

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Utilização de Redes Bipartidas para criação de um sistema de recomendação de produtos financeiros

Gabriela Passos

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Gabriela Passos

Utilização de Redes Bipartidas para criação de um sistema de recomendação de produtos financeiros

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestra – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria.
VERSÃO REVISADA

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. Francisco Aparecido Rodrigues

USP – São Carlos
Junho de 2020

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

P289u PASSOS, GABRIELA
Utilização de Redes Bipartidas para criação de um
sistema de recomendação de produtos financeiros /
GABRIELA PASSOS; orientador FRANCISCO APARECIDO
RODRIGUES. -- São Carlos, 2020.
56 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Mestrado Profissional em Matemática, Estatística
e Computação Aplicadas à Indústria) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2020.

1. . I. RODRIGUES, FRANCISCO APARECIDO, orient.
II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Gabriela Passos

**Use of Bipartite Networks to create a recommendation
system for financial products**

Dissertation submitted to the Institute of Mathematics
and Computer Sciences – ICMC-USP – in accordance
with the requirements of the Professional Master's
Program in Mathematics Statistics and Computing
Applied to Industry, for the degree of Master in Science.
FINAL VERSION

Concentration Area: Mathematics, Statistics and
Computing

Advisor: Prof. Dr. Francisco Aparecido Rodrigues

USP – São Carlos
June 2020

À todos que dedicam sua carreira em prol da ciência e tecnologia.

AGRADECIMENTOS

Primeiramente eu agradeço ao Prof. Dr. Francisco Aparecido Rodrigues pela orientação, sendo sempre muito paciente e otimista em relação a tudo. À Profa. Dra. Cibele Russo pela amizade, pela força e conselhos. À Universidade de São Paulo pela formação.

À minha mãe, por sempre ficar deslumbrada e orgulhosa com meus estudos. Aos meus irmãos Matheus e Heloísa pela existência.

Aos amigos da Yoga que me ajudavam a descontraír e meditar todas as semanas, aliviando todo o estresse.

As amigas da graduação, que sempre que podíamos estávamos juntas para rirmos e nos apoiarmos. Aos colegas de turma do MECAl por todos os momentos de alegria e desesperos que passamos juntos.

À Anita que mesmo muito longe, estava ali toda semana para rirmos das nossas escolhas, brindarmos nossas conquistas e nos apoiarmos nas derrotas. As amigas que a Natura me deu por toda a sabedoria, crescimento profissional e descontrações durante todos os momentos.

Não poderia ficar de fora dessa lista meus primos Thiago, Caio e Davi que estão sempre compartilhando experiências.

Agradeço também aos gestores que passaram em minha vida profissional que me ajudaram no meu crescimento, em especial ao Francisco Pigato que me incentivou no início dessa jornada e ao Tácio Valente pela confiança.

Por último, mas nunca menos importante, ao Vinícius que foi fundamental para a conclusão dessa etapa, que estava comigo em todos os momentos, que segurou todas as pontas, que me transmitiu paciência e que fez meus dias mais felizes e descomplicados, uma grande parte dessa dissertação também é dele.

“Eu entrego, confio, aceito e agradeço.”

Hermógenes

RESUMO

PASSOS, G. **Utilização de Redes Bipartidas para criação de um sistema de recomendação de produtos financeiros**. 2020. 56 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Com a nova era digital, nota-se que a troca de informação é praticamente instantânea, tornando cada vez mais imprevisível o padrão de comportamento dos clientes. Desta forma, os sistemas de recomendação são fortemente utilizados no mundo, pois entender o cliente e ofertar experiências cada vez mais relevantes faz-se necessário. Com isso, surge a demanda de se criar um algoritmo eficiente que lide corretamente com os novos comportamentos de consumo.

Uma maneira de direcionar efetivamente os produtos e serviços das empresas é o sistema de recomendação, que ajuda a melhorar o potencial de consumo e com isso a receita. Estes são também interessantes para o cliente, pois eles apresentam possibilidades baseadas em seus interesses, diminuindo assim o número de objetos indesejados a serem ofertados. Uma forma eficiente de se construir um sistema é encontrando os padrões de consumo dos clientes e desse modo identificar os interesses corretos de cada consumidor.

Um conjunto com usuários e objetos, onde cada usuário coleta/compra um certo objeto, é o necessário para se criar um sistema de recomendação. A teoria de redes complexas pode ser utilizada para descrever e modelar tais sistemas. Em particular, para sistemas de recomendação é comum utilizar projeções de redes bipartidas para resolver o problema. Nessa dissertação, serão abordadas formas de recomendação de produtos utilizando uma projeções assimétrica.

Com a criação de um sistema de recomendação com a projeção de rede bipartida é possível comparar o método com outras formas de recomendação, como uma recomendação aleatória e a regra de associação. Quando o sistema é testado em momentos distintos é possível verificar que as recomendações feitas através da projeção possuem uma acurácia melhor que os outros sistemas.

Palavras-chave: Redes Complexas, Redes Bipartidas, Projeção de redes bipartidas, Recomendação.

ABSTRACT

PASSOS, G. **Use of Bipartite Networks to create a recommendation system for financial products**. 2020. 56 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

In this digital era, information exchange time is seen as almost instantaneous, which in turn makes the behavior patterns relatively unpredictable. This way, recommending systems are being more and more used in business to provide a better understanding of each customer and offer each one of them a more relevant experience. There is a demand, then, for an algorithm that can deal correctly with the new consumer patterns.

A new way to effectively direct products and services from companies to customers is to utilize a system, which helps to improve upselling and income. These systems are also interesting from a customer perspective, as they present possibilities based on previous interests, diminishing the suggestion of unwanted objects. Efficient systems identify consuming patterns and make a suggestion based on them for each customer individually.

A set of users and objects, where each user collects / buys a given object, is a recommendation system. Complex networks theory can be used to describe and model such systems. In particular, for recommending systems it is common to utilize bipartite networks to solve this problem. In this work, recommending products using bipartite networks will be studied and applied.

After designing and implementing a recommending system using projections on a bipartite network, it is possible to compare this method with other types of recommenders, such as random recommenders and association rule. When the system is tested in distinct moments, it is verifiable that the recommendations made using bipartite network projections have a better accuracy than the others.

Keywords: Complex Networks, bipartite networks, bipartite network projection, recommendation.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação de uma rede bipartida. Fonte: Figura elaborada pela autora. .	22
Figura 2 – Distribuição da quantidade de itens por cesta, sendo ramos de atividade à esquerda e produtos financeiros à direita. Fonte: Figura elaborada pela autora a partir de dados reais.	24
Figura 3 – Representação de amostras de redes reais analisadas nesse trabalho. Rede de produtos financeiros (à esquerda) e rede de ramos de atividade (à direita). Fonte: Figura elaborada pela autora a partir de dados reais.	25
Figura 4 – Representação de um grafo simples com sua respectiva matriz adjacente. Fonte: Figura elaborada pela autora.	27
Figura 5 – Representação de um grafo simples. Os círculos representam os nós do grafo e as linhas (arestas) ligando os círculos representam as conexões. Fonte: Figura elaborada pela autora.	28
Figura 6 – Representação da cidade de Königsberg e as 7 pontes que ligavam a cidade. Fonte: Charge de Simon Kneebone, sobre as pontes de Königsberg.	29
Figura 7 – Representação de um grafo simples onde ilustramos o grau (k) e o grau médios dos vizinhos (knn). Os valores nesse caso para o vértice 1 (em verde) são $k_1 = 3$. Fonte: Figura elaborada pela autora.	31
Figura 8 – Ilustração de um grafo bipartido de relação entre 6 crianças e 3 festas. Fonte (WASSERMAN; FAUST, 1994), página 301.	32
Figura 9 – Rede bipartida ponderada. Fonte: (VALEJO <i>et al.</i> , 2017)	32
Figura 10 – Exemplo de uma rede bipartida (1) e a representação da projeção em U (2) e em O (3). Fonte: Adaptação de (ZHOU <i>et al.</i> , 2007)	33
Figura 11 – Representação de sistema de recomendação por filtragem colaborativa. Fonte: (ROLIM <i>et al.</i> , 2017)	37
Figura 12 – Ilustração do processo de alocação de recursos em rede bipartida. Os três superiores são nós O e os quatro inferiores são nós U. O processo consiste em: primeiro, o recurso vai de O para U ($1 \rightarrow 2$) e, em seguida, retorna para O ($2 \rightarrow 3$). Fonte: Adaptação de (ZHOU <i>et al.</i> , 2007).	39
Figura 13 – Distribuição de itens por cesta: à esquerda a distribuição dos produtos financeiros e à direita a distribuição por ramos de atividade. Fonte: Figura elaborada pela autora a partir de dados reais.	42
Figura 14 – Distribuição da quantidade de itens por cesta de produtos financeiros. Fonte: Figura elaborada pela autora a partir de dados reais.	43

Figura 15 – Distribuição de ramos de atividade por safra. Fonte: Figura elaborada pela autora.	44
Figura 16 – Distribuição de clientes com a quantidade de ramos por safra. Fonte: Figura elaborada pela autora.	44
Figura 17 – Sistema RBP: Distribuição da primeira recomendação de ramo de atividade por safra. Fonte: Figura elaborada pela autora.	46
Figura 18 – Sistema SA: Distribuição da primeira recomendação de ramo de atividade por safra. Fonte: Figura elaborada pela autora.	46
Figura 19 – Sistema SF: Distribuição da frequência de ramo de atividade por safra. Fonte: Figura elaborada pela autora.	47
Figura 20 – Comparação das simulações: à esquerda temos a comparação dos 3 sistemas por safra e à direita o resultado geral. Fonte: Figura elaborada pela autora. .	47
Figura 21 – Sistema RBP: Distribuição da frequência de produtos recomendados. Fonte: Figura elaborada pela autora.	48
Figura 22 – Sistema SF: Distribuição da frequência de produtos recomendados. Fonte: Figura elaborada pela autora.	48
Figura 23 – Sistema SA: Distribuição da frequência de produtos recomendados. Fonte: Figura elaborada pela autora.	48
Figura 24 – Comparação das simulações. Fonte: Figura elaborada pela autora.	49

LISTA DE TABELAS

Tabela 1 – Representação de uma matriz entre uma rede de crianças e festas	31
Tabela 2 – Base de dados com 3 itens e 5 transações	36
Tabela 3 – Base de dados	42
Tabela 4 – Métricas de transação de clientes por safra	43
Tabela 5 – Métricas de transação de clientes por safra	45

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Objetivo	23
1.2	Contribuições	23
1.3	Organização dos capítulos	25
2	CONCEITOS E MÉTODOS	27
2.1	Breve histórico	29
2.2	Principais métricas de redes	30
2.2.1	<i>Grau</i>	30
2.3	Redes Bipartidas	30
2.3.1	<i>Projeção simples</i>	33
3	SISTEMA DE RECOMENDAÇÃO	35
3.1	Regra de associação	35
3.2	Filtragem colaborativa	36
3.3	Regra por frequência	37
3.4	Recomendação por Redes Bipartidas	38
3.4.1	<i>Projeção ponderada</i>	38
4	RESULTADOS E DISCUSSÕES	41
4.1	Descritiva dos dados reais	41
4.2	Simulação dos sistemas recomendação	44
4.2.1	<i>Recomendação de ramos de atividade</i>	45
4.2.2	<i>Recomendação de produtos/serviços</i>	46
5	CONCLUSÃO E PERSPECTIVAS FUTURAS	51
	REFERÊNCIAS	53

INTRODUÇÃO

Nas últimas décadas, temos observado uma popularização do uso de mídias digitais e socais, o que tem gerado uma quantidade de informação maior do que indivíduos podem processar e assimilar, gerando o conceito *information overload*. Com isso, os sistemas de recomendação têm se tornado cada vez mais necessários, como em recomendação de livros (LINDEN; SMITH; YORK, 2003), filmes (MILLER *et al.*, 2003), vendas e assim por diante. Pensando nesse dinamismo de sempre oferecer para o clientes experiências cada vez mais atrativas, os sistemas de recomendação são usados para sugerir produtos mais adequados para os clientes. Nesse trabalho, consideramos de um sistema de recomendação de produtos financeiros, de modo a auxiliar clientes na tomada de decisões com respeito às suas necessidades.

Podemos separar os métodos de recomendação disponíveis em duas categorias distintas: (i) métodos qualitativos e (ii) quantitativos, cada qual com diferentes graus de exatidão relativa em previsões de longo e curto prazos, conforme o nível de sofisticação utilizado e a base lógica (dados históricos, opiniões de especialistas, ou estudos) da qual a recomendação é derivada (BALLOU, 2009).

Os métodos quantitativos baseiam-se em pesquisas ou técnicas comparativas com o objetivo de produzir estimativas objetivas. Ballou et al. (BALLOU, 2009) ressalta ainda, que as informações referentes aos fatores que moldam o resultado de uma previsão quantitativa não são flexíveis, tampouco subjetivas. A opção de utilização dos métodos qualitativos pelos responsáveis em definir as estratégias de competitividade e previsão, dá-se, em geral, ou pela falta de informação técnica e familiaridade com os métodos quantitativos, ou pela sensação de controle sobre o processo preditivo (SANDERS; MANRODT, 1994).

Como as empresas possuem muitos dados de transações e cadastros de clientes, foi possível criar uma metodologia, voltada para o marketing, que permite um profundo conhecimento entre o cliente e possibilidade do mercado. Esse conceito é conhecido como *Customer Relationship Management*, ou CRM, que relaciona a necessidade de cada cliente, mostrando quais são os

melhores serviços e produtos (ZENONE, 2007).

Geralmente, para isso é utilizada uma técnica, baseada em regra de negócio ou um modelo matemático (RESNICK; VARIAN, 1997). Nesse trabalho em especial trataremos de recomendação de produtos e ramos financeiros (seguros, empréstimos, ramos de lojas, etc) através de redes complexas. Redes complexas são um tipo de estrutura de dados. Como exemplos de redes complexas podemos citar redes sociais, interações biológicas e o cérebro humano. No caso de sistemas de recomendação, podemos usar redes complexas.

Quando uma rede pode ser dividida em dois subconjuntos, esse novo grafo é chamado de bipartido. Muitos sistemas podem ser modelados como redes bipartidas (HOLME *et al.*, 2003), como por exemplo a rede metabólica (JEONG *et al.*, 2000) que é composta de substâncias químicas e reações químicas, rede de citação entre pesquisadores e trabalhos de pesquisa (GUSTAFSSON; HANCOCK; CÔTÉ, 2014), rede de cliente-produto para um sistema de recomendação (LOPS; GEMMIS; SEMERARO, 2011), etc.

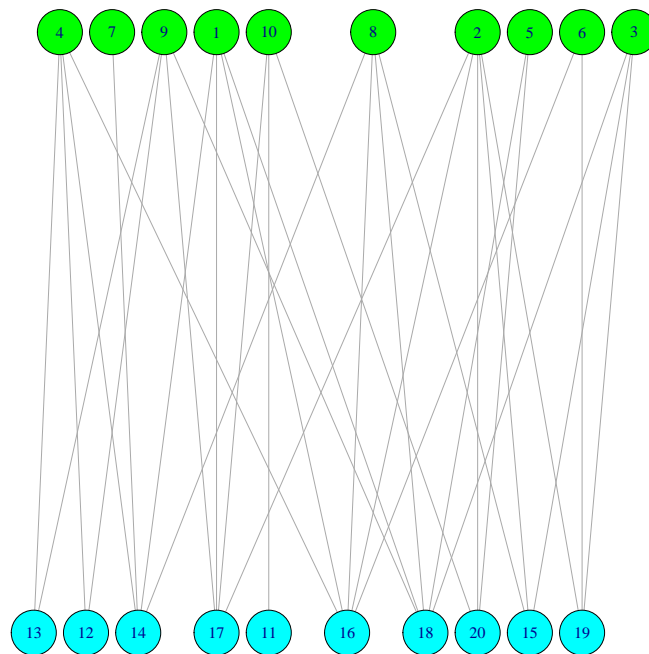


Figura 1 – Representação de uma rede bipartida. Fonte: Figura elaborada pela autora.

Trabalhos recentes analisaram (LIND; GONZALEZ; HERRMANN, 2005)-(ESTRADA; RODRIGUEZ-VELAZQUEZ, 2005) e modelaram (OHKUBO; TANAKA; HORIGUCHI, 2005) redes bipartidas. Para representar diretamente as relações entre os dois conjuntos de vértices na rede, a rede bipartida é comprimida pela sua projeção (ZHOU *et al.*, 2007), sendo a representação mais simples a forma não ponderada (NEWMAN, 2001)-(BARABÁSI *et al.*, 2002), ou seja, sem considerar a frequência com que conexões entre pares de vértices ocorrem. No entanto, para melhor representar a estrutura da rede, já que a projeção é sempre menos informativa do que a rede bipartida original, é necessário usar-se o grafo bipartido considerando os pesos na projeção. Uma maneira direta para realizar essa tarefa é ponderar diretamente as conexões pelo número de vezes que os usuários coletam o mesmo objeto (LI *et al.*, 2007). Esses pesos, mesmo que simples de calcular, são muito mais informativos do que uma projeção não ponderada.

1.1 Objetivo

Nesse dinamismo de melhores serviços e personalização da oferta, um sistema de recomendação é altamente necessário. E, este consiste em um grupo de usuários e objetos, onde cada usuário coletou/comprou alguns objetos. Se os usuários só puderem comprar objetos (sem classificá-los), o sistema de recomendações pode ser descrito por uma matriz de adjacência a_{ij} ($n \times m$, n usuários por m objetos). Uma suposição razoável é que os objetos que foram comprados são os que os usuários gostam/precisam, e um algoritmo de recomendação tem como objetivo prever as opiniões pessoais do usuário em relação a objetos ainda não visualizados por ele. Geralmente, o maior interesse é verificar se o usuário gostaria ou não do produto em questão. (ZHOU *et al.*, 2007).

Observando todas essas possibilidades através das projeções de redes bipartidas estamos propondo trazer essa análise para o universo corporativo, onde muito pouco se utiliza de redes complexas. Percebendo essa oportunidade de mostrar que pode-se criar sistemas de recomendação de produtos através de outras metodologias foi o que motivou esse trabalho. Nesse trabalho, analisamos a construção de um sistema de recomendação para clientes de uma instituição de produtos financeiros, através de projeções de redes bipartidas (ZHOU *et al.*, 2007) e comparamos esse sistema com outros métodos de recomendação, como regra de associação (AGRAWAL RAKESH; SWAMI, 1993) ou recomendação aleatória.

1.2 Contribuições

A escolha de um produto/serviço em uma gama de alternativas é uma dificuldade para o cliente. O trabalho nas empresas com grande quantidade de informação, permitiu uma evolução computacional das análises, proporcionando recomendações mais precisas que uma recomendação humana (RESNICK; VARIAN, 1997). Por essa razão, as empresas estão cada vez mais

utilizando os sistemas de recomendação como tecnologia para sanar o problema relacionado com as escolhas.

Em instituições financeiras menores, essas recomendações acabam sendo realizadas de forma aleatória. Com isso, muitas pessoas engajam-se apenas em único produto. A dificuldade de entender o produto como um todo, ou até mesmo a falta de familiaridade com o processo, gera grandes números de clientes que não usufruem de todas as oportunidades de serviço disponibilizadas. Quando pensa-se em instituição financeira, um dos primeiros produtos que vem a mente é o cartão de crédito. Um cliente engajado no produto é aquele que usa o cartão em todos os momentos, para qualquer tipo de compra. Dessa forma, a medida que ele experimenta novos ramos de consumo, mais familiarizado e satisfeito ele está com o produto. Por isso, grande parte do início do ciclo de vida do cliente é voltado para o seu aprendizado e rentabilização, onde a empresa explica para ele todas as possibilidades que o produto oferece, bem como estimula-o para que se gaste cada vez mais e com mais frequência utilizando o cartão. Outro produto menos trivial de se pensar, mas que existe também em instituições financeiras, são os seguros ou serviços que o cliente pode contratar, sendo este mais difícil de introduzir o cliente, pois geralmente é aceito quando há real necessidade do serviço. Porém, é de suma importância para a empresa que cada vez mais produtos sejam contratados e vistos como essenciais e importantes para o cliente.

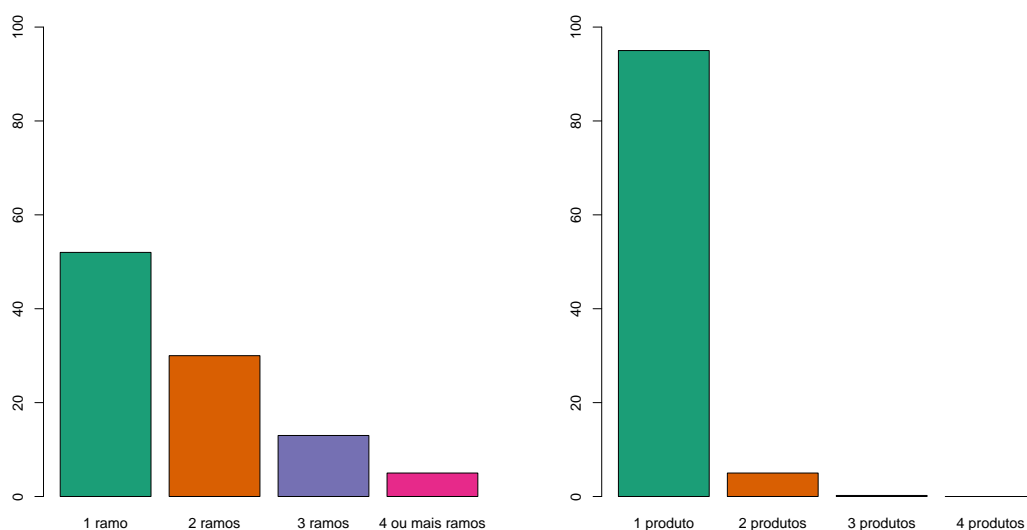


Figura 2 – Distribuição da quantidade de itens por cesta, sendo ramos de atividade à esquerda e produtos financeiros à direita. Fonte: Figura elaborada pela autora a partir de dados reais.

Na empresa analisada nesse trabalho, tanto aumentar a experiência com o cartão de crédito quando a venda de serviços é um grande desafio, pois quando analisa-se a primeira semana de experiência com o produto que pode ser usado em 10 diferentes ramos, aproximadamente 50% das transações na primeira semana foi em apenas 1 ramo de atividade (Figura 2 - gráfico à esquerda). Já quando estuda-se os produtos financeiros, 95% dos clientes possuem apenas um

produto do catálogo (Figura 2 - gráfico à direita), o que é um resultado esperado, dado que o catálogo é restrito e o desafio é conseguir que o cliente experimente mais o produto e conheça outros produtos do catálogo.

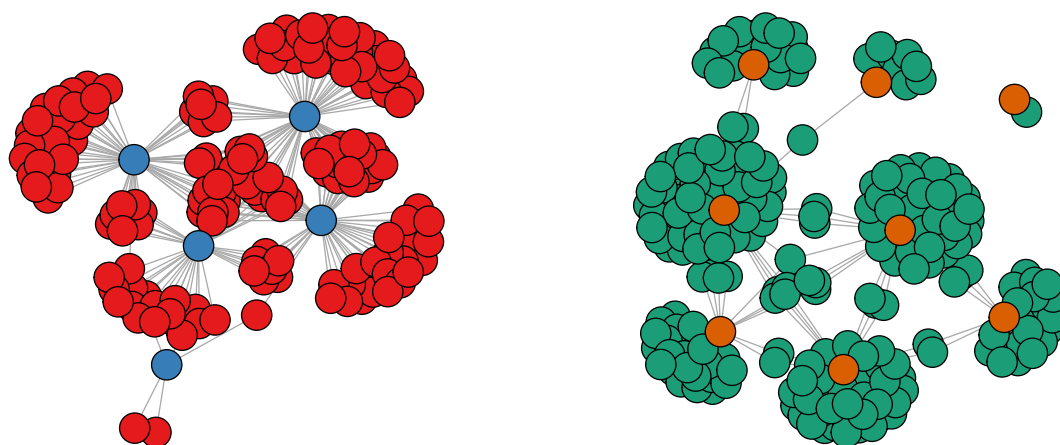


Figura 3 – Representação de amostras de redes reais analisadas nesse trabalho. Rede de produtos financeiros (à esquerda) e rede de ramos de atividade (à direita). Fonte: Figura elaborada pela autora a partir de dados reais.

Pensando na melhoria da venda dos produtos e sua maior distribuição, propomos um sistema de recomendação através da metodologia de redes complexas, mas especificamente redes bipartidas (Figura 3). Serão simulados sistemas de diferentes tipos para comparar-se com o método proposto, como sistemas de formas aleatória e através de regras de frequência.

1.3 Organização dos capítulos

A organização desta dissertação está da seguinte forma: No capítulo 2 iremos descrever conceitos sobre redes complexas, redes bipartidas e suas propriedades. A importância e os diferentes tipos de recomendação são abordados no Capítulo 3. No Capítulo 4 serão apresentados e discutidos os resultados dos processos descritos nos capítulos anteriores, assim como o desenho e forma de simulação do experimento. Por fim, no Capítulo 5 serão apresentadas as conclusões das análises que foram desenvolvidas e as perspectivas de melhoras do modelo apresentado.

CONCEITOS E MÉTODOS

Redes complexas são formadas por elementos discretos, chamados vértices, que se conectam por arestas. Ou seja, a estrutura de uma rede pode ser representada por um grafo. Na Figura 5 é apresentado um grafo simples, onde os círculos são os nós e as linhas, as arestas. O grafo pode ser representado por uma matriz de adjacência A_{uo} , sendo que um elemento $A_{uo} = 1$ se existe ligação entre os vértices u e o , ou $A_{uo} = 0$, caso contrário. A Figura 4 ilustra um grafo e sua respectiva matriz de adjacência. Exemplos de redes complexas incluem redes sociais, interações biológicas, e nosso cérebro.

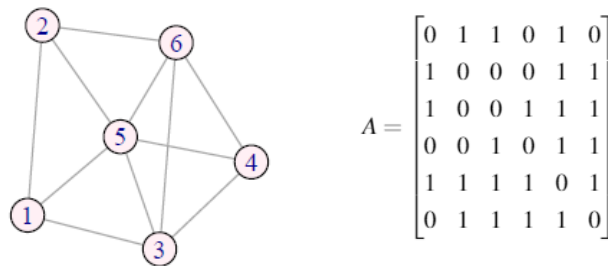


Figura 4 – Representação de um grafo simples com sua respectiva matriz adjacente. Fonte: Figura elaborada pela autora.

Motivados por entender os diversos tipos de sistema complexos reais o estudo de redes foi expandindo-se, conjuntamente com a grande quantidade de base de dados sendo gerados em diversos meios, tais como internet, dados médicos e de engenharia. A Internet talvez seja a rede complexa mais popular. Esta rede é constituída por roteadores ou sistemas autônomos,

que representam os vértices, ligados por conexões físicas, que representam as arestas (PASTOR-SATORRAS; VÁZQUEZ; VESPIGNANI, 2001), (FALOUTSOS; FALOUTSOS; FALOUTSOS, 1999). Padrões de amizades entre indivíduos (RAPOPORT; HORVATH, 1961), relações entre empresas (DAVIS; GREVE, 1997), e redes de contatos sexuais (LILJEROS *et al.*, 2001) são típicos exemplos de conexões de pessoas ou grupos de indivíduos dispostos em uma rede social.

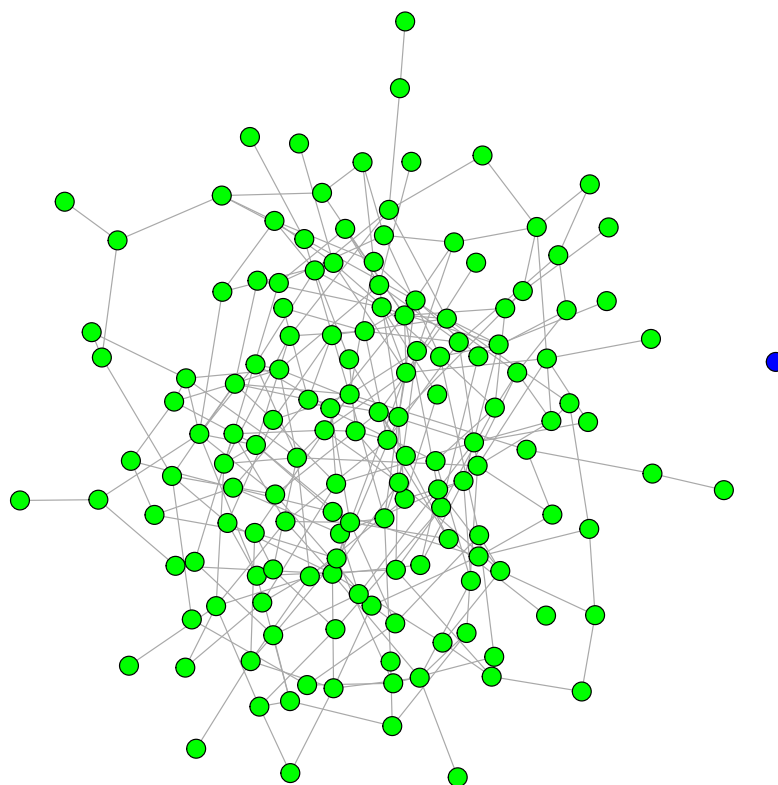


Figura 5 – Representação de um grafo simples. Os círculos representam os nós do grafo e as linhas (arestas) ligando os círculos representam as conexões. Fonte: Figura elaborada pela autora.

2.1 Breve histórico

A teoria dos grafos surgiu em 1736 quando Leonhard Euler resolveu o problema das sete Pontes de Königsberg, que consistia em sanar uma dúvida que os moradores da cidade levantaram: existia a possibilidade de um morador em uma determinada ilha ou margem do rio, atravessar todas as pontes uma única vez e voltar para a mesma parte da cidade? Para isso, é importante entender que a cidade de Königsberg era dividida por um rio e havia um total de sete pontes ligando os diferentes pontos da cidade como ilustrado na Figura 6. Tal problema permaneceu sem solução até o ano de 1736, quando o matemático Leonard Euler demonstrou que tal travessia seria impossível. A estrutura matemática usada por Euler para resolver o problema ficou conhecida como grafo. Então, em 1736 a área dos grafos surgiu, a partir de uma curiosidade sobre o trabalho de Euler é que ele acabou encontrando problemas reais que dependiam apenas do padrão das conexões entre seus elementos para serem resolvidos (ESTRADA, 2012).

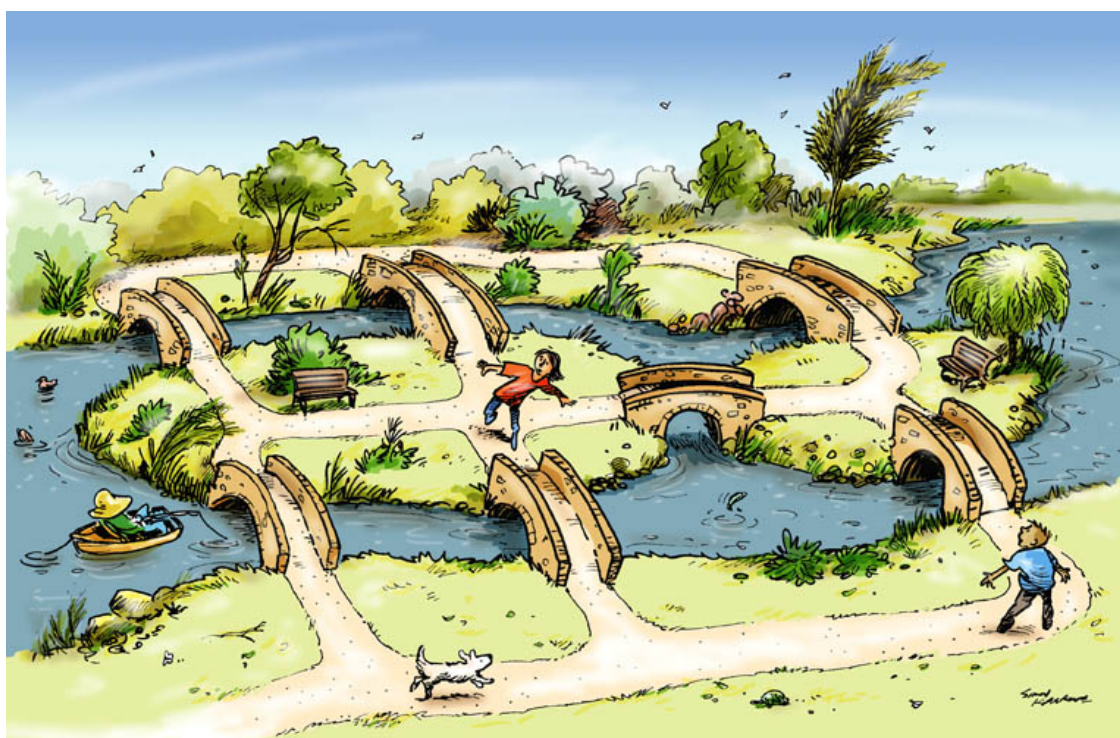


Figura 6 – Representação da cidade de Königsberg e as 7 pontes que ligavam a cidade. Fonte: Charge de Simon Kneebone, sobre as pontes de Königsberg.

Tempos depois, muitos trabalhos foram realizados, onde a topologia dos grafos foram exploradas e novos conceitos introduzidos, assim como diferentes tipos de grafos (BIGGS; LLOYD; WILSON, 1986). O estudo dos grafos aleatórios marcam o início da próxima grande mudança na teoria dos grafos e o modelo mais conhecido foi o de Erdős e Rényi. Um dos aspectos interessantes dos grafos aleatórios é que tais grafos são definidos em termos de probabilidades, sendo que suas propriedades podem ser obtidas com o uso de estatística. Muitos outros estudiosos se dedicaram a estudar grafos, com o aumento da capacidade computacional e a acessibilidade de

dados, foi possível a montagem de redes reais de diversos sistemas. Com essa nova possibilidade, as diferenças das redes aleatórias e regulares ficou clara e por isso, foram criados vários modelos e medidas que possibilitam a observação da complexidade presente nas redes.

2.2 Principais métricas de redes

Na teoria de redes foram desenvolvidas diversas medidas locais e globais. Visualmente a extração de informação de uma rede é muito limitada, por essa razão faz-se necessário algumas medidas topológicas que descrevam as propriedades da rede.

2.2.1 Grau

Uma das medidas mais comuns que quantifica a centralidade de um nó i é o grau, ou seja, mede o número de conexões de cada vértice k_i e dado por:

$$k_i = \sum_{j=1}^N A_{ij} \quad (2.1)$$

Onde $\sum_{j=1}^N A_{ij}$ é a soma dos elementos da linha i e N é o número de vértices. O grau é uma medida fundamental para as comparações de vértices, por exemplo para identificar os nós que possuem muitas conexões, denominados por *hubs*, quanto maior o grau, a importância do vértice é maior, em relação ao número de conexões. Essa medida também permite quantificar outras propriedades da rede, como a densidade de conexões que é medida pelo grau médio e descrita em 2.2 e M é o número de arestas. Já a probabilidade de se selecionar um vértice com grau k , $P(k=k)$ e está descrita em 2.3 onde N_k é o número de vértices com grau k . A Figura 7 ilustra o conceito de grau.

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{1}{N} \underbrace{\sum_{i=1}^N \sum_{j=1}^N A_{ij}}_{2M} \quad (2.2)$$

$$P(k = k) = P(k) = \frac{N_k}{N} \quad (2.3)$$

2.3 Redes Bipartidas

Na teoria dos grafos, um grafo bipartido é quando os vértices podem ser divididos em dois conjuntos disjuntos, ou seja, tem-se um conjunto $U = \{u_1, u_2, \dots, u_m\}$ e $O = \{o_1, o_2, \dots, o_n\}$, onde as ligações ocorrem apenas entre U e O . Para entender melhor a relação entre os objetos no grafo bipartido, (WASSERMAN; FAUST, 1994) descreve de forma hipotética uma festa infantil, onde o grafo é composto por 6 crianças e 3 festas diferentes (Figura 8), nota-se que as linhas

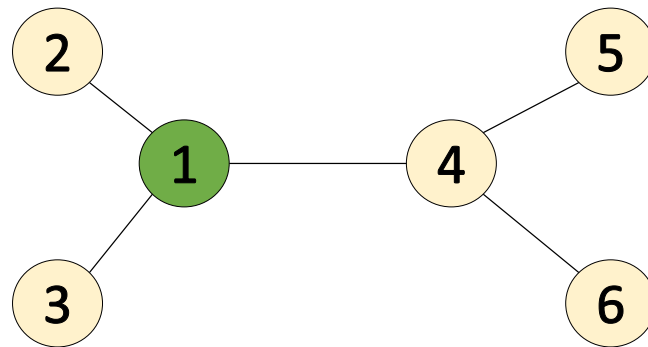


Figura 7 – Representação de um grafo simples onde ilustramos o grau (k) e o grau médios dos vizinhos (k_{nn}). Os valores nesse caso para o vértice 1 (em verde) são $k_1 = 3$. Fonte: Figura elaborada pela autora.

ligam apenas crianças e festas. Um grafo bipartido também pode ser escrito por uma matriz de adjacência, que é uma forma de representar as ligações direcionais ou não entre os usuários usando uma matriz numérica. A matriz é preenchida com 1 na célula (i,j) se a linha da criança é ligada a coluna da festa e 0 caso contrário. Utilizando como exemplo a rede de festas e crianças, a matriz criada seria dada pela Tabela 1, onde as colunas são as festas e as linhas os convidados. Existe uma vantagem de se armazenar os dados dessa forma, pois fica codificado o fato de um elemento ter conexão com um objeto. Também dessa forma é possível verificar a existência de um isolamento na rede (onde um elemento não possui conexão com nenhum objeto), esse formato também ajuda nas análises estatísticas. Uma desvantagem de usar esse formato de dados é que ele pode ocupar muito espaço (DENNY, 2014).

Tabela 1 – Representação de uma matriz entre uma rede de crianças e festas

Convidado	Festa 1	Festa 2	Festa 3
Allison	1	0	1
Drew	0	1	0
Eliot	0	1	1
Keith	0	0	1
Ross	1	1	1
Sarah	1	1	0

Uma rede bipartida sempre pode ser escrita como uma rede unipartida, denominada também como redes homogêneas, onde existe somente um tipo de vértice, ignorando a partição de U . Na literatura, existem métodos de ponderação que permitem descrever de forma realista suas relações topológicas, pretendendo minimizar a perda de informações. Duas redes homogêneas ponderadas ou não podem ser por exemplo, derivadas de uma rede bipartida. Esse processo é denominado de projeção unipartida (CHUERUBIM; SILVA, 2019).

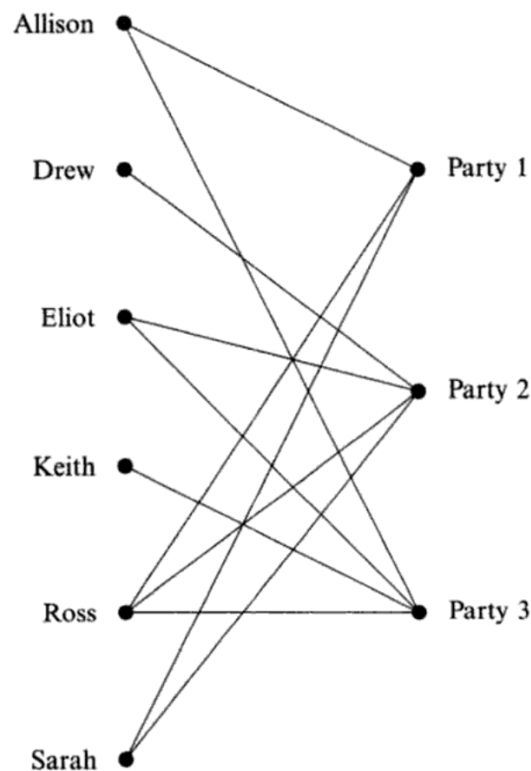


Figura 8 – Ilustração de um grafo bipartido de relação entre 6 crianças e 3 festas. Fonte ([WASSERMAN; FAUST, 1994](#)), página 301.

Ao fixar-se dois vértices pode-se analisar o número de vizinhos comuns entre eles, bem como a soma dos pesos das arestas entre os vértices e seus respectivos vizinhos em comum ([VALEJO *et al.*, 2017](#)). A Figura 9, ilustra a criação de uma aresta ponderada entre os vértices 4 e 5 com base em seus vizinhos comuns em uma projeção unipartida ponderada.

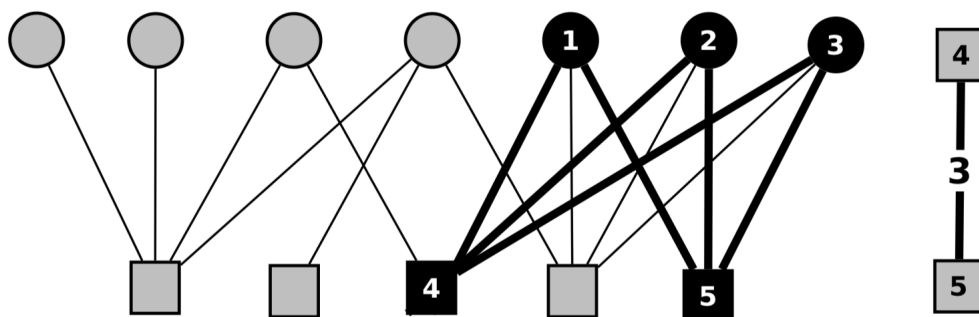


Figura 9 – Rede bipartida ponderada. Fonte: ([VALEJO *et al.*, 2017](#))

Muitos sistemas podem ser descritos como redes bipartidas ([HOLME *et al.*, 2003](#)), como por exemplo a rede metabólica ([JEONG *et al.*, 2000](#)) que é composta de substâncias químicas e

reações químicas, rede de citação entre pesquisadores e trabalhos de pesquisa (GUSTAFSSON; HANCOCK; CÔTÉ, 2014), rede de cliente-produto para um sistema de recomendação (LOPS; GEMMIS; SEMERARO, 2011), entre outros.

Para compactar informações sobre redes bipartidas, o método de projeção é comumente utilizado (ZHOU *et al.*, 2007).

O método mais simples de projeção visa projetar a rede bipartida original em uma rede não ponderada, desconsiderando-se as características da rede bipartida. Com essa desconsideração a projeção perde muitas informações importantes entre as conexões feitas e pode fazer com que estruturas diferentes de redes bipartidas tenham a mesma representação, por isso, utilizar algum método de ponderação faz-se necessário.

2.3.1 Projeção simples

Para enxergar diretamente as relações entre os conjuntos em uma rede bipartida, a rede é comprimida pela sua projeção (ZHOU *et al.*, 2007). A projeção U é a criação de uma rede contendo apenas nós de U, onde a conexão entre os nós de U existe se eles possuem um elemento de O em comum. A Figura 10 - 2 e a Figura 10 - 3 mostram a projeção de U e O respectivamente.

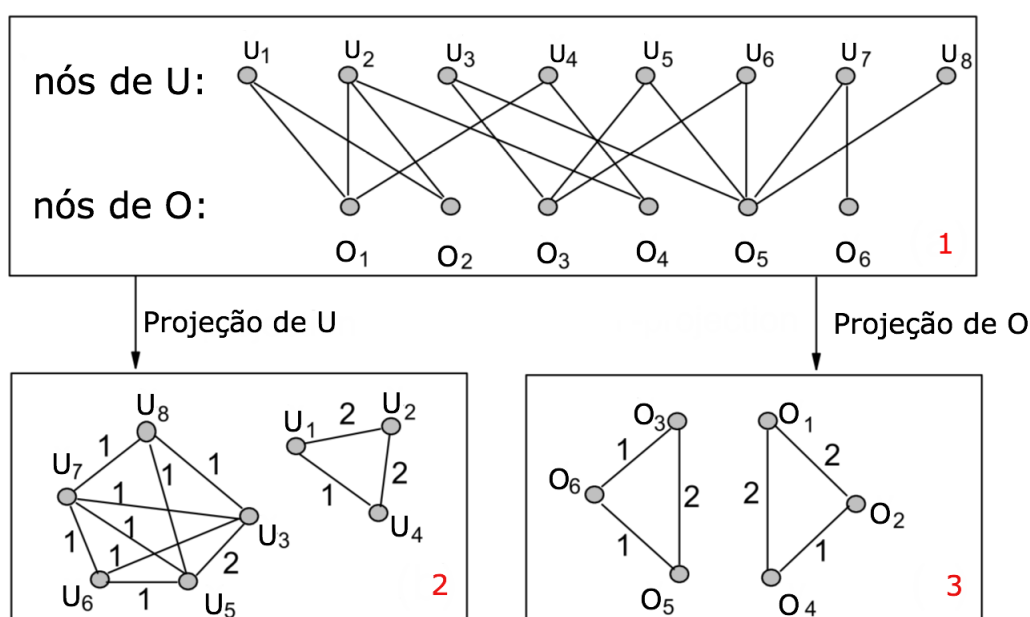


Figura 10 – Exemplo de uma rede bipartida (1) e a representação da projeção em U (2) e em O (3). Fonte: Adaptação de (ZHOU *et al.*, 2007)

A forma mais simples de realizar a projeção de uma rede bipartida é de forma não ponderada (NEWMAN, 2001)-(BARABÁSI *et al.*, 2002), ou seja, não considerar a frequência. A perda de informação nesse tipo de projeção é inevitável, mesmo que algumas propriedades

topológicas possam ser qualitativamente obtidas a partir desta versão não ponderada. Por exemplo, se dois usuários tiverem comprado mais de 50 objetos, e apenas um objeto em comum, conclui-se que os usuários possuem comportamento de compra distintos. No entanto, na projeção de usuários não ponderada, esses dois casos têm exatamente a mesma representação gráfica.

Para melhor refletir a estrutura da rede, já que a projeção é sempre menos informativa do que a rede bipartida original, é necessário se usar o grafo bipartido para quantificar os pesos na projeção. Uma maneira é pesar diretamente pelo número de vezes que os usuários coletaram o mesmo objeto (LI *et al.*, 2007). A Figura 10 - 2 e 3 mostra os pesos dados nas projeções de U e O respectivamente. Esses pesos, mesmo que simples de calcular, são muito mais informativos do que uma projeção não ponderada.

A grande questão é como pesar essas arestas da projeção. A ponderação hiperbólica é uma outra técnica usada para ponderar a projeção, onde em alguns casos, o uso de ponderação simples pode não ser muito esclarecedor. As redes de colaboração científica é um exemplo, pois nessas redes espera-se que dois cientistas cujos nomes aparecem em um documento, juntamente com muitos outros co-autores, se conheçam menos do que dois que eram os únicos autores de um artigo. Dessa forma, foi proposto pesar as arestas inversamente de acordo com o número de afiliações comuns no conjunto vizinho. Isso é mais facilmente alcançado pela introdução de um fator de escala $1/(n - 1)$ na contagem simples, o que enfraquece o link entre nós com correspondências comuns mais populares.

Quando a projeção da rede é feita, pode-se utilizar alguma metodologia para ponderar as arestas de forma que as informações contidas na rede original sejam preservadas. A ponderação simples é um exemplo disso, onde utiliza-se um peso entre as arestas decorrente do número de vezes que a conexão em comum ocorre. Esse método, aplicado na Figura 10 - 2 por exemplo, mostra que os nós U_1 , U_2 e U_4 se conectam com pesos diferentes. Voltando para a Figura 10 - 1, verifica-se que esses mesmos nós possuem os objetos O_1 e O_2 em comum, porém como o U_4 só tem ligação com o O_1 , enquanto os outros nós possuem as duas ligações, os pesos aplicados as arestas são distintos.

Em alguns casos, como mostra (FAN *et al.*, 2007) essa ponderação simples pode não ser a melhor forma de ponderação da projeção da rede bipartida. Por isso, será abordado a seguir um método de projeção ponderado.

SISTEMA DE RECOMENDAÇÃO



No ambiente atual em que vivemos, em que há uma grande quantidade de informação disponível, as recomendações personalizadas tornam-se um método importante para recuperação e descoberta de informação sobre os indivíduos. Quando essas informações se combinam a qualidade da experiência do cliente acaba mais satisfatória (DAVIDSON *et al.*, 2010). Nos últimos anos, muitos métodos de recomendação tem sido propostos, como por exemplo os modelos de agrupamento (BREESE; HECKERMAN; KADIE, 1998) e classificação (BILLSUS; PAZZANI, 1998), filtragem colaborativa ((GOLDBERG *et al.*, 1992)), regras de associação ((LIN; ALVAREZ; RUIZ, 2000)), dentre outros. Uma recomendação é baseada em informações de objetos coletados/comprados pelos indivíduos e a saída do sistema é um conjunto de produtos que o consumidor provavelmente gostaria de ter (SHAYA *et al.*, 2010).

Nesse capítulo vamos descrever alguns dos muitos sistemas de recomendação utilizados atualmente.

3.1 Regra de associação

A Regra de associação (*Association Rules*) é muito conhecida para resolver problemas de análise da cesta de compra, tal técnica baseia-se em reconhecer os itens mais frequentes nas cestas, onde o resultado obtido será como: SE compra cerveja TAMBÉM compra fralda.

Estabelecido por (AGRAWAL RAKESH; SWAMI, 1993), a técnica possui duas etapas: primeiramente cria-se um conjunto de itens frequentes, depois gera-se as regras de associação. Os conceitos de suporte e confiança são utilizados para assegurar que os resultados são válidos. O suporte (3.1) de um conjunto é definido como a proporção de transações da base de dados que contém esse conjunto. A confiança (3.2) é o percentual de transações que possuem exatamente a regra, no nosso exemplo seria todas as transações que contém CERVEJA e FRALDA. Outro conceito utilizado é o *lift* é razão do suporte observado que são esperados se X e Y são

independentes.

$$sup(X) = \frac{|\{y \in Y; X \subseteq y\}|}{|Y|} \quad (3.1)$$

$$Conf(X \Rightarrow Y) = \frac{sup(X \cup Y)}{sup(X)} \quad (3.2)$$

$$Lift(X \Rightarrow Y) = \frac{sup(X \cup Y)}{sup(X).sup(Y)} \quad (3.3)$$

Suponha que exista uma tabela de transação como a 2, usando as definições feitas podemos calcular as o suporte, confiança e o lift de $cerveja \Rightarrow fralda$.

Tabela 2 – Base de dados com 3 itens e 5 transações

ID	Cerveja	Fralda	Vinagre
10001	1	1	0
10002	0	0	1
10003	1	0	1
10004	1	1	1
10005	0	1	0

$Conf(Cerveja \Rightarrow Fralda) = 0.4/0.6 \simeq 66\%$, ou seja, 66% para transações que contém cerveja a regra está correta. Já o $Lift(Cerveja \Rightarrow Fralda) = 0.4/0.6 * 0.6 \simeq 1.1$.

3.2 Filtragem colaborativa

A filtragem colaborativa é um dos métodos mais populares de análise de preferências comuns em um grupo de pessoas. Esse algoritmo é baseado na predição entre os gostos e interesses de um usuário com o de outros usuários. Ou seja, o fundamento dessa técnica está na troca de conhecimentos e gostos comuns entre indivíduos. Para usar essa técnica em recomendação, os itens que serão sugeridos são filtrados através das avaliações que os clientes deram sobre os objetos. Essa preferência avalia o comportamento e as similaridades de interesses entre indivíduos (Figura 11).

Sabendo que é necessário um conjunto de notas de usuários aos itens, o banco de dados para análise deve conter essa classificação dos objetos em uma escala numérica. Com isso, é possível encontrar medidas de similaridade entre dois indivíduos, X e Y. Shardanand e Maes (SHARDANAND; MAES, 1995) propõem o uso do coeficiente de correlação de Pearson, r_{XY} , para esse fim. Após quantificar a similaridade entre perfis, esta é utilizada para sugerir objetos particularizados para casa usuário. Identifica-se todas as similaridades superiores a

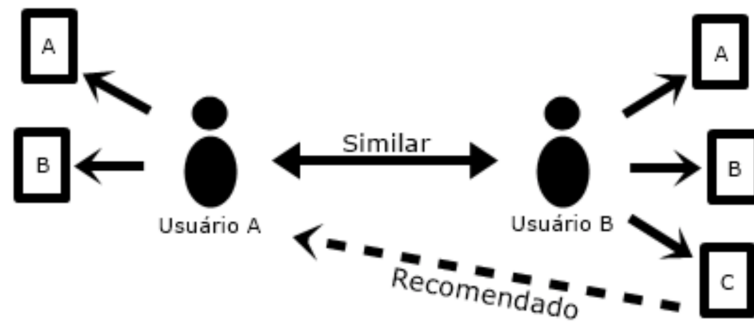


Figura 11 – Representação de sistema de recomendação por filtragem colaborativa. Fonte: (ROLIM *et al.*, 2017)

um limite t determinado, e calcula-se as previsões como a média ponderada das classificações dos usuários semelhantes para o item, em que o peso é a similaridade calculada (BILLSUS; PAZZANI, 1998). Esta previsão é calculada como em 3.4, onde X_A é uma classificação a ser prevista para o usuário X no item A , W é o conjunto dos avaliadores do objeto A e r_{XY} é a correlação entre os usuários X e Y (RESNICK *et al.*, 1994).

$$X_A = \bar{X} + \frac{\sum_{Y \in W} (Y_A - \bar{Y})}{\sum_{Y \in W} |r_{XY}|} \quad (3.4)$$

Mesmo que esses algoritmos de previsão que utilizam correlação possuam um bom desempenho, eles apresentam algumas limitações. (BILLSUS; PAZZANI, 1998) descreve alguns problemas específicos encontrados: primeiro, a correlação entre dois perfis de usuários só pode ser calculada com base nos itens que ambos os usuários avaliaram. Se os usuários puderem escolher entre milhares de itens para avaliar, é provável que a sobreposição de itens classificados entre dois usuários seja pequena em muitos casos. Portanto, muitos dos coeficientes de correlação calculados são baseados em apenas algumas observações. Segundo, a abordagem de correlação induz um modelo global de similaridades entre usuários, em vez de modelos separados para classes de classificações (por exemplo, classificação positiva versus avaliação negativa). Terceiro, e talvez mais importante, dois usuários só podem ser semelhantes se houver sobreposição entre os itens classificados, ou seja, se os usuários não classificarem nenhum item comum, seus perfis de usuário não poderão ser correlacionados.

3.3 Regra por frequência

Também conhecido como recomendação pelo mais popular, o algoritmo basicamente é uma frequência dos produtos.

Após computar todas as frequências, basta recomendar para o cliente o produtos que ele ainda não comprou e que seja um dos objetos mais populares. Ele é bastante utilizado quando

não se tem muitos dados ou muitos produtos, então baseia-se nas compras existentes e elencam os produtos (GRUS, 2018).

3.4 Recomendação por Redes Bipartidas

Há pouco tempo, sistemas de recomendação baseados em rede têm demonstrado um desempenho melhor que o método de filtragem colaborativa, contudo a maior parte dos algoritmos centralizou-se em casos não ponderados. Porém, quando a ponderação não é feita, há perda de informação, o que dificulta a classificação de um produto preferido pelo usuário (LIU; SHANG; CHEN, 2009).

3.4.1 Projeção ponderada

A projeção ponderada com pesos assimétricos proposta por (ZHOU *et al.*, 2007), mostra como se determinar o peso da projeção O, w_{ij} , onde esse peso pode ser considerado como a importância do nó i no sentido de j . Para isso, volta-se a representação bipartida para derivar a expressão analítica de w_{ij} . Dado que a rede bipartida original não é ponderada, o valor de um nó O qualquer deve ser igualmente distribuído para os seus vizinhos em U. Analogamente, o valor em qualquer nó U deve ser igualmente distribuído para seus vizinhos O (ZHOU *et al.*, 2007). A Figura 12 - 1 mostra os nós de O que recebem inicialmente os pesos x , y e z . O processo de alocação de valores consiste em duas etapas; primeiro de O para U, depois de U para O. A representação desses passos podem ser vistos na Figura 12 - 2 e Figura 12 - 3, respectivamente. No final do processo tem-se os valores nos nós de O denotados por x' , y' e z' , tais valores podem ser obtido como em 3.5:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 11/18 & 1/6 & 5/18 \\ 1/9 & 5/12 & 5/12 \\ 5/18 & 5/12 & 4/9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (3.5)$$

A matriz de valores descritos na Equação 3.5 é a matriz de adjacência ponderada, onde o elemento na i -ésima linha e j -ésima coluna representa a fração do valor que o j -ésimo nó O transferiu para o i -ésimo nó O.

O valor do l -ésimo nó de U na projeção quando o valor de O flui para U será dado por:

$$f(u_l) = \sum_{i=1}^n \frac{a_{il}f(o_i)}{k(o_i)}, \quad (3.6)$$

onde $f(o_i) \geq 0$, $k(o_i)$ é o grau de o_i e a_{il} é a matriz de adjacência $n \times m$, onde E é conjunto de

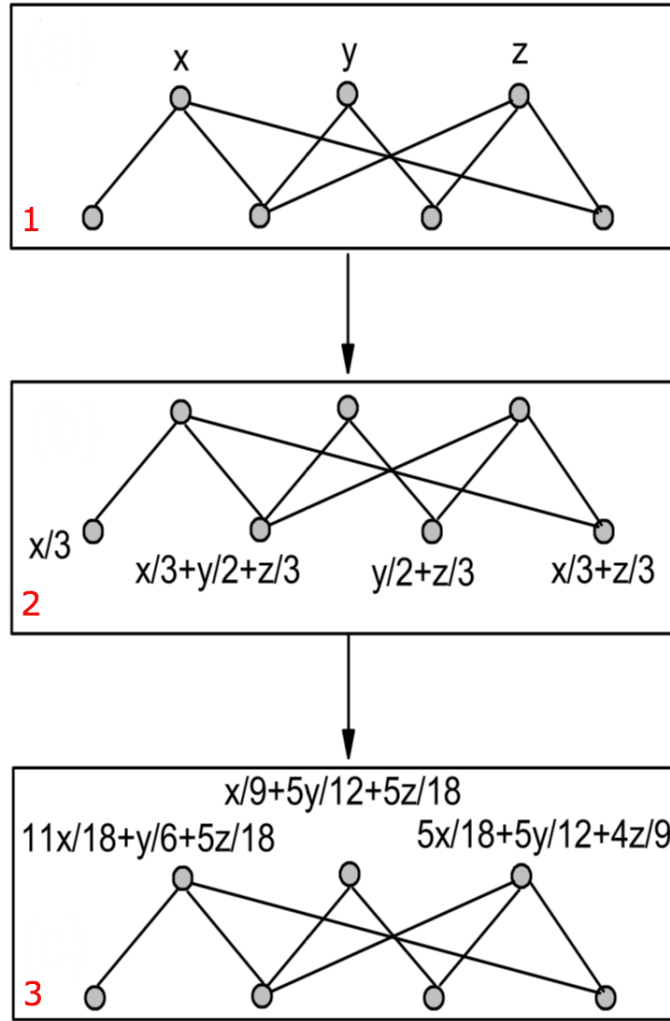


Figura 12 – Ilustração do processo de alocação de recursos em rede bipartida. Os três superiores são nós O e os quatro inferiores são nós U. O processo consiste em: primeiro, o recurso vai de O para U ($1 \rightarrow 2$) e, em seguida, retorna para O ($2 \rightarrow 3$). Fonte: Adaptação de (ZHOU *et al.*, 2007).

arestas e é dada por:

$$a_{il} = \begin{cases} 1, & \text{se } o_i u_l \in E \\ 0, & \text{c.c.} \end{cases} \quad (3.7)$$

O valor de u_i quando retornado no próximo passo para O é escrito como:

$$f'(o_l) = \sum_{l=1}^m \frac{a_{il} f(u_l)}{k(u_l)} = \sum_{l=1}^m \frac{a_{il}}{k(u_l)} \sum_{j=1}^n \frac{a_{jl} f(o_j)}{k(o_j)}, \quad (3.8)$$

que pode ser escrito em função dos pesos como:

$$f'(o_l) = \sum_{j=1}^m w_{ij} f(o_j) \quad (3.9)$$

A matriz W é a matriz da projeção de O ponderada. Chega-se então na conclusão de que diferentemente da matriz de adjacência, a matriz W não é simétrica, ou seja,

$$\frac{w_{ij}}{k(o_j)} = \frac{w_{ji}}{k(o_i)} \quad (3.10)$$

Ao final do processo, elenque para cada cliente u_i todos os objetos não comprados e ordene pelo maior valor de $f'(o_i)$. Os produtos com maior valor de f' serão recomendados.

RESULTADOS E DISCUSSÕES

Neste capítulo, um estudo de simulação numérica é apresentado para comparar a eficiência do sistema de recomendação por rede bipartida ponderada, com os métodos de regras de associação e aleatória. A proposta da comparação será avaliada pela acurácia dos métodos a posteriori. A partir de agora denotaremos o sistema de recomendação por rede bipartida como RBP, o sistema de frequência por SF e o aleatório por SA. Os dados serão analisados em safras diferentes, ou seja, separamos os dados em dois momentos diferentes, cada safra é composta por um período específico do mês de análise, com início e fim iguais. Os ramos de atividade citados são locais de uso de cartão de crédito, como por exemplo ramo alimentação é composto por mercados, supermercados, padarias, etc. Os ramos (R1, R2,..., R10) não serão identificados no trabalho, mas cada um representa um conjunto de estabelecimentos que o cliente pode utilizar o cartão. Os produtos financeiros são compostos por seguros de vida, seguros do cartão, seguros para a mulher, planos de saúde e odontológico. Eles também não serão especificados nos resultados e todos serão denotados por códigos (A, B, C, D e E).

4.1 Descritiva dos dados reais

Para comparar os métodos de recomendação propostos no trabalho foram simulados sistemas com os dados reais presentes em quatro safras distintas (quando analisamos o ramo de experiência do produto); e uma única safra, quando analisamos os serviços (denotaremos essas safras de análise como Momento 1). Como o intuito é verificar a eficácia da recomendação, os resultados foram analisados posteriormente (em safras no momento seguinte ao Momento 1 e essa safra denotaremos por Momento 2) e comparados aos acontecimentos reais. O armazenamento do conjunto de dados está disposto em uma tabela. A Tabela 3 representa o banco de dados de forma simplificada, onde existe uma coluna com o ID do cliente e colunas com os itens coletados, então é possível verificar que na cesta do cliente 10001 possui apenas o item A, já na cesta do cliente 10002, possui os itens B e C.

Tabela 3 – Base de dados

ID	Produto ₁	Produto ₂	Produto ₃	Produto ₄	Produto ₅
10001	A	-	-	-	-
10002	B	C	-	-	-
10003	A	C	E	-	-

O primeiro conjunto de dados refere-se aos produtos financeiros vendidos no momento 1. Como descrito no Capítulo 1, existe uma grande dificuldade em introduzir serviços e produtos para o cliente. Pelo fato de alguns produtos serem ofertados prioritariamente, por ser de mais fácil acesso, mais barato ou de utilidade mais específica, uma grande quantidade de clientes possuem apenas um produto do catálogo. Esse viés na escolha sem critério do produto a ser ofertado para o cliente, faz com que a base de dados possua um perfil de clientes com produtos semelhantes, enquanto outros produtos não são vendidos. A Figura 13 apresenta a distribuição dos produtos mais vendidos (à esquerda), sendo que 50% dos produtos vendidos são os chamados de A, um produto mais barato, enquanto alguns produtos mais elaborados como o C e D, não são vendidos com facilidade.

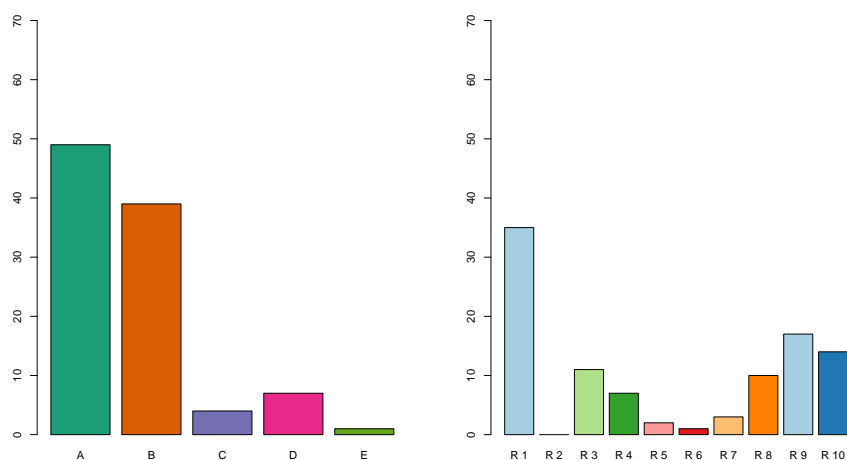


Figura 13 – Distribuição de itens por cesta: à esquerda a distribuição dos produtos financeiros e à direita a distribuição por ramos de atividade. Fonte: Figura elaborada pela autora a partir de dados reais.

Quando se analisa a cesta do cliente, conclui-se o que foi dito anteriormente, isto é, que a maioria dos clientes possuem apenas um produto por cesta. A Figura 14 apresenta essa distribuição para a safra analisada de clientes com produtos financeiros.

O segundo conjunto de dados refere-se aos ramos de atividade utilizados no momento 1 de cada safra. A Figura 15 mostra a distribuição dos ramos por safra, onde se verifica que não há mudanças significativas da distribuição, ou seja, o R1 sempre é o mais utilizado, seguido

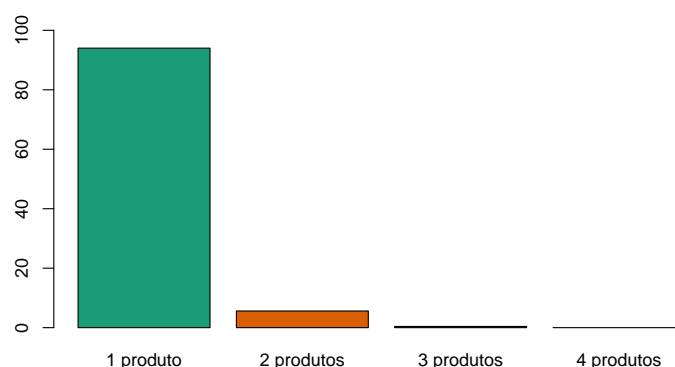


Figura 14 – Distribuição da quantidade de itens por cesta de produtos financeiros. Fonte: Figura elaborada pela autora a partir de dados reais.

pelo R9, R10 e R8 em todas as safras analisadas, isto mostra que o perfil de compras dos clientes é constante e a maioria tem a primeira experiência no Ramo 1. Esse ramo é um dos que classificamos como essenciais, onde 44% das primeiras transações são realizadas nele. Porém, como descrito anteriormente, o esperado era que a partir da primeira experiência com o produto o cliente introduzisse outros ramos ao seu consumo. No entanto, quando verificamos a Figura 16, percebe-se que mais da metade dos clientes utilizaram o produto em apenas um ramo. A Tabela 4 contém algumas métricas das transações por safra analisada. Em todas as safras a média de transações realizadas em uma semana em ramos distintos foi de 1.7, o que volta a confirmar o que foi discutido no Capítulo 1, isto é, que os clientes precisam se engajar cada vez mais no produto, sendo que uma forma disso ocorrer é que eles experimentem outros ramos.

Tabela 4 – Métricas de transação de clientes por safra

Safra	Média	Mínimo	Máximo
Safra 1	1.7	1	7
Safra 2	1.7	1	6
Safra 3	1.7	1	7
Safra 4	1.7	1	7

Pensando na estrutura de redes, podemos descrever algumas métricas. A Tabela 5 mostra os valores de tais métricas, como grau médio e assortatividade, para o conjunto de dados de produtos e para cada uma das safras do banco de dados. Pode-se verificar que o grau médio e a assortatividade das safras dos ramos são muito parecidos e todas as redes são classificadas como dissortativas.

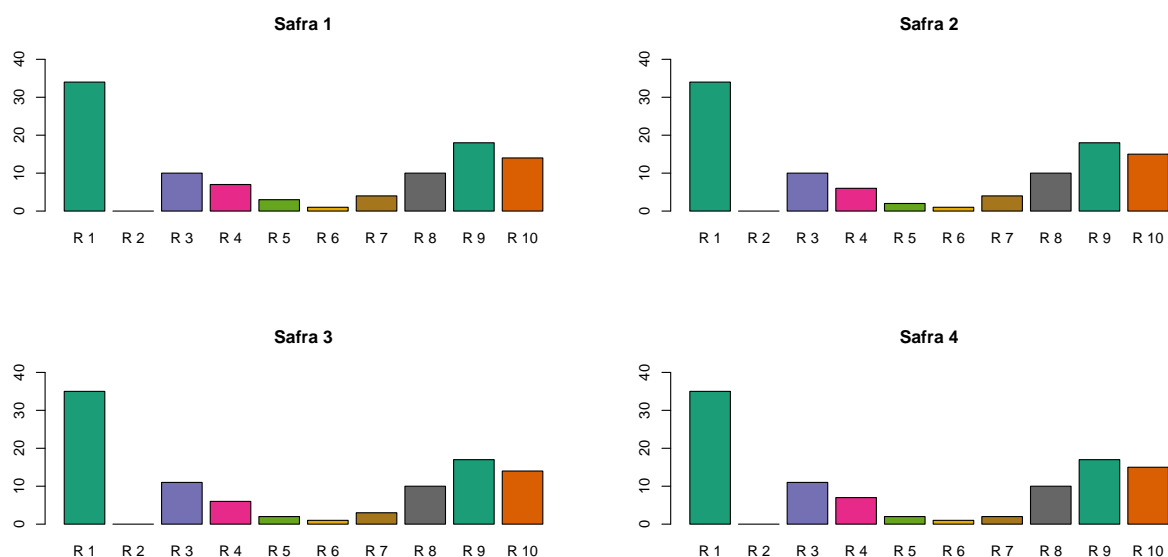


Figura 15 – Distribuição de ramos de atividade por safra. Fonte: Figura elaborada pela autora.

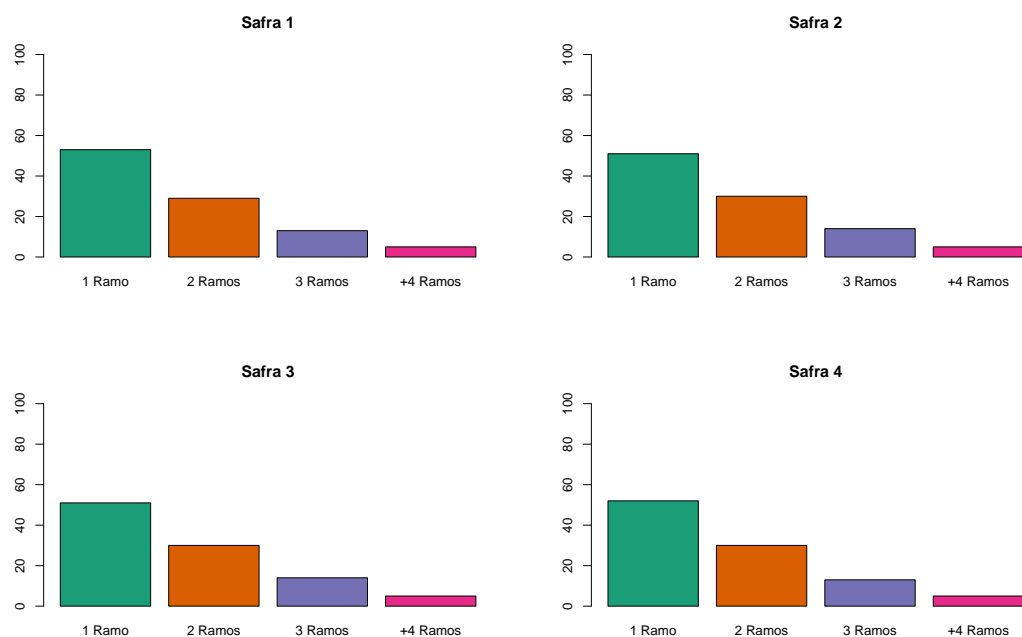


Figura 16 – Distribuição de clientes com a quantidade de ramos por safra. Fonte: Figura elaborada pela autora.

4.2 Simulação dos sistemas recomendação

Após a análise inicial que mostra a concentração de um único produto/ramo por cliente, realizou-se a simulação dos sistemas propostos e pôde-se verificar a diferença entre a quantidade de recomendações por clientes. A RBP recomenda todos os produtos não adquiridos para todos

Tabela 5 – Métricas de transação de clientes por safra

Dados	$\langle k \rangle$	Q
Produtos	2.1	-0.81
Ramos - Safra 1	3.44	-0.607
Ramos - Safra 2	3.47	-0.611
Ramos - Safra 3	3.48	-0.598
Ramos - Safra 4	3.41	-0.59

os clientes. Quando se faz a recomendação aleatória para se recomendar todos os produtos que o cliente ainda não coletou, é necessário incluir algumas regras além do simples fato de ser aleatório. Porém, pensando em proporção de produtos que podem ser ofertados, o SA oferta todos os produtos praticamente na mesma proporção, o que já era esperado. Por outro lado, isso não é o que ocorre quando analisamos o SF, uma vez que apenas os produtos mais populares vão aparecer prioritariamente.

4.2.1 Recomendação de ramos de atividade

Nessa seção iremos apresentar os resultados das simulações envolvendo o banco de dados com compras em diferentes ramos.

O mais desejado em um sistema de recomendação é uma proporção entre os produtos recomendados, podendo ser ramos de atividade ou produtos financeiros, em que o melhor cenário é tal que exista a possibilidade de todos os produtos serem recomendados. As Figuras 17, 18 e 19 nota-se que a RBP recomenda todos os produtos apenas em um cenário com várias recomendações. Como estamos analisando apenas a primeira recomendação, percebemos que a maior concentração de recomendações encontram-se em 4 ou 5 ramos distintos. O SA recomenda todos os produtos, como já era esperado. No entanto, quando analisamos o SF, verificamos que o método não recomenda todos os produtos, uma vez que para isso seria necessário analisar todas as possibilidades de cestas que foram elencadas. Na Figura 19, por exemplo, é apresentada 15 cestas com maior frequência e mesmo assim alguns produtos não apareceram, como o R2. A apresentação da Figura 19 é um pouco diferente dos demais gráficos, isso porque o algoritmo do SF produz cestas com maior suporte, com isso a ordem dos produtos são alteradas e algumas cestas acabam apresentando mais de um item.

Como descrito no início, o conjunto de dados foi analisado em dois momentos distintos, e após a simulação poderia se comparar as recomendações propostas com as compras reais e dessa forma compor um indicador de acurácia do modelo, em que quanto maior o indicador, mais acurado.

Assim sendo, a Figura 20 mostra essa comparação, onde é possível verificar que a curva do RBP é mais acurada que os outros sistemas em todas as safras analisadas, bem como na

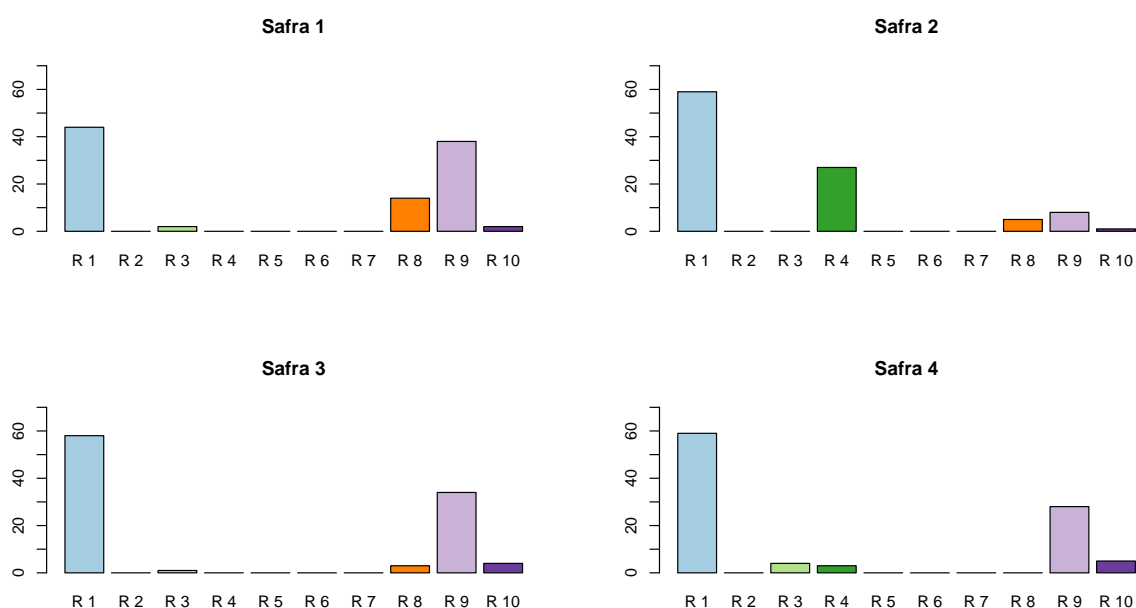


Figura 17 – Sistema RBP: Distribuição da primeira recomendação de ramo de atividade por safra. Fonte: Figura elaborada pela autora.

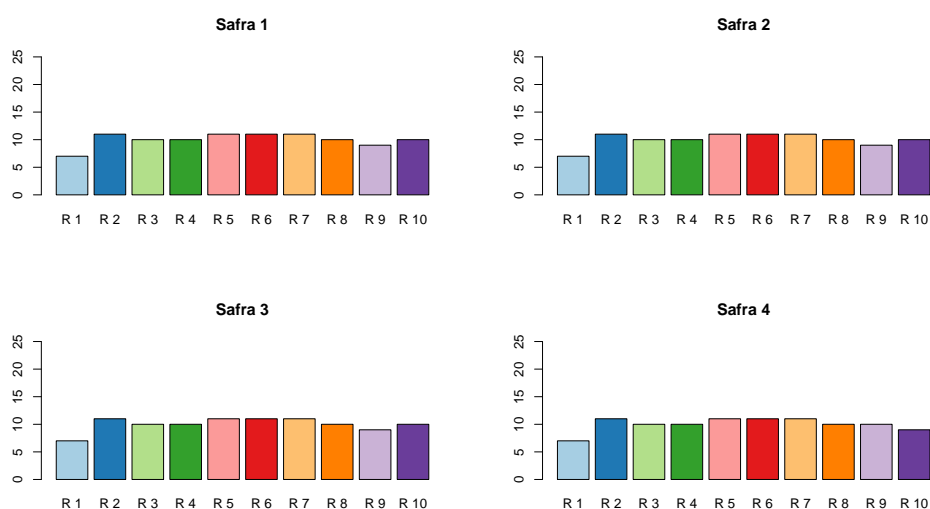


Figura 18 – Sistema SA: Distribuição da primeira recomendação de ramo de atividade por safra. Fonte: Figura elaborada pela autora.

compilação geral. O SF fica em segundo e o SA é o que menos acerta em todos os cenários.

4.2.2 Recomendação de produtos/serviços

Nessa seção iremos apresentar os resultados das simulação envolvendo o banco de dados com compras de produtos ou serviços.

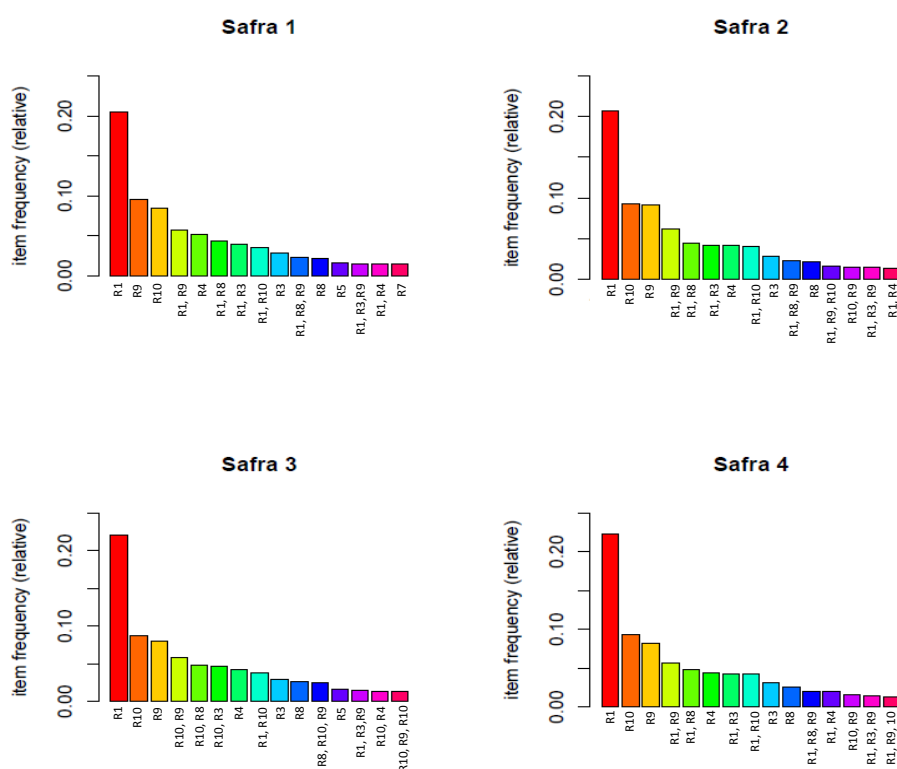


Figura 19 – Sistema SF: Distribuição da frequência de ramo de atividade por safra. Fonte: Figura elaborada pela autora.

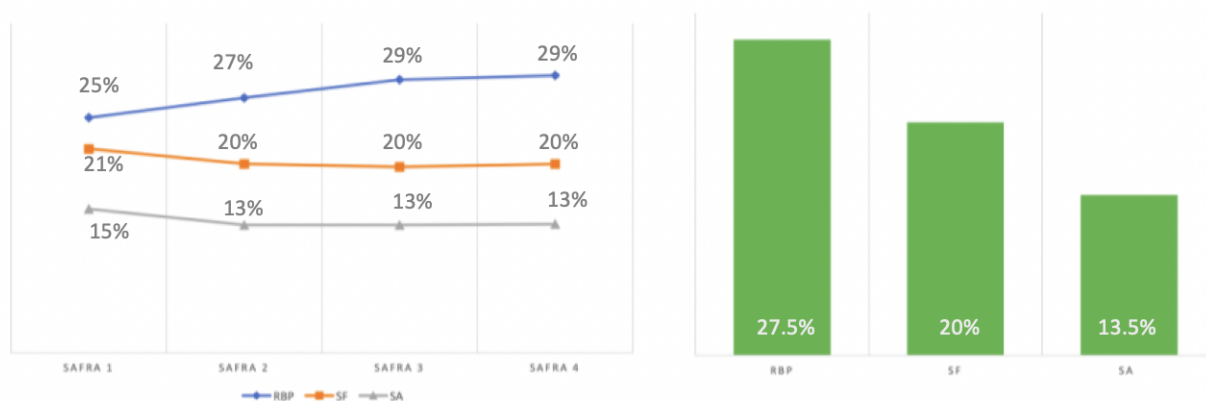


Figura 20 – Comparação das simulações: à esquerda temos a comparação dos 3 sistemas por safra e à direita o resultado geral. Fonte: Figura elaborada pela autora.

Diferentemente da simulação anterior, onde o conjunto de dados não era de uma safra única, quando se verifica as Figuras 21, 22 e 23, pode notar-se que o RBP recomenda todos os produtos. Porém, alguns produtos com menor intensidade e que possui a maior concentração em 2 itens distintos. O SA recomenda todos os produtos como já era esperado. Quando analisamos o SF, verifica-se que ele não recomenda todos itens, uma vez que para isso seria necessário novamente analisar todas as possibilidades de cestas que foram elencadas.

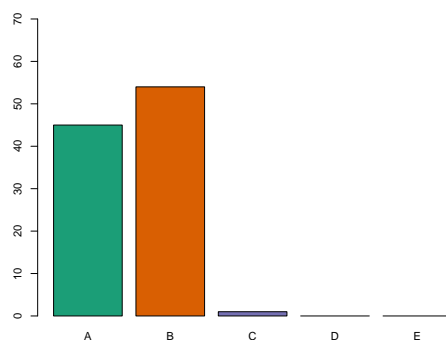


Figura 21 – Sistema RBP: Distribuição da frequência de produtos recomendados. Fonte: Figura elaborada pela autora.

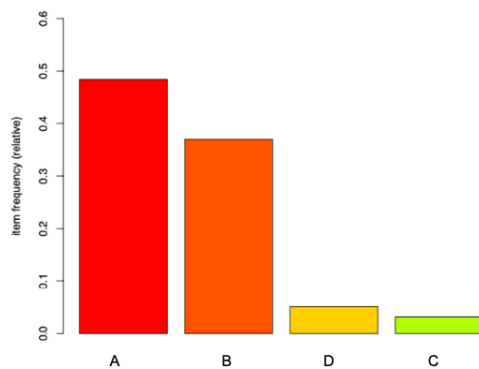


Figura 22 – Sistema SF: Distribuição da frequência de produtos recomendados. Fonte: Figura elaborada pela autora.

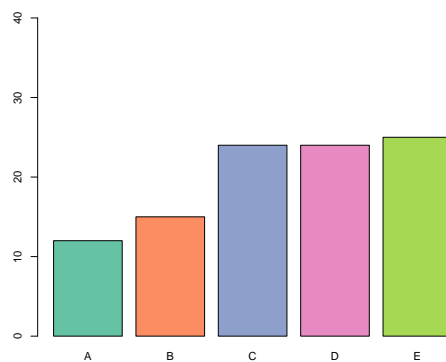


Figura 23 – Sistema SA: Distribuição da frequência de produtos recomendados. Fonte: Figura elaborada pela autora.

Esse conjunto de dados também foi analisado em dois momentos distintos, onde após a simulação pôde-se comparar os resultados com as compras reais e dessa forma levantar um

indicador de acurácia. Sabendo disso, a Figura 24 traz exatamente essa comparação, onde é possível verificar que o RBP é superior aos outros sistemas. O SA fica em segundo lugar e o SF é o que menos acerta em todos os cenários.

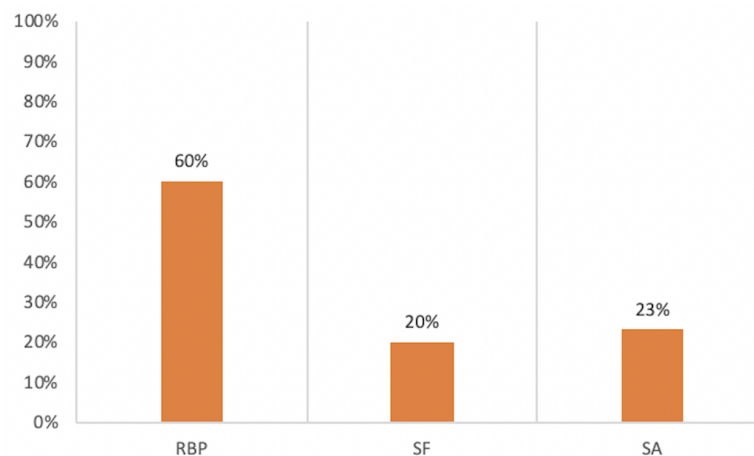


Figura 24 – Comparação das simulações. Fonte: Figura elaborada pela autora.

CONCLUSÃO E PERSPECTIVAS FUTURAS

Nesse trabalho utilizamos os conceitos de redes complexas aplicados a dados reais para simular sistemas de recomendação para uma instituição financeira. Como discutido durante todo o trabalho, os sistemas de recomendação tem um papel fundamental no contato, fidelização e rentabilização do cliente.

No Capítulo 2 foi apresentada uma introdução ao conceito de redes, bem como algumas propriedades. Vários sistemas de recomendação já foram propostos na literatura, um dos mais populares é a filtragem colaborativa (SARWAR *et al.*, 2000), porém existem outras formas de recomendação. No Capítulo 3 foram introduzidos alguns sistemas de recomendação existentes no mercado e descrito com mais detalhes a recomendação utilizando redes bipartidas, mas especificamente a projeção ponderada de redes bipartidas proposta por (ZHOU *et al.*, 2007) que foi o sistema escolhido para essa dissertação.

Para comparar o método proposto, usamos mais duas formas de recomendação e verificamos o comportamento deles em uma amostra após o treino dos modelos. No Capítulo 4 descrevemos os dados e os resultados, mostrando que o sistema por redes complexas é superior aos outros métodos, além de gerar maior número de recomendações e ofertas para todos os clientes. Um ponto negativo a se destacar do sistema de recomendação utilizando redes bipartidas é que ele é muito demorado quando rodado para um conjunto de dados muito grandes, se ampliado para sistemas com mais produtos disponíveis para serem ofertados é necessário um ambiente robusto para que sua performance seja bem executada.

As simulações como descrita também no Capítulo 4 mostraram que os dados reais foram divididos em dois momentos para serem comparados, porém o ideal era realizar campanhas que podem amplificar esses resultados e realizar, assim, ofertas direcionadas para cada tipo de público, esse tipo de análise será feito futuramente. Outra perspectiva futura será realizar uma análise de agrupamento inicial para melhorar a oferta dos clientes baseando-se também no perfil de cada grupo de clientes, que talvez possa melhorar ainda mais a performance do modelo.

REFERÊNCIAS

AGRAWAL RAKESH, T. I.; SWAMI, A. Mining association rules between sets of items in large databases. **International Conference Management of Data**, SIGMOD-93, v. 22, n. 2, p. 207–216, 1993. Citado nas páginas 23 e 35.

BALLOU, R. H. **Gerenciamento da Cadeia de Suprimentos-: Logística Empresarial**. [S.l.]: Bookman Editora, 2009. Citado na página 21.

BARABÁSI, A.-L.; JEONG, H.; NÉDA, Z.; RAVASZ, E.; SCHUBERT, A.; VICSEK, T. Evolution of the social network of scientific collaborations. **Physica A: Statistical mechanics and its applications**, Elsevier, v. 311, n. 3-4, p. 590–614, 2002. Citado nas páginas 23 e 33.

BIGGS, N.; LLOYD, E. K.; WILSON, R. J. **Graph Theory, 1736-1936**. [S.l.]: Oxford University Press, 1986. Citado na página 29.

BILLSUS, D.; PAZZANI, M. J. Learning collaborative information filters. In: **Icml**. [S.l.: s.n.], 1998. v. 98, p. 46–54. Citado nas páginas 35 e 37.

BREESE, J. S.; HECKERMAN, D.; KADIE, C. Empirical analysis of predictive algorithms for collaborative filtering. In: MORGAN KAUFMANN PUBLISHERS INC. **Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence**. [S.l.], 1998. p. 43–52. Citado na página 35.

CHUERUBIM, M. L.; SILVA, I. d. Possibilidades em inteligência artificial na detecção de padrões e previsão de acidentes em rodovias. 2019. Citado na página 31.

DAVIDSON, J.; LIEBALD, B.; LIU, J.; NANDY, P.; VLEET, T. V.; GARGI, U.; GUPTA, S.; HE, Y.; LAMBERT, M.; LIVINGSTON, B. *et al.* The youtube video recommendation system. In: **ACM. Proceedings of the fourth ACM conference on Recommender systems**. [S.l.], 2010. p. 293–296. Citado na página 35.

DAVIS, G. F.; GREVE, H. R. Corporate elite networks and governance changes in the 1980s. **American journal of sociology**, The University of Chicago Press, v. 103, n. 1, p. 1–37, 1997. Citado na página 28.

DENNY, M. Social network analysis. **Institute for Social Science Research, University of Massachusetts, Amherst**, v. 13, 2014. Citado na página 31.

ESTRADA, E. **The structure of complex networks: theory and applications**. [S.l.]: Oxford University Press, 2012. Citado na página 29.

ESTRADA, E.; RODRIGUEZ-VELAZQUEZ, J. A. Subgraph centrality in complex networks. **Physical Review E, APS**, v. 71, n. 5, p. 056103, 2005. Citado na página 23.

FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the internet topology. In: **ACM. ACM SIGCOMM computer communication review**. [S.l.], 1999. v. 29, n. 4, p. 251–262. Citado na página 28.

FAN, Y.; LI, M.; ZHANG, P.; WU, J.; DI, Z. The effect of weight on community structure of networks. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 378, n. 2, p. 583–590, 2007. Citado na página 34.

GOLDBERG, D.; NICHOLS, D.; OKI, B. M.; TERRY, D. Using collaborative filtering to weave an information tapestry. **Communications of the ACM**, Association for Computing Machinery, Inc., v. 35, n. 12, p. 61–71, 1992. Citado na página 35.

GRUS, J. **Data Science do zero: Primeiras regras com o Python**. [S.l.]: Alta Books Editora, 2018. Citado na página 38.

GUSTAFSSON, H.; HANCOCK, D. J.; CÔTÉ, J. Describing citation structures in sport burnout literature: A citation network analysis. **Psychology of Sport and Exercise**, Elsevier, v. 15, n. 6, p. 620–626, 2014. Citado nas páginas 22 e 33.

HOLME, P.; LILJEROS, F.; EDLING, C. R.; KIM, B. J. Network bipartivity. **Physical Review E**, APS, v. 68, n. 5, p. 056107, 2003. Citado nas páginas 22 e 32.

JEONG, H.; TOMBOR, B.; ALBERT, R.; OLTVAI, Z. N.; BARABÁSI, A.-L. The large-scale organization of metabolic networks. **Nature**, Nature Publishing Group, v. 407, n. 6804, p. 651, 2000. Citado nas páginas 22 e 32.

LI, M.; WU, J.; WANG, D.; ZHOU, T.; DI, Z.; FAN, Y. Evolving model of weighted networks inspired by scientific collaboration networks. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 375, n. 1, p. 355–364, 2007. Citado nas páginas 23 e 34.

LILJEROS, F.; EDLING, C. R.; AMARAL, L. A. N.; STANLEY, H. E.; ÅBERG, Y. The web of human sexual contacts. **Nature**, Nature Publishing Group, v. 411, n. 6840, p. 907, 2001. Citado na página 28.

LIN, W.; ALVAREZ, S. A.; RUIZ, C. Collaborative recommendation via adaptive association rule mining. **Data Mining and Knowledge Discovery**, v. 6, p. 83–105, 2000. Citado na página 35.

LIND, P. G.; GONZALEZ, M. C.; HERRMANN, H. J. Cycles and clustering in bipartite networks. **Physical review E**, APS, v. 72, n. 5, p. 056127, 2005. Citado na página 23.

LINDEN, G.; SMITH, B.; YORK, J. Amazon. com recommendations: Item-to-item collaborative filtering. **IEEE Internet computing**, IEEE, n. 1, p. 76–80, 2003. Citado na página 21.

LIU, J.; SHANG, M.; CHEN, D. Personal recommendation based on weighted bipartite networks. In: IEEE. **2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery**. [S.l.], 2009. v. 5, p. 134–137. Citado na página 38.

LOPS, P.; GEMMIS, M. D.; SEMERARO, G. Content-based recommender systems: State of the art and trends. In: **Recommender systems handbook**. [S.l.]: Springer, 2011. p. 73–105. Citado nas páginas 22 e 33.

MILLER, B. N.; ALBERT, I.; LAM, S. K.; KONSTAN, J. A.; RIEDL, J. Movielens unplugged: experiences with an occasionally connected recommender system. In: ACM. **Proceedings of the 8th international conference on Intelligent user interfaces**. [S.l.], 2003. p. 263–266. Citado na página 21.

NEWMAN, M. E. The structure of scientific collaboration networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 98, n. 2, p. 404–409, 2001. Citado nas páginas 23 e 33.

OHKUBO, J.; TANAKA, K.; HORIGUCHI, T. Generation of complex bipartite graphs by using a preferential rewiring process. **Physical Review E**, APS, v. 72, n. 3, p. 036120, 2005. Citado na página 23.

PASTOR-SATORRAS, R.; VÁZQUEZ, A.; VESPIGNANI, A. Dynamical and correlation properties of the internet. **Physical review letters**, APS, v. 87, n. 25, p. 258701, 2001. Citado na página 28.

RAPOPORT, A.; HORVATH, W. J. A study of a large sociogram. **Behavioral science**, Wiley Online Library, v. 6, n. 4, p. 279–291, 1961. Citado na página 28.

RESNICK, P.; IACOVOU, N.; SUCHAK, M.; BERGSTROM, P.; RIEDL, J. Grouplens: an open architecture for collaborative filtering of netnews. In: ACM. **Proceedings of the 1994 ACM conference on Computer supported cooperative work**. [S.l.], 1994. p. 175–186. Citado na página 37.

RESNICK, P.; VARIAN, H. R. Recommender systems. **Communications of the ACM**, ACM, v. 40, n. 3, p. 56–58, 1997. Citado nas páginas 22 e 23.

ROLIM, V.; FERREIRA, R.; COSTA, E.; CAVALCANTI, A.; DIONÍSIO, M. Um estudo sobre sistemas de recomendação de recursos educacionais. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. [S.l.: s.n.], 2017. v. 6, n. 1, p. 724. Citado nas páginas 15 e 37.

SANDERS, N. R.; MANRODT, K. B. Forecasting practices in us corporations: Survey results. **Interfaces**, 1994. Citado na página 21.

SARWAR, B.; KARYPIS, G.; KONSTAN, J.; RIEDL, J. Analysis of recommender algorithms for ecommerce. In: **ACM E-Commerce Conference, Minneapolis, MN**. [S.l.: s.n.], 2000. p. 158–167. Citado na página 51.

SHARDANAND, U.; MAES, P. Social information filtering: Algorithms for automating "word of mouth". In: CITESEER. **Chi**. [S.l.], 1995. v. 95, p. 210–217. Citado na página 36.

SHAYA, S. A.; MATHESON, N.; SINGARAYAR, J. A.; KOLLIAS, N.; BLOOM, J. A. **Intelligent performance-based product recommendation system**. [S.l.]: Google Patents, 2010. US Patent 7,809,601. Citado na página 35.

VALEJO, A.; FERREIRA, V.; FILHO, P. G.; OLIVEIRA, M. C. F. de; LOPES, A. de A. One-mode projection-based multilevel approach for community detection in bipartite networks. In: **SIMBig**. [S.l.: s.n.], 2017. p. 101–108. Citado nas páginas 15 e 32.

WASSERMAN, S.; FAUST, K. **Social network analysis: Methods and applications**. [S.l.]: Cambridge university press, 1994. v. 8. Citado nas páginas 15, 30 e 32.

ZENONE, L. C. **CRM-Customer Relationship Management: gestão do relacionamento com o cliente e a competitividade empresarial**. [S.l.]: Novatec Editora, 2007. Citado na página 22.

ZHOU, T.; REN, J.; MEDO, M.; ZHANG, Y.-C. Bipartite network projection and personal recommendation. **Physical Review E**, APS, v. 76, n. 4, p. 046115, 2007. Citado nas páginas 15, 23, 33, 38, 39 e 51.

