos críticos associados à garantia e manutenção da segurança da informação corporativa.

2. Segurança da Informação à luz da Norma ABNT ISO 27002

Segundo a norma ABNT NBR ISO/IEC 27002:2013 [ABNT ISO 27002 2013], para se alcançar o que pode ser considerado “Segurança da Informação”, faz-se necessária a “implementação de um conjunto adequado de controles, incluindo políticas, processos, procedimentos, estrutura organizacional e funções de software e hardware”. Para tanto, “estes controles precisam ser estabelecidos, implementados, monitorados, analisados criticamente e melhorados, quando necessário, para assegurar que os objetivos do negócio e a segurança da informação da organização são atendidos”.

Essa norma [ibid.] também recomenda que:

“O valor da informação vai além das palavras escritas, números e imagens: conhecimento, conceitos, ideias e marcas são exemplos de formas intangíveis da informação. Em um mundo interconectado, a informação e os processos relacionados, sistemas, redes e pessoas envolvidas nas suas operações são informações que, como outros ativos importantes, têm valor para o negócio da organização e, conseqüentemente, requerem proteção contra vários riscos.”.

A norma ABNT ISO 27002 [2013] define como Ativo “objeto de ameaças, tanto acidentais como deliberadas, enquanto que os processos, sistemas, redes e pessoas têm vulnerabilidades inerentes”.

A definição e classificação da informação é um requisito mandatório para determinação do nível de proteção das informações. Devido à ocorrência de incidentes de segurança nas empresas, as informações e os ativos informacionais carecem cada vez mais de proteção. Esta requer a implantação de medidas de segurança da informação que vai além de tecnologias computacionais, de modo a assegurar suas propriedades. Para Bunker [2012], embora exista uma grande quantidade de tecnologias da informação destinadas à proteção e segurança dos ativos de informação, não é suficiente para assegurar a efetividade dos mecanismos e instrumentos de segurança da informação.

Os grandes pilares de segurança da informação, conforme preceitua a norma ABNT ISO 27002 [2013], são:

- Confidencialidade – garante que a informação só será acessada por pessoas autorizadas.
- Integridade – garante que a informação alcance seu destino sem algum tipo de modificação.
- Disponibilidade – garante que informações solicitadas devem estar disponíveis para acesso a qualquer momento, de acordo com a requisição do usuário.

A Norma ABNT ISO 27002 [2013] tem como objetivo

Fornecer diretrizes para práticas de gestão de segurança da informação e normas de segurança da informação para as organizações, incluindo a seleção, implementação e o gerenciamento de controles, levando em

consideração os ambientes de risco da segurança da informação da organização.

Encontra-se distribuída em 14 seções de controles de segurança da informação, totalizando 35 objetivos de controles e 114 controles [ibid.] e está circunscrita na família de normas ABNT NBR ISO/IEC 27000, que trata da gestão de segurança da informação.

3. Metodologia

Com o objetivo de mapear e analisar a real situação de uma organização quanto ao cumprimento dos preceitos de segurança da informação, utilizando como parâmetro os 35 objetivos de controle da norma ABNT ISO 27002 [2013], um estudo de caso simples foi empreendido no âmbito de uma empresa, tomando como base a observação das práticas corporativas que contribuem positiva ou negativamente para a gestão de segurança da informação.

O estudo iniciou-se com uma breve pesquisa bibliográfica para balizar a aplicação da ABNT ISO 27002 [2013], buscando a compreensão de suas definições e requisitos. Em seguida, escolheu-se de forma deliberada uma organização de médio porte que atua no segmento de prestação de serviços em tecnologia da informação, com aproximadamente 90 colaboradores diretos, situada na cidade de Goiânia. Nesse contexto organizacional foi promovido um diagnóstico com base na medição do nível de maturidade de cada prática adotada em termos de controles previstos na norma ABNT NBR ISO/IEC 27002:2013. O processo observacional e de análise e discussão consumiu um tempo de aproximadamente 10 meses, concluído no primeiro semestre de 2018.

Para contemplar a avaliação das seções previstas na norma ABNT ISO 27002 [2013], pontos focais na empresa foram selecionados para poder responder as questões propostas no instrumento de coleta, produzindo artefatos de registro de acompanhamento, contendo o preenchimento de checklists de requisitos. Além de anotar o grau de maturidade, que varia de 0 (não existe) a 2 (definido), os artefatos refletiram as seções daquela norma. Por fim, as análises foram feitas considerando todo o arcabouço informacional e os critérios de avaliação adotados.

Cada artefato produzido consistia de uma série de perguntas diretas, correspondendo ao instrumento de coleta principal, com as respectivas respostas, apresentando o seguinte gabarito:

- ‘SIM’, ‘NÃO’ e ‘NÃO SE APLICA (NA)’, para cada controle avaliado;
- breve resumo de controles que possam suportar, ou já suportam, o controle levantado;
- breve resumo das dificuldades/impossibilidades que possam vir a atender, ou já atendem o requisito;
- se existe ou não a intenção/projeto de implantar uma política/controle sobre o que foi questionado, caso não o tenha;
- o nível de maturidade daquele controle, o qual é classificado como 0 para ‘Não Existe’, 1 para ‘Inicial’ e 2 para ‘Definido’.

4. Resultado e Discussão

O grau de maturidade mapeado no checklist está centrado no diagnóstico dos controles associados aos objetivos de controle previstos na ABNT ISO 27002 [2013]. O grau de maturidade é atribuído de acordo com as respostas obtidas em cada domínio. Ao final é contabilizada a quantidade de cada nível e essa informação é incluída na planilha ‘Nível de Maturidade’, como ilustrado na Figura 1.

Domínios de Segurança de Informação	Total de Perguntas	Nível de Maturidade			
		Não se aplica	0 - Não existente	1 - Inicial	2 - Definido
Seção 5 - Política de Segurança	32				
Seção 6 - Organização da Segurança da Informação	33				
Seção 7 - Segurança em recursos humanos	33				
Seção 8 - Gestão de ativos	44				
Seção 9 - Controle de acesso	66				
Seção 10 - Criptografia	21				
Seção 11 - Segurança física e do ambiente	88				
Seção 12 - Segurança nas operações	92				
Seção 13 - Segurança nas comunicações	34				
Seção 14 - Aquisição, desenvolvimento e manutenção de sistemas	59				
Seção 15 - Relacionamento na cadeia de suprimento	57				
Seção 16 - Gestão de incidentes de segurança da informação	33				
Seção 17 - Aspectos da segurança da informação na gestão da continuidade do negócio	12				
Seção 18 - Conformidade	40				

Figura 1. Nível de Maturidade [Adaptado de ABNT ISO 27002 2013]

Com a obtenção de todas as respostas, pode-se classificar o nível de maturidade de cada seção, mapeando assim de forma ampla as principais conformidades (e não conformidades) em segurança da informação naquela organização.

A título de ilustração, o score obtido nas duas primeiras seções está representado nas Figuras 2 e 3. A primeira Seção analisada foi a ‘Seção 5 – Política de segurança da informação’, na qual foi realizado um total de 32 perguntas e apenas duas tiveram respostas ‘SIM’, como mostra a Figura 2. As perguntas que resultaram nessas respostas foram: ‘A direção apoia as diretrizes deste documento?’ e ‘A organização adota práticas de Backup?’.

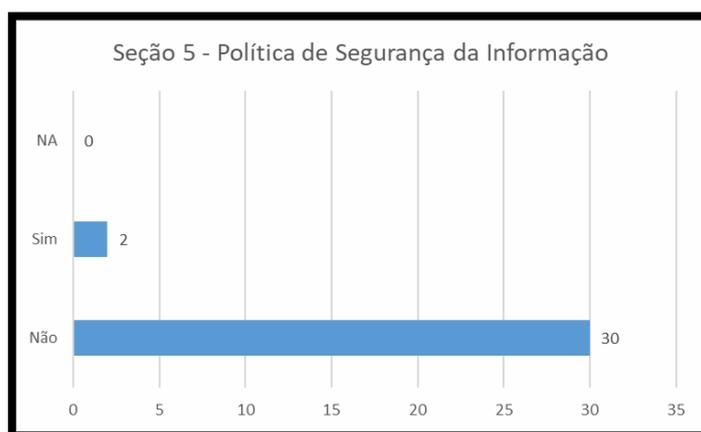


Figura 2. Respostas quanto aos controles da Seção 05 – ABNT ISO 27002

A predominância de respostas ‘Não atende’ aos controles da Seção 05 é agravada pela inexistência de uma Política de Segurança da Informação. Em seguida analisou-se a ‘Seção 06 – Organização da Segurança da Informação’, cujo resultado está sintetizado na Figura 3.

As demais seções passaram pelo mesmo processo de avaliação, culminando com o diagnóstico de maturidade sintetizado na Figura 4. Vale ressaltar que os níveis de maturidade obtidos refletem o estágio atual em termos de práticas adotadas por parte dos colaboradores, das medidas de controle empregadas e/ou institucionalizadas, bem como os hábitos e costumes que caracterizam a cultura de segurança da informação no âmbito da unidade de análise.

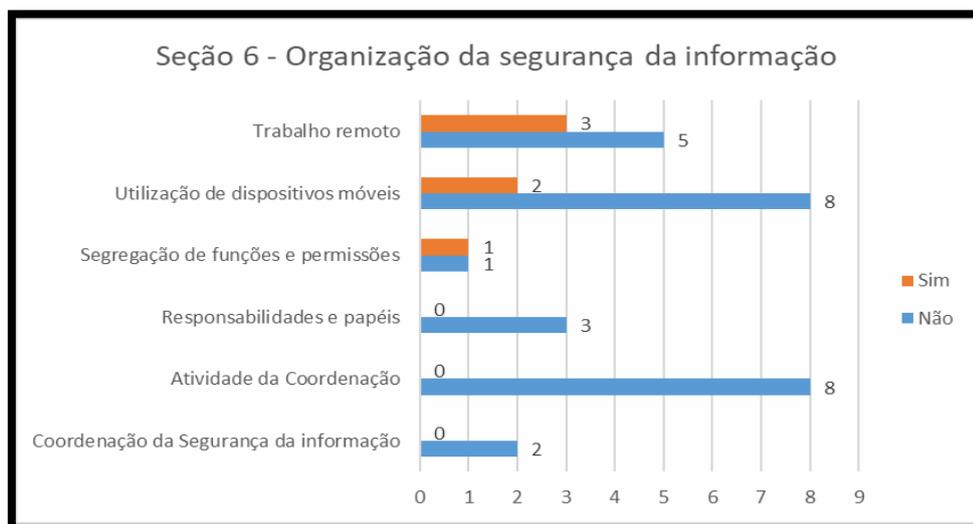


Figura 3. Respostas quanto aos controles da Seção 06 – ABNT ISO 27002 [2013]

A organização não possui nenhum procedimento ou prática definida que diz respeito às Seções ‘Seção 8 – Gestão de ativos’, ‘Seção 10 – Criptografia’ e ‘Seção 15 – Relacionamento na cadeia de suprimentos’. Por outro lado, a ‘Seção 11 – Segurança física e do ambiente’ e ‘Seção 12 – Segurança nas operações’ foram as que mais tiveram representatividade quanto ao nível de maturidade ‘Inicial’. As demais seções se limitaram há poucos controles no nível ‘Inicial’, conforme ilustrado na Figura 4.

O nível de maturidade ‘Inicial’ sugere que a empresa possui algumas práticas de uso nas diretrizes abordadas, mas, não necessariamente dispõe de políticas específicas para implementá-las. Por fim, há controles que não se aplicam à organização observada, a exemplo do que ocorre com algumas diretrizes contidas na ‘Seção 11 – Segurança física e do ambiente’, ‘Seção 12 – Segurança nas operações’ e ‘Seção 18 – Conformidade’.

Observou-se também que, apesar de ser uma empresa de médio porte e com taxas de crescimento crescentes nos últimos anos, não existia uma preocupação iminente quanto às ameaças até então identificadas, seja porque não havia orçamento previsto para segurança da informação, seja porque ainda não há nenhuma prática no nível de maturidade ‘definido’.

A sensação de falta de consciência e incompreensão por parte dos colaboradores da empresa pode ser percebida quando o termo segurança da informação é constantemente confundido com “segurança do ambiente”, “segurança elétrica”, ou mesmo “segurança corporativa”.

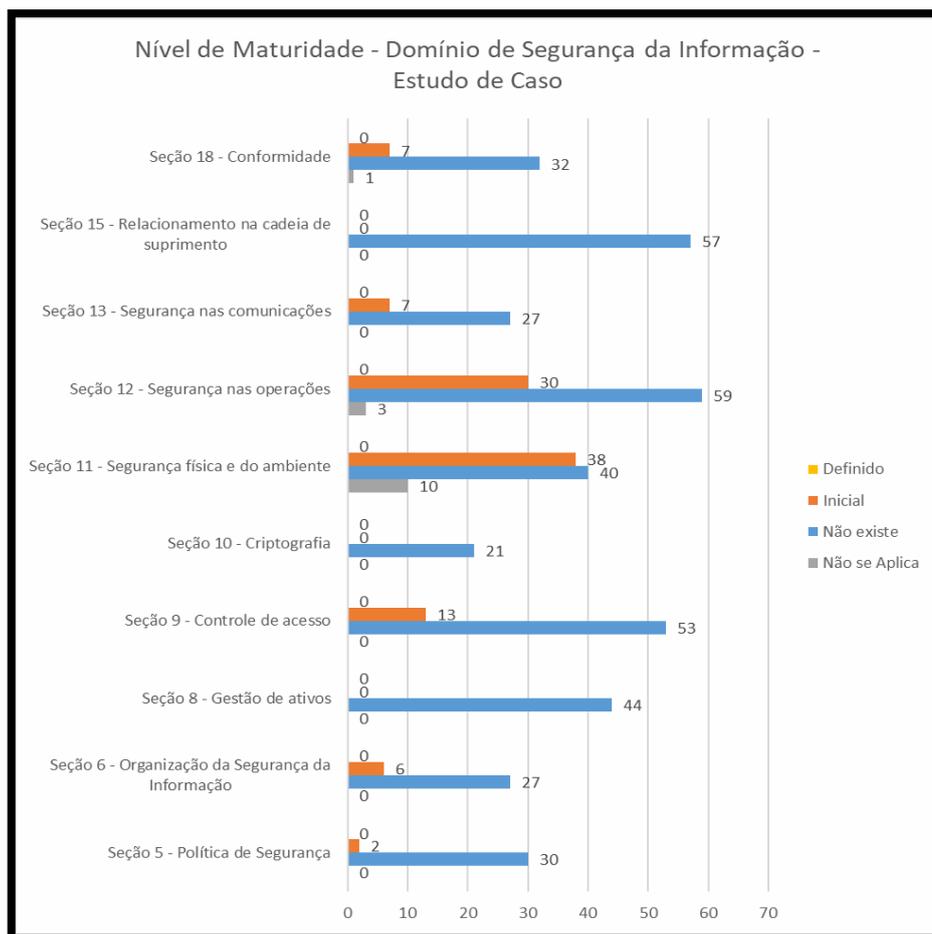


Figura 4. Nível de Maturidade – Estudo de caso [O Autor 2018]

4.1. Discussão

A avaliação das práticas e das medidas adotadas no âmbito da organização estudada, à luz dos controles recomendados pela ABNT ISO 27002 [2013], permitiu revelar um quadro de graves inconformidades, vulnerabilidades e incompreensões no que tange à segurança da informação.

Considerando que a função básica da área de segurança da informação é a proteção dos ativos de informação, por meio da minimização dos riscos a níveis aceitáveis [FERREIRA AND ARAÚJO 2008] e considerando que as impressões alcançadas na unidade de análise, fruto das observações participante e da aplicação dos instrumentos de coleta junto aos pontos focais da organização, permitiu elucidar o estado de coisas que orbitam os sistemas e as tecnologias de informação, uma série de conjecturas pode ser extraída, ainda que a confirmação esteja fora da alçada deste estudo:

- a implementação de controles por meio de processos para a proteção das informações é uma das formas de prover segurança aos sistemas de informação;
- o passo que antecede qualquer medida de controle de segurança da informação diz respeito à realização da classificação da informação, de modo a proteger as mais críticas;
- é preciso definir papéis, responsabilidades e responsabilização no âmbito da gestão de segurança da informação para iniciar o processo de implantação de segurança da informação na organização;

- os usuários de sistemas de informação tendem a não observar as práticas de segurança da informação por desconhecimento ou desinteresse;
- os profissionais de TI tendem a priorizar medidas associadas exclusivamente a tecnologias da informação.

Ainda que as constatações acerca do nível de maturidade das práticas e controles de segurança da informação estejam condicionadas a um modelo de referência reconhecido e consagrado pela norma ABNT ISO 27002 [2013], assumir as suposições retrocitadas implica na adoção de um modelo de ciclo de vida para implantação de um Sistema de Gestão de Segurança da Informação (SGSI) que contemple as várias fases que devem ser executadas continuamente para propiciar melhorias e aumento gradativo do nível de maturidade da gestão de segurança da informação. O Quadro 1 apresenta uma proposta de modelo de ciclo de vida para implantar o SGSI, inspirada na norma ABNT NBR ISO/IEC 27001:2013, que trata da estruturação de um SGSI.

Quadro 1. Ciclo de Vida para Implantação do SGSI

Fase I - Plan	Fase II - Do	Fase III - Check	Fase IV - Act
<ul style="list-style-type: none"> ○ Estruturação do SGSI ○ Plano Diretor de Segurança ○ Diagnóstico de Segurança ○ Avaliação, Tratamento dos Riscos e Seleção dos Controles de Segurança ○ Declaração de Aplicabilidade (<i>Statement of Applicability</i>) 	<ul style="list-style-type: none"> ○ Política Corporativa de Segurança da Informação ○ Classificação da Informação ○ Plano de Continuidade dos Negócios e de TI ○ Treinamento e Conscientização ○ Implementação dos Controles Especificados na Declaração de Aplicabilidade 	<ul style="list-style-type: none"> ○ Monitoração dos Controles de Segurança ○ Gestão de Incidentes ○ Revisão do nível de risco residual ○ Auditoria Interna do SGSI 	<ul style="list-style-type: none"> ○ Implementação de melhorias ○ Ações Corretivas e Preventivas ○ Comunicação das Ações e Resultados para Alta Administração e Partes Interessadas ○ Assegurar que as Melhorias foram Implementadas e Atenderam as Expectativas

Fonte: Ferreira e Araújo [2008]

O SGSI envolve preliminarmente a identificação de fatores críticos de sucesso e o mapeamento de dificuldades para o processo de implantação da gestão da segurança da informação, com base na experiência da própria organização [FERREIRA E ARAÚJO 2008].

5. Considerações Finais

A pesquisa permitiu diagnosticar as práticas e os procedimentos de segurança da informação adotados no âmbito organizacional, em conformidade com objetivos de controle da ABNT ISO 27002 [2013]. A empresa, objeto de análise deste estudo, revelou que os controles observados apresentam predominantemente baixo nível de maturidade, refletindo as vulnerabilidades e ameaças quanto aos preceitos de confidencialidade, integridade e disponibilidade.

Ainda que incipiente, os resultados obtidos sugerem que diversas abordagens e estratégias de implantação são passíveis de serem adotadas naquele contexto, desde que observem minimamente as suposições verificadas nesta pesquisa. Além disso, os artefatos produzidos revelam um quadro de conformidades (e inconformidades) em termos de segurança da informação, que poderão contribuir para a definição da estratégia de implantação de um sistema de gestão de segurança da informação, já que alguns pressupostos devem ser observados para que não haja ambiguidades e interpretações equivocadas por parte dos gestores que estarão à frente desse processo.

Esta pesquisa terá continuidade no escopo do projeto de conclusão de curso do primeiro autor, cujas hipóteses aqui levantadas subsidiarão a nova etapa do trabalho, pautado pelo sistema de gestão de segurança da informação, de modo a ampliar os horizontes de avaliação para a proposta do plano de ação a ser elaborado, levando em consideração aspectos de viabilização da aplicação no contexto corporativo. Nesse sentido, apropriar-se da norma ABNT NBR ISO/IEC 27001:2013 e de outras referências que abordam o sistema de gestão de segurança da informação é um dos requisitos-chaves para a sequência do trabalho.

6. Referências

- ABNT ISO 27002. (2013). ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR/ISO/IEC 27002:2013 tecnologia da informação – técnicas de segurança – código de prática para controles de segurança da informação.
- BARBOSA, R. R. (1997). Monitoração ambiental: uma visão interdisciplinar. *Revista de Administração*, São Paulo: v.32, n.4, p. 42-53, out./dez.
- BEAL, A. (2005). Segurança da informação: princípios e melhores práticas para a proteção dos ativos de informação nas organizações. São Paulo: Atlas.
- BUNKER, G. (2012). “Technology is not enough: taking a holistic view for information assurance”. In: *Information Security Technical Report*, (17): 19-25.
- FERREIRA, F. N. F.; ARAÚJO, M. T. (2008). Política de Segurança da Informação – Guia Prático para Elaboração e Implementação, Editora Ciência Moderna, Rio de Janeiro.
- FURNELL, S.; THOMSON, K. L. (2009). From Culture to disobedience: recognising the varying user acceptance of IT security. *Computer Fraud & Security*, (2), 5-10.
- LAUDON, Kenneth C.; LAUDON, Jane P. (1999). Sistema da Informação com Internet. Rio de Janeiro: LTC.
- LORENS, E. M. (2007). Aspectos normativos da Segurança da Informação: um modelo de cadeia de regulamentação, Dissertação de Mestrado, Universidade de Brasília, Brasília-DF, Brasil.
- OLIVEIRA, W. (2001). Segurança da Informação. Florianópolis: Visual Books.
- SANTOS, D. L. R.; SILVA, R. M. S. (2012). Segurança da Informação: a Norma ISO/IEC 27000 e ISO/IEC 27001. Trabalho de Segurança de Informação do MCI 2012/2013. Universidade do Porto, Faculdade de Engenharia.
- SMITH, A.; JONES, B. (1999). On the complexity of computing. In *Advances in Computer Science*, pages 555–566. Publishing Press.
- SPANCESKI, F. R. (2004). Política de Segurança da Informação: desenvolvimento de um modelo voltado para instituições de ensino. Monografia do Trabalho de Conclusão de Curso em Sistemas de Informação.

TURBAN, E.; MCLEAN, E.; WETHERBE, J. (2004). Tecnologia da informação para gestão. Transformado os negócios da economia digital. [S.l.: s.n.]

Alianças Defensivas Globais em Grafos Roda e em Prismas Complementares

Rafael de Castro Freitas¹, Elisângela Silva Dias¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Câmpus Samambaia – 74690-900 – Goiânia – GO – Brasil

{rafaelcastro, elisangela}@inf.ufg.br

Abstract. *Graphs are very useful and it helps solving problems. A relatively new that has been studied is alliances in graphs. In this paper, it is studied the defensive alliances, which can be used as a mathematical model in web communities, social medias, chain food, etc. It was found results in wheel graphs and complementary prisms of complete graphs.*

Resumo. *Grafos são de grande utilidade para auxiliar na solução de problemas. Um tema relativamente novo é o de alianças em grafos. Neste trabalho, são estudadas as alianças defensivas, que podem ser utilizadas como modelo matemático em comunidades web, redes sociais, cadeias alimentares, etc. Foram encontrados resultados em grafos roda e em prismas complementares de grafos completos.*

1. Introdução

Grafos podem ser utilizados para modelar e facilitar a visualização de diversas situações da vida real. Por exemplo, pode-se utilizar um grafo para encontrar os menores caminhos para atravessar de um ponto da cidade ao outro. Isso faz com que a área de Teoria dos Grafos seja bem explorada, pois a sua aplicação é extensa em diversas áreas, como a de negócios, tráfego, distribuição de produtos, entre outros.

Um conceito recente e que está sendo bastante estudado é o de alianças em grafos, que foi introduzido por Kristiansen, Hedetniemi e Hedetniemi em 2003 e publicado em 2004 no artigo *Alliances in Graphs* [Hedetniemi and Kristiansen 2004]. As principais classificações de alianças são: defensivas, ofensivas e poderosas. As alianças defensivas, por exemplo, servem de modelo matemático para vários problemas teóricos e práticos em diversas áreas do conhecimento, como a de comunidades *web*, redes sociais, cadeias alimentares, estrutura de dados, etc. O que faz com que muitos pesquisadores [Shafique 2004, Brigham and Hedetniemi 2009, Bermudo and Yero 2010, Dias 2012, Yero and Rodríguez-Velázquez 2013, Silva 2015, Gonçalves 2016] estejam levantando estudos nesse tema é que o cálculo do número da aliança em grafos pertence à classe de problemas NP-completos [Shafique 2004].

Além disso, também é um problema complicado encontrar uma das alianças em um grafo qualquer. Quando o grafo é pequeno, a tarefa é simples. Entretanto, a complexidade aumenta à medida que o grafo cresce, e também quando trabalhamos com alianças globais, que afetam todo o grafo. Sendo assim, o objetivo desta pesquisa é alcançar novos resultados teóricos em alianças defensivas globais nos grafos roda e em prismas complementares.

2. Definições

Para uma melhor compreensão do trabalho, é necessário entender algumas definições. Grafos são estruturas que contêm conjuntos de vértices e arestas, tal que estes conjuntos são diferentes de vazio, podem ser finitos ou infinitos e são disjuntos. Denotamos o grafo G por $G = (V(G), E(G))$, tal que $V(G)$ é o conjunto de vértices e $E(G)$ é o conjunto de arestas, que consiste em um conjunto de pares de vértices. Quando duas arestas nunca conectam os mesmos vértices, o grafo é chamado de *grafo simples*. Dizemos que um grafo é *não orientado* se $uv = vu$. Caso contrário, se $uv \neq vu$, temos um grafo *orientado*. Um *conjunto dominante* (CD) é um subconjunto $D \subseteq V(G)$, tal que cada vértice pertencente a $V(G) \setminus D$ é adjacente a pelo menos um vértice em D . Neste trabalho, considera-se grafos simples, não orientados e *finitos*, que são os grafos que possuem um conjunto finito de vértices. Tem-se que $|V(G)| = n$ e $|E(G)| = m$.

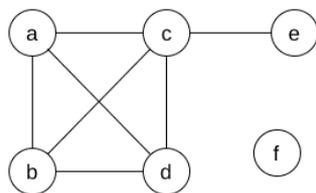


Figura 1. Exemplo de um grafo G qualquer.

Sejam u e v dois vértices em um grafo G , eles são ditos *adjacentes* ou *vizinhos* se forem *extremidades* de uma aresta a . Diz-se também que a conecta os vértices u e v . Se uma aresta a conecta dois vértices distintos (u, v) , julga-se que $a = uv$ e $a \in E(G)$. A *vizinhança* entre os vértices de um grafo pode ser aberta ou fechada. *Vizinhança aberta*, denotada por $N(v)$, ocorre quando considera-se apenas os vizinhos de um vértice v . *Vizinhança fechada*, denotada por $N[v]$, ocorre quando considera-se a vizinhança de v mais ele próprio, ou seja, v também é incluso no conjunto.

O *grau de um vértice* v , denotado por $g(v)$, é o número de arestas que incidem neste vértice. Se um vértice possui grau zero ele é chamado de vértice *isolado*, sendo assim ele também não possui nenhum vizinho. Um vértice é dito *pendente* se ele possuir grau igual a 1. Analisando o grafo G na Figura 1, tem-se que $g(a) = 3$, $g(e) = 1$ e $g(f) = 0$. Então, o vértice e é pendente e o vértice f é isolado.

Muitas vezes é possível se deparar com grafos complexos e extensos, mas talvez será necessário trabalhar somente com uma parte dele. É possível trabalhar apenas com uma parte do grafo original, desde que as extremidades das arestas remanescentes sejam mantidas. Um grafo menor obtido a partir de outro é chamado de *subgrafo*. Um subgrafo de G é um grafo $G' = (V(G'), E(G'))$ se e somente se $V(G') \subseteq V(G)$ e $E(G') \subseteq E(G)$. Denota-se ainda que um subgrafo *induzido* de G consiste em um subconjunto V' não vazio, tal que $V' \subseteq V(G)$, e E' consiste de todas as arestas de $E(G)$, cujo os extremos pertencem ao V' . Na Figura 2, é possível ver um grafo com um de seus possíveis subgrafos.

Grafos podem ser representados de várias maneiras, gerando estruturas diferentes para cada uma delas. Quando isso ocorrer, define-se *isomorfismo de grafos*. Sejam $G_1 = (V(G_1), E(G_1))$ e $G_2 = (V(G_2), E(G_2))$ dois grafos. Eles serão *isomorfos* se e somente

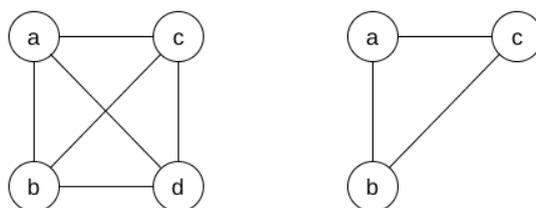


Figura 2. Grafo e um de seus subgrafos.

se existir uma função bijetora f de $V(G_1)$ para $V(G_2)$, tal que se a e b são vértices e são adjacentes em G_1 , então $f(a)$ e $f(b)$ são adjacentes em G_2 para todo a e b em $V(G_1)$. Em outras palavras, dois grafos são isomorfos se os vértices destes mesmos grafos forem biunívocos, preservando a relação de adjacência. Portanto, nos grafos simples, a relação de isomorfismo só será dada se o número de vértices, número de arestas e os graus dos vértices forem iguais nos dois grafos. Entretanto, não há garantia de que todos os grafos são isomorfos baseando-se apenas nestas premissas. Não se conhece ainda um conjunto útil de invariantes para determinar isomorfismo entre grafos simples [5]. Na Figura 3, é possível observar que os vértices a, b, c, d, e e f em G_1 correspondem aos vértices z, w, x, y, u e v em G_2 , respectivamente. Assim sendo, tem-se que $f(a) = z, f(b) = w, f(c) = x, f(d) = y, f(e) = u, f(f) = v$.



Figura 3. Exemplo de isomorfismo entre um grafo G_1 e um G_2 .

Caminho, denotado por P_n , é uma sequência de $n - 1$ arestas entre um vértice inicial v_0 e um vértice final v_n . O caminho pode ser um *ciclo*, denotado por C_n , se o início e o fim for no mesmo vértice, ou seja, se $v_0 = v_n$. *Caminho simples* é um caminho que não contém a mesma aresta mais de uma vez. Pode se observar ver que os vértices 1, 5, 6, 3 no grafo da Figura 4, formam um caminho simples de comprimento 3, pois $\{1, 5\}, \{5, 6\}$ e $\{6, 3\}$ são arestas de G . Observe que os vértices 4, 5, 3, 1 não formam um caminho, pois $\{5, 3\}$ não é uma aresta. Tem-se ainda que os vértices 1, 2, 5, 1 formam um ciclo de comprimento 3, pois $\{1, 2\}, \{2, 5\}$ e $\{5, 1\}$ são arestas e este caminho começa e termina em 1.

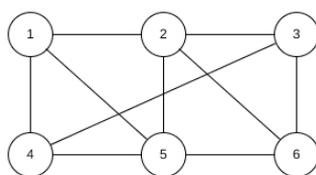


Figura 4. Um grafo G simples.

Um grafo é *completo*, denotado por K_n , se quaisquer dois de seus vértices distin-

tos são adjacentes. Na Figura 5, tem-se um exemplo de um grafo completo. Uma *clique* é um subconjunto de vértices C , tal que $C \subseteq V(G)$, sendo que para cada dois vértices $x, y \in C$, existe uma aresta conectando-os. Em outras palavras, uma clique é um subgrafo induzido de G , que é um grafo completo.

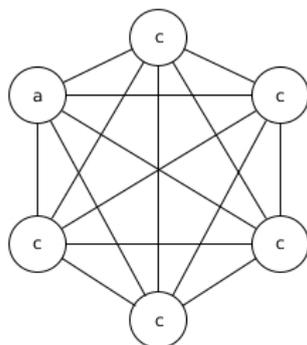


Figura 5. Um grafo K_6 .

Um grafo roda, denotado por W_n , é um ciclo com um vértice central. Observe que este vértice central domina todos os demais. Na Figura 6, temos um exemplo de um grafo roda.

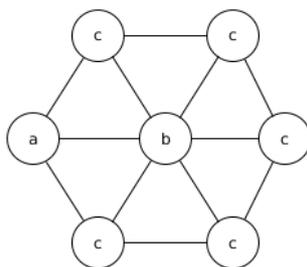


Figura 6. Um grafo W_7 .

O conceito de alianças em grafos é relativamente novo, tendo surgido em 2003. De modo geral, o conceito de aliança pode ser definido como um conjunto de pessoas que possuem interesses em comum que se unem para um ataque ou para uma defesa. Na teoria dos grafos, esta definição é semelhante. *Aliança* em grafos, de modo bem simples, é um conjunto de vértices com algumas propriedades coletivas. Assim como na vida real, há alguns tipos diferentes de alianças em grafos, a saber: alianças defensivas, que podem ser fortes e/ou globais; as alianças ofensivas, que também podem ser fortes e/ou globais; e as alianças poderosas, que podem ser globais.

Seja $S \subseteq V(G)$. S é uma *aliança defensiva* (AD) se e somente se $\forall v \in S, |N[v] \cap S| \geq |N[v] \setminus S|$. Em outras palavras, todo vértice $v \in S$ tem no máximo um vizinho a mais em $V(G) \setminus S$ do que em S . Na Figura 8, é possível ver um exemplo de aliança defensiva.

Uma *aliança defensiva forte* (ADF) ocorre se e somente se $\forall v \in S, |N[v] \cap S| > |N[v] \setminus S|$. Neste caso, a inequação é estrita e todo vértice $v \in S$ é fortemente defendido, ou seja, um vértice no conjunto S tem mais vizinhos em S , incluindo ele próprio, do que em $V(G) \setminus S$. Na Figura 7, os vértices destacados formam uma aliança defensiva forte.

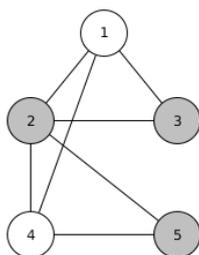


Figura 7. Exemplo de aliança defensiva forte e aliança poderosa.

Uma aliança defensiva é *global* (ADG) quando todo vértice em $V(G)\setminus S$ é adjacente a pelo menos um membro da aliança S , tornando S um conjunto dominante. Pode-se ver um exemplo na Figura 8 de uma aliança defensiva global.

Uma *aliança defensiva forte global* (ADFG) ocorre quando todos os vértices em $V(G)\setminus S$ de uma ADF é adjacente a pelo menos um membro da aliança S , tornando o conjunto S dominante.

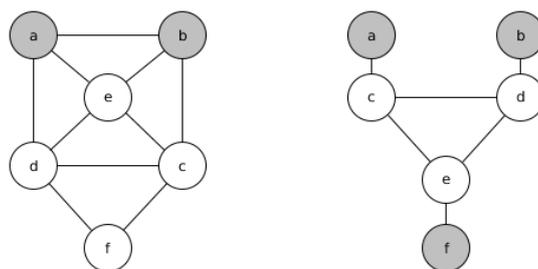


Figura 8. Exemplo de aliança defensiva e aliança defensiva global.

Se $S \subseteq V(G)$ e S é um conjunto não vazio de vértices, tem-se a ocorrência de uma *aliança ofensiva* (AO) se e somente se $\forall v \in \partial S, |N[v] \cap S| \geq |N[v] \setminus S|$. Em outras palavras, todo vértice na fronteira de S tem pelo menos um vizinho a mais em S do que em $V(G)\setminus S$. Na Figura 9, é possível ver um exemplo de aliança ofensiva.

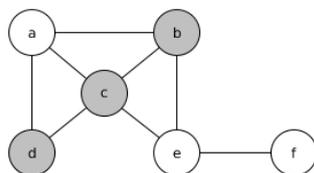


Figura 9. Exemplo de uma aliança ofensiva.

Uma *aliança ofensiva global* (AOG) segue a mesma definição da ADG, ou seja, uma AOG ocorre quando todo vértice em $V(G)\setminus S$ é adjacente a pelo menos um membro de S , tornando S um conjunto dominante. Na Figura 10, os vértices marcados formam uma AOG em um grafo cúbico.

Uma aliança ofensiva é *forte* (AOF) se $v \in \partial S, |N[v] \cap S| > |N[v] \setminus S|$. Neste caso, a inequação também é estrita e o conjunto é fortemente defendido. Desse modo,

pode-se dizer que toda aliança ofensiva forte é uma aliança ofensiva, todavia, nem toda aliança ofensiva é uma aliança ofensiva forte. Segue um exemplo de aliança ofensiva forte na Figura 10.

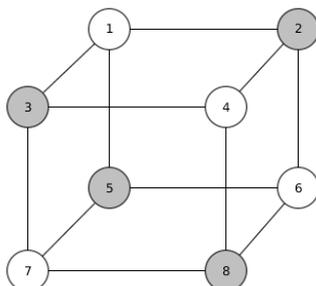


Figura 10. Exemplo de aliança ofensiva global e aliança ofensiva forte.

Na Figura 7, tem-se a formação de uma aliança defensiva e ofensiva no conjunto $S = \{2, 3, 5\}$, pois os vértices possuem mais vizinhos em S do que fora de S , levando em consideração a vizinhança fechada de vértice. Nessas condições, quando uma aliança é defensiva e ofensiva ao mesmo tempo, ela é definida de *aliança poderosa* (AP).

Uma *aliança poderosa global* (APG) é definida quando há um conjunto dominante, ou seja, quando todo vértice em $V(G) \setminus S$ é adjacente a pelo menos um membro da aliança S . Na Figura 11, o conjunto $S = \{1, 2, 5, 6, 7, 8, 11, 12\}$ é tanto uma aliança poderosa, quanto um conjunto dominante.

O número mínimo da k -aliança defensiva em um grafo é denotado por $a_k(G)$; da AD é denotado por $a(G)$; da ADF é denotado por $\hat{a}(G)$; da ADG é denotado por $\gamma_a(G)$; da ADFG é denotado por $\gamma_{\hat{a}}(G)$; da AO é denotado por $a_o(G)$; da AOF é denotado por $\hat{a}_o(G)$; da AOG é denotado por $\gamma_o(G)$; da AP é denotado por $a_p(G)$; da APG é denotado por $\gamma_{a_p}(G)$.

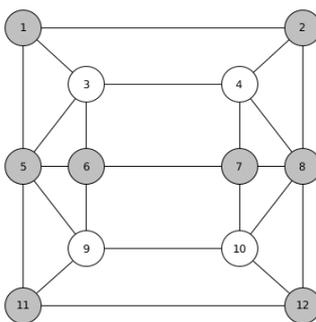


Figura 11. Exemplo de aliança poderosa global.

2.1. Produto Cartesiano de Grafos

O *produto cartesiano* $G_1 \times G_2$ de dois grafos G_1 e G_2 resulta em um outro grafo com um conjunto de vértices $V(G_1 \times G_2) = V(G_1) \times V(G_2)$. Os vértices (v_1, v_2) e (u_1, u_2) são vizinhos se seguir a condição de exclusividade:

1. ou $[v_1 u_1 \text{ e } v_2 = u_2]$;

2. ou $[v_2u_2$ e $v_1 = u_1]$.

Segue um exemplo de produto cartesiano na Figura 12.

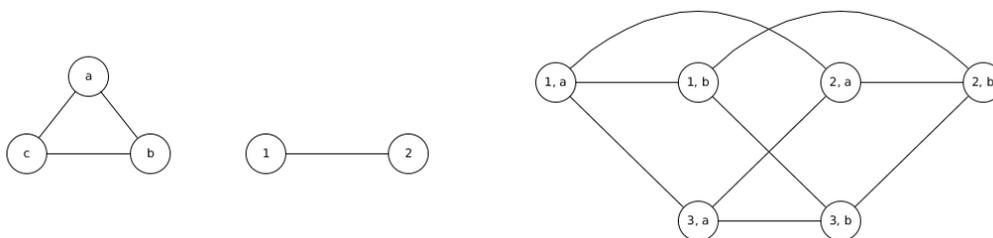


Figura 12. Produto cartesiano $C_3 \times K_2$.

2.2. Prisma complementar

Denota-se por $\overline{G} = (V(\overline{G}), E(\overline{G}))$ o *complemento* de um grafo G em que ambos possuem o mesmo conjunto de vértices e \overline{G} possui um conjunto de arestas complementares. Em outras palavras, se uma aresta uv existir em G , os vértices u e v não são vizinhos em \overline{G} e vice-versa.

O *prisma complementar*, denotado por $G\overline{G}$, é o emparelhamento perfeito de um grafo G com o seu complemento. O grafo resultante é a união disjunta de $G \cup \overline{G}$, emparelhando os vértices correspondentes com suas arestas complementares. Pode-se ver o grafo prisma complementar $C_4\overline{C_4}$ na Figura 13.

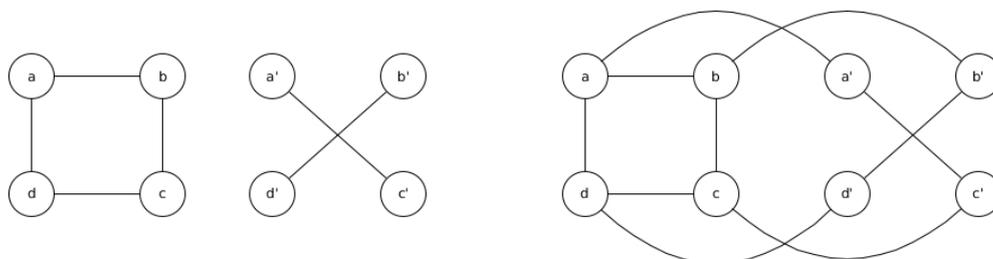


Figura 13. Grafo prisma complementar $C_4\overline{C_4}$.

3. Resultados

Foi encontrado que para grafos cordais, o número da aliança ofensiva global é um problema que se encontra na classe *NP-completa* [Dourado and Szwarcfiter 2014].

Para grafos roda, foi provado os seguintes resultados.

Teorema 1. Para um grafo roda W_n , com $n \geq 4$, tem-se que:

$$\gamma_a(W_n) = \begin{cases} \lfloor \frac{n}{2} \rfloor + 1, & \text{se } n \equiv 3 \pmod{4}, \\ \lfloor \frac{n}{2} \rfloor, & \text{caso contrário.} \end{cases}$$

Demonstração. Seja $V_e(W_n)$ o subconjunto de vértices externos de W_n , $v \in V_e(W_n)$, e S o conjunto que forma ADG, com $S \subseteq V_e$. Como o grafo roda é 3-regular, exceto pelo vértice central, cada v possui 3 vizinhos, dois vértices externos adjacentes mais o vértice central, que é comum a todos. Pela definição de aliança defensiva, sabe-se que para todo vértice $v \in S$, $|N[v] \cap S| \geq |N[v] \setminus S|$. Ou seja, cada um dos vértices $v \in V_e$ precisa de pelo menos mais um vizinho para que seja formado uma AD.

Note que os vértices externos do grafo roda formam um grafo ciclo C_{n-1} , já que desconsidera-se o vértice central. Neste contexto, é possível concluir que a cardinalidade do conjunto dominante para o grafo W_n é o mesmo do grafo C_n , isto é, $\gamma(W_n) = \lfloor \frac{n+1}{3} \rfloor$. Para determinar o conjunto dominante, não é necessário a exigência de ter pelo menos mais um vizinho da AD, uma vez que cada v consegue dominar dois vizinhos, e também o vértice central. Tendo isto em vista, que toda ADG também é um conjunto dominante, é preciso de no mínimo $\lfloor \frac{n}{2} \rfloor$ vértices no grafo, dispostos em $\lfloor \frac{n}{2} \rfloor$ pares de vértices. Para grafos em que $n \equiv 3 \pmod{4}$, é preciso acrescentar mais um vértice a um dos pares de vértices, totalizando $\lfloor \frac{n}{2} \rfloor + 1$, visto que não é possível obter um número exato de pares v . \square

Teorema 2. Para um grafo roda W_n , com $n \geq 6$, tem-se que:

$$\gamma_a(W_n) = \begin{cases} 2 \cdot \gamma(W_{n-\lfloor \frac{\gamma_a(W_{n-2})}{2} \rfloor}) - 1, & \text{se } n \equiv 2 \pmod{4}, \\ 2 \cdot \gamma(W_{n-\lfloor \frac{\gamma_a(W_{n-2})}{2} \rfloor}), & \text{caso contrário.} \end{cases}$$

Demonstração. Seja W'_n um subgrafo induzido de W_n formado pelo subconjunto de vértices externos de W_n , então $V'_e(W'_n)$. Observe que W'_n é o grafo C_n . Portanto, o conjunto dominante de $\gamma(W'_n) = \gamma(C_n)$. O conjunto dominante de um grafo roda é dado por $\gamma(W_n) = 1$, uma vez que o vértice central consegue dominar todos os vértices, isso é trivial. Porém, se for considerado o conjunto dominante de W_n , não se obtém o valor mínimo de uma ADG todas as vezes, isto porque deve se considerar o vértice central mais a metade dos vértices externos para que seja formada uma ADG, obtendo $\gamma_a(W_n) = 1 + \lfloor \frac{|V(W_{n-1})|}{2} \rfloor$.

Seja D o conjunto dominante de W'_n , e seja S o conjunto que forma ADG em W'_n . Como foi dito no Teorema 1, um vértice $v \in D$ consegue dominar 2 vizinhos mais o vértice central. Entretanto, isso não é suficiente para formar uma ADG. Ao pegar a metade do conjunto S de dois grafos anteriores de W_n e subtrair o resultado da quantidade n de vértices, será obtido o grafo anterior com um n equivalente. Tendo o conjunto dominante desse grafo anterior em mãos, é possível obter uma ADG adicionando mais um vértice em cada $v \in D$, formando pares. Observe que ao fazer isso, a quantidade de vértices é dobrada. Esta é uma construção que garante uma ADG mínima.

Em outras palavras, dado um W_n qualquer, a metade dos vértices do conjunto que forma ADG de dois grafos anteriores, $\frac{\gamma_a(W_{n-2})}{2}$, menos o próprio n , $\frac{\gamma_a(W_{n-2})}{2} - n = x$, retorna o grafo anterior equivalente ao W_n , representado por W_{n-x} . Assim, o conjunto S é obtido ao dobrar a quantidade de vértices do conjunto dominante de W_{n-x} , obtendo a fórmula $2 \cdot \gamma(W_{n-\lfloor \frac{\gamma_a(W_{n-2})}{2} \rfloor})$.

Para grafos com $n \equiv 2 \pmod{4}$ vértices, ao dobrar o conjunto D do grafo anterior equivalente, a quantidade de vértices ultrapassa uma unidade, já que sempre obteremos um número par de vértices. Porém, o conjunto S deixa de ser mínimo. Como foi demonstrado no Teorema 1 ao retirar um vértice, deixando o vértice isolado junto a um par, ainda é possível obter uma ADG, mínima, no grafo W_n . Por esta razão, tem-se que $2 \cdot \gamma(W_{n-\lfloor \frac{\gamma_a(W_{n-2})}{2} \rfloor}) - 1$. \square

Teorema 3. Para o grafo $K_n \overline{K_n}$, com $n \geq 2$, tem-se que: $\gamma_a(K_n \overline{K_n}) = n$.

Demonstração. Seja G o grafo $K_n \overline{K_n}$. Observe que G é também um grafo split, já que o conjunto de vértices $V(K_n)$ forma uma clique, e o conjunto $V(\overline{K_n})$ forma um conjunto independente. No subgrafo $\overline{K_n}$ tem-se um conjunto de vértices folhas, ou seja, cada $v \in V(\overline{K_n})$ tem apenas um vizinho, que são os vértices pertencentes à clique. Então, para se ter uma ADG, é preciso pegar todo o conjunto $V(\overline{K_n})$, obtendo os n vértices.

No subgrafo K_n todos os vértices são fortemente conectados. Sendo assim, é possível obter uma ADG com $\lceil \frac{n}{2} \rceil$ vértices. Porém, esse conjunto não é dominante, uma vez que cada $v \in V(K_n)$ possui uma folha. Por conta disto, metade do grafo sempre ficará descoberta. Portanto, para se obter uma ADG no subgrafo $V(K_n)$, também é preciso de n vértices. \square

4. Conclusões

No decorrer deste trabalho, foram estudados alguns grafos cordais e prismas complementares, identificando o comportando da aliança defensiva (AD) e aliança defensiva global (ADG) nestas classes de grafos. Foi possível, por exemplo, encontrar uma relação entre conjunto dominante (CD) e ADG, obtendo assim outro resultado equivalente nos grafos rodas.

Após a atualização do atual estado da arte, notou-se que existem poucos ou nenhum resultado em alguns tipos de alianças para determinadas classes de grafos, o que acredita-se que é um problema inexplorado.

Ao estudar a AD e ADG nos grafos cordais e produto de grafos, pode-se perceber que para os grafos roda e prisma complementar do grafo completo é mais fácil de se encontrar um padrão e de obter uma ADG. Ao passo que outros grafos são mais complicados de se enxergar este mesmo padrão, já que os vértices encontram-se mais dispersos. Isto foi observado ao estudar o produto $P_n \overline{P_n}$. Foi obtido resultado para ADG neste grafo, mas não foi possível definir uma fórmula concisa como as apresentadas na Seção 3. Dito isto, como trabalho futuro, espera-se obter esta fórmula para o grafo em questão, além de explorar outros tipos de alianças nos grafos cordais e em outras classes, tais como produto cartesiano, produto lexicográfico e também prismas complementares de outros grafos.

Referências

- Bermudo, S., Rodríguez-Velázquez, J. A., Sigarreta, J. M. and Yero, I. G. (2010). On global offensive k-alliances in graphs. *Applied Mathematics Letters*, pages 1454 – 1458.
- Brigham, R. C., Dutton, R. D., Haynes, T. W. and Hedetniemi, S. T. (2009). Powerful alliances in graphs. *Discrete Mathematics*, Elsevier, pages 2140 – 2147.

- Chang, C. W., Chia, M. L., Hsu, C. J., Kuo, D. and Lai, L. L. (2012). Global defensive alliances of trees and Cartesian product of paths and cycles. *Discrete Applied Mathematics*, pages 160:479 – 487.
- Dias, E. S. (2012). *Alianças defensivas em grafos*, volume 1. Editora UFG, 1st edition.
- Dourado, M. C., Faria, L., Pizaña, M. A., Rautenbach, D. and Szwarcfiter, J. L. (2014). On defensive alliances and strong global offensive alliances. *Discrete Applied Mathematics*, pages 163:136 – 141.
- Favaron, O., Fricke, G., Goddard, W., Hedetniemi, S. M., Hedetniemi, S. T., Kristiansen, T., Laskar, R. C. and Skaggs, R. D. (2004). Offensive Alliances in Graphs. *Discussiones Mathematicae*, Graph Theory, pages 24:263 – 275.
- Goddard, W. and Henning, M. (2013). Independent domination in graphs: A survey and recent results. *Discrete Mathematics*, pages 313:839 – 854.
- Gonçalves, B. C. (2016). Alianças defensivas globais em alguns produtos de grafos. Technical report, Instituto de Informática, Universidade Federal de Goiás, Goiânia.
- Haynes, T. W.; Henning, M. A. and van der Merwe, L. C. (2008) Domination and total domination in complementary prisms. *J. Combin. Optimization*, Springer Netherlands.
- Hedetniemi, S., Hedetniemi, S. and Kristiansen, P. (2004). Alliances in graphs. *Journal of Combinatorial Mathematics and Combinatorial Computing*, pages 157 – 177.
- Rosen, K. H. (2009). *Matemática Discreta e Suas Aplicações*. McGraw-Hill, 6th edition, pages 589 – 675.
- Shafique, K. H. (2004). *Partitioning a Graph in Alliances and its Application to Data Clustering*. PhD thesis, School of Computer Science, University of Central Florida, Orlando.
- Silva, L. R. S. da (2015). *Sobre Alianças Defensivas e Ofensivas Globais em Alguns Produtos de Grafos e Grafos Simpliciais*. PhD thesis, Instituto de Informática, Universidade Federal de Goiás. Goiânia.
- Yero, I. G. and Rodríguez-Velázquez, J. A. (2013). *Defensive alliances in graphs: a survey*. PhD thesis, Departamento de Matemáticas, Escuela Politécnica Superior de Algeciras, Universidad de Cádiz. Espanha.

Análise de Desempenho de Sistemas MIMO Utilizando Codificação Alamouti e Diversidade na Recepção MRC

Vagner L. C. Silva¹, Flávio H. T. Vieira¹

¹Escola de Engenharia Elétrica, Mecânica e de Computação
Universidade Federal de Goiás (UFG) – Goiânia – GO – Brasil

vagnerbas13@gmail.com, flavio@emc.ufg.br

Abstract. *In this work, the behavior of the BER (Bit Error Rate) as a function of the SNR (Signal Noise Ratio) for MIMO (Multiple Input Multiple Output) systems is analysed considering different spatial diversity schemes: Alamouti coding and maximum ratio combining (MRC). Furthermore, it is analysed in this article the operation of a MIMO communication system combining the two spatial diversity schemes. Through simulations, it is shown that the MRC technique provides lower error rates than the Alamouti coding. Furthermore, by combining the Alamouti code with the MRC method, it is possible to improve the BER of the system.*

Resumo. *Neste trabalho, é analisado o comportamento da BER (Bit Error Rate) em função da SNR (Signal Noise Ratio) para sistemas MIMO (Multiple Input Multiple Output) considerando diferentes esquemas de diversidade espacial: codificação Alamouti e combinação por máxima relação (MRC-Maximum Ratio Combining). Além disso, neste artigo é analisado a operação de um sistema de comunicação MIMO combinando os dois esquemas de diversidade espacial. Por meio dos resultados de simulações, mostra-se que a técnica MRC apresenta taxas de erro menores que a codificação Alamouti. Além disso, ao combinar o código Alamouti com o método MRC é possível melhorar a BER do sistema.*

1. Introdução

Devido a demanda por taxas de transmissão mais elevadas e canais de comunicação mais confiáveis, os sistemas com múltiplas antenas de transmissão e recepção, também chamados de sistemas MIMO, se apresentam como uma forma de alcançar tais requisitos [Joshi et al. 2016]. Dentre as vantagens destes sistemas, encontra-se que em ambientes com limitação de banda ou de potência de transmissão, seu desempenho e confiabilidade podem ser superiores aos dos sistemas SISO (*Single Input Single Output*) [Haykin and Moher 2005].

Em termos de melhoria da confiabilidade do canal de comunicação, técnicas de diversidade espacial permitem reduzir o efeito do desvanecimento do canal por múltiplos percursos, uma das principais causas para o aumento da taxa de erro no sistema [Joshi et al. 2016]. As técnicas de diversidade espacial consistem na transmissão e/ou recepção de múltiplas versões de um sinal, por meio de múltiplas antenas de transmissão e/ou recepção.

Nesse contexto, pode-se classificar as técnicas existentes em três tipos: diversidade na recepção, diversidade na transmissão e diversidade na transmissão e recepção. O

sistema pode apresentar múltiplas antenas na recepção (diversidade na recepção), múltiplas antenas na transmissão (diversidade na transmissão) ou em ambos (diversidade na transmissão e recepção). Estes sistemas podem ser chamados de SIMO (*Single Input Multiple Output*), MISO (*Multiple Input Multiple Output*) e MIMO, respectivamente.

As técnicas de diversidade na recepção realizam a combinação dos sinais recebidos em cada antena receptora. Na diversidade na transmissão, os sinais são codificados antes da transmissão, passam pelo canal de comunicação e, por fim, são decodificados no receptor. Além disso, é possível combinar as técnicas, fazendo uso de múltiplas antenas na transmissão e na recepção. Na Figura 1 são ilustradas tais configurações.

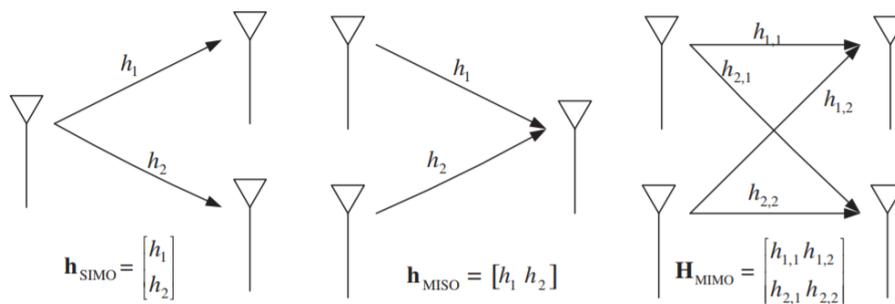


Figura 1. Exemplos de configurações de antenas. [Cho et al. 2010]

Dentre as técnicas de diversidade na recepção, destacam-se: combinação por seleção (*SC-Selection Combining*), combinação por máxima relação (*MRC*), combinação de ganho igual (*EGC-Equal Gain Combining*) entre outras [Haykin and Moher 2005]. Já na diversidade na transmissão, podem ser citadas: codificação de Alamouti, codificação complexa ortogonal generalizada, codificação de treliça entre outras [Cho et al. 2010]. Neste trabalho, serão analisadas as técnicas *MRC* e codificação de Alamouti.

Alguns estudos foram desenvolvidos na tentativa de avaliar a BER do sistema *MRC* combinado com Alamouti, entre estes estudos citam-se: [Anitha et al. 2015] onde é avaliado a taxa de erro para modulações *BPSK*, *QAM* e *PAM*; em [Singh and Rao 2017] é realizada uma análise considerando as modulações *BPSK*, *QPSK*, *4-DQPSK*, *16-QAM* e *4-FSK* e em [Joshi et al. 2016] avalia-se extensivamente um sistema Alamouti para modulações *BPSK*, *QPSK*, *8PSK* e *16QAM*.

Deste modo, o presente trabalho busca investigar a taxa de erro de bit (*BER*) para um sistema de comunicação *MIMO* utilizando as técnicas *MRC*, Alamouti e um sistema combinando Alamouti com *MRC*. Para isto, o trabalho está organizado da seguinte forma: na seção 2, descreve-se o modelo matemático de um canal *MIMO*. Na seção 3, detalha-se o comportamento do combinador *MRC*. Na seção 4, avalia-se o comportamento da codificação de Alamouti. Na seção 5, são mostradas algumas simulações, juntamente com seus resultados. Por fim, na seção 6 são apresentadas as conclusões sobre os resultados do trabalho.

2. Modelo de Sistema MIMO

Sejam N_r o número de antenas receptoras, N_t o número de antenas transmissoras, $h_{i,j}(t)$ o ganho entre a antena receptora i e transmissora j , x_j o sinal transmitido pela j -ésima

antena, y_i o sinal recebido pela i -ésima receptora e z_i o ruído captado pela i -ésima antenna receptora. Onde $i = 1, 2, \dots, N_r$ e $j = 1, 2, \dots, N_t$. Como pode ser visto na Figura 2.

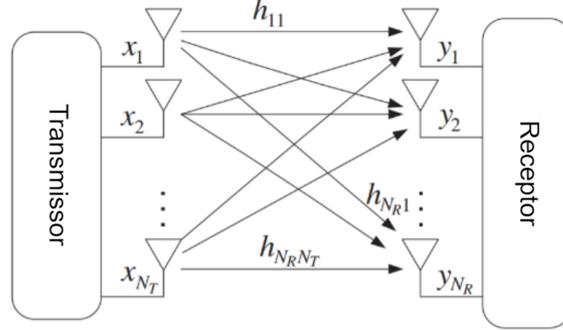


Figura 2. Modelo matemático de um sistema MIMO. [Cho et al. 2010]

Portanto, o sinal recebido pela i -ésima antenna, em função dos sinais transmitidos, é dado por:

$$y_i(t) = \sum_{j=1}^{N_t} h_{i,j}(t)x_j(t) + z_i(t) \quad (1)$$

Considerando que o canal seja invariante no tempo, ao menos durante o período de transmissão de um símbolo, então é possível considerar $h_{i,j}(t) = h_{i,j}$. Agrupando em notação matricial todos os sinais recebidos, obtém-se:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_r} \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,N_t} \\ h_{2,1} & h_{2,2} & \dots & h_{2,N_t} \\ \vdots & \vdots & & \vdots \\ h_{N_r,1} & h_{N_r,2} & \dots & h_{N_r,N_t} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_t} \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{N_r} \end{bmatrix} \quad (2)$$

Fazendo $\mathbf{x} = [x_1, \dots, x_{N_t}]^T$, $\mathbf{y} = [y_1, \dots, y_{N_r}]^T$, $\mathbf{z} = [z_1, \dots, z_{N_r}]^T$ e $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_{N_r}]^T$, onde $\mathbf{h}_i = [h_{i,1}, \dots, h_{i,N_t}]$. Então, a Equação 2, se reduz a:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \quad (3)$$

Para simplificar a análise apresentada nas próximas seções, considera-se que o ruído do canal seja gaussiano (AWGN-Additive White Gaussian Noise), ou seja $z_i \sim N(0, N_0)$. Além disso, é considerado um canal com desvanecimento plano em frequência e distribuído de acordo com a distribuição de Rayleigh, ou seja $h_{i,j} = \alpha_{i,j}e^{j\theta_{i,j}}$, com $\alpha_i \sim Rayleigh(\sigma)$.

Ao longo do texto, um sistema MIMO com N_t antenas de transmissão e N_r antenas de recepção, poderá ser chamado de sistema MIMO $N_t \times N_r$.

3. Combinação por Máxima Relação (MRC)

A técnica de combinação por máxima relação (MRC) é caracterizada pela presença de uma antenna transmissora e múltiplas antenas receptoras (SIMO-Single Input Multiple

Output), logo $N_t = 1$ e N_r é livre, como pode ser visto na Figura 3. O espaçamento entre as antenas deve ser escolhido de forma a garantir que cada ramo de recepção seja mutuamente independente. O MRC é composto por dois blocos funcionais: um conjunto de receptores e um combinador linear.

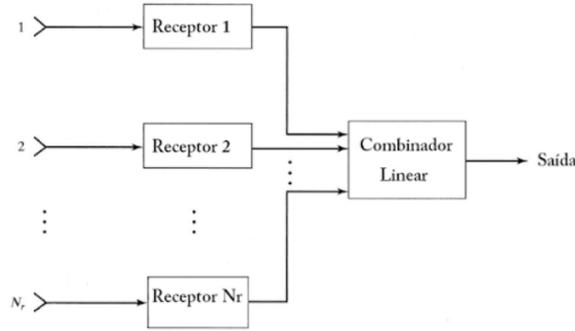


Figura 3. Estrutura do receptor MRC. [Haykin and Moher 2005]

Cada antena receptora recebe uma versão do sinal transmitido e envia para o combinador linear. O combinador linear atribui pesos para cada ramo receptor e ao final soma a contribuição de cada ramo, formando o sinal combinado. A escolha dos pesos de cada ramo é feita com o objetivo de maximizar a SNR do sistema.

Como $N_t = 1$, então será omitido o índice j do modelo de canal. Portanto, o sinal recebido em cada antena receptora é dado por:

$$y_i = \alpha_i e^{j\theta_i} x + z_i \quad (4)$$

Seja a_i o peso atribuído pelo combinador linear ao i -ésimo ramo receptor. Portanto, a saída do combinador linear é dada por:

$$y_{mrc} = \sum_{i=1}^{N_r} a_i y_i = \sum_{i=1}^{N_r} a_i [\alpha_i e^{j\theta_i} x + z_i] = x \sum_{i=1}^{N_r} a_i \alpha_i e^{j\theta_i} + \sum_{i=1}^{N_r} a_i z_i \quad (5)$$

Com base na Equação 5, observa-se que o sinal recebido pode ser decomposto em dois termos, um representando a contribuição do sinal transmitido e outro representando a contribuição do ruído. Assim, a SNR esperada do combinador linear é dada por [Haykin and Moher 2005]:

$$SNR_{mrc} = \frac{\mathbf{E} \left[\left| x \sum_{i=1}^{N_r} a_i \alpha_i e^{j\theta_i} \right|^2 \right]}{\mathbf{E} \left[\left| \sum_{i=1}^{N_r} a_i z_i \right|^2 \right]} = \left(\frac{E}{N_0} \right) \frac{\mathbf{E} \left[\left| \sum_{i=1}^{N_r} a_i \alpha_i e^{j\theta_i} \right|^2 \right]}{\mathbf{E} \left[\left| \sum_{i=1}^{N_r} a_i \right|^2 \right]} \quad (6)$$

Por meio da Equação 6, é possível derivar a SNR instantânea, como pode ser visto em [Haykin and Moher 2005]:

$$\gamma_{mrc} = \left(\frac{E}{N_0} \right) \frac{\left| \sum_{i=1}^{N_r} a_i \alpha_i e^{j\theta_i} \right|^2}{\left| \sum_{i=1}^{N_r} a_i \right|^2} \quad (7)$$

Deste modo, é possível encontrar a solução ótima para o problema. Utilizando a inequação de Cauchy-Schwarz tem-se um limite superior para a SNR instantânea (Equação 7), ou seja [Haykin and Moher 2005]:

$$\gamma_{mrc} \leq \left(\frac{E}{N_0} \right) \frac{\sum_{i=1}^{N_r} |a_i|^2 \sum_{i=1}^{N_r} |\alpha_i e^{j\theta_i}|^2}{\sum_{i=1}^{N_r} |a_i|^2} \leq \left(\frac{E}{N_0} \right) \sum_{i=1}^{N_r} \alpha_i^2 \quad (8)$$

Portanto, γ_{mrc} é máximo quando a Inequação 8 se torna uma igualdade, isso acontece quando os parâmetros do combinador linear são proporcionais a $(\alpha_i e^{j\theta_i})^*$, logo $a_i = c(\alpha_i e^{j\theta_i})^*$ define os parâmetros ótimos para o MRC, sendo c uma constante de proporcionalidade, cujo valor não interfere em γ_{mrc} [Haykin and Moher 2005]. Utilizando os parâmetros obtidos, a γ_{mrc} é igual a:

$$\gamma_{mrc} = \left(\frac{E}{N_0} \right) \sum_{i=1}^{N_r} \alpha_i^2 = \sum_{i=1}^{N_r} \gamma_i \quad (9)$$

4. Codificação de Alamouti

A codificação de Alamouti é um tipo de codificação espaço temporal de bloco (STBC-*Space Time Block Code*) [Cho et al. 2010]. Neste tipo de codificação, o sinal passa por uma pré-codificação, com o objetivo de simplificar a decodificação no receptor [Cho et al. 2010].

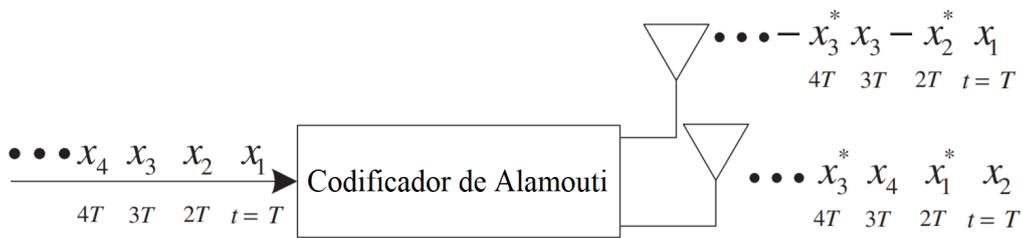


Figura 4. Ilustração da codificação de Alamouti. [Cho et al. 2010]

O processo de pré-codificação do sinal, é ilustrado na Figura 4. No instante t são transmitidos os símbolos x_1 e x_2 , respectivamente, pelas antenas 1 e 2. No instante seguinte $t+T$, são transmitidos os símbolos $-x_2^*$ e x_1^* pelas antenas 1 e 2, respectivamente. Devido ao atraso de propagação do sinal, a recepção ocorre no instante t' , então o sinal é armazenado. No instante seguinte $t' + T$ o segundo sinal chega ao receptor, então inicia-se o processo de decodificação dos símbolos, após a realização da decodificação, os símbolos $y_{t'}$ e $y_{t'+T}$ são extraídos [Cho et al. 2010].

Levando em consideração as contribuições de desvanecimento e deslocamento de fase para cada símbolo, adicionando o ruído captado pelo receptor e considerando que o

canal é estacionário durante a transmissão dos dois símbolos. Então, os sinais recebidos nos instantes t' e $t' + T$ são representados, respectivamente, pela Equação 10:

$$\begin{aligned} y_{t'} &= \alpha_1 e^{j\theta_1} x_1 + \alpha_2 e^{j\theta_2} x_2 + z_1 \\ y_{t'+T} &= -\alpha_1 e^{j\theta_1} x_2^* + \alpha_2 e^{j\theta_2} x_1^* + z_2 \end{aligned} \quad (10)$$

Reformulando as Equações 10, obtém-se:

$$\begin{bmatrix} y_{t'} \\ y_{t'+T}^* \end{bmatrix} = \begin{bmatrix} \alpha_1 e^{j\theta_1} & \alpha_2 e^{j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{-j\theta_1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad (11)$$

Multiplicando ambos os lados da Equação 11 pela hermitiana da matriz de ganho do canal, visto na Equação 11, encontra-se:

$$\begin{bmatrix} \alpha_1 e^{-j\theta_1} & \alpha_2 e^{-j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{-j\theta_1} \end{bmatrix} \begin{bmatrix} y_{t'} \\ y_{t'+T}^* \end{bmatrix} = (\alpha_1^2 + \alpha_2^2) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 e^{-j\theta_1} & \alpha_2 e^{-j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{-j\theta_1} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad (12)$$

Com isso, define-se a saída do decodificador Alamouti e o vetor de contribuição de ruído, respectivamente, por:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 e^{-j\theta_1} & \alpha_2 e^{-j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{-j\theta_1} \end{bmatrix} \begin{bmatrix} y_{t'} \\ y_{t'+T}^* \end{bmatrix}, \quad \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 e^{-j\theta_1} & \alpha_2 e^{-j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{-j\theta_1} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad (13)$$

Combinando as Equações 12 e 13 obtém-se:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = (\alpha_1^2 + \alpha_2^2) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \text{ou, } y_j = (\alpha_1^2 + \alpha_2^2)x_j + v_j \quad \text{com } j = 1, 2 \quad (14)$$

Portanto, é possível perceber que a interferência entre antenas foi eliminada, pois o símbolo x_2 foi removido do sinal y_1 e o símbolo x_1 foi eliminado de y_2 , logo, os sinais podem ser decodificados independentemente [Cho et al. 2010]. Esse resultado mostra que o processamento no receptor foi simplificado, pois necessita de apenas N_t operações para decodificar os sinais transmitidos [Cho et al. 2010].

5. Simulações e Resultados

Nesta seção, são avaliados as taxas de erro, por meio de simulações, dos esquemas de combinação MRC $1 \times N_r$, codificação Alamouti 2×1 e com a combinação das duas formas, esquema Alamouti $2 \times N_r$. Os parâmetros de simulação considerados foram: variação da SNR de 0dB a 20dB; utilização de modulação BPSK (Binary Phase Shift Keying); estimação perfeita do canal no receptor e um total de 10Mbytes de dados transmitidos.

Conforme abordado em [Anitha et al. 2015], um sistema com codificação Alamouti combinado com MRC apresenta melhor confiabilidade para cenários com modulação BPSK. Essa característica nos motivou a utilizar esta modulação, com o objetivo de realizar a comparação entre os resultados. Além disso, uma vez que estamos assumindo estimação perfeita do canal no receptor, a análise se torna mais simples, permitindo, portanto, avaliar a taxa de erro para variações no número de antenas no sistema.

5.1. Avaliação do Esquema MRC

Com o objetivo de avaliar a redução da BER com o aumento do número de antenas receptoras, realizou-se uma simulação com vários sistemas MRC. Utilizou-se as configurações 1x2, 1x4, 1x6 e 1x8. Na Figura 5, nota-se que conforme aumenta-se o número de antenas receptoras, mais ingreme é a inclinação da curva BER, o que indica um melhor desempenho.

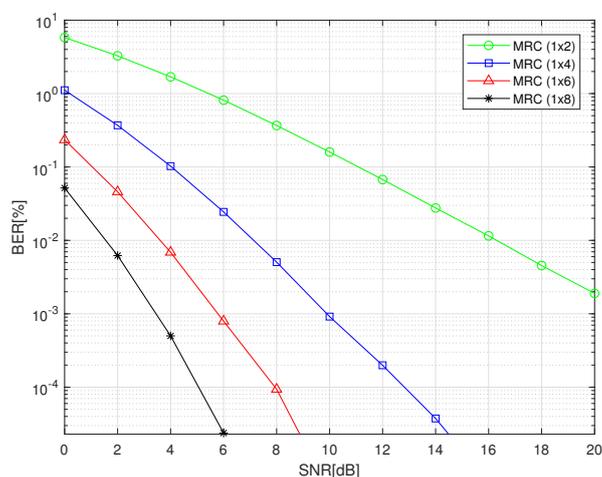


Figura 5. Avaliação da BER do esquema MRC para diferentes quantidades de antenas receptoras.

Quando a $BER = 10^{-2}\%$, observa-se que o esquema MRC 1x8 apresenta menor SNR, enquanto que o esquema MRC 1x2 apresenta a maior SNR. Esse resultado indica que é necessário um canal de melhor qualidade para que o sistema MRC 1x2 atinja a mesma BER do MRC 1x8. Na Tabela 1 são mostrados os valores aproximados de SNR para os sistemas analisados, considerando $BER = 10^{-2}\%$.

	MRC 1x2	MRC 1x4	MRC 1x6	MRC 1x8
SNR	16dB	8dB	4dB	2dB

Tabela 1. Valores aproximados de SNR para $BER = 10^{-2}\%$.

5.2. Comparação entre Codificação Alamouti e SISO Sem Codificação

Realizando uma comparação entre um sistema com codificação Alamouti e um sistema SISO, é possível perceber que as taxas de erro do sistema Alamouti, em geral, são menores do que a do sistema SISO, como pode ser visto na Tabela 2 e na Figura 6. Além disso,

Sistema	0dB	5dB	10dB	15dB	20dB
Alamouti 2x1	23.020	6.532	1.105	0.148	0.014
SISO	14.640	6.532	2.326	0.791	0.248

Tabela 2. Comparação da BER com diferentes valores de SNR.

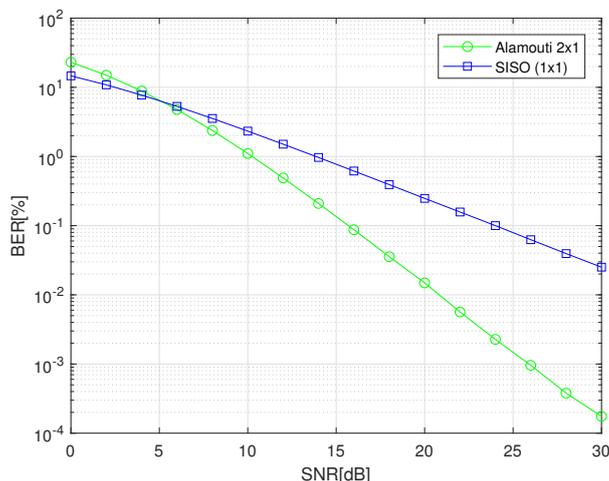


Figura 6. Comparação entre esquema Alamouti e sistema SISO

nota-se que para valores menores do que 5dB de SNR a BER do sistema SISO é inferior a do sistema com codificação de Alamouti.

Os resultados de BER apresentados na Tabela 2, para o sistema Alamouti, se aproximam dos apresentados em [Joshi et al. 2016], reforçando, portanto, os resultados apresentados.

5.3. Avaliação do Esquema Alamouti com MRC

Com o objetivo de avaliar a BER do esquema Alamouti com MRC (Configuração $2 \times N_r$), realizou-se comparações com o combinador MRC (Configuração $1 \times N_r$). Como pode ser visto na Figura 7, foram realizadas comparações entre os esquemas Alamouti 2x1, MRC 1x2, Alamouti 2x4 e MRC 1x4.

Ao comparar o sistema MRC (1x2) com o Alamouti (2x1), Figura 7, percebe-se que a BER do MRC é sempre inferior a do Alamouti. Pela Tabela 3, nota-se que fixando a BER em $10^{-2}\%$, o sistema com MRC (1x2) consegue prover tal valor, mesmo em um canal com SNR menor, comparado ao esquema Alamouti (2x1).

	MRC (1x2)	Alamouti (2x1)	MRC (1x4)	Alamouti e MR (2x4)
SNR	16dB	20dB	7dB	6dB

Tabela 3. Valores aproximados de SNR para BER = $10^{-2}\%$.

Além disso, ao adicionar mais antenas na recepção, nota-se que o desempenho do sistema Alamouti (2x4) se torna melhor do que o do MRC (1x4), a partir de uma SNR de aproximadamente 3dB. Com isso, nota-se que a combinação da técnica Alamouti

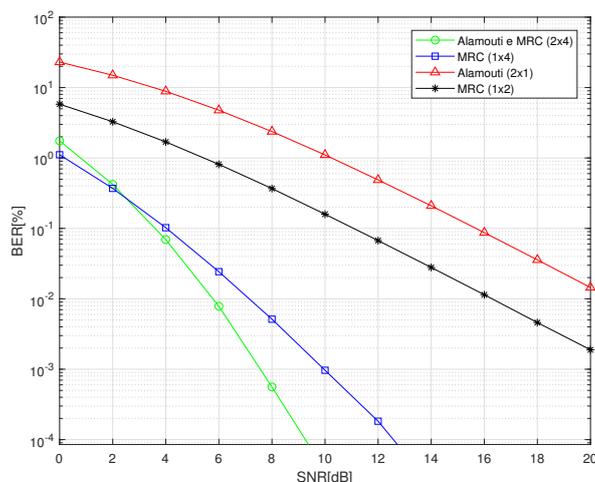


Figura 7. Comparação entre MIMO Alamouti com MRC e MRC simples

com MRC melhora a confiabilidade do canal, como observado em [Singh and Rao 2017]. Não obstante, como pode ser constatado na Tabela 3, o aumento do número de antenas receptoras faz com que o sistema mantenha, mais facilmente, uma determinada taxa de erro objetivo, mesmo para um canal com pior qualidade.

6. Conclusão

Foi analisado o desempenho de algumas configurações de sistemas MIMO, visando melhorar a confiabilidade do canal, por meio de técnicas de diversidade espacial. Analisou-se a codificação de Alamouti, a combinação MRC e a junção da codificação de Alamouti com a combinação MRC.

Por meio das simulações, foi possível constatar que a codificação de Alamouti 2x1 apresenta valores de BER, em geral, mais elevados quando comparado com o MRC 1x2, mas quando comparado com um sistema SISO, se mostrou em vantagem.

Além disso, foi possível perceber que o desempenho do sistema com MRC melhora conforme se aumenta o número de antenas receptoras, sendo esse resultado reforçado por meio de simulação. Por fim, notou-se que a utilização conjunta do MRC com a codificação Alamouti permitiu melhorar o desempenho do sistema comparado ao sistema com MRC sem codificação Alamouti.

Para trabalhos futuros propõem-se a realização de uma análise de desempenho considerando estimação imperfeita do canal no receptor, para verificar o impacto do erro na determinação dos parâmetros do MRC. Além disso, será proposto a avaliação do desempenho da codificação Alamouti com MRC em um sistema MIMO multi-usuário.

Referências

Anitha, M., Priya, R. S., Lakshmi, M. R., and Vijayalakshmi, S. (2015). Mimo system performance using various modulations under different channels with stbc, zf and mrc. *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*.

- Cho, Y. S., Kim, J., Yang, W. Y., and Kang, C.-G. (2010). *MIMO-OFDM Wireless Communications With MATLAB*. John Wiley & Sons Ltd.
- Haykin, S. and Moher, M. (2005). *Modern Wireless Communications*. Pearson Prentice Hall, international edition.
- Joshi, A., Airee, N., and Shukla, B. (2016). Extensive performance analysis of stbc system with and without perfect csi in various fading channels. *2016 International Conference on Signal Processing and Communication (ICSC)*.
- Singh, S. and Rao, S. A. P. (2017). Performance analysis of diversity in mimo wireless system. *2017 International Conference on Signal Processing and Communication (ICSPC)*.

Sistema Integrado Educação Digital

Leidiane B. P. Rodrigues¹, Raimunda D. dos Santos²

Instituto de Informática – Universidade Federal de Goiás (UFG)

Caixa Postal 74690-900– Goiânia – GO – Brasil

leidianebeatrizpassosrodrigues@gmail.com, raidelfinoufg@gmail.com

***Abstract.** This article has as objective to describe a program that will be created for help a project that is release in higher education institution in Goiânia's city , this program in general help in 1)organization, 2)frequency, 3)learning and facility control and management about contents taught in the project. As well as the the academic system of Universidade Federal de Goiás (UFG), where the project is realised, the system that will be developed it's' like one existing in the same higher institution SIGAA, but how the people that participate in project not can have no relation with university, they can not use the system, in this way it is necessary to create a independent system for they use it.*

***Resumo.** Este artigo tem como objetivo descrever um sistema que será criado para ajudar a gerir o projeto, que é realizado em uma instituição de ensino superior na cidade de Goiânia. O referido sistema, em linhas gerais, será imprescindível para: 1) organizar a documentação do projeto; 2) controlar a frequência dos alunos; 3) organizar os conteúdos ministrados nas aulas do projeto. Assim como o sistema acadêmico da Universidade Federal de Goiás (UFG), onde o projeto é realizado. O sistema que desenvolveremos é similar ao um existente na mesma, o SIGAA, mas como os participantes dos projeto não podem ter vínculo com a instituição de ensino, eles não podem usá-lo sistema da mesma, por isso faz -se necessário criar um sistema independente para eles usarem.*

1. Introdução

A tecnologia presente no mundo contemporâneo fez com que as pessoas mudassem as formas de produzir um trabalho, de registrar um documento, de se comunicar com uma pessoa, de fazer uma entrevista de emprego e, principalmente, aprender algo novo. É com base nessas mudanças que o projeto EDUCAÇÃO DIGITAL:: Políticas, Leitura, Produção Textual, Identidade e Letramento Digital com trabalhadoras/es terceirizadas/os da UFG, doravante EDUCAÇÃO DIGITAL, foi criado e tem como principal objetivo ensinar tanto trabalhadores/as terceirizados/as Universidade Federal de Goiás, doravante UFG, quanto membros da comunidade externa a usarem as novas e as velhas tecnologias, seja como ferramenta de trabalho, de gerenciamento de despesas, como, por exemplo uma planilha, como veículo de informação diária, ou, até mesmo como veículo facilitador da formação pessoal, de exercício da cidadania, de comunicação.

Com o intuito de situar melhor o leitor, falaremos agora um pouco sobre o projeto, para o qual desenvolveremos o referido sistema acadêmico. Trata-se de um projeto de extensão e de pesquisa, estruturado em três módulos e tem como base da ementa o ensino de informática básica e avançada, ensinadas nos módulos I e II. Cada módulo tem duração de um semestre letivo, conforme o calendário acadêmico da instituição onde a pesquisa é realizada, no nosso caso a UFG. Cabe ao módulo III o ensino de noções básicas de Espanhol como Língua Estrangeira, a partir do uso dos recursos tecnológicos ensinados nos módulos anteriores, tais como: pesquisas a jornais e revistas online, consulta a bibliotecas virtuais, gramáticas e dicionários online, ferramentas do Google Drive, dentre outros recursos.

Faz-se necessário dizer que o projeto EDUCAÇÃO DIGITAL obteve uma grande demanda de alunos, durante o primeiro ano em que foi executado e o ato de gerenciar todas as matrículas, tarefas entregues e trabalhos, sem uma plataforma onde os alunos pudessem ter acesso para realizar especificamente essas atividades tornou-se complicado tanto para os monitores, quanto para a coordenação que acompanhava e orientava esses alunos.

Foi através da observação dessa dificuldade de correção das atividades e de gerenciamento dos alunos e das turmas que percebemos a necessidade de criarmos uma plataforma de apoio para as atividades realizadas pelos alunos. Com o referido sistema nós vamos cadastrar todos os participantes do curso, inserir atividades avaliativas, slides de aulas, acompanhamento de desempenho e frequência no curso.

A partir da implementação desse sistema o projeto dará um grande avanço tecnológico, além de incentivar os monitores, alunos dos cursos da área da computação, que estão aptos a desenvolverem essa tarefa, colocando em prática os conhecimentos adquiridos nas aulas de graduação. Dessa forma uma grande contribuição será deixada para a primeira etapa do projeto e poderá ser aprimorada, conforme a necessidade do usuário.

2. Metodologia

Para realizar este trabalho foi necessário fazer um estudo de caso para verificar a aplicabilidade do produto em desenvolvimento, visto que a necessidade do público alvo é a de um sistema onde os responsáveis consigam gerenciar o curso, por meio da plataforma. Para isso, vários sistemas similares foram estudados tais como o Moodle, o Sigaa e o Google sala de aula. Dentre esses sistemas citados, dois são restritos aos alunos da universidade e o Google sala de aula não atende a todas as necessidades do projeto, tais como: prazo para entrega de atividades, chamada, progresso no curso entre outras. Durante o período de observação e de produção da lista de requisitos, foram feitas análises e comparações de ferramentas adequadas para o desenvolvimento do sistema. Solicitamos a opinião de desenvolvedores experientes na área, pesquisamos as ferramentas, analisamos a versão e o suporte ao usuário da mesma e analisamos também as ferramentas dos sistemas similares citados acima para assim encontrar as ferramentas mais adequadas para o desenvolvimento. Os passos desse processo de desenvolvimento estão fundamentados no modelo Agile ou (modelo ágil), que define etapas fundamentais do processo de forma que o tempo de desenvolvimento fique menor quando comparado com o modelo tradicional (cascata), que ao invés da próxima

etapa começar somente após a etapa anterior terminar, o modelo ágil aceita ideias, reformulações, e correções necessárias sem que todos os requisitos sejam analisados novamente [Pressman 2006].

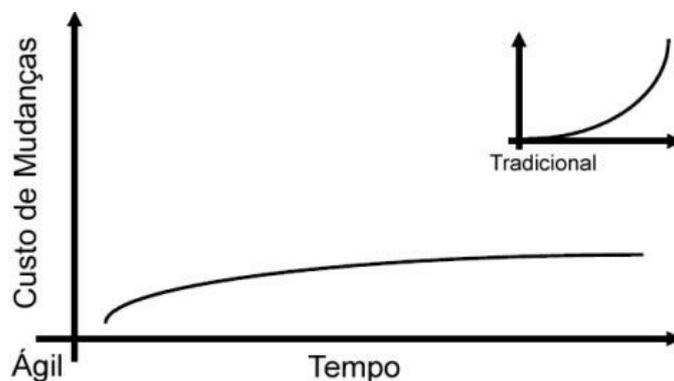


Figura 1. Custo de alterações entre o desenvolvimento Ágil e o Tradicional

3. Desenvolvimento

3.1. Análise de viabilidade

Para o estudo de análise de viabilidade podemos levantar três pontos: 1) viabilidade técnica, 2) econômica e 3) operacional [Sommerville 2000]. Todos esses pontos foram analisados pela equipe responsável pelo projeto, por desenvolvedores que também são monitores do projeto, por professores especialistas na área e por outros desenvolvedores de sistema, além de ter sido estudado em conjunto com o cliente em questão, a coordenação do projeto. Dessa forma, concluímos na análise de viabilidade que, para a viabilidade técnica as funções não possuem restrições que possam comprometer o desempenho do sistema. Quanto à análise de custo, os gastos levantados para desenvolvimento do sistema são considerados baixos e possíveis, e a viabilidade operacional mostra que sistema pode ser implementado sem maiores problemas e em tempo hábil.

Em relação à viabilidade econômica, segue a estimativa na tabela 1 e a viabilidade operacional na tabela 2.

Tabela 1. Viabilidade econômica

Viabilidade Econômica	Custos
Licença do Proto.io	R\$ 202,78/mês
Lucidchart - Modelagem UML	R\$ 41,18/mês
UOLhost - Hospedagem de site com plataforma Java e Tomcat	R\$ 79,90/ano

Tabela 2. Viabilidade operacional

Viabilidade Operacional	Prazo
Levantamento de requisitos	2 meses
Análise de requisitos	2 meses
Organização/aprendizado da equipe desenvolvedora	2 meses
Desenvolvimento do módulo 1	6 meses
Desenvolvimento do módulo 2	6 meses
Desenvolvimento do módulo 3	4 meses
Teste do sistema	1 mês
Total	1 ano e 11 meses

3.2. Modelagem UML

Para mostrar melhor os estudos feitos para o desenvolvimento do sistema, foi utilizado o diagrama UML (*Unified Modeling Language*) ou em português (*Linguagem de Modelagem Unificada*). Para modelagem foi utilizado o diagrama de caso de uso, que mostra o sistema do ponto de vista do usuário e o digrama de classes para agrupar, de maneira mais simples os objetos em comum no sistema [DEVMEDIA 2012].

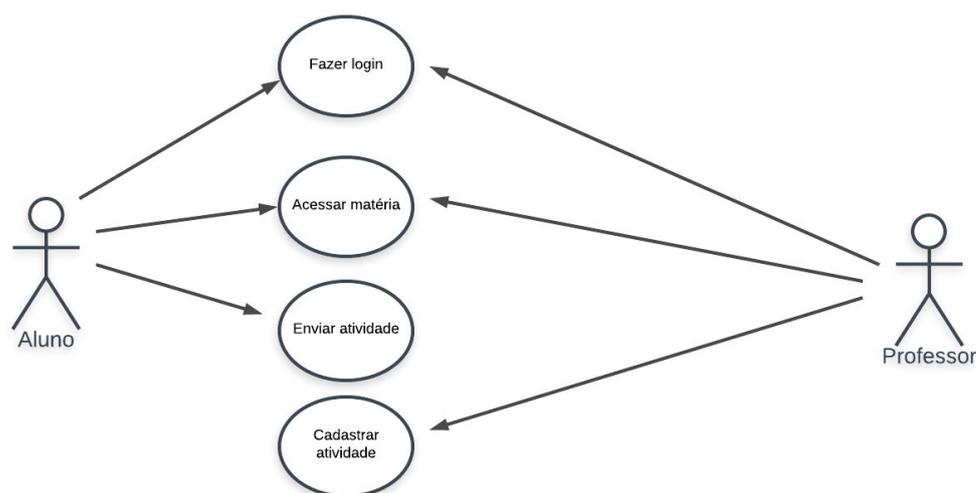


Figura 2. Diagrama de caso de uso

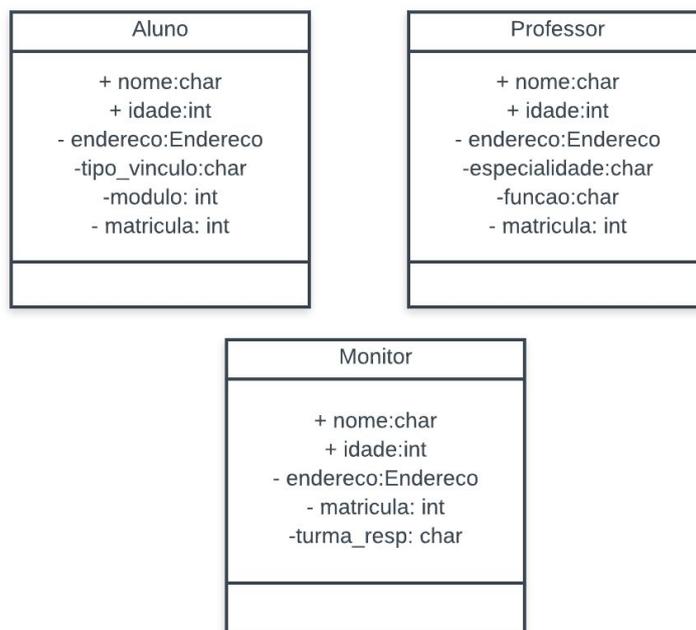


Figura 3. Diagrama de classe

3.3. Framework Vue.js

O Vue.js é framework progressivo para a construção de interfaces de usuários, além disso, Vue foi projetado para ser incremental, ao contrário dos outros framework, que, na maioria das vezes são monolíticos, ou seja, todos os seus módulos não estão em um único sistema [VUE 2018]. Além disso o Vue possui uma biblioteca principal onde seu foco é na camada view do modelo MVC (*Model-View-Controller*), que é um padrão de arquitetura que separa a representação da informação do usuário. Também é possível, através do Vue utilizar outras bibliotecas. Uma das grandes vantagens desta framework é a velocidade, além da escalabilidade e da simplicidade.

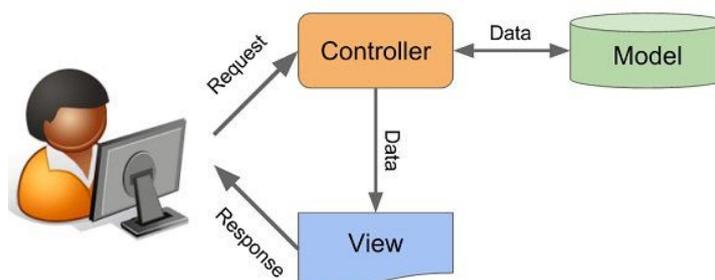


Figura 4. Modelo MVC

O modelo MVC pode ser utilizado por qualquer aplicação. Para entendê-lo melhor, segue uma descrição abaixo:

- Model : contém lógica de negócios;

- View : processa os dados do usuário e transmite os comandos do usuário para o Controller;
- Controller: interage com o Model e produz dados para o View.

3.4. Spring Boot

O Spring Boot é um projeto que visa acelerar e facilitar o processo de configuração e publicação de uma determinada aplicação. Ele favorece as convenções de configuração, basta você dizer qual o tipo da sua aplicação: web, segurança entre outros [SPRING 2018]. Ele é utilizado para fazer um back end robusto e além de ser bom para a segurança da sua aplicação.

3.5. Banco de dados

O banco de dados a ser utilizado no sistema será o MySQL, que é um sistema gerenciador de banco de dados e utiliza a linguagem SQL (*Structure Query Language*) para gerenciar o banco de dados [Elmasri 2005]. Além disso, escolhemos o banco de dados MySQL, por ser uma aplicação gratuita, popular e simples para gerenciar os dados.

3.6. Servidor web Tomcat 7.0

O servidor web Tomcat é um servidor de código aberto baseado em Java. Ele é indicado para aplicações web, principalmente as que fazem uso das tecnologias Spring Boot, além disso o Tomcat é uma aplicação bastante estável [DEVMEDIA 2013]. E usaremos a versão 7.0 por ainda ter suporte disponível ao usuário.

3.7. IDE Eclipse para Spring

Para desenvolver o sistema utilizaremos a interface de desenvolvimento Eclipse para java. A IDE possui todas as ferramentas necessárias para o desenvolvimento em linguagem java, a qual a mesma já contém pacotes java para o desenvolvimento de uma aplicação web [ECLIPSE 2018].

3.8. Licença

O sistema que será desenvolvido tem como objetivo ajudar projetos semelhantes ao EDUCAÇÃO DIGITAL, que, futuramente possam necessitar de um sistema de gerenciamento, assim como o nosso projeto hoje necessita. Sendo assim, não temos como objetivo o comércio do sistema, por isso optamos pela licença de Atribuição-NãoComercial-CompartilhaIgual CC BY-NC-SA [CREATIVE COMMONS 2018], esta licença permite que a pessoa interessada faça alterações para fins não comerciais, que o crédito atribuído ao trabalho continue sendo do desenvolvedor que o licenciou a primeira vez e que sua distribuição seja feita sobre os mesmos termos idênticos ao desta licença .

4. Protótipos

Alguns protótipos de tela do sistema já foram desenvolvidas para melhor compreensão dos desenvolvedores e o cliente do sistema. Os protótipos foram criados através da ferramenta Proto.io, que é uma plataforma online para desenvolvimento de protótipos.



Figura 5. Protótipo tela de login

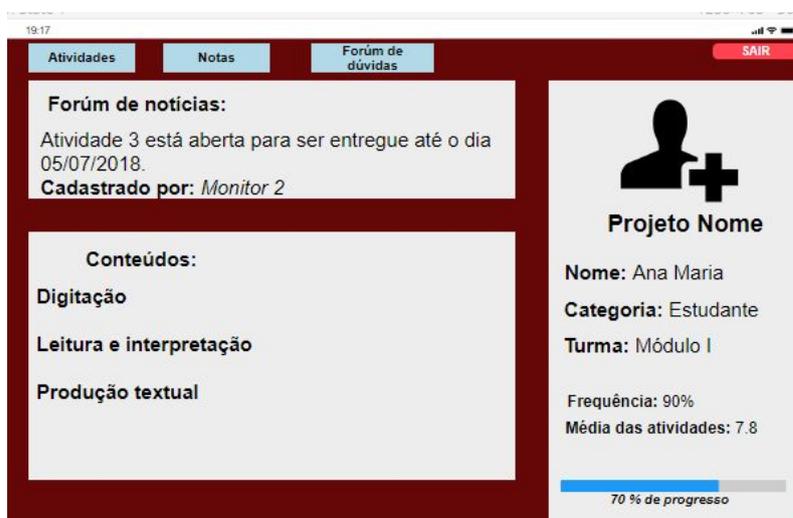


Figura 6. Protótipo tela principal

5. Conclusão

Considerando todos os aspectos aqui apresentados, podemos concluir que o sistema em questão estará pronto para entrar em fase de desenvolvimento e deve ser concluído em até um ano. Por se tratar de um sistema restrito, cuja finalidade é a de atender somente a um projeto específico, seu desenvolvimento não acarretará grandes atrasos, nem implicará em grandes prejuízos, enquanto não for desenvolvido de uma só vez. Nesse mesmo sentido, o pretendemos entregar o sistema em módulos, até que todo o sistema

esteja completamente desenvolvido. Faz - se necessário ressaltar a importância que essa ferramenta terá tanto para a instituição de ensino que abriga o projeto EDUCAÇÃO DIGITAL, quanto para a sua equipe, além disso, futuramente, outros outros projetos poderão querer usufruir do sistema, precisando somente de algumas modificações. O intuito do sistema é ajudar os colaboradores e alunos que irão usar os recursos que ele oferecerá.

6. Referências:

DEVMEDIA. O que é UML e diagramas de caso de uso: Introdução à prática UML.

Disponível em

<https://www.devmedia.com.br/o-que-e-uml-e-diagramas-de-caso-de-uso-introducao-pratica-a-uml/23408>. (2012).

Pressman, R. S. Engenharia de Software. 6º ed. — Porto Alegre: McGrawHill, (2006).

Sommerville, I. Engenharia de Software. 6º ed. — São Paulo: Addison Wesley, (2000).

Sommerville, I.; Engenharia de Software. 9. ed. — São Paulo: Pearson Prentice Hall, (2011).

Carvalho, A. Engenharia de Software. 1º ed. — Campinas: Ed. da Unicamp, (2001).

DEVMEDIA. Orientações básicas na elaboração de um diagrama de classes.

Disponível em

<https://www.devmedia.com.br/orientacoes-basicas-na-elaboracao-de-um-diagrama-d-e-classes/37224>. (2016).

DEVMEDIA. Processos de Software. Disponível em

<https://www.devmedia.com.br/processos-de-software/21977>. (2011).

REACT. A JavaScript library for building user interfaces. Disponível em

<https://reactjs.org/>. (2018).

ECLIPSE. Eclipse IDE for Java EE Developers. Disponível em

<http://www.eclipse.org/downloads/packages/eclipse-ide-java-ee-developers/mars2>. (2018).

DEVMEDIA. Introduzido o servidor de aplicações Apache Tomcat. Disponível em

<https://www.devmedia.com.br/introduzindo-o-servidor-de-aplicacao-apache-tomcat/27939>. (2013).

VUEJS. The progressive JavaScript framework. Disponível em <https://vuejs.org/>. (2018).

SPRING. Spring framework 5.0. Disponível em <https://spring.io/>. (2018).

Elmasri, R.; Navathe, S. B.; Sistemas de Banco de Dados. 4º ed. São Paulo: Pearson-Addison-Wesley. (2005).

Date, C. J.. Introdução a Sistemas de Banco de Dados. 8º ed. — Rio de Janeiro: Elsevier Editora. (2004).

CREATIVE COMMONS. Atribuição-NãoComercial-CompartilhaIgual CC BY-NC-SA. Disponível em <https://creativecommons.org/licenses/by-nc-sa/3.0/br/>. (2018).

PROTO.IO. Disponível em <https://proto.io/>. (2018) .

Uma Plataforma Flexível de Computação de Alto Desempenho para Solução de Problemas de Sensoriamento Remoto

Leandro L. Parente¹, Roberto U. Paiva¹, Claudinei O. Santos¹, Evandro C. Taquary¹, Laerte G. Ferreira¹

¹Lapig – Laboratório de Processamento de Imagens e Geoprocessamento (UFG)
CAMPUS II Samambaia - Cx. POSTAL 131 - Goiânia - GO / CEP: 74001-970

{leal.parente, robertourzed, claudineisantosnx, evandro.taquary, lapig.ufg}@gmail.com.br

Abstract. *The increasing imaging capacity of the terrestrial surface, through governmental and private satellites, is producing an unprecedented remote sensing data amount. With ever better spatial and temporal resolutions, the transformation of this data into scientifically relevant information demands high-performance computing platforms. The objective of this work is to present a computational infrastructure that enables the practice of High Performance Computing to optimize the solutions of remote sensing problems. To this end, a flexible parallel and distributed computing platform is being implemented and tested in solving problems in this area. In this work we present the results and viability of the platform by running BFAST algorithm, used to detect changes in time series analysis. The results showed that the platform is promising, and can optimize solutions of remote sensing problems at low cost and little programming effort.*

Resumo. *O aumento da capacidade de imageamento da superfície terrestre, por meio de satélites governamentais e privados, está produzindo um volume de dados sem precedentes para a área de sensoriamento remoto. Com resoluções espaciais e temporais cada vez melhores, a transformação destes dados em informações cientificamente relevantes demanda plataformas de computação de alto desempenho. O objetivo deste trabalho é apresentar uma infraestrutura computacional que viabilize a prática da Computação de Alto Desempenho para acelerar a solução de problemas de sensoriamento remoto. Para tanto, uma plataforma flexível de computação paralela e distribuída está sendo implementada e testada na resolução de problemas desta área. No presente trabalho, apresentamos os resultados e viabilidade da plataforma executando o algoritmo BFAST, utilizado para detecção de mudanças em análises de séries temporais. Os resultados mostraram que a plataforma é promissora, e pode otimizar a solução de problemas da área de sensoriamento remoto a baixo custo e pouco esforço de programação.*

1. Introdução

O aumento da capacidade de imageamento da superfície terrestre, por meio de satélites governamentais [ROY et al., 2014] e privados [HOUBORG, R.; MCCABE, M. F. A., 2018], está produzindo um volume de dados sem precedentes para a área de sensoriamento remoto [HANSEN et al. 2012]. Com resoluções espaciais (*i.e.* correspondência espacial de um pixel em relação a uma porção do território) e temporais (*i.e.* quantidade de tempo necessária para

revisitar um mesmo local) cada vez melhores, a transformação destes dados em informações cientificamente relevantes demanda plataformas de computação de alto desempenho. Atualmente o Google Earth Engine, uma plataforma de *cloud-computing* capaz analisar dados de sensoriamento remoto em escala planetária, se coloca como uma alternativa para atender esta demanda, entretanto seus usuários precisam adequar suas rotinas de processamento para os paradigmas de programação e linguagens suportadas pela plataforma [GORELICK et al., 2017]. Esta característica dificulta a utilização de algoritmos já disponíveis em diversos frameworks pela comunidade científica em geral [Ma et al., 2015].

Diante disto, este trabalho buscou desenvolver uma plataforma flexível de computação de alto desempenho, específica para problemas de sensoriamento remoto, capaz de permitir a execução paralela de rotinas de processamento já implementadas. A rotina de processamento utilizada como estudo de caso foi o BFAST (*Breaks For Additive Seasonal and Trend*), um algoritmo de detecção de mudanças, que decompõe uma série temporal de pixels em três componentes (tendência, sazonalidade, e ruído), e está implementando em R [VERBESSELT et al., 2010]. A rotina de detecção de mudanças do BFAST utiliza um processo iterativo, por meio de ajuste de um modelo linear por partes, o que torna o algoritmo computacionalmente oneroso e justifica sua utilização na plataforma proposta. Sua execução ocorreu sobre dados MODIS, com resolução espacial de 250m e resolução temporal de 16 dias [FRIEDL et al., 2010], obtidos sobre o Bioma Cerrado - com uma extensão total de 2.045.000 km² - nos últimos 17 anos.

2. Ambientes de Testes

A execução da rotina de processamento ocorreu em em dois ambiente: em um servidor em torre, para a execução sequencial, e na plataforma proposta por este trabalho, para execução paralela e distribuída, com o intuito de obter resultados de performance.

2.1. Ambiente de teste sequencial

O servidor em torre utilizado para o teste de processamento sequencial possui as seguintes especificações de *hardware*:

- Placa mãe: *Dell PowerEdge T420 3015M*
- Processador: 2 processadores Intel Xeon CPU E5-2430 v2 - 2.50 Ghz
- Memória: 22GB
- Armazenamento: *HD 3TB*

2.2. Ambiente de teste paralelo/distribuído

Os testes de processamento em paralelo e distribuído foram realizados utilizando um grid com 60 nós (ver figura 1), no qual o servidor mestre encarregou-se de distribuir a mesma imagem do sistema operacional Debian 9, via NFS (*Network File System*), TFTP (*Trivial File Transfer Protocol*) e boot PXE (*Preeboot Execution Environment*), garantindo um ambiente idêntico e homogêneo para todos os nós. A tecnologia *Wake on Lan* foi utilizada para ligar e desligar os nós do *grid* e o tráfego de rede entre os nós ocorreu via rede *gigabit*. As especificações de *hardware* dos nós do *grid* são:

- Placa mãe: Intel modelo 525mw
- Processador: Intel Atom D525 (*Dual-Core*) - 1.80 Ghz
- Memória: 4GB
- Armazenamento: Partição de armazenamento *GlusterFS* - 4TB

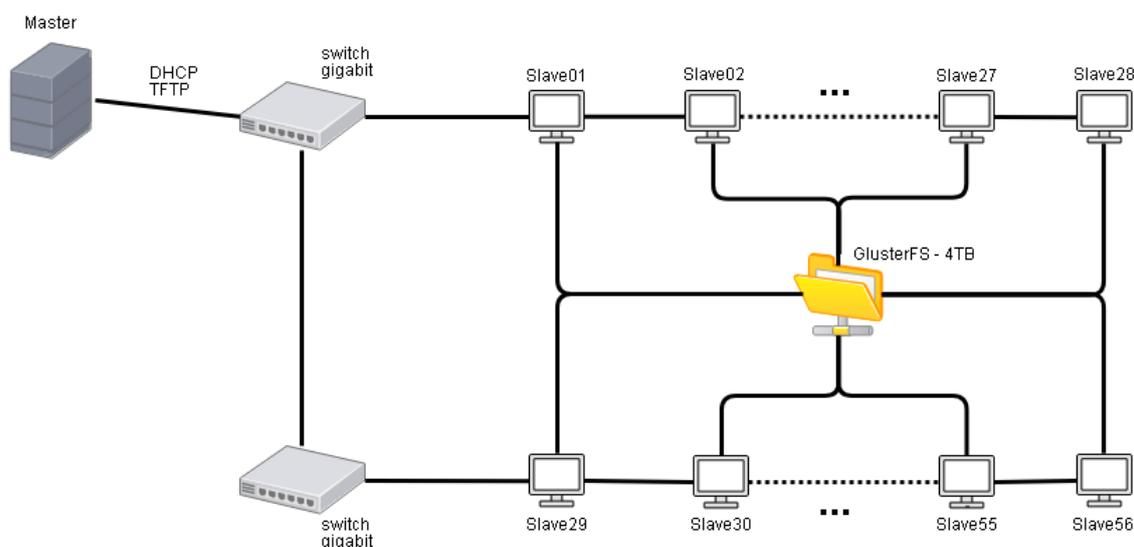


Figura 1. Arquitetura da plataforma, na qual o servidor Master encarregou-se de distribuir a mesma imagem de sistema operacional, com o intuito de garantir um ambiente idêntico e homogêneo, para todos os nós. O GlusterFS utilizou uma estratégia distribuída e replicada para armazenar e compartilhar os dados na plataforma.

Para armazenamento e compartilhamento de dados, foi utilizado um sistema de arquivos distribuído, em 4 nós via *GlusterFS*. O *GlusterFS* é uma tecnologia de armazenamento que permite, a partir de vários volumes hospedados em diferentes servidores, a construção de um sistema de arquivos em rede replicado e distribuído [DAVIES, A.; ORSARIA, A., 2013]. Um *benchmark* de performance entre o *GlusterFS*, *Red Hat® Ceph Storage* e *Hadoop Distributed File System* foi utilizado como critério de escolha de sistema de arquivo distribuído [TESTING, 2018]. O *GlusterFS* dispõe de diversas estratégias de alocação dos dados, a partição de armazenamento da arquitetura optou por uma estratégia distribuída e replicada, devido seu melhor desempenho de leitura de arquivos e garantias de confiabilidade e escalabilidade [RED HAT, 2018].

2.3. Estratégia de Paralelização

O problema em questão é “trivialmente paralelizável” [HERLIHY, 2011], sendo possível atribuir uma tarefa de processamento a cada série temporal, para serem executadas simultânea e independentemente, sem necessidade de intercomunicação. Desta forma, num cenário ideal, poderíamos associar cada série a um núcleo de processamento e, em tese, realizar o processamento de todas as séries temporais com o mesmo tempo de processamento de uma única. Nesse sentido os testes foram realizados atribuindo blocos de séries-temporais a tarefas de processamento que foram executadas concorrentemente nos nós do cluster. Ao todo foram processados 1.820 blocos com 27.300 séries temporais com 414 observações cada. O tempo total de processamento foi aproximadamente 6 dias, com consumo máximo de CPU em 56 nós. O processamento foi paralelizado via Slurm (Simple Linux Utility for Resource Management), um distribuidor de tarefas com funcionalidades de gerenciamento de recursos dos nós, tal como alocação de memória e núcleos de processamento [YOO, 2003].

3. Resultados

O tempo total de execução do BFAST, incluindo leitura/escrita dos dados e processamento, foi

~303 horas, ou 12,6 dias, em um ambiente de testes multicore e ~152 horas, ou 6,3 dias, no ambiente de testes do cluster. Algumas considerações em relação aos ambientes de testes devem ser discutidas. Foi utilizado um servidor de médio custo para a realização dos testes no ambiente multicore, uma máquina com um sistema de arquivos local, sem tempo de rede para leitura e escrita dos dados, com 12 processadores físicos (24 núcleos no total) e tempos de *clock* e arquitetura mais robusta do que as máquinas do *grid*. O *grid*, no entanto, possui máquinas de baixo consumo e custo, com arquiteturas mais antigas e simples, fazendo uso de um ambiente de armazenamento compartilhado na rede. Mesmo diante deste contexto, a plataforma se mostrou mais eficiente em termos de performance.

4. Conclusão

Este trabalho desenvolveu uma plataforma inicial de computação de alto desempenho na qual foi possível utilizar algoritmos e rotinas de processamento prontas, sem nenhuma necessidade de re-implementação, a qual permitiu diminuir pela metade o tempo de execução do algoritmo BFAST. As tecnologias utilizadas pela plataforma permitem aumentar a escalabilidade de processamento, sendo relativamente simples incluir outros nós de processamento, homogêneos e/ou heterogêneo na arquitetura. Os resultados mostraram a possibilidade desta plataforma tirar proveito do poder de processamento latente de equipamentos *commodities*, a um custo relativamente baixo e com muito pouco esforço de programação. Não obstante, a arquitetura proposta também pode ser implantada em equipamentos mais robustos, tanto quando em ambientes de *cloud-computing*. Como ideia para trabalhos futuros, pode-se pensar em criar containers de aplicação (ex. *Docker*) a fim de dinamizar a implantação da plataforma em qualquer infraestrutura computacional, aumentando ainda mais sua flexibilidade e operação em diversos ambientes computacionais.

Referências

- [1] ROY, D.P.; WULDER, M.A.; LOVELAND, T.R.; WOODCOCK, C.E; ALLEN, R.G.; ANDERSON, M.C.; HELDER, D.; IRONS, J.R.; JOHNSON, D.M.; KENNEDY, R.; SCAMBOS, T.A.; SCHAAF, C.B.; SCHOTT, J.R.; SHENG, Y.; VERMOTE, E.F.; BELWARD, A.S.; BINDSCHADLER, R.; COHEN, W.B.; GAO, F.; HIPPLE, J.D.; HOSTERT, P.; HUNTINGTON, J.; JUSTICE, C.O.; KILIC, A.; KOVALSKYY, V.; LEE, Z.P.; LYMBURNER, L.; MASEK, J.G.; MCCORKEL, J.; SHUAI, Y.; TREZZA, R.; VOGELMANN, J.; WYNNE, R.H.; ZHU, Z. Landsat-8: Science and product vision for terrestrial global change research. *Remote sensing of Environment*, v. 145, p. 154-172, 2014.
- [2] HOUBORG, R.; MCCABE, M. F. A. Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sensing of Environment*, v. 209, p. 211-226, 2018.
- [3] HANSEN, M. C.; LOVELAND, T. R. A review of large area monitoring of land cover change using Landsat data. *Remote sensing of Environment*, v. 122, p. 66-74, 2012.

- [4] GORELICK, N.; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S.; THAU, D.; MOORE, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, v. 202, p. 18-27, 2017.
- [5] Ma, Y.; Wu, H.; Wang, L.; Huang, B.; Ranjan, R.; Zomaya, A.; Jie, W. Remote sensing big data computing: challenges and opportunities. *Future Generation Computer Systems*, v. 51, p. 47-60, 2015.
- [6] VERBESSELT, J., HYNDMAN, R., NEWNHAM, G., CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. *Remote sensing of Environment*, v. 114, n. 1, p. 106-115, 2010.
- [7] FRIEDL, M. A.; SULLA-MENASHE, D.; TAN, B.; SCHNEIDER, A.; RAMANKUTTY, N.; SIBLEY, A.; HUANG, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote sensing of Environment*, v. 114, n. 1, p. 168-182, 2010.
- [8] DAVIES, A., & ORSARIA, A. Scale out with GlusterFS. *Linux Journal*, v. 2013, n. 235, p. 1, 2013.
- [9] TESTING OF SEVERAL DISTRIBUTED FILESYSTEMS (HDFS , CEPH AND GLUSTERFS) FOR SUPPORTING THE HEP EXPERIMENTS ANALYSIS. SEMANTICSCHOLAR. Disponível em:
<[https://www.semanticscholar.org/paper/Testing-of-several-distributed-filesystems-\(HDFS-C-Donvito-Marzulli/776dc2f7cdeb32274d0c0e5fecf7ef9a13fcdf13](https://www.semanticscholar.org/paper/Testing-of-several-distributed-filesystems-(HDFS-C-Donvito-Marzulli/776dc2f7cdeb32274d0c0e5fecf7ef9a13fcdf13)>. Acesso em: 15 ago. 2018.
- [10] RED HAT GLUSTER STORAGE TECHNICAL PRESENTATION. Red Hat. Disponível em:
<<https://www.redhat.com/cms/managed-files/Red%20Hat%20Gluster%20Storage%20Technical%20Deck%281%29.pdf>>. Acesso em: 15 ago. 2018.
- [11] HERLIHY, M., & SHAVIT, N. The art of multiprocessor programming. Morgan Kaufmann, 2011.
- [12] YOO, A. B., JETTE, M. A., GRONDONA, M. Slurm: Simple linux utility for resource management. In: *Workshop on Job Scheduling Strategies for Parallel Processing*. Springer, Berlin, Heidelberg, 2003. p. 44-60.

FGSSJOIN: UM ALGORITMO PARALELO BASEADO EM GPU PARA O PROBLEMA DA JUNCTÃO POR SIMILARIDADE EM BASES DE DADOS TEXTUAIS

Rafael David Quirino¹, Wellington Santos Martins², Leonardo Andrade Ribeiro³

Instituto de Informática
Universidade Federal de Goiás
email: rafaelquirino@inf.ufg.br, wellington@inf.ufg.br

Abstract. Similarity joins are essential operations in systems that perform data integration, cleaning, and mining. However, getting all pairs of similar records in a database can be a computationally expensive operation. Most state-of-the-art solutions are based on inherently sequential algorithms. In this work we propose a parallel algorithm to solve the problem, taking advantage of the power of many-core architectures. Our solution uses filtering and divide-and-conquer techniques, achieving good scalability. The experiments showed significant increases in performance compared to the fastest algorithms in the literature.

Resumo. Junções por similaridade são operações básicas e essenciais em sistemas que realizam integração, limpeza e mineração de dados. Entretanto, obter todos os pares de registros similares em um Banco de Dados pode ser uma operação muito cara computacionalmente. A maioria das soluções do estado-da-arte se baseiam em algoritmos inerentemente sequenciais. Neste trabalho propomos um algoritmo paralelo para resolver o problema, aproveitando o poder das arquiteturas many-core. Nossa solução utiliza técnicas de filtragem e de divisão-e-conquista, alcançando boa escalabilidade. Os experimentos mostraram aumentos significativos de performance em comparação com os algoritmos mais rápidos da literatura.

Palavras-chave: Processamento Avançado de Consultas, Computação de Alto Desempenho, Junção Paralela por Similaridade de Conjuntos, GPU.

1 INTRODUÇÃO

Existem vários formatos de aplicação da operação de Junção por Similaridade de Conjuntos, mas a utilizada neste trabalho, Auto Junção de Todos os Pares (All-pairs Self Join), consiste na operação de obtenção de uma lista de todos os pares de registros de uma única relação cuja pontuação de similaridade, de acordo com alguma Função de Similaridade baseada em Conjuntos, seja maior ou igual a um limiar escolhido (por exemplo 70%). Escolhemos trabalhar dados textuais, utilizando nos experimentos bases de dados reais e frequentemente utilizadas na literatura, para fins de comparação com os algoritmos já existentes.

Determinar a exata pontuação de similaridade entre objetos complexos, particularmente em junções onde todos os pares precisam ser comparados, é uma operação frequentemente

¹Aluno

²Orientador

³Co-orientador

cara, mesmo para os algoritmos do estado-da-arte. No caso de dados textuais, o problema da "maldição da dimensionalidade" se torna muito aparente, uma vez que as representações de dados textuais tendem a ser esparsas e de alta dimensionalidade. Nesse contexto, funções de similaridade baseadas em conjuntos são bastante atrativas pois os predicados envolvendo tais funções podem ser equivalentemente representadas em termos de restrições de intersecção. Consequentemente, o problema pode ser reduzido à identificação de todos os pares com intersecção suficiente, dado um certo limiar.

Neste cenário fica claro a necessidade de soluções paralelas para o problema. Nesse trabalho nós apresentamos um algoritmo paralelo, de granularidade fina, para resolver o problema da Junção por Similaridade de Conjuntos, bem como uma implementação baseada em GPU, que permite a utilização de técnicas de filtragem por prefixo consagradas na literatura. O algoritmo paralelo proposto se baseia em estratégias de particionamento e divisão-e-conquista que garantem o processamento de bases de dados volumosas com tempos de execução inferiores aos dos melhores algoritmos do estado-da-arte. As principais contribuições desse trabalho são:

- Um algoritmo paralelo para o problema da Junção por Similaridade de Conjuntos.
- Uma implementação baseada em GPU para o algoritmo proposto.
- Um trabalho experimental extenso, com bases de dados padrões na literatura relacionada.

2 JUNÇÃO PARALELA E O ALGORITMO FGSSJOIN

2.1 Indexação

Nos algoritmos do estado-da-arte, as listas do índice invertido são criadas durante o processo de filtragem, o que os torna inerentemente algoritmos sequenciais: conjuntos são sequencialmente testados contra o índice e o estado das listas em uma iteração depende do seu estado na iteração anterior. Para resolver este problema, precisamos criar todo o índice invertido estaticamente antes da fase de filtragem. Desta forma, podemos realizar sondagens de forma independente, porque o índice está sempre completo. Portanto, precisamos de um algoritmo paralelo eficiente para criar o índice invertido. Para isso, precisamos concatenar todos os elementos de todos os conjuntos em um vetor que chamamos de E (Entries). Seja $e \in E$ uma entrada que contenha três campos: o conjunto ao qual ele pertence, uma *feature* (equivalente a um qgram de algum conjunto) e sua informação posicional. Assim, os conjuntos são reduzidos a um vetor de tuplas (s_i, f_i, p_i) . Isso será importante na fase de filtragem posicional. Ao criar as entradas para o algoritmo de índice invertido, nós apenas adicionamos features do midprefix ao vetor E . Então, computamos um vetor de contagem, que conta a ocorrência de cada feature e, em seguida, executa uma soma de prefixos para obter os índices iniciais de cada lista no índice invertido. O algoritmo 1 mostra a estratégia paralela usada para criar todo o índice invertido na memória. Observe que V representa nosso dicionário (vocabulário).

2.2 Filtragem

Com todo o índice invertido armazenado na memória, podemos configurar cada processador para executar uma sondagem (processar um conjunto) em relação ao índice, usando apenas features do prefixo médio do conjunto. Um problema importante é como acumular as pontuações (intersecções) entre todos os pares de conjuntos. Uma vez que podemos potencialmente ter todos os pares como sendo candidatos (especialmente para limiares mais baixos), precisamos de "buckets" para todos os pares possíveis de conjuntos na coleção. Então, criamos uma matriz de $n \times n$ elementos chamada b (os compartimentos para conter as pontuações parciais), onde n é o número de conjuntos em nossa coleção). Ele conterá a intersecção parcial (pontuações parciais) entre conjuntos sondados e indexados, pois apenas partes dos conjuntos (os elementos do prefixo) são usados nesta etapa. Depois de realizar a fase de filtragem, precisamos compactar o conteúdo da tabela de pontuação para obter apenas as pontuações positivas, ou seja, os pares candidatos. Então, criamos um vetor $n \times n$ adicional, o vetor *compacted_buckets*. Após a compactação, esta matriz conterá os índices das pontuações positivas na tabela. Para reduzir os requisitos de memória (vetor $n \times n$) particionamos a base de dados em blocos e, portanto, podemos processar bases de qualquer tamanho. A figura 1 ilustra o estado da memória após a fase de filtragem. Observe que cada linha na primeira matriz representa um conjunto que está sendo consultado contra o índice, enquanto cada coluna representa um conjunto indexado.

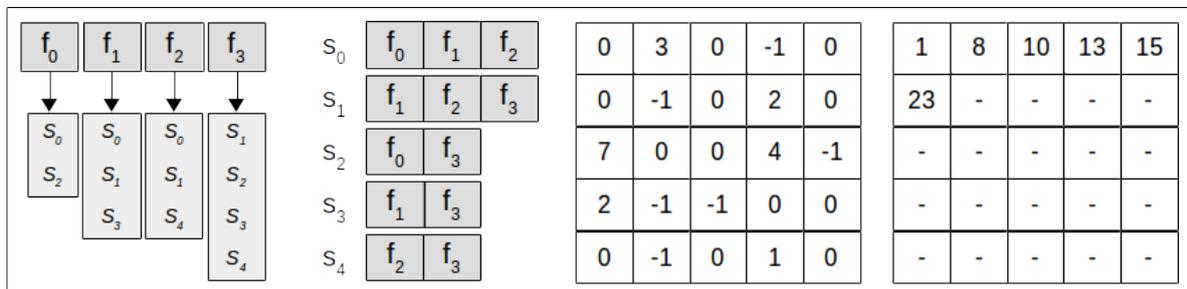


Figura 1: Memória da GPU após a fase de filtragem, com índice invertido, conjuntos, a matriz b (buckets) de pontuações parciais e sua matrix correspondente compactada (com os índices para os elementos em b que possuem valores positivos), respectivamente. Uma vez que temos cinco conjuntos neste exemplo, nossas tabelas são 5×5 . Os números na primeira tabela são os pontuações parciais após a fase de filtragem. Números negativos (-1) representam $-\infty$. A segunda tabela é a versão compactada da primeira; Seus valores (os índices absolutos dos elementos selecionados da primeira tabela) representam os pares candidatos e são fornecidos a fase de verificação.

2.3 Verificação

A fase de verificação, que finalmente produz o resultado final, pode ser processada trivialmente em paralelo. Ele consiste simplesmente em realizar o cálculo da pontuação restante em cada

par candidato para verificar se existe intersecção suficiente para qualificá-lo como um par similar. Isso pode ser feito facilmente em paralelo, fazendo com que cada processador execute a verificação em um par candidato. Podemos usar a pontuação parcial para reduzir um pouco os cálculos de intersecção. Ao comparar a última feature dos prefixos dos dois conjuntos em um par candidato, podemos iniciar o cálculo de intersecção na posição da feature com o menor ID em seu próprio conjunto, e desde o início no outro conjunto do par, com o valor inicial sendo a pontuação parcial do par candidato, já que qualquer correspondência com features do prefixo neste conjunto já foi calculada na fase de filtragem. Em cada etapa do cálculo de intersecção, também testaremos se as features restantes são suficientes para atingir o limiar, marcando aqueles que não são. Aqueles não marcados por este processo formam o conjunto resultado, os pares semelhantes de acordo com o limiar e a função de similaridade.

3 EXPERIMENTOS

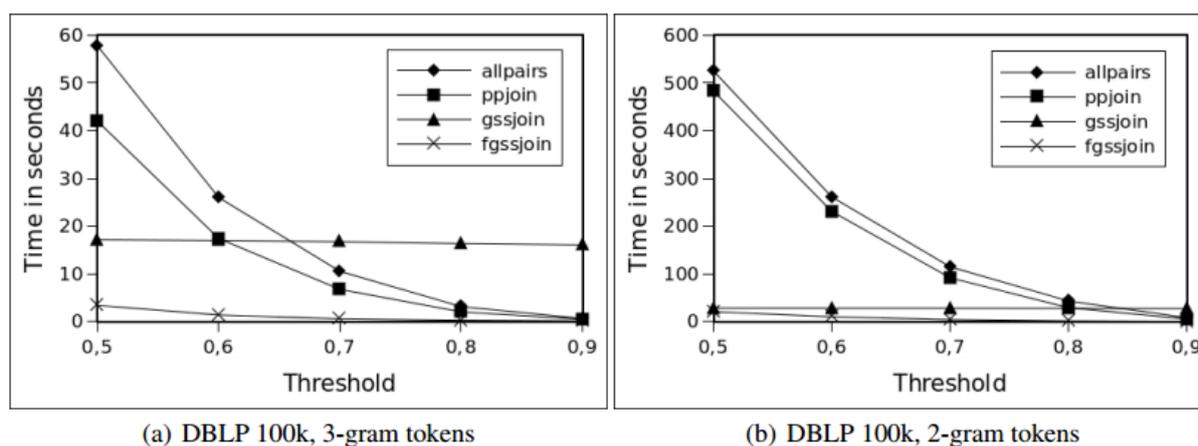


Figura 2: Tempos de execução para a base DBLP, 100k registros, com conjuntos 2-gram e 3-gram, para vários valores de limiar.

Testamos dois algoritmos sequenciais de referência, *allpairs* [Bayardo et al 2007] e *ppjoin* [Xiao et al 2011], bem como um algoritmo massivamente paralelo *gssjoin* [Ribeiro et al 2016] e o nosso algoritmo *fgssjoin*.

Nossos experimentos foram executados em uma máquina equipada com dois Intel Xeon E5-2620, cada um com 6 núcleos de processamento (12 threads em hiper-threading) e 20MB de memória cache, 16GB de memória RAM e 4 Nvidia GTX Titan Black, cada uma com 2880 núcleos de processamento e 6 GB de memória, apesar de só utilizarmos uma GTX Titan Black para nossa implementação paralela.

Utilizamos dois bancos de dados: DBLP (uma coleção de títulos de artigos de ciência da computação e autores), com registros de 100k e IMDB (uma coleção de títulos de filmes, programas de TV, etc.), com 300k registros, que são bases de dados populares em trabalhos na área. Nós pré-processamos os conjuntos de dados, removendo acentos e pontuação e definindo todos

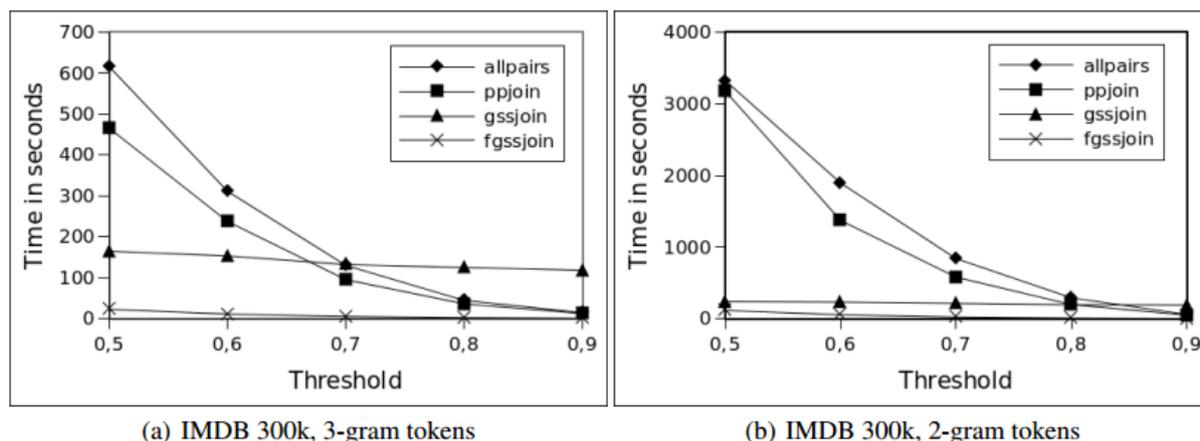


Figura 3: Tempos de execução para a base IMDB, 300k registros, com conjuntos 2-gram e 3-gram, para vários valores de limiar.

os caracteres em caixa baixa. Ambos os conjuntos de dados foram tokenizados com 2-grams e 3-grams. Realizamos nossos experimentos variando os valores de limiar de 0,5 a 0,9, em incrementos de 0,1. Os algoritmos de comparação foram executados no processador Xeon, exceto gssjoin, que também foi executado no GTX Titan Black.

Nosso algoritmo alcançou acelerações consideráveis de até 25x em relação ao algoritmos de referência na literatura. As melhores velocidades foram alcançadas quando os conjuntos de dados foram tokenizados com 2-grams. Claro que os 2-grams menores requerem mais poder computacional, uma vez que há menos combinações de 2 caracteres e isso leva a mais pares candidatos na fase de filtragem, elevando a carga para a fase de verificação. Além disso, esperamos que nosso algoritmo se torne ainda melhor à medida que o tamanho do conjunto de dados cresce.

Referências

- [1] Leonardo Andrade Ribeiro e Theo Härder, Generalizing Prefix Filtering to Improve Set Similarity Joins, *Information Systems*, pp. 62-78, (2011).
- [2] Chuan Xiao e Wei Wang e Xuemin Lin e Jeffrey Xu Yu e Guoren Wang, Efficient Similarity Joins for Near-duplicate Detection, *TODS*, pp. 15, (2011).
- [3] Roberto J. Bayardo e Yiming Ma e Ramakrishnan Srikant, Scaling up All Pairs Similarity Search, *WWW*, pp. 131-140, (2007).
- [4] Sidney R. Junior e Rafael D Quirino e Leonardo Andrade Ribeiro e Wellington S Martins, gSSJoin: a GPU-based Set Similarity Join Algorithm, *SBBD*, pp. 64-75, (2016).
- [5] Rafael D Quirino e Sidney R. Junior e Leonardo Andrade Ribeiro e Wellington S Martins, fgssjoin: A GPU-based Algorithm for Set Similarity Joins, *19th ICEIS*, pp. 152-161, (2017).

A Fast Word2Vec implementation on manycore architectures for Text Representation and its applications

Leonardo A. Amorim, Mateus F. Freitas, Chayner C. Barros, Wellington S. Martins

¹ Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74.001-970 – Goiânia – GO – Brazil

{leonardoafonso,mateusfreitas,chaynercordeiro,wellington}@inf.ufg.br

Abstract. *Word embedding has made possible to work with semantics in any application that works with a text document. Through algorithms that implement this technique, such as Word2Vec, it is possible to discover the similarity between words, paragraphs and even whole documents. However, generating word embedding still has a high computational cost. Some researches have been proposed parallel algorithms in recent years to deal with this problem, but gains in performance have ranged from 2 to 20 times compared to the original implementations. Manycore architectures have been able to scale algorithms in a more performative way. Since the accuracy of the word embeddings generation depends on a large amount of data (Big Data), it is necessary that new scalable parallel algorithms are developed to deal with this large amount of data (billions of words). Scalable parallel algorithms development is one the most complex and difficult tasks so, in this work, we focus on exploiting parallelism in text representation with applications. We work with Word2Vec for text representation.*

1. Introduction

The representation of words by vectors has become state-of-the-art in classification tasks. This is mainly due to the growth of Word2Vec that generates compact representations of words to express the semantic relationships through vector operations. The Word2Vec has been used in various fields of computing such as Sentiment Analysis, Word Clustering, Machine Translation, Named entity recognition and Automated Program Repair [Amorim and et al 2018]. The use of Word2Vec in Information Retrieval may be promising because it adds semantics in document classification and this can enhance the text classification task.

The major Word2Vec bottlenecks are inherent iterative nature of Stochastic Gradient Descent (SGD) algorithm and between the hidden layer and output layer (or Softmax layer). The time taken to generate the word embeddings makes the experiments in the areas cited severely affected by delays in the results. This delay can force the use of small vocabulary and, as a consequence, can impact the accuracy of the generation of word embeddings. There are parallel algorithms for the SGD in manycore architectures, but they do not have good strong scalability. The scalability of the parallel algorithm is important because every day we have hardware with more cores and taking advantage of this capability is a major challenge. In this context, there are two open questions that are going to guide this research:

- How to exploit parallelism in text representation in an efficient and scalable way?

- How to apply the parallel proposals in the context of Automated Program Repair application?

The work is organized as follows. Section 2 provides background to Word embedding model. Section 3 summarizes existing approaches to improving Word2Vec performance. Section 4 presents new ideas to explore parallelism in Word2Vec. Finally, in Section 5 we present our conclusions.

2. Background

Word embedding is the collective name for a set of language modeling and feature learning techniques in natural language processing where words or phrases from the vocabulary are mapped to vectors of real numbers in a low-dimensional space relative to the vocabulary size (“continuous space”). Methods to generate this mapping include neural networks, dimensionality reduction on the word co-occurrence matrix [Levy and Goldberg 2014], probabilistic models, and explicit representation in terms of the context in which words appear. Word and phrase embeddings, when used as the underlying input representation, have been shown to boost the performance of NLP tasks such as syntactic parsing, sentiment analysis, automatic document classification. Current methods for generating word embeddings fall into two categories: count-based methods [Blei et al. 2003] (e.g. Latent Semantic Analysis), and predictive methods (e.g. neural probabilistic language models). In this work, we focus on Word2Vec, which is a predictive method with moderate computational costs that permits the processing of large datasets.

Word2Vec learns continuous word embeddings from plain text in an entirely unsupervised way. The model assumes the Distributional Hypothesis [Harris 1954], which states that words that appear in the same contexts tend to share semantic meaning. This allows us to estimate the probability of two words occurring close to each other. With Word2Vec, a neural network is trained with streams of n -grams of words so as to predict the n -th word, given words $[1, \dots, n - 1]$ or the other way around. The output is a matrix of word vectors or context vectors respectively. Since the neural network used has a simple linear hidden layer, the intensity of correlation between words is directly measured by the inner product between word embeddings. Such a shallow neural network might not produce as precise distributed representations as deep neural networks on relatively small datasets, but they can process data much more efficiently. When operating on large datasets, this approach usually produces better results than using a more sophisticated algorithm [Mikolov et al. 2013].

The two neural network models used by Word2Vec are known as continuous bag-of-words (CBOW) and Skip-gram. Rather than predicting a word conditioned on its predecessor, as in a traditional bi-gram language model, the CBOW model predicts the current word based on the context, while the Skip-gram model predicts the neighborhood words given the current word. The training of these models has to compute normalization terms, which has complexity $\mathcal{O}(|\mathcal{V}|)$, where $|\mathcal{V}|$ is the vocabulary size. To reduce this high computational cost, Word2Vec uses fast training algorithms: hierarchical softmax and negative sampling. Hierarchical softmax makes use of a Huffman tree representation of the vocabulary, which saves calculations, at the potential loss of some accuracy. On the other hand, negative sampling avoids using the words observed next to one another in

the training data as positive examples, and instead sample random words from the corpus and present to the network as negative examples. Word2Vec remains a popular choice for building word vectors due to their efficiency and simplicity. Although these algorithms are widely used, generating word embeddings are still too costly, which impacts negatively on the time for conducting experiments in both Information Retrieval and Machine Learning applications.

3. Related Works

The first state-of-the-art algorithms for Word embeddings including Word2vec have been parallelized for multi-core CPU architectures but are based on vector-vector operations that are memory-bandwidth intensive and do not efficiently use computational resources. The original Word2Vec was implemented this way.

To reduce generation time for word vectors, [Liu, 2014] implemented the optimization CBOW model in CUDA. He made a in-warp approach to CUDA architecture. A WARP works with one word and updates the hidden layer of the artificial neural network in shared memory. Each WARP of thread updates the weights for one word of the sentence. He uses one in-warp to handle one data (one word), and 32 threads manage the parallelism in one data in in-warp [Liu 2014].

In [Ji et al. 2016], it was shown that the reuse of various data structures in the algorithm through the use of minibatching, allows one to express the problem using matrix multiply operations producing good strong scalability. In [Taylor et al. 2016], it was explored an unconventional training method to train networks without gradient descent steps.

In [Gupta and Khare 2017], they present BlazingText, a high performance distributed word2vec implementation that leverages massive parallelism provided by modern GPUs. They exploit GPU parallelism to discover the right balance between throughput and accuracy. This work replicated the matrices and processed disjoint parts of sentences in each GPU. Their work uses one thread block per sentence. Since the corpus is too large to fit in the GPU memory, this work uses disk stream to the GPU and several sentences are batch transferred to decrease the cost of data transfer between CPU and GPU. And at regular intervals, their algorithm produces a combination of matrix weights (one coming from each GPU) and transmits the new weights (new matrices) for each GPU to continue its iteration. According to the authors, there was no significant fall in accuracy with multi-GPU implementation. Their proposed implementation achieves near-linear scalability across multiple GPUs.

In [Shim et al. 2017], they propose a fast approximation method of a softmax function with a very large vocabulary using singular value decomposition (SVD). SVD-softmax targets fast and accurate probability estimation of the topmost probable words during inference of neural network language models. The proposed method transforms the weight matrix used in the calculation of the output vector by using SVD. The approximate probability of each word can be estimated with only a small part of the weight matrix using a few large singular values and the corresponding elements for most of the words.

In [Simonton and Alaghband 2017], the authors propose new methods to increase the speed of the Word2Vec on multi-core shared memory CPU systems, and on modern

NVIDIA GPUs with CUDA. They perform this on multi-core CPUs by batching training operations to increase thread locality and to reduce accesses to shared memory. They propose new heterogeneous NVIDIA GPU CUDA implementations of both the skip gram hierarchical softmax and negative sampling techniques that utilize shared memory registers and in-warp shuffle operations for maximized performance.

Grave et al. propose an approximate strategy to train neural network based language models over vast vocabularies efficiently. Their approach, called adaptive softmax, avoids the linear dependency on the vocabulary size by exploiting the unbalanced word distribution to form clusters that explicitly minimize the expectation of computational complexity. Their approach further reduces the computational cost by employing the specificities of modern architectures and matrix-matrix vector operations [Grave et al. 2016].

The speedup value of the parallel state-of-art versions ranges from 2 to 20 times. Therefore, we will present in the next section some ideas that have the potential to improve the scalability of Word2Vec.

4. New ideas to explore parallelism in Word2Vec

To gain more performance and scalability of Word2Vec in many-core architectures, we speculate that the following experiments can accomplish promising results:

- Regarding implementation in Multi-GPU, we can develop new schemes to distribute the gradient for Artificial Neural Network used by Word2Vec. These schemes can be inspired by articles of Deep Learning in Distributed Systems [Dean et al. 2012];
- Search for a data structure that better exploits the features of manycore architectures (that allows exploring more parallelism i.e without leaving idle processors) in order to replace the Huffman tree in the Hierarchical Softmax;
- Group words that are most likely to co-occur in order to generate vectors for word classes [Shen et al. 2017]. As a consequence, the number of vectors that need to be updated will decrease. It is a similar effect when reducing matrices using SVD. This idea explores the use of K-Nearest Neighbors algorithm and uniform sampling to approximate the softmax function and achieve $\mathcal{O}(\sqrt{V})$ runtime¹;
- Explore matrix batch multiplication operations for CBOW Architecture in Multi-GPU.

5. Conclusion

The proposals of the related works can be summarized in reducing the weight matrices of the Word2Vec model; perform matrix multiplication operations rather than vector-matrix; perform operations on batch matrices in order to hide latency; to exploit low-level GPU resources (operations using in-warp) in order to increase scalability and reduce memory access. Although parallel implementations for Word2Vec or Softmax have had good performance gains, scalability can still be improved. Therefore, we expect to acquire novel and faster Word2Vec algorithm to generate word embeddings able to surpass current state of the art baselines.

¹<http://cs231n.stanford.edu/reports/2017/pdfs/130.pdf>

References

- Amorim, L. and et al (2018). A new word embedding approach to evaluate potential fixes for automated program repair. In *The International Joint Conference on Neural Networks*. IEEE 2018.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.
- Dean, J., Corrado, G. S., Monga, R., Chen, K., Devin, M., Le, Q. V., Mao, M. Z., Ranzato, M., Senior, A., Tucker, P., Yang, K., and Ng, A. Y. (2012). Large scale distributed deep networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS’12, pages 1223–1231, USA. Curran Associates Inc.
- Grave, E., Joulin, A., Cissé, M., Grangier, D., and Jégou, H. (2016). Efficient softmax approximation for gpus. *CoRR*, abs/1609.04309.
- Gupta, S. and Khare, V. (2017). Blazingtext: Scaling and accelerating word2vec using multiple gpus. In *Proceedings of the Machine Learning on HPC Environments, MLHPC’17*, pages 6:1–6:5, New York, NY, USA. ACM.
- Harris, Z. (1954). Distributional structure. *Word*, 10(23):146–162.
- Ji, S., Satish, N., Li, S., and Dubey, P. (2016). Parallelizing word2vec in shared and distributed memory.
- Levy, O. and Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27*, pages 2177–2185. Curran Associates, Inc.
- Liu, R. (2014). Unpublished work, retrieved from the author’s github web page <https://libraries.io/github/fengchenhpc> on 2018-08-15.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *CoRR*, abs/1310.4546.
- Shen, Y., Tan, S., Pal, C. J., and Courville, A. C. (2017). Self-organized hierarchical softmax. *CoRR*, abs/1707.08588.
- Shim, K., Lee, M., Choi, I., Boo, Y., and Sung, W. (2017). Svd-softmax: Fast softmax approximation on large vocabulary neural networks. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 5463–5473. Curran Associates, Inc.
- Simonton, T. M. and Alagband, G. (2017). Efficient and accurate word2vec implementations in gpu and shared-memory multicore architectures. In *2017 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–7.
- Taylor, G., Burmeister, R., Xu, Z., Singh, B., Patel, A., and Goldstein, T. (2016). Training neural networks without gradients: A scalable admm approach.

Impulsionando Árvores Extremamente Aleatórias Enscadas em Paralelo para Classificação de Textos

Julio C. B. Pires¹, Wellington S. Martins², Daniel X. de Sousa³

¹Instituto Federal Goiano (IF Goiano) – Campus Urutaí
Urutaí – GO – Brazil

²Instituto de Informática (INF) – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brazil

³Instituto Federal de Goiás (IFG) – Campus Anápolis
Anápolis – GO – Brazil

julio.pires@ifgoiano.edu.br, wellington@inf.ufg.br,
daniel.sousa@ifg.edu.br

Abstract. *The present work proposes the parallelization of BERT, an algorithm that combines boosting with bagging of extremely randomized trees to make automatic classification of textual datasets. Using high dimensionality sets can make the construction of classifiers an onerous task. Parallelism combined with graphics cards can overcome this challenge, since they offer a high processing power, the execution time can decrease considerably.*

Resumo. *O presente trabalho tem como proposta a paralelização do BERT, um algoritmo que combina boosting com bagging de árvores extremamente aleatórias para fazer classificação automática de conjuntos de dados textuais. Usar conjuntos de alta dimensionalidade pode tornar a construção dos classificadores uma tarefa onerosa. O paralelismo aliado às placas gráficas pode contornar esse desafio, uma vez que elas oferecem um alto poder de processamento, o tempo de execução pode diminuir consideravelmente.*

1. Introdução

Com a grande quantidade de informação disponível na Web e o seu rápido crescimento, surge a necessidade da organização e extração de padrões úteis dessa massa de dados [Campos et al. 2017]. Neste sentido, estão as estratégias do aprendizado de máquina, como a classificação, que constrói modelos através de vários exemplos para prever rótulos de classe para novos exemplos. Construir classificadores para execução sequencial na era do *Big Data* pode se tornar inviável, uma vez que a quantidade massiva de dados precisa de um grande esforço de processamento. Usar programação de alto desempenho em unidades de processamento gráfico (*GPUs*) pode acelerar o processo de aprendizado, ou seja, a computação dos classificadores se torna menos onerosa.

A maior parte da informação da web está em linguagem natural, classificar estes textos é uma forma de compreender melhor esses dados. Assim, o presente trabalho aborda a classificação automática de textos, que na sua maioria, são conjuntos de dados massivos e ruidosos. Um dos algoritmos de maior sucesso para resolver problemas de classificação é o Florestas Aleatórias (FAs), porém sua performance não é a mesma em

face à ruídos nos dados [Campos et al. 2017]. Para resolver esse problema pode ser usado o algoritmo Árvores Extremamente Aleatórias (AEAs), uma solução baseada no FAs, que adiciona um pouco mais de aleatoriedade na parte da construção das árvores de decisão. Levando tudo isso em consideração, a proposta está em paralelizar o algoritmo **BERT** (*Boosted Extremely Randomized Trees*) [Campos et al. 2017] para melhoria na velocidade de execução.

O algoritmo **BERT** usa *boosting* [Mitchell e Frank 2017], um procedimento que cria vários classificadores de forma iterativa (Figura 1). São atribuídos pesos para as amostras do conjunto de dados, e a cada iteração (treino, modelo e teste) esses pesos são modificados. Assim, as amostras mais difíceis de classificar (erros) recebem um peso maior e conseqüentemente são o foco da iteração subsequente. Cada iteração constrói um grupo de AEAs (modelo). Além disso, as árvores de cada grupo são construídas usando outro procedimento, o *bagging*, onde acontece uma amostragem com substituição do conjunto de treinamento original (Figura 3). Cada classificador é representado por uma árvore binária, que é construída de cima para baixo através da divisão dos dados de treinamento. Outra estratégia do **BERT** está em atualizar apenas os pesos das amostras que não são usadas na construção (*out-of-bag*) [Salles et al. 2015] da árvore. Ao término das iterações, todos os grupos de cada iteração formam o classificador final (predição).

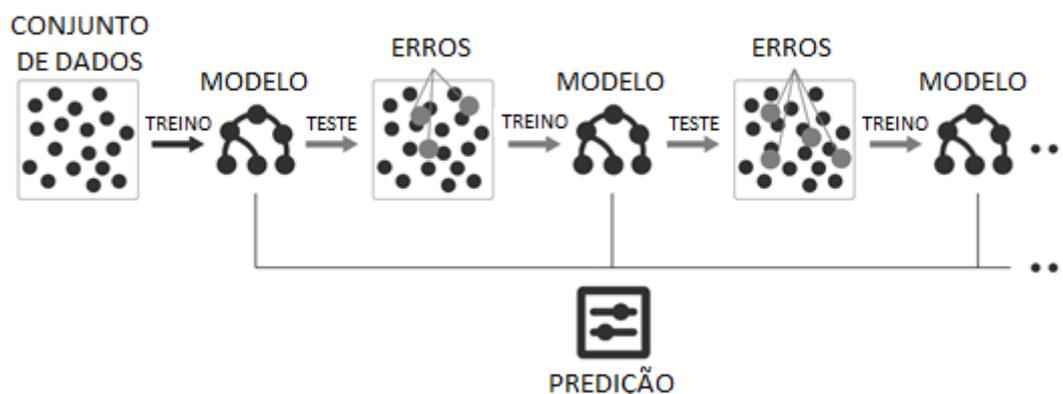


Figura 1. Esquema de *boosting*. Adaptado de [Alonso 2017].

2. Proposta

O volume massivo de dados usado no treinamento demanda computação de alto desempenho para deixar o processamento mais rápido. Placas gráficas são uma alternativa barata em comparação à clusters ou servidores para ganho de desempenho, mas não se deve ignorar a alta complexidade de codificação e projeto cuidadoso para maximizar a eficiência e ocupação de memória. Assim, maximizar a eficiência de execução na *GPU* é um tanto quanto difícil e requer um projeto refinado, onde o programador precisa conhecer a arquitetura de comunicação das memórias e os fluxos de execução das tarefas de uma *GPU*. Esses hardwares possuem memórias organizadas de forma hierárquica e oferecem paralelismo na forma da arquitetura *SIMD*, onde uma única instrução é executada sobre muitos dados diferentes. Essas placas são otimizadas para o uso do paralelismo massivo em uma grande quantidade de dados (*throughput*) para esconder a latência. Existem duas plataformas principais de programação para *GPUs*, a *OpenCL* (*Khronos*) e a *CUDA* (*NVIDIA*). A última é mais utilizada e já fornece várias bibliotecas otimizadas e por isso é a escolhida para este projeto. Os programas executados nessas placas são chamados de *kernel*. Eles são processados em grades de blocos de *threads*

(Figura 2), ou seja, *threads* são agrupadas em blocos e blocos são agrupados em grades. *Threads* são executadas em grupos de 32 (*warp*) [Jansson et al. 2014].

Existem algumas soluções paralelas que usam árvores de decisão, FAs e AEAs [Cano 2018], mas ainda não se conhece uma paralelização do algoritmo *BERT*. Várias estratégias foram usadas para construir algoritmos baseados em árvores na GPU, como a construção do FAs em nível de árvore [Grahn et al. 2011], implementação do FAs e o AEAs em nível de nó [Jansson et al. 2014] e construção de nível por nível [Marron et al. 2014]. A proposta deste trabalho consiste em usar soluções paralelas eficientes para resolver problemas de classificação de textos. Essa solução melhorada poderá ser usada em problemas de classificação em geral, análise de sentimentos e categorização de tópicos, por exemplo. Fazer um projeto cuidadoso para maximizar o desempenho do algoritmo pode reduzir o tempo e escalar a execução do algoritmo.

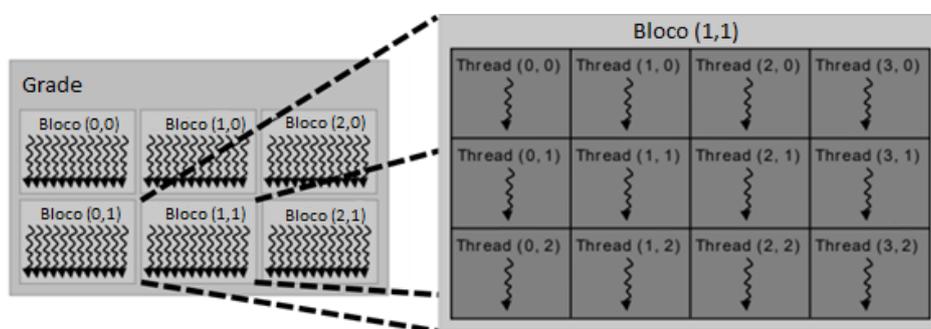


Figura 2. Uma Grade de Blocos de *Threads*. Adaptado de [NVIDIA 2018].

3. Metodologia

A metodologia deste trabalho comporta as etapas de fundamentação teórica, construção do algoritmo sequencial, projeto e implementação do algoritmo paralelo, análise dos resultados, otimização do algoritmo e documentação. Para o referencial teórico foi feita uma pesquisa sobre classificação, árvores de decisão, e os algoritmos FAs e AEAs. Além disso, foi feito um estudo dos procedimentos de *bagging* e de *boosting* (Figura 3). Além de utilizar alguns livros base, a maioria dos artigos foram encontrados em diversos repositórios científicos. Depois do estudo dos trabalhos relacionados ao tema da pesquisa foi feita a construção do algoritmo sequencial.

A próxima etapa consiste da implementação completa da versão paralela. Essa versão será executada em uma *GPU* para comparação com a versão sequencial. Para isso será usada a métrica de *speedup*, que é a razão do tempo do algoritmo sequencial pelo tempo do algoritmo paralelo [Navarro et al. 2014]. Depois da execução dos algoritmos sequencial e paralelo, será feita a colheita dos resultados para avaliação de desempenho. Depois o código será analisado em busca de erros e maiores oportunidades de ganho de tempo. Por fim será a documentação do trabalho na forma de um artigo.

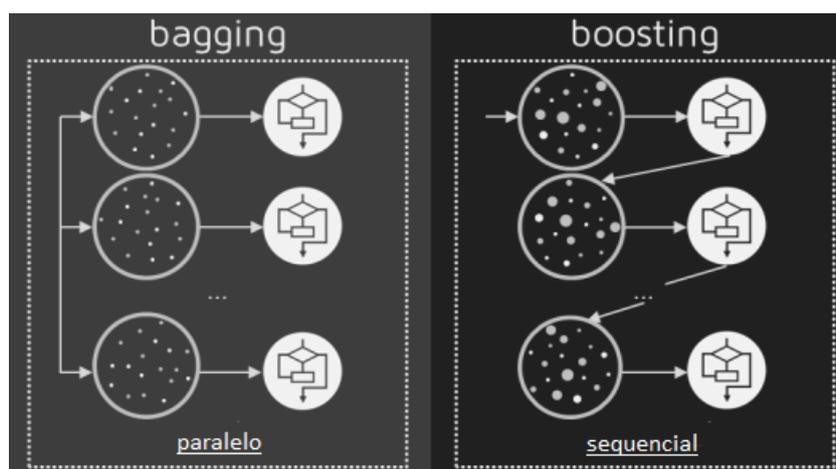


Figura 3. Diferença entre *bagging* e *boosting*. Adaptado de [Garrido 2016].

4. Resultados Preliminares

Após o levantamento bibliográfico, análise de códigos sequencias, estudo de estratégias de códigos paralelos, entendimento de algoritmos descritos em vários artigos e implementação própria do algoritmo sequencial, a próxima etapa foi a implementação de partes do algoritmo paralelo. Várias versões paralelas foram projetadas e implementadas, algumas não alcançaram os resultados desejados e outras apresentaram problemas de instabilidade. Atualmente, o foco está no afinamento e complemento do código paralelo, uma vez que a parte do *bagging* já está pronta. Os experimentos foram conduzidos na preparação de 8 amostragens em 200 iterações. Os resultados iniciais com conjuntos relativamente medianos (4uni e acm) mostraram um bom *speedup* (Tabela 1) e validam uma parte da proposta. A próxima etapa consiste em criar de fato todas as árvores.

Tabela 1. Comparação entre as execuções sequencial e paralela

Conjunto de Dados	Atributos	Amostras	Classes	Tempo Sequencial	Tempo Paralelo	Speedup
4uni	4.196	6.624	7	86,2368 s	1,3986 s	61,8 x
acm	59.991	26.546	10	1384,2556 s	20,3990 s	67,9 x

5. Considerações

A classificação de textos é utilizada para organizar informações e tem algumas aplicações, como análise de sentimentos e categorização de tópicos. Essa tarefa lida com conjuntos de dados ruidosos e de alta dimensionalidade, o que pode diminuir a eficácia de um classificador e tornar o custo computacional oneroso. A utilização de conjuntos de treino grandes demanda execução de forma rápida, o que é indispensável na era do *Big Data*. Desse modo, neste trabalho foi proposto a paralelização do algoritmo **BERT** utilizando placas gráficas visando ganhos de velocidade. Por enquanto, fazendo apenas o *bagging* das árvores com conjuntos medianos a abordagem melhora consideravelmente o tempo de execução. Isso é importante, uma vez que os dados usados para classificação de texto podem ser grandes. Além disso, mais e mais informação nova fica disponível a cada dia na imensidão da Web. Crucial é fazer uso dessa informação para construir soluções

inteligentes que executem em tempo hábil. O próximo passo pode estar na implementação paralela do **BERT** usando o esquema das Florestas Aleatórias Profundas [Zhou and Feng 2017]. Cada floresta é organizada em cascata e as entradas da próxima floresta dependem da floresta anterior.

Referências

- Alonso, M. J. (2017) “Introduction to Boosted Trees”, <https://blog.bigml.com/2017/03/14/introduction-to-boosted-trees>, September.
- Campos, R., Canuto, S., Salles, T., de Sá, C. C., and Gonçalves, M. A. (2017). Stacking bagged and boosted forests for effective automated classification. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 105–114, New York, NY, USA. ACM.
- Cano, A. (2018). A survey on graphic processing unit computing for large-scale data mining. *Wiley Interdisc. Rev.: Data Mining and Knowledge Discovery*, 8(1).
- Garrido, A. P. (2016) “What is the difference between Bagging and Boosting?”, <https://quantdare.com/what-is-the-difference-between-bagging-and-boosting>, September.
- Grahn, H., Lavesson, N., Lapajne, M. H., Slat, D. (2011). CudaRF: A CUDA-based implementation of random forests. In *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications*, AICCSA. 95-101.
- Jansson, K., Sundell, H., and Boström, H. (2014). gpurf and gpuert: Efficient and scalable gpu algorithms for decision tree ensembles. In *Proceedings of the 2014 IEEE International Parallel & Distributed Processing Symposium Workshops*, IPDPSW '14, pages 1612–1621, Washington, DC, USA. IEEE Computer Society.
- Marron, D., Bifet, A., and Morales, G. D. F. (2014). Random forests of very fast decision trees on gpu for mining evolving big data streams. In *Proceedings of the Twenty-first European Conference on Artificial Intelligence*, ECAI'14, pages 615–620, Amsterdam, The Netherlands, The Netherlands. IOS Press.
- Mitchell, R. and Frank, E. (2017). Accelerating the xgboost algorithm using gpu computing. *PeerJ Computer Science*, 3:e127.
- Salles, T., Gonçalves, M., Rodrigues, V., and Rocha, L. (2015). Broof: Exploiting out-of-bag errors, boosting and random forests for effective automated classification. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 353–362, New York, NY, USA. ACM.
- Navarro, C., Hitschfeld-Kahler, N., and Mateu, L. (2014). A Survey on Parallel Computing and its Applications in Data-Parallel Problems Using GPU Architectures. *Communications in Computational Physics*, 15(2), 285-329.
- NVIDIA. (2018) “Cuda C Programming Guide”, <https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html>, September.
- Zhou, Z.-H. and Feng, J. (2017). Deep forest: Towards an alternative to deep neural networks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, IJCAI-17, pages 3553–3559.

Análise Custo x Desempenho em relação às Instâncias Preemptáveis em Provedores Públicos de Nuvem

Amanda L. D., Eduardo C. R. C., Jonas P. S., Matheus O. V., Aletéia P. F. Araújo

¹Departamento de Ciência da Computação – Universidade de Brasília (UnB)
Brasília – DF – Brazil

{amandadantas19,themustellafuro,matheusov,aleteia.araujo}@gmail.com,
eduardo.calandrini@hotmail.com

Abstract. *This article makes use of Google Cloud and Amazon Web Service Cloud preemptible machines running a workload made of queries on a MongoDB collection. The cost and average execution time, besides other metrics, were used to form an analysis of cost and computing between both providers.*

Resumo. *Este trabalho propõe um estudo de caso sobre os provedores Google Cloud e Amazon Web Service Cloud, executando uma carga de trabalho baseada em consultas no banco de dados MongoDB. Foram comparados custos e tempo médio de execução, dentre outras métricas, para apresentar uma análise entre custo e desempenho de máquinas preemptáveis nos dois provedores.*

1. Introdução

Como a utilização de nuvens computacionais está cada vez mais consolidada no mercado, diversos usuários estão interessados em estratégias mais econômicas para utilizar o serviço de maneira adequada às suas necessidades.

Nesse cenário, instâncias preemptáveis (*Spot Instances*) são oferecidas por alguns provedores de nuvem como uma alternativa menos custosa em comparação a instâncias dedicadas. Este tipo de serviço oferece ao consumidor a execução de sua carga de processamento (*workload*) por um preço menor em detrimento do tempo total de execução. Este trabalho busca simular um mesmo cenário em diferentes máquinas preemptáveis dos provedores Amazon Web Service Cloud (AWS EC2) e Google Cloud, trazendo uma comparação entre desempenho e custo das mesmas ao realizar cargas de trabalho em um banco de dados NoSQL, o MongoDB.

2. Contexto Teórico

2.1. Amazon Web Service Cloud

O Amazon Web Service Cloud foi o primeiro provedor de nuvem a disponibilizar um serviço de instâncias preemptáveis em 2009, utilizando em sua precificação um sistema de leilão. Um preço de mercado, chamado preço *spot*, é definido pela AWS e define se a execução do usuário será preemptada ou não¹.

¹https://docs.aws.amazon.com/pt_br/AWSEC2/latest/UserGuide/concepts.html

2.2. Google Cloud Platform

Diferente da Amazon, a Google Cloud precifica suas instâncias preemptáveis de maneira fixa, cuja execução pode ser quase 80% menor que o preço da instância dedicada correspondente. As instâncias Google, porém, são executadas apenas por 24 horas sendo preemptadas ao fim desse período. As instâncias preemptáveis são recursos finitos da Google Compute Engine então nem sempre estão disponíveis. Também não são cobertas por nenhum dos contratos a nível de serviço oferecidos pela Google².

3. Metodologia

3.1. Carga de Trabalho

O experimento envolveu consultas a um banco de dados NoSQL, o MongoDB. Cada máquina preemptável utilizada nos testes teve seu ambiente preparado para receber dados referentes ao pagamento de benefícios do programa Bolsa Família no ano de 2017. Esse conjunto de dados foi exportado e reunido em uma coleção que totaliza 15.59 GBs³. A carga de trabalho aplicada consiste em uma série de consultas e processamento sobre as amostras em questão [Neto et al. 2017].

Como *workload* de testes foi definida a contagens de registros no banco onde a informação “Valor Pago” do registro consiste em um valor inteiro específico, sendo este variado para cada execução (entre 39 e 998, os valores mínimo e máximo registrados no banco). Visando monitorar o comportamento das instâncias preemptáveis durante longos períodos de execução, uma rotina automatizou a execução destas consultas. Aplicando a mesma rotina em diferentes máquinas, o desempenho de cada instância foi analisado a partir da comparação do tempo de resposta entre as execuções realizadas (vide seção a seguir).

3.2. Arquitetura de Hardware

O experimento teve como base o teste de máquinas preemptáveis disponibilizadas pelos serviços gratuitos para testes na região de *Oregon - USA*. Definido este escopo, apenas uma máquina configuração de instância *spot* foi oferecida pela AWS em sua licença gratuita. A Google Cloud, em contrapartida, disponibilizou maior flexibilidade de configuração para as máquinas preemptáveis. Com isso, o experimento contou com uma instância *spot* na AWS e 3 máquinas preemptáveis da Google Cloud.

Uma das principais dificuldades encontradas na realização do trabalho foi a limitação de configurações de hardware imposta pela AWS. Apenas um tipo de máquina *spot* foi disponibilizado no plano gratuito da provedora para a zona avaliada, habilitando pouca flexibilidade para testes de diferentes cenários. Com esta adversidade, apenas uma máquina Amazon foi utilizada no experimento. Além destas, foram utilizadas 3 máquinas preemptáveis da Google Cloud, cujas especificações seguem nesta subseção. Todas as máquinas estão localizadas na região de *Oregon - USA*.

Utilizando a Amazon EC2, foi criada uma máquina virtual do tipo t1.micro (1 vCPU / 0.615 GiB) com 30 GB de disco magnético e sistema operacional Ubuntu Server 16.04 (ami-db710fa3). A instância Google 1 foi criada como base no tipo f1-micro (1

²<https://cloud.google.com/compute/docs/instances/preemptible>

³<http://www.portaldatransparencia.gov.br/download-de-dados/bolsa-familia-pagamentos>

vCPUj / 0,6 GB de memória) com 30 GB de armazenamento magnético do tipo disco permanente padrão e sistema operacional Ubuntu 16.04 (ubuntu-1604-xenial-v20180612).

A instância Google 2 foi criada com tipo n1-standard-1 (1 vCPUj / 3,75 GB de memória) com disco similar à máquina Google 1. Portanto, a principal diferença na configuração destas máquinas é a quantidade de memória disponível. Por fim, a instância Google 3 foi criada com as mesmas configurações da Google 2, mas com um disco SSD em vez do disco magnético da anterior.

4. Resultados

4.1. Custos

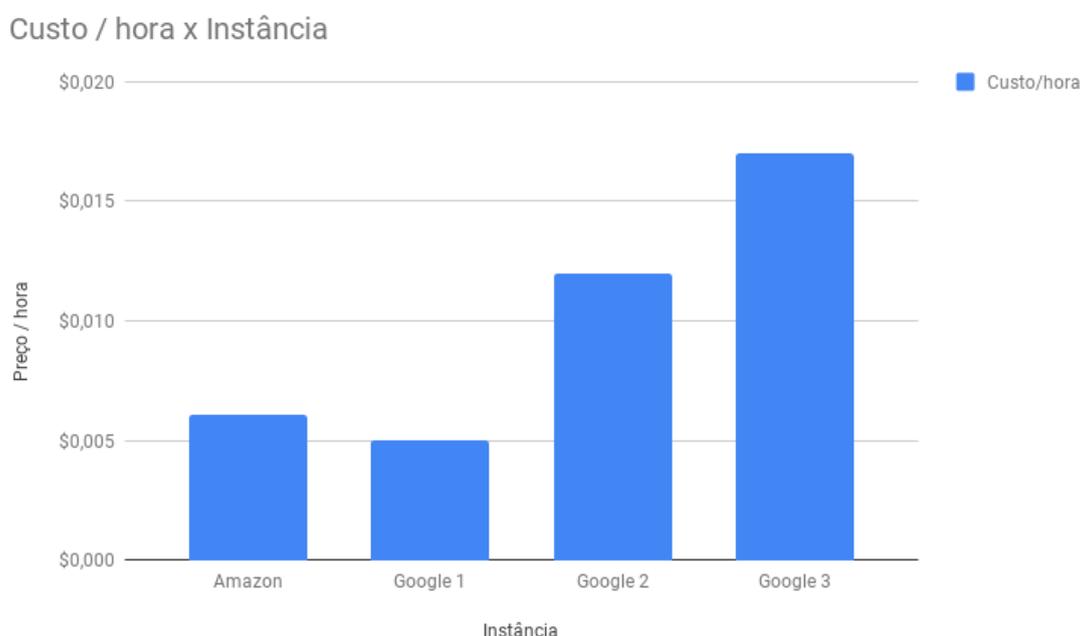


Figura 1. Custo das Instâncias por hora.

O valor cobrado pela execução do *workload* na nuvem Amazon é fragmentado em custo de processamento da instância (que totalizaram \$0.002 por hora), manutenção do disco da máquina (\$0.002 por hora) e operações de entrada e saída (\$0.0021 por hora). Além do valor total cobrado pela execução do *workload* na nuvem Amazon (\$0.0061 por hora), o gráfico da Figura 1 apresenta os preços exercidos pelas instâncias Google utilizadas, onde as operações de entrada e saída são gratuitas. Em uma leitura do gráfico fica evidente que as instâncias Google 2 e 3, com hardware mais potente, são consideravelmente mais caras (\$0.012 e \$0.017, respectivamente).

4.2. Execução dos *workloads*

As Figuras 2, 3, 4 e 5 apresentam o tempo de execução das consultas realizadas no MongoDB de cada instância, dados estes obtidos a partir dos logs construídos pela rotina de automação nas mesmas. No eixo X está representada a data e horário de início das execuções, enquanto o eixo Y indica o tempo da execução da consulta.

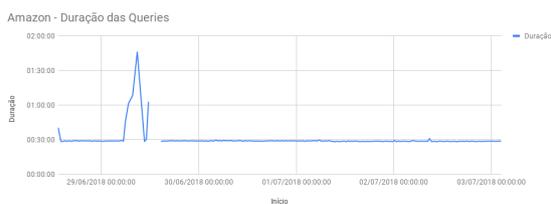


Figura 2. Execução Amazon



Figura 3. Execução Google 1

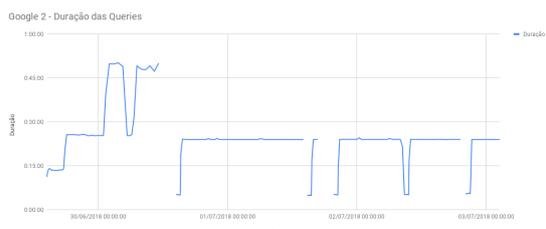


Figura 4. Execução Google 2



Figura 5. Execução Google 3

Uma primeira característica notável no comportamento da instância Amazon é sua estabilidade em relação às instâncias Google, dentre as máquinas com armazenamento em discos magnéticos. Além disso, a característica de execução contínua fornecida pela AWS possibilitou que amostras mais completas fossem obtidas na mesma. O único período de interrupção da coleta de amostras foi causado por problemas no servidor MongoDB, o que não caracteriza preempção ou indisponibilidade da instância. O gráfico na Figura 2 apresenta um tempo de execução estável de aproximadamente 28 minutos, exceto em períodos de anormalidade da instância.

Alguns intervalos de tempo nos gráficos da Google Cloud não geraram amostras uma vez que a Google Cloud força a preempção das máquinas a cada 24hrs, exigindo que as máquinas fossem reiniciadas. Esta característica impactou em amostras reduzidas para comparações, mas que, ainda sim, geraram resultados relevantes.

O gráfico da Figura 3 demonstra o tempo de execução da instância Google 1 - de preço semelhante à instância Amazon - e apresenta desempenho consideravelmente melhor que o provedor concorrente nas primeiras horas de execução. Entretanto, o tempo de execução de cada consulta se tornou cada vez maior à medida que o tempo de vida da instância aumentou até o ponto de perder esta eficiência comparativa. Uma observação notável é que, uma vez reiniciadas, as instâncias retomam um melhor desempenho, até decair eventualmente.

A instância Google 2 foi utilizada no experimento imaginando a possibilidade da ausência de memória disponível consistir em um gargalo e causar essa perda de performance. Suas configurações foram muito semelhantes à instância Google 1, exceto pelo aumento de memória em relação a primeira. Em análise do gráfico presente na Figura 4, é possível perceber que o tempo de execução para o *workflow* tem desempenho muito similar à instância anterior, quando analisada em longo prazo. Com isso, a hipótese de gargalo devido a memória disponível foi descartada. Além disso, é relevante pontuar que esta instância foi a única que sofreu uma preempção não prevista, registrada antes de se completar a 4ª hora de execução da mesma (iniciada entre 14:00 e 15:00 de 01/07/2018 e

preemptada entre 17:00 e 18:00 do mesmo dia).

Imaginando que o gargalo do *workflow* pudesse estar nas operações de disco, foi criada uma terceira instância Google muito similar a Google 2. Com a análise do gráfico na Figura 5, podemos observar que há uma queda considerável no tempo de execução da instância em relação às demais, indiferente de provedor analisado. A única diferença entre as máquinas Google 2 e Google 3 está na configuração de disco, onde a terceira instância utiliza um SSD enquanto a anterior conta com um disco magnético. Com esta constatação, surgiu a hipótese de o principal gargalo do *workflow* em questão estar vinculado às operações de disco.

5. Conclusões

Diante do exposto, é possível concluir que os resultados obtidos entre os provedores avaliados não indicam um provedor com vantagem competitiva absoluta. Os resultados sugerem que as máquinas preemptáveis Google são mais adequadas para projetos com maior tolerância à preempção e menor tempo de duração, pois a instância sempre será interrompida pelo menos uma vez no período de 24 horas. O preço fixo por máquina Google e a gratuidade sobre operações em disco tornam a precificação do mesmo mais previsível, possibilitando melhor planejamento de uso da instância.

Em contrapartida, a Amazon apresentou menor variação no tempo de resposta das consultas realizadas. Por fim, a instância Amazon tem uma execução consideravelmente mais estável, uma vez que um lance alto o suficiente garante a estabilidade da máquina a longo prazo.

Uma evidente limitação no escopo do trabalho realizado consiste no uso de um único *workload* de testes, o que oferece um resultado limitado de performance, restrito a operações de contagem no banco de dados MongoDB. Constatada essa limitação, sugere-se que trabalhos futuros realizem o experimento em diferentes *workloads*, como processamento de sequenciamento genético e codificação de sinais.

Além disso, procedimentos parecidos podem ser aplicados à máquinas com diferentes configurações, incluindo outros provedores de máquinas preemptáveis. Estas análises poderiam investigar a hipótese apresentada quanto ao gargalo, aqui associado às limitações de leitura e escrita em discos magnéticos (HDD).

Referências

- Neto, A. D., Fernandes, E. P., Holanda, M., Ribeiro, E., and Carvalho, R. N. (2017). MongoDB performance analysis: A comparative study between stand-alone and shared cluster deployments with open data from brazilian bolsa familia program. In *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6.

DStream: Plataforma de coleta, processamento e armazenamento de streams de dados de sensoriamento remoto

Sávio Oliveira¹, Vagner Rodrigues¹, Laerte Ferreira², Claudinei Santos², Wellington Martins¹

¹Instituto de Informática - Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Quadra D, Câmpus Samambaia
131 - CEP 74001-970 - Goiânia - GO - Brasil

²LAPIG - CAMPUS II Samambaia - Cx. POSTAL 131
CEP: 74001-970 - Goiânia - GO - Brasil

{savio.teles, vagner}@gogeo.io, {lapig.ufg, claudineisantosnx}@gmail.com,
wellington@inf.ufg.br

Abstract. *Currently, one of the best applications has been in identifying transformations on Earth surface, which is changing at an unprecedented rate. Hence, great efforts have been made to create solutions to identify these changes by processing large volumes of data (Big Data) continuously generated by several sources (streaming). The goal of this paper is to explore the parallel architectures to propose a new distributed solution for the collect, storing, indexing and processing of time series of remote sensing images.*

Resumo. *Atualmente, uma das maiores aplicações de séries temporais tem sido na identificação de transformações da superfície do planeta Terra, que está mudando a uma taxa sem precedentes. Por isso, grandes esforços tem sido feitos para criar soluções para identificar estas mudanças processando grandes volumes de dados (Big Data) gerados continuamente por diversas fontes (streaming). O objetivo deste trabalho é explorar as arquiteturas paralelas para propor uma nova solução distribuída para coleta, armazenamento, indexação e processamento de séries temporais de imagens de sensoriamento remoto.*

1. Introdução

A superfície do planeta Terra está mudando a uma taxa sem precedentes, no qual ecossistemas de florestas diminuem a uma velocidade alarmante e áreas urbanas e agrícolas expandem em volta do espaço natural. Análise das séries temporais de imagens de sensoriamento remoto tem se tornado indispensáveis na identificação destas mudanças.

Na classe de problemas computacionalmente custosos, a análise de séries temporais é um dos problemas com maior demanda de poder computacional [Rakthanmanon et al. 2013] e podem se beneficiar das arquiteturas paralelas. O grande volume de dados, com alta dimensionalidade, sendo gerados em grande quantidade a todo momento traz desafios únicos para explorar análise de séries temporais no cenário de *Big Data* [Fan et al. 2014]. Mais especificamente na área de sensoriamento remoto, podemos dizer que estamos vivendo uma era do *Big Remote Sensing Data* [Chi et al. 2016].

As soluções paralelas centralizadas (MultiCore e ManyCore), com memória compartilhada, não são capazes de processar séries temporais no contexto *Big Data*

[Sottile et al. 2009] e, por isso, diversas soluções paralelas com memória distribuída tem sido propostas para processar esse grande volume de dados de séries temporais [Chen et al. 2014]. As melhores soluções paralelas para *Big Data* utilizam ambas arquiteturas de memórias (compartilhada e distribuída) provendo flexibilidade para otimizar o paralelismo em aplicações para extrair o máximo de eficiência [Sottile et al. 2009].

No contexto de *Big Remote Sensing Data*, não foram encontrados algoritmos eficientes para processamento de streams de dados de séries temporais de imagens de sensoriamento remoto, sendo um dos grandes desafios e oportunidades da área [Ma et al. 2015]. Foi proposta a plataforma DistSensing [de Oliveira et al. 2017] para processamento de séries temporais, mas que não permite trabalhar com *streams* de dados. O objetivo deste trabalho é propor uma plataforma, denominada DStream, para coletar, armazenar e processar os dados de *Big Remote Sensing Data* que chegam a todo momento de diversas fontes. Na Seção 2 apresentamos a descrição desta plataforma, com a avaliação sendo apresentada na Seção 3 e as conclusões na Seção 4.

2. DStream

O processamento de *streams* de dados deve ser realizado no momento em que os dados chegam ao sistema com uma baixa latência de resposta. Para atender a estes requisitos, este trabalho apresenta uma solução, denominada DStream, para o processamento de *streams* de dados de séries temporais de imagens de sensoriamento remoto. A arquitetura dessa solução, apresentada na Figura 1, é constituída por módulos de coleta, transformação e armazenamento dos dados de séries temporais. Os dados são coletados de diversas fontes que sejam capazes de enviar dados históricos de imagens de sensoriamento remoto.

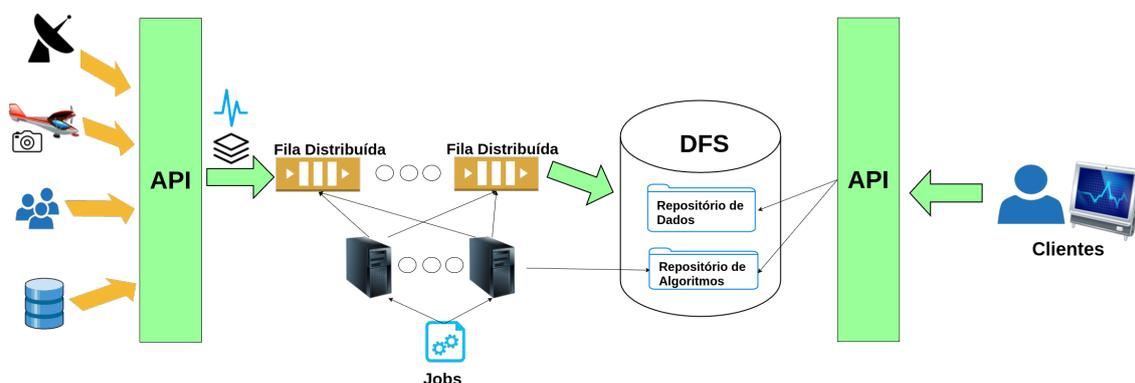


Figura 1. Arquitetura do Sistema de Processamento de *Stream* de Dados de Séries Temporais

A fila de entrada é distribuída e visível a todos os servidores do *cluster*, em que qualquer servidor é capaz de recuperar um objeto da fila e processar. Este objeto recuperado contém um bloco com um conjunto de *pixels* e cada *pixel* carrega a série histórica dos seus valores de NDVI. Esse bloco é processado utilizando os algoritmos de análise de séries temporais armazenados no sistema de arquivos distribuído (DFS). Esses algoritmos são implementados por um especialista em alguma área, que deseja realizar análises sobre as séries temporais de imagens de sensoriamento remoto. Para isso, o cliente deve submeter um código para a API da DStream, que armazena este código no DFS para ser então

utilizado no processamento dos blocos de *pixels*. Esses códigos devem seguir um padrão, de receber como entrada a série temporal de um *pixel* e gerar uma saída no formato texto.

Cada *job* no sistema é executado por vários servidores do *cluster*, cabendo a plataforma garantir a tolerância a falhas, o controle de fluxo e o balanceamento dos *jobs* entre os servidores. Cada objeto enviado na fila distribuída é recuperado por um *job* executado em uma máquina. Esse *job* é responsável por recuperar cada algoritmo de análise de séries temporais de imagens de sensoriamento remoto armazenados no DFS e executar estes algoritmos sobre o objeto contendo um bloco de séries temporais de vários *pixels*. Para cada série temporal de um *pixel*, os algoritmos são executados e a saída enviada para a fila de saída do sistema, contendo o identificador do *pixel* e do algoritmo como chave e o texto retornado por cada algoritmo como valor.

Esta arquitetura, na qual objetos são recuperados de uma fila de entrada, processados e enviados para uma fila de saída, permite que outros *jobs* sejam integrados facilmente ao sistema, bastando adicionar filas distribuídas e *jobs* para consumir dados destas filas. Por exemplo, suponha que o sistema receba os dados brutos das imagens e precise construir as séries temporais calculando o índice de vegetação NDVI de cada *pixel* ao longo do tempo. Antes de enviar os dados para a fila de entrada de processamento dos algoritmos de análise, o stream de dados pode ser enviado para a fila de entrada para o cálculo do NDVI, para depois ser enviado para a fila de processamento dos algoritmos.

Quando os dados chegam a fila de saída do sistema, cabe a um *job* recuperar estes dados e armazenar o resultado no DFS. As aplicações clientes devem ficar constantemente consultando o sistema para recuperar o resultado do processamento dos seus algoritmos. Esses resultados são obtidos do DFS pela API e retornados ao cliente do sistema.

3. Avaliação

Para validar o DStream foi realizada uma avaliação com o algoritmo BFAST, muito utilizado na análise de séries temporais de sensoriamento remoto, e três variações do mesmo em relação ao tempo de resposta e eficácia na detecção de *breakpoints* em uma série temporal de imagens de sensoriamento remoto. O algoritmo BFAST tem como objetivo estimar os pontos de *breakpoints* de uma série temporal, ou seja, os pontos em que a tendência da série temporal muda dentro de um determinado período.

O sistema DStream foi instalado em uma máquina com processador Intel Core i7-2670QM (2.2GHz, 6Mb Cache, 4 núcleos, 8 Threads) e 6GB de memória RAM. Foram realizados testes com 72.361 séries temporais de *pixels*, em que cada série temporal possui 394 valores de NDVI ao longo do tempo de um *pixel*. Para cada uma das 72.361 séries temporais utilizadas no experimento, um especialista da área de sensoriamento remoto do Laboratório de Processamento de Imagens e Geoprocessamento (LAPIG)¹ detectou manualmente, através da visualização de imagens de satélite em diferentes períodos, os pontos ao longo do tempo onde houveram mudanças no tipo de uso de solo para cada *pixel*. Essas séries temporais foram enviadas para a plataforma, simulando um *stream* de dados históricos de imagens de sensoriamento remoto.

Como pode ser visto na Figura 2, o BFAST apresentou tempo de resposta em média de aproximadamente 4 segundos para cada série temporal. Foram implementadas

¹<https://www.lapig.iesa.ufg.br/>

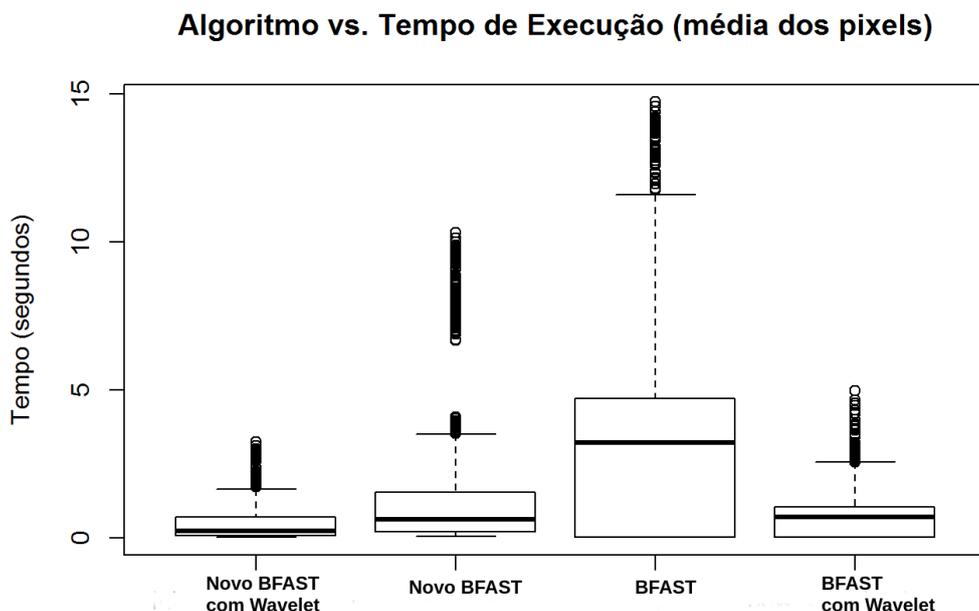


Figura 2. Gráfico de comparação entre as propostas de variação do BFAST.

três variações do BFAST para reduzir o tempo de execução. A primeira variação utiliza Wavelets (“BFAST com Wavelet” na legenda do gráfico da Figura 2), que foi utilizada para reduzir o tamanho da série temporal de 394 valores pela metade, ou seja, para 197 valores, eliminando grande quantidade de ruídos da série temporal. O algoritmo BFAST foi então aplicado sobre essa série temporal de 197 valores e o tempo de execução caiu de 4 segundos para aproximadamente 1,5 segundos. Em relação a eficácia, a utilização de Wavelet reduziu o número de falsos positivos do BFAST, ou seja, reduziu os pontos incorretos no tempo onde o BFAST determina que ocorreram mudanças no tipo de uso de solo através da identificação de *breakpoints*. Isto aconteceu porque com a utilização de Wavelets, o número de ruídos foi reduzido e, assim, mudanças bruscas nos valores do NDVI foram desprezados. Essas mudanças bruscas podem acontecer devido a diversos fenômenos, como a presença de nuvens, que impede que o satélite colete com acurácia suficiente o valor de NDVI do *pixel*. O BFAST foi melhor na identificação dos pontos corretos de mudanças no tipo de uso do solo, pois o uso de Wavelets removeu da série temporal além dos ruídos alguns valores importantes para identificação dos *breakpoints*.

A segunda variação (“Novo BFAST” na legenda do gráfico da Figura 2) modifica o algoritmo do BFAST para utilizar como variável explicativa da regressão linear um valor constante igual a 1, enquanto no algoritmo BFAST a variável explicativa é o tempo. Desta forma, o tempo não tem tanta influência na identificação de *breakpoints*. Esta variação teve desempenho melhor que o BFAST, mas pior que o algoritmo “BFAST com Wavelet”. Esta segunda variação conseguiu identificar um maior número de *breakpoints* que as outras técnicas e, por isso, teve uma taxa maior de acertos e de erros.

Uma terceira variação (“Novo BFAST com Wavelet” na legenda do gráfico da Figura 2) executa a modificação do BFAST proposta na segunda variação sobre a série temporal depois de aplicada a Transformada de Wavelet. Ela teve o menor tempo de resposta entre todas as técnicas avaliadas, aproximadamente 1 segundo, tendo uma acurácia média na identificação de falsos positivos e verdadeiros *breakpoints*. Em relação aos fal-

sos positivos teve uma acurácia pior apenas que o “BFAST com Wavelet”. Para encontrar os verdadeiros pontos de *breakpoints* teve uma acurácia pior que o BFAST e a segunda variação do BFAST, denominado “Novo BFAST”.

4. Conclusão

O grande volume de dados não estruturados, com alta dimensionalidade, sendo gerados em grande quantidade a todo momento coloca o problema de análise de séries temporais no contexto de *Big Data* [Fan et al. 2014]. Para que seja possível explorar o potencial do *Big Data*, este trabalho propõe a plataforma DStream para processamento distribuído de stream de dados de séries temporais, com a execução do algoritmo de análise de séries temporais utilizando várias variações do algoritmo BFAST para validar o casamento da arquitetura da plataforma com o problema de análise de séries temporais.

A plataforma DStream foi capaz de processar as análises de séries temporais com as variações do algoritmo BFAST, processando as séries temporais como *stream* de dados. Assim, a arquitetura pode ser validada em relação ao problema de executar *streams* de dados de sensoriamento remoto. Em trabalhos futuros, serão realizadas avaliações com maior número de máquinas, mas a arquitetura proposta para a plataforma DStream permite que mais máquinas sejam adicionadas tendo uma redução no tempo de resposta final da análise da série temporal.

Referências

- Chen, M., Mao, S., and Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2):171–209.
- Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., and Zhu, Y. (2016). Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104(11):2207–2219.
- de Oliveira, S. S. T., de Castro Cardoso, M., dos Santos, W., Costa, P., do Sacramento Rodrigues, V. J., and Martins, W. S. (2017). Distsensing: A new platform for time series processing in a distributed computing environment. *Revista Brasileira de Cartografia*, 69(5).
- Fan, J., Han, F., and Liu, H. (2014). Challenges of big data analysis. *National science review*, 1(2):293–314.
- Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., and Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, 51:47–60.
- Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., and Keogh, E. (2013). Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(3):10.
- Sottile, M. J., Mattson, T. G., and Rasmussen, C. E. (2009). *Introduction to concurrency in programming languages*. CRC Press.

Index of Authors

- Abreu, Marcos, [65](#)
Alencar, Wanderley, [131](#)
Almeida, Gabriel, [145](#)
Alves, Douglas, [201](#)
Alves, Paulo, [159](#)
Alves, Ricardo, [25](#)
Amorim, Leonardo, [337](#)
Araujo, Aleteia, [349](#)
- Barros, Chayner, [337](#)
Bittar, Thiago, [187](#)
Borges, Vinicius, [39](#), [53](#), [65](#), [77](#)
- Cardoso, Kleber, [39](#), [65](#), [77](#)
Carvalho, Luana, [237](#)
Cassiano, Kátia, [11](#)
Cordeiro, Douglas, [11](#)
Costa, Eduardo, [349](#)
- Dantas, Amanda, [349](#)
Dias, Elisângela, [145](#), [297](#)
Duarte, Luciano, [249](#)
- Faria, Vitor, [89](#)
Fernandes, Deborah, [215](#)
Fernandes, Marcio, [215](#)
Ferreira, Laerte, [327](#), [359](#)
Fleury, Leonardo, [53](#)
Freitas, Mateus, [337](#)
Freitas, Rafael, [297](#)
- Jesus, Pétersson, [277](#)
Júnior, Divino, [53](#)
Júnior, Nivaldo, [263](#)
- Lima, Eliomar, [201](#), [249](#), [263](#), [287](#)
Lobato, Luanna, [187](#)
- Martins, Wellington, [333](#), [337](#), [343](#), [359](#)
Moraes, Fausto, [39](#)
Moreira, Waldir, [39](#)
Moura, Cícero, [173](#)
- Nascimento, Hugo, [131](#)
- Oliveira, Antonio, [39](#), [53](#), [65](#), [77](#)
Oliveira, Carlos, [263](#)
Oliveira, Fábio, [131](#)
Oliveira, Sávio, [359](#)
- Paiva, Roberto, [327](#)
Parente, Leandro, [327](#)
Patto, Vinicius, [237](#)
Pinto, Leizer, [77](#)
Pires, Julio, [343](#)
- Quirino, Rafael, [333](#)
- Ribeiro, Leonardo, [333](#)
Rocha, Flávio, [103](#)
Rocha, Yan, [287](#)
Rodrigues, Leidiane, [317](#)
Rodrigues, Lucas, [77](#)
Rodrigues, Vagner, [359](#)
- Salvini, Rogerio, [117](#)
Santos, Claudinei, [327](#), [359](#)
Santos, Larissa, [187](#)
Santos, Raimunda, [317](#)
Severino, Sílvio, [225](#)
Silva, Alessandro, [277](#)
Silva, Jurandir, [117](#)
Silva, Leonardo, [103](#)
Silva, Marcelo, [39](#)
Silva, Nádia, [225](#)
Silva, Patrícia, [187](#)
Silva, Rodolfo, [215](#)
Silva, Salatyel, [103](#)
Silva, Vagner, [307](#)
Soares, Jonas, [349](#)
Sobrinho, Francisco, [159](#)
Sousa, Daniel, [343](#)
Sousa, Pablo, [65](#)
Spíndola, Guilherme, [89](#)
- Taddeo, Leandro, [159](#)
Taquary, Evandro, [327](#)
- Velasco, Gislainy, [237](#)
Vieira, Flávio, [307](#)
Vieira, Matheus, [349](#)
- Xavier, Otávio, [173](#)