

| CURSO DE CIÊNCIA DA COMPUTAÇÃO – TCC |                      |
|--------------------------------------|----------------------|
| ( X ) PRÉ-PROJETO    ( ) PROJETO     | ANO/SEMESTRE: 2020.2 |

## USO DE REDES COMPLEXAS PARA ANÁLISE DE CONSUMO EM ECOMMERCE

Diogo Warmeling

Aurélio Faustino Hoppe – orientador

### 1 INTRODUÇÃO

Com o crescimento dos acessos à internet e a disseminação de dispositivos móveis, os comércios eletrônicos (e-commerce) vêm crescendo de forma exponencial nos últimos anos e, com a pandemia do COVID-19, o crescimento foi ainda maior, tornando-se o método mais comum de compra para diversos setores da economia brasileira (TERRA, 2020). Segundo os dados da 42ª edição do Webshoppers, um estudo sobre e-commerces do Brasil elaborado semestralmente pela Ebit e Elo, o crescimento nas vendas do primeiro semestre de 2020 foi de 47%, caracterizando-se como dobro dos registrados nos últimos anos. Isso, segundo Lunardi (2018), aconteceu devido a entrada de novos adeptos a esse meio de compra, resultando em um crescimento de 40%, chegando a 41 milhões de pessoas.

Brito (2020) destaca que esse crescimento não ficou ~~restringido-restrito~~ somente aos *marketplaces*, mas também atingiu muitos comércios eletrônicos de pequeno e médio porte, dos quais, a maior parte não aproveita o grande potencial que a análise dos dados de acessos de seus clientes fornece. Tais análises possibilitam entender o comportamento dos usuários para mantê-los mais tempo em seus *sites*, e conseguir recomendar o produto desejado antes da desistência da compra, assim como auxiliar os clientes a encontrarem as melhores opções disponíveis, aumentando a taxa de efetivação da compra por parte dos usuários. No entanto, a grande maioria dos *sites* apenas disponibilizam listas de produtos semelhantes e associados ao acesso do usuário, desconsiderando o seu perfil, que pode estar seguindo um padrão de acessos.

Segundo Santos (2019) existem diversas pesquisas sobre recomendações em plataformas de comércio eletrônico, que em sua maioria fazem cálculo de similaridade entre entidades (compradores e produtos) para recomendar novos produtos a um potencial comprador. Dentre elas, encontram-se uma grande quantidade de técnicas e algoritmos disponíveis pensadas para análise de dados tais como sistemas de recomendações, algoritmos de aprendizado de máquina e redes complexas, que permitem encontrar relações e agrupar itens, analisar o fluxo de navegação ou o comportamento dos usuários. Tais estruturas são baseadas em grafos, sendo composto por um conjunto de nós, ligados por arestas e que possui uma estrutura topológica não trivial. Elas podem representar diversos aspectos do mundo, como modelar pessoas e suas conexões nas redes sociais, páginas da *web* e seu fluxo de acesso, entre outros inúmeras possibilidades (BARABÁSI, 2003).

Zhong *et al.* (2014) afirmam que um sistema de recomendação baseado em grafos é mais eficiente do que um modelo baseado em similaridade entre nós (perfis similares de usuários, por exemplo). Diante desta afirmação, este trabalho tem como objetivo gerar redes baseadas no fluxo de navegação dos usuários de comércios eletrônicos, buscando entender e contribuir com o aumento das efetivações de compra.

#### 1.1 OBJETIVOS

Este trabalho tem como objetivo gerar redes baseadas no fluxo de navegação dos usuários de comércios eletrônicos, buscando entender e contribuir com o aumento das efetivações de compra.

Os objetivos específicos são:

- utilizar a rede para realizar a geração de grupos de produtos para recomendação;
- analisar a rede, calculando métricas e identificando correlações entre a navegação e as compras;
- identificar os maiores pontos de saídas do e-commerce a partir da rede gerada;
- integrar a ferramenta a API Suaview.

### 2 TRABALHOS CORRELATOS

Neste capítulo são apresentados três trabalhos que apresentam semelhanças com o trabalho proposto. A seção 2.1 discorre sobre um sistema de recomendação baseada em agrupamentos (CARVALHO, 2016). A seção 2.2 apresenta um algoritmo para recomendação de relacionamentos em redes sociais (SILVA, 2010). Por fim, a seção 2.3 aborda a construção de um *webcrawler* e a identificação de grupos de produtos comprados em conjunto (SANTOS, 2019).

Formatado: Fonte: 10 pt

## 2.1 RECOMENDAÇÃO BASEADA EM MODULARIDADE

Carvalho (2016) desenvolveu um sistema para gerar recomendações de produtos com base em agrupamentos, utilizando algoritmos de redes complexas, que precisava ser performático e ter baixa taxa de erro. Para isso, a autora utilizou o *LensKit framework* que dá suporte a sistemas de recomendação.

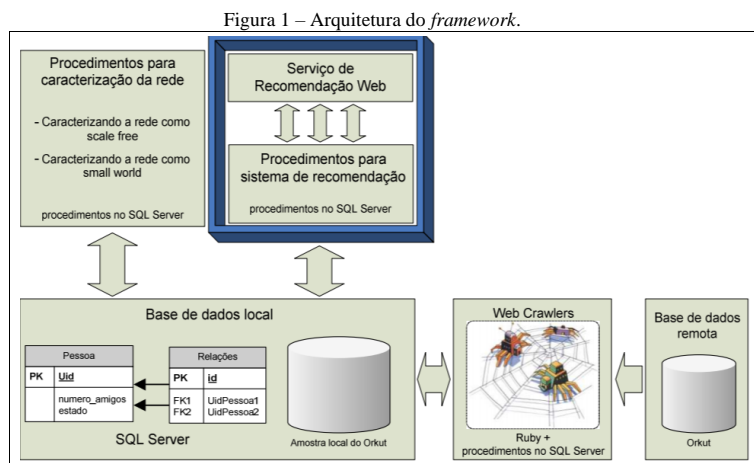
Segundo Carvalho (2016) foram utilizadas as bases Movie Lens 100K, MovieTweatings, Movie Lens 1M, BookCrossing e Jester. As três primeiras bases associam usuários a filmes. Já a quarta base associa os usuários a livros e a quinta base faz a associação dos usuários com piadas, todas com as qualificações dadas pelos usuários aos itens. A partir delas, foram executados os algoritmos Louvain com as modularidades de Newman e Girvan e Agrupamento com Movimento de Vértices (AMV) com a modularidade Suzuki-Wakita.

Com base nos experimentos, Carvalho (2016) optou pela utilização do método **Agrupamento com Movimento de Arestas (AMA)**, pois apresentou os melhores tempos de treinamento e predição. Na análise, a autora gerou 6 grupos de itens e 6 grupos de usuários a partir da base Movie Lens 100K utilizando o algoritmo AMV, encontrando alguns padrões, como: um grupo com uma quantidade grande de itens, porém poucas arestas, significando que são filmes pouco avaliados; mulheres assistiram mais os filmes dos gêneros infantil e de animação, além de assistirem poucos filmes de ação. Carvalho (2016) também notou que o grupo que contém em sua maioria usuários entre os 20 e 40 anos assistem mais filmes de suspense e romance. Já o grupo que contém em sua maioria usuários acima dos 60 anos assistem mais filmes de ação e comédia.

Carvalho (2016) conclui que redes complexas e sistemas de recomendações podem gerar bons resultados, já a hipótese de que quanto maior a métrica de modularidade, menor a taxa de erro, se provou verdadeiro em sua maioria, a não ser na utilização do algoritmo AMV com a modularidade Suzuki-Wakita. Como trabalhos futuros, Carvalho (2016) aponta a necessidade da criação de um algoritmo que encontre a quantidade ideal de grupos, para que não haja a necessidade da passagem de parâmetros, assim como realize a reorganização da rede nas inserções de vértices e faça recomendações quando não for encontrada ligações.

## 2.2 RECOMENDAÇÃO DE RELACIONAMENTOS EM REDES SOCIAIS BASEADA EM GRAFOS

Silva (2010) desenvolveu um sistema de recomendação de relacionamentos em redes sociais, utilizando grafos. O autor propôs um método híbrido utilizando algoritmo genético e dados topológicos, utilizando a rede social Orkut como base. Para o desenvolvimento do método foi construído um *framework* com 4 módulos: *webcrawlers*, responsáveis pela captura das informações; Base de dados local, responsável pelo armazenamento das informações; Procedimentos para caracterização da rede, responsável pela geração dos grafos; Serviço de Recomendação Web, responsável por buscar dos grafos o grupo de itens para recomendação. Esses módulos são representados pela Figura 1.



Fonte: Silva (2010).

A partir da Figura 1, pode-se perceber que existe um módulo referente a base de dados, utilizando SQL Server, contendo duas tabelas, uma tabela *Pessoa* representando os nós e a tabela *Relações* representando os vértices do grafo. Já o módulo *webcrawlers* é responsável pela busca dos dados da rede social Orkut, salvando os dados encontrados na base de dados local. O *crawler* utiliza um usuário como nó raiz, realizando uma busca em largura

**Comentado [MH1]:** No parágrafo anterior este algoritmo não era apresentado como uma opção. Na frase seguinte você volta a citar AMV. Então fica confuso qual é a linha que a autora seguiu.

para acessar os amigos e fazer os mapeamentos. Foram utilizadas duas bases, uma com usuários brasileiros e outra de indianos. O módulo de caracterização da rede foi desenvolvido em T-SQL a partir de cálculos e constatações das redes *Scale Free* e *Small World*. Por fim, o módulo de serviço de recomendação, também escrito em T-SQL, executa os cálculos de índices, a ponderação e a partir disso, gera as recomendações finais. Esse processo passa por duas etapas, a filtragem e ordenação. A filtragem é responsável por limitar a quantidade de vértices para a ordenação. Neste sentido foi utilizado o conceito de *clustering coefficient*, alcançando os vértices com dois saltos. Já a ordenação utiliza um mecanismo de *auto-ajuste* para regular os pesos. Nele, são utilizados 3 índices, o primeiro é composto pela quantidade de amigos em comum. O segundo se refere a densidade do conjunto formado no primeiro índice. No terceiro é a densidade do conjunto formado pelos adjacentes dos itens. A partir desses índices, gera-se a ponderação e a recomendação.

Silva (2010) realizou dois experimentos. No primeiro foi executado a recomendação para dois conjuntos de dados, executando várias vezes ajustando os pesos manualmente com variações pequenas, até que se atingisse o menor valor para a função de otimização. Os melhores resultados obtidos foram um posicionamento médio de 513 e 272 para a sub-rede brasileira e indiana respectivamente, dentro do conjunto de candidatos. Para avaliar a eficiência, o autor utilizou seu próprio usuário, chegando em uma taxa de 75% de acertos nas 20 primeiras recomendações e 50% nas 100 primeiras recomendações.

No segundo experimento Silva (2010) realizou a adição do algoritmo genético e uma validação cruzada com a estratégia *Friend of friend* (FOF). Os dados utilizados vieram da rede social Oro-Aro do C.E.S.A.R, dos quais foram enviadas 10 recomendações para um grupo de 70 usuários, onde 31 realizaram os testes. As recomendações tiveram uma porcentagem de 77.69% de aceitação, 5% superior a estratégia FOF.

Silva (2010) conclui que no primeiro experimento os resultados obtidos foram satisfatórios, porém o método de validação é falho por não possibilitar a reprodução em um ambiente externo. Ele também escreve que a caracterização da rede foi essencial para a elaboração do mecanismo de recomendação. Como melhorias Silva (2010) sugere a adição de outros índices na ordenação para melhorar a ponderação e a implementação de um módulo de algoritmo genético para aprimorar a otimização assim como a realização de testes com outras localizações ou a utilização de alguma API para a extração de dados no lugar do *webcrawler*, pois isso limita apenas ao uso de redes sociais, sendo que qualquer alteração no site pode ocasionar falha de extração dos dados.

### 2.3 ANÁLISE DA REDE DE PRODUTOS COMPRADOS EM CONJUNTO NO COMÉRCIO ELETRÔNICO

Santos (2019), tinha como objetivo a construção de um *webcrawler* para obter uma base de dados, assim como a construção de uma rede de produtos comprados em conjunto para analisar e identificar correlações entre produtos e validar hipóteses, como, produtos melhores avaliados serem *hubs* de rede e validar se somente as métricas de rede são o suficiente para o treinamento do modelo.

Para montar a base de dados foi criado um *webcrawler*, utilizando o *selenium*, para capturar os produtos comprados em conjunto do e-commerce da Amazon. A busca é iniciada com um produto selecionado manualmente, e posteriormente é executado de forma automática como uma busca em largura pelos outros produtos comprados em conjunto. Ainda segundo Santos (2019) o *crawler* foi desenvolvido em PHP com a execução em *Bash* para fazer execuções em paralelo. Após a extração, os dados são enviados separados em dois arquivos *csv* para uma aplicação em python que utiliza a biblioteca *NetworkX* para processar os dados e gerar as redes, que são salvas em um *dataframe* do *pandas*, para gerar as previsões de estimativa de preço e as previsões de ligação entre nós. Para criar as estimativas de preços foram utilizados o *Regressor Random Forest* e o coeficiente Gini na montagem das árvores com maiores relevâncias. O resultado obtido pode ser visualizado na Figura 2, onde pode-se ver uma linha tracejada verde que representa o preço previsto sendo igual ao preço real. Já os pontos azuis representam as predições, os que se localizam acima possuem preço previsto superior ao real, e os que se localizam abaixo tem preço previsto inferior ao real.

Para a previsão de ligação entre nós foram utilizados todos os atributos no treinamento e a coluna *existe\_ligacao\_com\_o\_1* como alvo e a biblioteca *Scikit-Learn* para o treinamento. Por se tratar de classificações binárias bastante desbalanceadas foi utilizada a matriz de confusão para fazer a análise dos resultados, analisando os falsos positivos e positivos negativos. Dos 768 itens processados, Santos (2019) obteve 634 identificados corretamente, 1 positivo negativo e 133 falsos positivos. Para uma segunda execução foi considerado a utilização da *Receiver Operating Characteristic Curve* (ROC) para identificar o ponto com maior precisão, um ponto onde não seja sensível demais onde terá muitos falsos positivos e nem específico demais, onde obterá muitos positivos falsos. Com a curva ROC ele obteve um limiar de probabilidade de 0,66, onde valores superiores eram considerados positivos e inferiores negativos. Após a aplicação do resultado da curva ROC, Santos (2019) obteve resultados consideravelmente superiores, reduzindo o número de falsos positivos de 133 para 93.

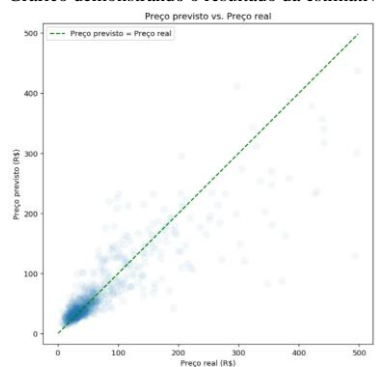
**Comentado [MH2]:** Ao longo do texto há várias frases muito longas. Aqui está um exemplo onde o parágrafo todo é uma única frase. Isso cansa o leitor. Procure dividir em frases menores, em que cada frase apresenta uma ideia.

**Comentado [MH3]:** Preço de que? De um produto? De um conjunto de produtos? De uma compra efetuada por um usuário? Por vários usuários? Explicar melhor.

**Formatado:** Fonte: Itálico

**Comentado [MH4]:** Melhorar a redação para deixar mais claro.

Figura 2 - Gráfico demonstrando o resultado da estimativa de preço.



Fonte: Santos (2019)

Santos (2019) conclui que *hubs* de rede nem sempre serão os mais bem avaliados, porém confirmou a correlação deles serem os mais bem posicionados nas listas de mais vendidos. Santos (2019) também confirmou a hipótese de que a utilização exclusiva dos atributos de métrica foram suficientes para obter melhores resultados em comparação com a *baseline* do modelo de previsão de ligações entre nós, ao contrário dos resultados obtidos na precificação dos produtos.

Santos (2019) sugere a análise da predição de relação entre todos os nós da rede, além de revisar os atributos utilizados no treinamento para poder otimizar o processo. Outro ponto é a integração com APIs de comércio eletrônico, para que não seja necessária a extração das informações com o *webcrawler*.

### 3 PROPOSTA DA FERRAMENTA

Neste capítulo são definidas as justificativas para a elaboração deste trabalho, assim como os requisitos funcionais, não funcionais e a metodologia que será aplicada no desenvolvimento.

#### 3.1 JUSTIFICATIVA

No Quadro 1 é apresentado um comparativo entre os trabalhos correlatos. As linhas representam as características e as colunas os trabalhos.

Quadro 1 – Comparativos entre os trabalhos correlatos.

| Características \ Correlatos  | Carvalho (2016)                             | Silva (2010)              | Santos (2019)                  |
|-------------------------------|---|---------------------------|--------------------------------|
| Obtenção dos dados            | API   | Webcrawler e API          | Webcrawler                     |
| Método de processamento       | LensKit                                     | T-SQL                     | Scikit Learn                   |
| Tipo de item analisado        | Avaliações de filmes                        | Amizades em redes sociais | Produtos comprados em conjunto |
| Usa grafo bipartido           | Sim   | Não                       | Não                            |
| Principal algoritmo / técnica | Agrupamento com Movimento de Vértices (AMV) | Algoritmo próprio         | Random Forest                  |

Fonte: elaborado pelo autor.

A partir do Quadro 1, observa-se duas estratégias de obtenção de dados, enquanto Carvalho (2016) utilizou uma API que forneciam as informações necessárias para a montagem das redes, Santos (2019) em seu trabalho criou um *webcrawler* que navega pela loja da Amazon e captura os dados para sua base. Já Silva (2010) utiliza ambas as técnicas, criando um *webcrawler* que captura informações da rede social Orkut para o primeiro experimento e utiliza uma API da C.E.S.A.R para o seu segundo experimento.

Carvalho (2016) utilizou grafos bipartidos, onde definiu nós do tipo usuário e nós do tipo filme para a sua análise, enquanto Silva (2010) e Santos (2019) não fizeram uso de tal particularidade dos grafos, já que no projeto de Silva (2019) somente foram mapeados os usuários e no projeto de Santos (2019) foram mapeados os itens. Quanto aos métodos de processamentos, pode-se observar que dois trabalhos utilizam as bibliotecas da linguagem Python para as predições e recomendações, Carvalho (2016) com o LensKit e Santos (2019) com o Scikit Learn,

enquanto Silva (2010) utilizou a linguagem T-SQL. ~~O~~ olhando para os principais algoritmos e técnicas, percebe-se que ~~enquanto~~ Carvalho (2016) utilizou ~~a~~ o algoritmo AMV para realizar o agrupamento. ~~e~~ Santos (2019) ~~utilizou~~ aplicou o Random Forest para a criação das estimativas de preços. ~~e~~ Silva (2010) decidiu criar seu próprio algoritmo utilizando uma série de técnicas e conceitos para as etapas de classificação, filtragem e ponderação.

Dessa forma, conclui-se que dentro dos trabalhos pesquisados não foi encontrado algum que tenha como objetivo gerar recomendações e análises com base nos acessos dos usuários nos sites de comércio eletrônico. Percebe-se também que nenhum trabalho teve em sua proposta a ideia de realizar testes e validações práticas com o usuário final. Sendo assim, o trabalho proposto se difere pois pretende realizar uma análise alternativa, através do uso dos conceitos de redes complexas, pensada no usuário e sua navegação, tendo como objetivo manter seu interesse por um período maior de tempo; e que o conduza com mais naturalidade ao item desejado, sendo testado e colocado em prática através de uma solução de e-commerce já implantada no mercado.

### 3.2 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

A ferramenta a ser desenvolvida deverá:

- identificar as páginas que são pontos de saídas (nós folhas) (Requisito Funcional – RF);
- prever as prováveis próximas páginas a serem acessadas pelo usuário (RF);
- gerar grupos de usuários a partir dos acessos (RF);
- identificar a qual grupo o usuário pertence (RF);
- ser implementado na linguagem de programação Python (Requisito Não Funcional – RNF);
- modelar o problema seguindo os princípios de redes complexas (RNF);
- ser integrado com a API Suaview (RNF).

**Comentado [MH5]:** Estes requisitos não são suficientes para atender os objetivos específicos.

### 3.3 METODOLOGIA

O trabalho será desenvolvido observando as seguintes etapas:

- levantamento bibliográfico: pesquisar e estudar sobre redes complexas, sistemas de recomendação e trabalhos correlatos;
- levantamento dos requisitos: baseando-se nas informações da etapa anterior, reavaliar os requisitos propostos para a aplicação;
- integração com API Suaview: implementar a estrutura que irá se comunicar com a API Suaview para a obtenção dos dados;
- definição das ferramentas para modelagem e armazenamento das redes: pesquisar e escolher as ferramentas mais apropriadas para a modelagem e armazenamento das redes;
- definição das técnicas de análise de redes: pesquisar e definir as técnicas/algoritmos que serão utilizadas para analisar redes complexas;
- modelagem da rede: a partir do item (d) modelar a estrutura da rede de forma a facilitar a análise das correlações de acordo com o problema proposto;
- implementação da rede: implementar as funções de análise da rede considerando os itens (d), (e) e (f), tendo a linguagem Python como base;
- testes: avaliar a performance assim como validar a aderência e eficiência da rede em relação aos resultados alcançados e os que realmente acontecerão no e-commerce em termos de saída ou efetivação de compras e recomendações de produtos.

As etapas serão realizadas nos períodos relacionados no Quadro 2.

Quadro 2 – Cronograma de atividades a serem realizadas

| etapas / quinzenas   | 2021 |   |      |   |      |   |      |   |      |   |
|--|------|---|------|---|------|---|------|---|------|---|
|  | fev. |   | mar. |   | abr. |   | maio |   | jun. |   |
|  | 1    | 2 | 1    | 2 | 1    | 2 | 1    | 2 | 1    | 2 |
| levantamento bibliográfico   |      |   |      |   |      |   |      |   |      |   |
| levantamento dos requisitos  |      |   |      |   |      |   |      |   |      |   |
| integração com API Suaview   |      |   |      |   |      |   |      |   |      |   |
| definição das ferramentas para modelagem e armazenamento das redes |      |   |      |   |      |   |      |   |      |   |
| definição das técnicas de análise de redes                         |      |   |      |   |      |   |      |   |      |   |
| modelagem da rede  |      |   |      |   |      |   |      |   |      |   |
| implementação da rede  |      |   |      |   |      |   |      |   |      |   |
| testes   |      |   |      |   |      |   |      |   |      |   |

Fonte: elaborado pelo autor.

#### 4 REVISÃO BIBLIOGRÁFICA

Este capítulo descreve brevemente sobre os assuntos que fundamentarão o estudo a ser realizado: redes complexas e sistemas de recomendação.

Segundo Nahal (2014) redes complexas são estruturas topológicas que não seguem um padrão regular. São baseadas em grafos, formadas por elementos (vértices) e conexões (arrestas). As redes complexas podem ser classificadas como redes aleatórias, redes livre de escalas, entre outras. As redes aleatórias têm suas conexões adicionadas com a mesma probabilidade para cada elemento, onde todos os elementos terão aproximadamente a mesma quantidade de conexões. Porém, com o avanço dos estudos, observou-se que redes do mundo real contêm propriedades bem distintas das redes aleatórias. De acordo com Ribeiro (2017) as redes livres de escala contêm uma distribuição de conectividade que segue a lei de potência que indica uma ausência de escala típica. Enquanto as redes aleatórias contêm um número fixo de vértices, as redes livres de escala permitem a adição de vértices após a geração da rede. Broido e Clauset (2019) questionam o uso equivocado das redes sem escalas, argumentando que em muitos casos a utilização de outros modelos de rede seriam melhor aplicados.

De acordo com Andrade (2018) os sistemas de recomendação como são um conjunto de algoritmos que utilizam técnicas de aprendizado de máquina e recuperação de informação gerando recomendações baseadas em algum tipo de filtragem. Os sistemas sendo os mais comuns são a filtragem colaborativa e a baseada em conteúdo. Bonin (2018) descreve os sistemas baseados em conteúdo como tendo foco nas propriedades dos itens, onde a similaridade será determinada a partir das semelhanças entre as propriedades dos itens. Já no sistema de filtragem colaborativa o foco se dá na relação entre os usuários e os itens. A similaridade é determinada pelos grupos de usuários que visualizam, compram ou avaliam os mesmos itens.

#### REFERÊNCIAS

- ANDRADE, Michelle H. S. **Entenda como funcionam os Sistemas de Recomendação**. [2017]. Disponível em: <https://www.igti.com.br/blog/como-funcionam-os-sistemas-de-recomendacao/>. Acesso em: 7 out. 2020.
- BARABÁSI, A.-L. **Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life**, Plume Books, 2003.
- BONIN, Morvana. **O que são Sistemas de Recomendação? Veja exemplos**. [2018]. Disponível em: <https://king.host.blog/2018/09/o-que-sao-sistemas-de-recomendacao/#:~:text=Sistemas%20de%20recomendação%20são%20técnicas,escutar%20ou%20quais%20notícias%20ler>. Acesso em: 25 set. 2020.
- BRITO, Carina. **E-commerce cresce em 2020 impulsionado pela pandemia do coronavírus**. [2020]. Disponível em: <https://revistapegn.globo.com/Banco-de-ideias/Varejo/noticia/2020/08/e-commerce-cresce-em-2020-impulsionado-pela-pandemia-do-coronavirus.html>. Acesso em: 26 set. 2020.
- BROIDO, Anna D.; CLAUSET, Aaron. Scale-free networks are rare. **Nature Communications**, [S.L.], v. 10, n. 1, p. 1-10, 4 mar. 2019. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41467-019-08746-5>. Disponível em: <https://www.nature.com/articles/s41467-019-08746-5>. Acesso em: 11 out. 2020.
- CARVALHO, Maria A. A. S. **Recomendação baseada em modularidade**. 2016. 126f. Tese (Doutorado em Ciência da Computação) - Curso de Pós-Graduação em Ciência da Computação, Universidade Federal de Pernambuco, Recife.
- LUNARDI, Guilherme. **12 dados que comprovam o crescimento do e-commerce no Brasil**. [2018]. Disponível em: <https://www.ecommercebrasil.com.br/artigos/12-dados-que-comprovam-o-crescimento-do-e-commerce-no-brasil/>. Acesso em: 25 set. 2020.
- NAHAL, Jessica. **Introdução a redes complexas**. [2014]. Disponível em: <https://blog.dp6.com.br/introdução-a-redes-complexas-df73b623d67f>. Acesso em: 2 out. 2020.
- RIBEIRO, Larissa F. **Redes sem Escala Típica: Visão Geral, Modelos Alternativos e Técnicas Computacionais**. 2017. 93f. Dissertação (Mestrado em Física) - Curso de Pós-Graduação em Física, Universidade Federal do Rio Grande do Norte, Natal.
- SANTOS, Rafael Joseph Pagliuca. **Análise da rede de produtos comprados em conjunto no comércio eletrônico**. 2019. 130f. Dissertação (Mestrado em Ciência da Computação e Matemática Computacional) - Curso de Pós-Graduação em Ciência da Computação, Universidade de São Paulo, São Paulo.
- SCHNAIDER, Amanda. **E-commerce cresce 47%, maior alta em 20 anos**. [2020]. Disponível em: <https://www.meioemensagem.com.br/home/marketing/2020/08/27/e-commerce-cresce-47-maior-alta-em-20-anos.html>. Acesso em: 21 set. 2020.
- SILVA, Natai B. **Recomendação de relacionamentos em redes sociais baseada em grafos**. 2010. 94f. Dissertação (Mestrado em Ciência da Computação) - Curso de Pós-Graduação em Ciência da Computação, Universidade Federal de Pernambuco, Recife.

TERRA, Tiago. **Comércio eletrônico cresce com pandemia do COVID-19, diz ABComm**. [2020]. Disponível em: <https://www.mundodomarketing.com.br/ultimas-noticias/38602/comercio-eletronico-cresce-com-pandemia-do-covid-19-diz-abcomm.html>. Acesso em: 21 set. 2020.

Zhong et al. Study on Directed Trust Graph Based Recommendation for E-commerce System. **International Journal of Computers, Communications and Control**. v. 9, n. 4, p. 510-523, jun. 2014.

ASSINATURAS

(Atenção: todas as folhas devem estar rubricadas)

Assinatura do(a) Aluno(a): \_\_\_\_\_

Assinatura do(a) Orientador(a): \_\_\_\_\_

Assinatura do(a) Coorientador(a) (se houver): \_\_\_\_\_

Observações do orientador em relação a itens não atendidos do pré-projeto (se houver):



## FORMULÁRIO DE AVALIAÇÃO – PROFESSOR AVALIADOR

Acadêmico(a): Diogo Warmeling

Avaliador(a): Marcel Hugo

| ASPECTOS AVALIADOS <sup>1</sup> |  | atende | atende parcialmente | não atende |
|---------------------------------|--|--------|---------------------|------------|
| ASPECTOS TÉCNICOS               | 1. INTRODUÇÃO<br>O tema de pesquisa está devidamente contextualizado/delimitado?   | X      |                     |            |
|                                 | O problema está claramente formulado?  | X      |                     |            |
|                                 | 1. OBJETIVOS<br>O objetivo principal está claramente definido e é passível de ser alcançado?   | X      |                     |            |
|                                 | Os objetivos específicos são coerentes com o objetivo principal?   | X      |                     |            |
|                                 | 2. TRABALHOS CORRELATOS<br>São apresentados trabalhos correlatos, bem como descritas as principais funcionalidades e os pontos fortes e fracos?                          | X      |                     |            |
|                                 | 3. JUSTIFICATIVA<br>Foi apresentado e discutido um quadro relacionando os trabalhos correlatos e suas principais funcionalidades com a proposta apresentada?             | X      |                     |            |
|                                 | São apresentados argumentos científicos, técnicos ou metodológicos que justificam a proposta?  | X      |                     |            |
|                                 | São apresentadas as contribuições teóricas, práticas ou sociais que justificam a proposta?   | X      |                     |            |
|                                 | 4. REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO<br>Os requisitos funcionais e não funcionais foram claramente descritos?   |        |                     | X          |
|                                 | 5. METODOLOGIA<br>Foram relacionadas todas as etapas necessárias para o desenvolvimento do TCC?  | X      |                     |            |
|                                 | Os métodos, recursos e o cronograma estão devidamente apresentados e são compatíveis com a metodologia proposta?   | X      |                     |            |
|                                 | 6. REVISÃO BIBLIOGRÁFICA (atenção para a diferença de conteúdo entre projeto e pré-projeto)<br>Os assuntos apresentados são suficientes e têm relação com o tema do TCC? | X      |                     |            |
|                                 | As referências contemplam adequadamente os assuntos abordados (são indicadas obras atualizadas e as mais importantes da área)?   | X      |                     |            |
| ASPECTOS METODOLÓGICOS          | 7. LINGUAGEM USADA (redação)<br>O texto completo é coerente e redigido corretamente em língua portuguesa, usando linguagem formal/científica?                            | X      |                     |            |
|                                 | A exposição do assunto é ordenada (as ideias estão bem encadeadas e a linguagem utilizada é clara)?  |        | X                   |            |

### PARECER – PROFESSOR AVALIADOR: (PREENCHER APENAS NO PROJETO)

O projeto de TCC ser deverá ser revisado, isto é, necessita de complementação, se:

- qualquer um dos itens tiver resposta NÃO ATENDE;
- pelo menos 5 (cinco) tiverem resposta ATENDE PARCIALMENTE.

**PARECER:** ( ) APROVADO ( ) REPROVADO

Assinatura: \_\_\_\_\_ Data: \_\_\_\_\_

<sup>1</sup> Quando o avaliador marcar algum item como atende parcialmente ou não atende, deve obrigatoriamente indicar os motivos no texto, para que o aluno saiba o porquê da avaliação.

| CURSO DE CIÊNCIA DA COMPUTAÇÃO – TCC |             |                      |
|--------------------------------------|-------------|----------------------|
| ( X ) PRÉ-PROJETO                    | ( ) PROJETO | ANO/SEMESTRE: 2020.2 |

## USO DE REDES COMPLEXAS PARA ANÁLISE DE CONSUMO EM E-COMMERCE

Diogo Warmeling

Aurélio Faustino Hoppe – orientador

### 1 INTRODUÇÃO

Com o crescimento dos acessos à internet e a disseminação de dispositivos móveis, os comércios eletrônicos (e-commerce) vêm crescendo de forma exponencial nos últimos anos e, com a pandemia do COVID-19, o crescimento foi ainda maior, tornando-se o método mais comum de compra para diversos setores da economia brasileira (TERRA, 2020). Segundo os dados da 42ª edição do Webshoppers, um estudo sobre e-commerces do Brasil elaborado semestralmente pela Ebit e Elo, o crescimento nas vendas do primeiro semestre de 2020 foi de 47%, caracterizando-se como dobro dos registrados nos últimos anos. Isso, segundo Lunardi (2018), aconteceu devido a entrada de novos adeptos a esse meio de compra, resultando em um crescimento de 40%, chegando a 41 milhões de pessoas.

Formatado: Fonte: 10 pt

Brito (2020) destaca que esse crescimento não ficou restringido somente aos *marketplaces*, mas também atingiu muitos comércios eletrônicos de pequeno e médio porte, dos quais, a maior parte não aproveita o grande potencial que a análise dos dados de acessos de seus clientes fornece. Tais análises possibilitam entender o comportamento dos usuários para mantê-los mais tempos em seus *sites*, e conseguir recomendar o produto desejado antes da desistência da compra, assim como auxiliar os clientes a encontrarem as melhores opções disponíveis, aumentando a taxa de efetivação da compra por parte dos usuários. No entanto, a grande maioria dos *sites* apenas disponibilizam listas de produtos semelhantes e associados ao acesso do usuário, desconsiderando o seu perfil, que pode estar seguindo um padrão de acessos.

Segundo Santos (2019), existem diversas pesquisas sobre recomendações em plataformas de comércio eletrônico, que em sua maioria fazem cálculo de similaridade entre entidades (compradores e produtos) para recomendar novos produtos a um potencial comprador. Dentre elas, encontram-se uma grande quantidade de técnicas e algoritmos disponíveis pensadas para análise de dados tais como sistemas de recomendações, algoritmos de aprendizado de máquina e redes complexas, que permitem encontrar relações e agrupar itens, analisar o fluxo de navegação ou o comportamento dos usuários. Tais estruturas são baseadas em grafos, sendo composto por um conjunto de nós, ligados por arestas e que possui uma estrutura topológica não trivial. Elas podem representar diversos aspectos do mundo, como modelar pessoas e suas conexões nas redes sociais, páginas da *web* e seu fluxo de acesso, entre outras inúmeras possibilidades (BARABÁSI, 2003).

Zhong *et al.* (2014) afirmam que um sistema de recomendação baseado em grafos é mais eficiente do que um modelo baseado em similaridade entre nós (perfis similares de usuários, por exemplo). Diante desta afirmação, este trabalho tem como objetivo gerar redes baseadas no fluxo de navegação dos usuários de comércios eletrônicos, buscando entender e contribuir com o aumento das efetivações de compra.

#### 1.1 OBJETIVOS

Este trabalho tem como objetivo gerar redes baseadas no fluxo de navegação dos usuários de comércios eletrônicos, buscando entender e contribuir com o aumento das efetivações de compra.

Os objetivos específicos são:

- utilizar a rede para realizar a geração de grupos de produtos para recomendação;
- analisar a rede, calculando métricas e identificando correlações entre a navegação e as compras;
- identificar os maiores pontos de saídas do e-commerce a partir da rede gerada;
- integrar a ferramenta a API Suaview.

Comentado [AS1]: Você vai utilizar uma página web ou setor de venda específico?

Comentado [AS2]: Que rede?

Comentado [AS3]: Explique qual rede

### 2 TRABALHOS CORRELATOS

Neste capítulo são apresentados três trabalhos que apresentam semelhanças com o trabalho proposto. A seção 2.1 discorre sobre um sistema de recomendação baseada em agrupamentos (CARVALHO, 2016). A seção 2.2 apresenta um algoritmo para recomendação de relacionamentos em redes sociais (SILVA, 2010). Por fim, a seção 2.3 aborda a construção de um *webcrawler* e a identificação de grupos de produtos comprados em conjunto (SANTOS, 2019).

## 2.1 RECOMENDAÇÃO BASEADA EM MODULARIDADE

Carvalho (2016) desenvolveu um sistema para gerar recomendações de produtos com base em agrupamentos, utilizando algoritmos de redes complexas, que precisava ser performático e ter baixa taxa de erro. Para isso, a autora utilizou o LensKit *framework* que dá suporte a sistemas de recomendação.

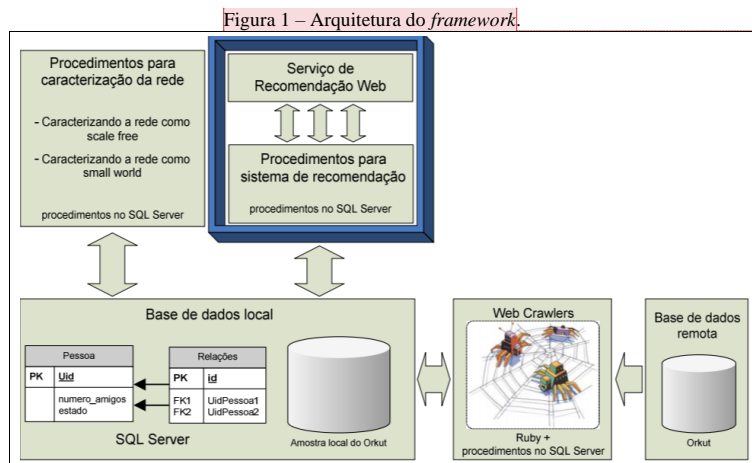
Segundo Carvalho (2016) foram utilizadas as bases de dados? Movie Lens 100K, MovieTweatings, Movie Lens 1M, BookCrossing e Jester. As três primeiras bases associam usuários a filmes. Já a quarta base associa os usuários a livros e a quinta base faz a associação dos usuários com piadas, todas com as qualificações dadas pelos usuários aos itens. A partir delas, foram executados os algoritmos Louvain com as modularidades de Newman e Girvan e Agrupamento com Movimento de Vértices (AMV) com a modularidade Suzuki-Wakita.

Com base nos experimentos, Carvalho (2016) optou pela a utilização do método Agrupamento com Movimento de Arestas (AMA), pois apresentou os melhores tempos de treinamento e predição. Na análise, a autora gerou 6 grupos de itens e 6 grupos de usuários a partir da base de dados? Movie Lens 100K utilizando o algoritmo AMV, encontrando alguns padrões, como, um grupo com uma quantidade grande de itens, porém poucas arestas, significando que são filmes pouco avaliados, mulheres assistiram mais os filmes dos gêneros infantil e de animação, além de assistirem poucos filmes de ação. Carvalho (2016) também notou que o grupo que contém em sua maioria usuário entre os 20 a 40 anos assistem mais filmes de suspense e romance. Já o grupo que contém em sua maioria usuários acima dos 60 anos assistem mais filmes de ação e comédia.

Carvalho (2016) conclui que redes complexas e sistemas de recomendações podem gerar bons resultados, já a hipótese de que quanto maior a métrica de modularidade, menor a taxa de erro, se provou verdadeiro em sua maioria, a não ser na utilização do algoritmo AMV com a modularidade Suzuki-Wakita. Como trabalhos futuros, Carvalho (2016) aponta a necessidade da criação de um algoritmo que encontre a quantidade ideal de grupos, para que não haja a necessidade da passagem de parâmetros, assim como realize a reorganização da rede nas inserções de vértices e faça recomendações quando não for encontrada ligações.

## 2.2 RECOMENDAÇÃO DE RELACIONAMENTOS EM REDES SOCIAIS BASEADA EM GRAFOS

Silva (2010) desenvolveu um sistema de recomendação de relacionamentos em redes sociais, utilizando grafos. O autor propôs um método híbrido utilizando algoritmo genético e dados topológicos, utilizando a rede social Orkut como base. Para o desenvolvimento do método foi construído um *framework* com 4 módulos. *webcrawlers*, responsáveis pela captura das informações; Base de dados local, responsável pelo armazenamento das informações; Procedimentos para caracterização da rede, responsável pela geração dos grafos; Serviço de Recomendação Web, responsável por buscar dos grafos o grupo de itens para recomendação. Esses módulos são representados pela Figura 1.



Fonte: Silva (2010).

A partir da Figura 1, pode-se perceber que existe um módulo referente a base de dados, utilizando SQL Server, contendo duas tabelas, uma tabela Pessoa representando os nós e a tabela Relações representando os vértices do grafo. Já o módulo *webcrawlers* é responsável pela busca dos dados da rede social Orkut, salvando os dados encontrados na base de dados local. O *crawler* utiliza um usuário como nó raiz, realizando uma busca em largura

Formatado: Inglês (Estados Unidos)

Comentado [AS4]: Frase longa. Rever.

Comentado [AS5]: Coloque o recurso de referência cruzada para figura/quadro/tabela. Faça isso em todo o texto.

Comentado [AS6]: - Tem estilo para legenda e fonte. Verifique todos do seu texto. Estão errados, sem estilo.

para acessar os amigos e fazer os mapeamentos. Foram utilizadas duas bases, uma com usuários brasileiros e outra de indianos. O módulo de caracterização da rede foi desenvolvido em T-SQL a partir de cálculos e constatações das redes *Scale Free* e *Small World*. Por fim, o módulo de serviço de recomendação, também escrito em T-SQL, executa os cálculos de índices, a ponderação e a partir disso, gera as recomendações finais. Esse processo passa por duas etapas, a filtragem e ordenação. A filtragem é responsável por limitar a quantidade de vértices para a ordenação. Neste sentido foi utilizado o conceito de *clustering coefficient*, alcançando os vértices com dois saltos. Já a ordenação utiliza um mecanismo de auto-ajuste para regular os pesos. Nele, são utilizados 3 índices, o primeiro é composto pela quantidade de amigos em comum. O segundo se refere a densidade do conjunto formado no primeiro índice. No terceiro é a densidade do conjunto formado pelos adjacentes dos itens. A partir desses índices, gera-se a ponderação e a recomendação.

Silva (2010) realizou dois experimentos, no primeiro foi executado a recomendação para dois conjuntos de dados, executando várias vezes ajustando os pesos manualmente com variações pequenas, até que se atingisse o menor valor para a função de otimização. Os melhores resultados obtidos foram um posicionamento médio de 513 e 272 para a sub-rede brasileira e indiana respectivamente, dentro do conjunto de candidatos. Para avaliar a eficiência, o autor utilizou seu próprio usuário, chegando em uma taxa de 75% de acertos nas 20 primeiras recomendações e 50% nas 100 primeiras recomendações.

No segundo experimento Silva (2010) realizou a adição do algoritmo genético e uma validação cruzada com a estratégia *Friend of friend* (FOF), os dados utilizados vieram da rede social Oro-Aro do C.E.S.A.R, dos quais foram enviadas 10 recomendações para um grupo de 70 usuários, onde 31 realizaram os testes. As recomendações tiveram uma porcentagem de 77.69% de aceitação, 5% superior a estratégia FOF.

Silva (2010) conclui que no primeiro experimento os resultados obtidos foram satisfatórios, porém o método de validação é falho por não possibilitar a reprodução em um ambiente externo. Ele também escreve que a caracterização da rede foi essencial para a elaboração do mecanismo de recomendação. Como melhorias Silva (2010) sugere a adição de outros índices na ordenação para melhorar a ponderação e a implementação de um módulo de algoritmo genético para aprimorar a otimização, assim como a realização de testes com outras localizações ou a utilização de alguma API para a extração de dados no lugar do *webcrawler*, pois isso limita apenas ao uso de redes sociais, sendo que qualquer alteração no site pode ocasionar falha de extração dos dados.

## 2.3 ANÁLISE DA REDE DE PRODUTOS COMPRADOS EM CONJUNTO NO COMÉRCIO ELETRÔNICO

Santos (2019), tinha como objetivo a construção de um *webcrawler* para obter uma base de dados, assim como a construção de uma rede de produtos comprados em conjunto para analisar e identificar correlações entre produtos e validar hipóteses, como, produtos melhores avaliados serem *hubs* de rede e validar se somente as métricas de rede são o suficiente para o treinamento do modelo.

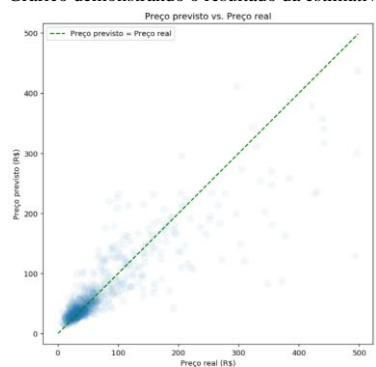
Para montar a base de dados foi criado um *webcrawler*, utilizando o *selenium*, para capturar os produtos comprados em conjunto do e-commerce da Amazon. A busca é iniciada com um produto selecionado manualmente, e posteriormente é executado de forma automática como uma busca em largura pelos outros produtos comprados em conjunto. Ainda segundo Santos (2019) o *crawler* foi desenvolvido em PHP com a execução em *Bash* para fazer execuções em paralelo. Após a extração, os dados são enviados separados em dois arquivos *csv* para uma aplicação em python que utiliza a biblioteca *NetworkX* para processar os dados e gerar as redes, que são salvas em um *dataframe* do *pandas*, para gerar as previsões de estimativa de preço e as previsões de ligação entre nós. Para criar as estimativas de preços foram utilizados o *Regressor Random Forest* e o coeficiente Gini na montagem das árvores com maiores relevâncias. O resultado obtido pode ser visualizado na Figura 2, onde pode-se ver uma linha tracejada verde que representa o preço previsto sendo igual ao preço real. Já os pontos azuis representam as predições, os que se localizam acima possuem preço previsto superior ao real, e os que se localizam abaixo tem preço previsto inferior ao real.

Para a previsão de ligação entre nós foram utilizados todos os atributos no treinamento e a coluna *existe\_ligacao\_com\_o\_1* como alvo e a biblioteca *Scikit Learn* para o treinamento. Por se tratar de classificações binária bastante desbalanceadas foi utilizada a matriz de confusão para fazer a análise dos resultados, analisando os falsos positivos e positivos negativos. Dos 768 itens processados, Santos (2019) obteve 634 identificados corretamente, 1 positivo negativo e 133 falsos positivos. Para uma segunda execução foi considerado a utilização da *Operating Characteristic Curve* (ROC) para identificar o ponto com maior precisão, um ponto onde não seja sensível demais onde terá muitos falsos positivos e nem específico demais, onde obteria muito positivos falsos. Com a curva ROC ele obteve um limiar de probabilidade de 0,66, onde valores superiores eram considerados positivos e inferiores negativos. Após a aplicação do resultado da curva ROC Santos (2019) obteve resultados consideravelmente superiores, reduzindo o número de falsos positivos de 133 para 93.

Comentado [AS7]: Frase longa. Rever. Não se faz parágrafo com uma única frase.

Comentado [AS8]: Frase confusa. Rever.

Figura 2 - Gráfico demonstrando o resultado da estimativa de preço.



Fonte: Santos (2019)

Santos (2019) conclui que *hubs* de rede nem sempre serão os mais bem avaliados, porém confirmou a correlação deles serem os mais bem posicionados nas listas de mais vendidos. Santos (2019) também confirmou a hipótese de que a utilização exclusiva dos atributos de métrica foram suficientes para obter melhores resultados em comparação com a *baseline* do modelo de previsão de ligações entre nós, ao contrário dos resultados obtidos na precificação dos produtos.

Santos (2019) sugere a análise da predição de relação entre todos os nós da rede, além de revisar os atributos utilizados no treinamento para poder otimizar o processo. Outro ponto é a integração com APIs de comércios eletrônicos, para que não seja necessária a extração das informações com o *webcrawler*.

### 3 PROPOSTA DA FERRAMENTA

Neste capítulo são definidas as justificativas para a elaboração deste trabalho, assim como os requisitos funcionais, não funcionais e a metodologia que será aplicada no desenvolvimento.

#### 3.1 JUSTIFICATIVA

No Quadro 1 é apresentado um comparativo entre os trabalhos correlatos. As linhas representam as características e as colunas os trabalhos.

Quadro 1 – Comparativos entre os trabalhos correlatos.

| Características               | Correlatos | Carvalho (2016)                             | Silva (2010)              | Santos (2019)                  |
|-------------------------------|------------|---|---------------------------|--------------------------------|
| Obtenção dos dados            |            | API   | Webcrawler e API          | Webcrawler                     |
| Método de processamento       |            | LensKit                                     | T-SQL                     | Scikit Learn                   |
| Tipo de item analisado        |            | Avaliações de filmes                        | Amizades em redes sociais | Produtos comprados em conjunto |
| Usa grafo bipartido           |            | Sim   | Não                       | Não                            |
| Principal algoritmo / técnica |            | Agrupamento com Movimento de Vértices (AMV) | Algoritmo próprio         | Random Forest                  |

Fonte: elaborado pelo autor.

A partir do Quadro 1, observa-se duas estratégias de obtenção de dados, enquanto Carvalho (2016) utilizou uma API que forneciam as informações necessárias para a montagem das redes, Santos (2019) em seu trabalho criou um *webcrawler* que navega pela loja da Amazon e captura os dados para sua base. Já Silva (2010) utiliza ambas as técnicas, criando um *webcrawler* que captura informações da rede social Orkut para o primeiro experimento e utiliza uma API da C.E.S.A.R para o seu segundo experimento.

Carvalho (2016) utilizou grafos bipartidos, onde definiu nós do tipo usuário e nós do tipo filme para a sua análise, enquanto Silva (2010) e Santos (2019) não fizeram uso de tal particularidade dos grafos, já que no projeto de Silva (2019) somente foram mapeados os usuários e no projeto de Santos (2019) foram mapeados os itens. Quanto aos métodos de processamentos, pode-se observar que dois trabalhos utilizam as bibliotecas da linguagem Python para as predições e recomendações, Carvalho (2016) com o LensKit e Santos (2019) com o Scikit Learn,

enquanto Silva (2010) utilizou a linguagem T-SQL. e Olhando para os principais algoritmos e técnicas, percebe-se que enquanto Carvalho (2016) utiliza o algoritmo AMV para realizar o agrupamento e Santos (2019) utilizou o Random Forest para a criação das estimativas de preços. Silva (2010) ~~decidiu criar~~ desenvolveu seu próprio algoritmo utilizando uma série de técnicas e conceitos para as etapas de classificação, filtragem e ponderação.

Dessa forma, conclui-se que dentro dos trabalhos pesquisados não foi encontrado algum que tenha como objetivo gerar recomendações e análises com base nos acessos dos usuários nos sites de comércio eletrônico. Percebe-se também que nenhum trabalho teve em sua proposta a ideia de realizar testes e validações práticas com o usuário final. Sendo assim, o trabalho proposto se difere pois pretende realizar uma análise alternativa, através do uso dos conceitos de redes complexas, pensada no usuário e sua navegação, tendo como objetivo manter seu interesse por um período maior de tempo, e que o conduza com mais naturalidade ao item desejado, sendo testado e colocado em prática através de uma solução de e-commerce já implantada no mercado.

**Comentado [AS9]:** Quais as contribuições teóricas, práticas ou sociais que justificam a proposta?

### 3.2 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

A ferramenta a ser desenvolvida deverá:

- identificar as páginas que são pontos de saídas (nós folhas) (Requisito Funcional – RF);
- prever as prováveis próximas páginas a serem acessadas pelo usuário (RF);
- gerar grupos de usuários a partir dos acessos (RF);
- identificar a qual grupo o usuário pertence (RF);
- ser implementado na linguagem de programação Python (Requisito Não Funcional – RNF);
- modelar o problema seguindo os princípios de redes complexas (RNF);
- ser integrado com a API Suaview (RNF).

### 3.3 METODOLOGIA

O trabalho será desenvolvido observando as seguintes etapas:

- levantamento bibliográfico: pesquisar e estudar sobre redes complexas, sistemas de recomendação e trabalhos correlatos;
- levantamento dos requisitos: baseando-se nas informações da etapa anterior, reavaliar os requisitos propostos para a aplicação;
- integração com API Suaview: implementar a estrutura que irá se comunicar com a API Suaview para a obtenção dos dados;
- definição das ferramentas para modelagem e armazenamento das redes: pesquisar e escolher as ferramentas mais apropriadas para a modelagem e armazenamento das redes;
- definição das técnicas de análise de redes: pesquisar e definir as técnicas/algoritmos que serão utilizadas para analisar redes complexas;
- modelagem da rede: a partir do item (d) modelar a estrutura da rede de forma a facilitar a análise das correlações de acordo com o problema proposto;
- implementação da rede: implementar as funções de análise da rede considerando os itens (d), (e) e (f), tendo a linguagem Python como base;
- testes: avaliar a performance assim como validar a aderência e eficiência da rede em relação aos resultados alcançados e os que realmente acontecerão no e-commerce em termos de saída ou efetivação de compras e recomendações de produtos.

As etapas serão realizadas nos períodos relacionados no Quadro 2.

Quadro 2 – Cronograma de atividades a serem realizadas

| etapas / quinzenas   | 2021 |   |      |   |      |   |      |   |      |   |
|--|------|---|------|---|------|---|------|---|------|---|
|  | fev. |   | mar. |   | abr. |   | maio |   | jun. |   |
|  | 1    | 2 | 1    | 2 | 1    | 2 | 1    | 2 | 1    | 2 |
| levantamento bibliográfico   |      |   |      |   |      |   |      |   |      |   |
| levantamento dos requisitos  |      |   |      |   |      |   |      |   |      |   |
| integração com API Suaview   |      |   |      |   |      |   |      |   |      |   |
| definição das ferramentas para modelagem e armazenamento das redes |      |   |      |   |      |   |      |   |      |   |
| definição das técnicas de análise de redes                         |      |   |      |   |      |   |      |   |      |   |
| modelagem da rede  |      |   |      |   |      |   |      |   |      |   |
| implementação da rede  |      |   |      |   |      |   |      |   |      |   |
| testes   |      |   |      |   |      |   |      |   |      |   |

Fonte: elaborado pelo autor.

#### 4 REVISÃO BIBLIOGRÁFICA

Este capítulo descreve brevemente sobre os assuntos que fundamentarão o estudo a ser realizado: redes complexas e sistemas de recomendação.

Segundo Nahal (2014) redes complexas são estruturas topológicas que não seguem um padrão regular, são baseadas em grafos, formadas por elementos (vértices) e conexões (arrestas). As redes complexas podem ser classificadas como redes aleatórias, redes livre de escalas, entre outras. As redes aleatórias têm suas conexões adicionadas com a mesma probabilidade para cada elemento, onde todos os elementos terão aproximadamente a mesma quantidade de conexões. Com o avanço dos estudos, observou-se que redes do mundo real contêm propriedades bem distintas das redes aleatórias. De acordo com Ribeiro (2017) as redes livres de escala contêm uma distribuição de conectividade que segue a lei de potência que indica uma ausência de escala típica, enquanto as redes aleatórias contêm um número fixo de vértices as redes livres de escala permitem a adição de vértices após a geração da rede. Broido e Clauset (2019) questionam o uso equivocado das redes sem escalas, argumentando que em muitos casos a utilização de outros modelos de rede seriam melhor aplicados.

De acordo com Andrade (2018) os sistemas de recomendação como um conjunto de algoritmos que utilizam técnicas de aprendizado de máquina e recuperação de informação gerando recomendações baseadas em algum tipo de filtragem, sendo as mais comuns, a colaborativa e a baseada em conteúdo. Bonin (2018) descreve os sistemas baseados em conteúdo como tendo foco nas propriedades dos itens, onde a similaridade será determinada a partir das semelhanças entre as propriedades dos itens. Já no sistema de filtragem colaborativa o foco se dá na relação entre os usuários e os itens, a similaridade é determinada pelos grupos de usuários que visualizam, compram, ou avaliam os mesmos itens.

#### REFERÊNCIAS

ANDRADE, Michelle H. S. **Entenda como funcionam os Sistemas de Recomendação**. [2017]. Disponível em: <https://www.igti.com.br/blog/como-funcionam-os-sistemas-de-recomendacao/>. Acesso em: 7 out. 2020.

BARABÁSI, A.-L. **Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life**, Plume Books, 2003.

BONIN, Morvana. **O que são Sistemas de Recomendação? Veja exemplos**. [2018]. Disponível em: <https://king.host/blog/2018/09/o-que-sao-sistemas-de-recomendacao/#:~:text=Sistemas%20de%20recomendação%20são%20técnicas,escutar%20ou%20quais%20notícias%20ler>. Acesso em: 25 set. 2020.

BRITO, Carina. **E-commerce cresce em 2020 impulsionado pela pandemia do coronavírus**. [2020]. Disponível em: <https://revistapegn.globo.com/Banco-de-ideias/Varejo/noticia/2020/08/e-commerce-cresce-em-2020-impulsionado-pela-pandemia-do-coronavirus.html>. Acesso em: 26 set. 2020.

BROIDO, Anna D.; CLAUSET, Aaron. Scale-free networks are rare. **Nature Communications**, [S.L.], v. 10, n. 1, p. 1-10, 4 mar. 2019. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41467-019-08746-5>. Disponível em: <https://www.nature.com/articles/s41467-019-08746-5>. Acesso em: 11 out. 2020.

CARVALHO, Maria A. A. S. **Recomendação baseada em modularidade**. 2016. 126f. Tese (Doutorado em Ciência da Computação) - Curso de Pós-Graduação em Ciência da Computação, Universidade Federal de Pernambuco, Recife.

LUNARDI, Guilherme. **12 dados que comprovam o crescimento do e-commerce no Brasil**. [2018]. Disponível em: <https://www.e-commercebrasil.com.br/artigos/12-dados-que-comprovam-o-crescimento-do-e-commerce-no-brasil/>. Acesso em: 25 set. 2020.

NAHAL, Jessica. **Introdução a redes complexas**. [2014]. Disponível em: <https://blog.dp6.com.br/introdução-a-redes-complexas-df73b623d67f>. Acesso em: 2 out. 2020.

RIBEIRO, Larissa F. **Redes sem Escala Típica: Visão Geral, Modelos Alternativos e Técnicas Computacionais**. 2017. 93f. Dissertação (Mestrado em Física) - Curso de Pós-Graduação em Física, Universidade Federal do Rio Grande do Norte, Natal.

SANTOS, Rafael Joseph Pagliuca. **Análise da rede de produtos comprados em conjunto no comércio eletrônico**. 2019. 130f. Dissertação (Mestrado em Ciência da Computação e Matemática Computacional) - Curso de Pós-Graduação em Ciência da Computação, Universidade de São Paulo, São Paulo.

SCHNAIDER, Amanda. **E-commerce cresce 47%, maior alta em 20 anos**. [2020]. Disponível em: <https://www.meioemensagem.com.br/home/marketing/2020/08/27/e-commerce-cresce-47-maior-alta-em-20-anos.html>. Acesso em: 21 set. 2020.

SILVA, Natai B. **Recomendação de relacionamentos em redes sociais baseada em grafos**. 2010. 94f. Dissertação (Mestrado em Ciência da Computação) - Curso de Pós-Graduação em Ciência da Computação, Universidade Federal de Pernambuco, Recife.

Comentado [AS10]: Referência não encontrada no texto.

TERRA, Tiago. **Comércio eletrônico cresce com pandemia do COVID-19, diz ABComm**. [2020]. Disponível em: <https://www.mundodomarketing.com.br/ultimas-noticias/38602/comercio-eletronico-cresce-com-pandemia-do-covid-19-diz-abcomm.html>. Acesso em: 21 set. 2020.

Zhong et al. Study on Directed Trust Graph Based Recommendation for E-commerce System. **International Journal of Computers, Communications and Control**. v. 9, n. 4, p. 510-523, jun. 2014.

Formatado: Inglês (Estados Unidos)



ASSINATURAS

(Atenção: todas as folhas devem estar rubricadas)

Assinatura do(a) Aluno(a): \_\_\_\_\_

Assinatura do(a) Orientador(a): \_\_\_\_\_

Assinatura do(a) Coorientador(a) (se houver): \_\_\_\_\_

Observações do orientador em relação a itens não atendidos do pré-projeto (se houver):

# FORMULÁRIO DE AVALIAÇÃO – PROFESSOR TCC I

Acadêmico(a): \_\_\_\_\_

Avaliador(a): \_\_\_\_\_

| ASPECTOS AVALIADOS <sup>1</sup> |  | atende | atende parcialmente | não atende |
|---------------------------------|--|--------|---------------------|------------|
| ASPECTOS TÉCNICOS               | 1. INTRODUÇÃO<br>O tema de pesquisa está devidamente contextualizado/delimitado?   | x      |                     |            |
|                                 | O problema está claramente formulado?  | x      |                     |            |
|                                 | 2. OBJETIVOS<br>O objetivo principal está claramente definido e é passível de ser alcançado?   |        | x                   |            |
|                                 | Os objetivos específicos são coerentes com o objetivo principal?   |        | X                   |            |
|                                 | 3. JUSTIFICATIVA<br>São apresentados argumentos científicos, técnicos ou metodológicos que justificam a proposta?  | x      |                     |            |
|                                 | São apresentadas as contribuições teóricas, práticas ou sociais que justificam a proposta?   |        | x                   |            |
| ASPECTOS METODOLÓGICOS          | 4. METODOLOGIA<br>Foram relacionadas todas as etapas necessárias para o desenvolvimento do TCC?  | x      |                     |            |
|                                 | Os métodos, recursos e o cronograma estão devidamente apresentados?  | x      |                     |            |
|                                 | 5. REVISÃO BIBLIOGRÁFICA (atenção para a diferença de conteúdo entre projeto e pré-projeto)<br>Os assuntos apresentados são suficientes e têm relação com o tema do TCC? | x      |                     |            |
|                                 | 6. LINGUAGEM USADA (redação)<br>O texto completo é coerente e redigido corretamente em língua portuguesa, usando linguagem formal/científica?                            | x      |                     |            |
|                                 | A exposição do assunto é ordenada (as ideias estão bem encadeadas e a linguagem utilizada é clara)?  | x      |                     |            |
|                                 | 7. ORGANIZAÇÃO E APRESENTAÇÃO GRÁFICA DO TEXTO<br>A organização e apresentação dos capítulos, seções, subseções e parágrafos estão de acordo com o modelo estabelecido?  | x      |                     |            |
|                                 | 8. ILUSTRAÇÕES (figuras, quadros, tabelas)<br>As ilustrações são legíveis e obedecem às normas da ABNT?  |        | x                   |            |
|                                 | 9. REFERÊNCIAS E CITAÇÕES<br>As referências obedecem às normas da ABNT?  | x      |                     |            |
|                                 | As citações obedecem às normas da ABNT?  | x      |                     |            |
|                                 | Todos os documentos citados foram referenciados e vice-versa, isto é, as citações e referências são consistentes?  |        | x                   |            |

## PARECER – PROFESSOR DE TCC I OU COORDENADOR DE TCC (PREENCHER APENAS NO PROJETO):

O projeto de TCC será reprovado se:

- qualquer um dos itens tiver resposta NÃO ATENDE;
- pelo menos 4 (quatro) itens dos **ASPECTOS TÉCNICOS** tiverem resposta ATENDE PARCIALMENTE; ou
- pelo menos 4 (quatro) itens dos **ASPECTOS METODOLÓGICOS** tiverem resposta ATENDE PARCIALMENTE.

**PARECER:** ( ) APROVADO ( ) REPROVADO

Assinatura: \_\_\_\_\_ Data: \_\_\_\_\_

<sup>1</sup> Quando o avaliador marcar algum item como atende parcialmente ou não atende, deve obrigatoriamente indicar os motivos no texto, para que o aluno saiba o porquê da avaliação.