Construção de Sistema de Recomendação de Produtos e Análise das Métricas das Regras de Associação para um *Marketplace Online* do Varejo de Alimentos

RESUMO

O consumo *online* no varejo de alimentos é um fenômeno sem retrocesso. Novos modelos de negócio estão surgindo no setor integrando o mundo físico ao mundo virtual. O presente estudo analisou o histórico das compras de consumidores de um *marketplace online* do varejo alimentar para a construção de um sistema de recomendação de produtos. O *marketplace*, uma empresa que exemplifica esses novos modelos de negócio que estão surgindo no varejo, é uma plataforma *online* que comercializa os produtos de estabelecimentos físicos parceiros mediante pagamento de um percentual (%) do valor gasto pelos consumidores. Foram encontradas 9.965 regras válidas de recomendação, categorizadas em 5 tipos, assim distribuídas: 2.348 regras no formato "1 para 1", 3.764 regras no formato "2 para 1", 2.904 regras no formato "3 para 1", 895 regras no formato "4 para 1" e 54 regras no formato "5 para 1". A análise das métricas de rede de associação das regras encontradas foi realizada a partir de técnicas de *Social Network Analysis* (SNA) de forma não ponderada e ponderada. Os resultados encontrados permitem de forma clara e robusta saber quais produtos são os mais relevantes para o objetivo de recomendação de produtos durante o processo de compra *online* dos consumidores. Por fim, foram apontadas algumas implicações gerenciais para o *marketplace* sobre a utilização do sistema de recomendação desenvolvido.

Palavras-chave: Sistema de Recomendação; *Social Network Analysis*; *Marketplace*.

1 INTRODUCÃO

O presente estudo objetivou construir, para um *marketplace online* do varejo alimentar, um sistema de recomendação de produtos e analisar as métricas da rede de associação das regras encontradas. O *marketplace* objeto deste estudo é uma plataforma *online* que faz parcerias com estabelecimentos físicos, como, por exemplo, supermercados, padarias, lojas de vinhos etc., para comercializar os produtos vendidos por esses estabelecimentos mediante um percentual (%) do valor gasto pelo consumidor.

É fácil compreender que fazer compras de gêneros alimentícios, de limpeza, dentre outros, não é algo que proporciona alegria para diversas pessoas quando precisam ir ao supermercado, pois elas preferem utilizar seu tempo fazendo outras atividades em suas vidas. Nesse sentido, para atender essa necessidade de compra dos consumidores, o *marketplace* que foi estudado contrata os denominados *shoppers*, que são os profissionais responsáveis por fazer e entregar as compras nas residências dos consumidores do *marketplace*. Foi com esse objetivo, juntando a falta de tempo e paciência das pessoas para ir ao supermercado fazer compras, que o *marketplace* iniciou sua operação na cidade de São Paulo e região. Destaca-se que algumas empresas, como o *marketplace* objeto desta pesquisa, já estão explorando o mercado *online* de venda de alimentos e entregando o que os consumidores mais querem: atendimento customizado, praticidade, facilidade de comparar *online* os preços entre estabelecimentos e tempo de entrega reduzido. Ao fazer suas compras no *marketplace* os consumidores podem solicitar, por exemplo, que a banana esteja madura, que o corte da carne seja fino, dentre outras opções de customização do atendimento.

Os estabelecimentos parceiros já reportaram ao *marketplace* que para comercializar produtos no mundo virtual é preciso ter competências que eles não possuem. Sendo assim, as parcerias estabelecidas entre os estabelecimentos e o *marketplace* tem sido profícua para todos os envolvidos, inclusive para os consumidores. Com absoluta certeza, este estudo apresenta um exemplo claro de como a evolução tecnológica está permitindo a criação de um novo modelo de negócio no varejo de alimentos, onde as barreiras entre o mundo físico e virtual são cada vez menores.

A operação do *marketplace* tem menos de três anos de existência e ele necessita nesse momento de um forte crescimento de vendas para garantir sua perenidade ao longo do tempo. Sendo assim, conforme as argumentações dos gestores da empresa, o crescimento de vendas se dará por diversos *drivers*, como, por exemplo, aumentar o número de produtos que cada consumidor compra, aumentar o ticket médio de cada consumidor, realizar campanhas de aquisição e retenção de consumidores de forma mais assertiva, dentre outros. Nesse sentido, diante do que já foi apresentado, enuncia-se a Questão de Pesquisa proposta nesse estudo: **Como recomendar os produtos aos consumidores do *marketplace online* do varejo alimentar com base no histórico de carrinhos?**

É válido ressaltar que a presente pesquisa se justifica tanto do ponto de vista científico quando do ponto de vista de negócios. Do ponto de vista científico, a pesquisa se justifica pela crescente migração das compras em lojas físicas do varejo alimentar para compras *online*. Grewal, Roggeveen e Nordfält (2017) apresentaram, em seu artigo sobre o futuro do varejo, que o cenário está em rápida evolução e que, apesar dos consumidores ainda tomarem a maioria das decisões de compra, a tecnologia, os novos modelos de negócio e modelos analíticos baseados em *machine learning* e *big data* sugerem que o processo de consumo está sofrendo atualmente consideráveis mudanças. Ainda segundo os autores, a tecnologia pode ajudar os varejistas a atingir os consumidores apropriados e permitir que os consumidores tomem decisões mais bem informadas sobre quais produtos consumir.

É notório que o avanço das operações de varejo do mundo físico para o virtual ainda possui muito espaço para crescer e muitos varejistas que só operam com lojas físicas cada vez mais precisarão se adequar a essa migração de consumo. No entanto, operar uma loja física é totalmente diferente de operar vendas *online*, e estas diferentes competências abrem espaço para as diversas parcerias que o *marketplace* deste estudo está conseguindo concretizar.

Um outro aspecto que deve ser considerado são as crescentes oportunidades e possibilidades advindas do uso de *big data* no varejo. Conforme Bradlow *et al*. (2017), as cinco principais dimensões de dados que podem ser utilizados na construção de modelos de *machine learning* no varejo são referentes a (1) clientes, (2) produtos, (3) tempo, (4) localização geoespacial e (5) canal. Com base nesse agrupamento proposto pelos autores, o presente estudo também se legitima cientificamente pelo fato de utilizar dados de algumas dessas dimensões e, principalmente, por permitir a operacionalização dos algoritmos de recomendação de forma eficiente (*real time*) e eficaz (aumentando potencialmente as vendas) pelo *marketplace*, o que, provavelmente, não seria possível de ser implementado com a mesma amplitude e potência diretamente pelos estabelecimentos físicos parceiros.

Do ponto de vista de negócios, especificamente para a operação do *marketplace* em estudo, esta pesquisa se justifica pelo fato de permitir uma possível melhoria da *performance* gerencial da empresa (crescimento da receita, crescimento do lucro, ações de marketing mais eficientes e eficazes etc.). Em outras palavras, acredita-se que os resultados obtidos pelo algoritmo utilizado para recomendação de produtos, se implementado pelos gestores da empresa, suportam, em conjunto com outras decisões estratégicas, o crescimento sustentável da operação. Ressalta-se que esta pesquisa se delimitou ao estudo de um modelo analítico para operação de um *marketplace* específico, ou seja, é uma pesquisa de caráter inteiramente *ad hoc*.

Por fim, apresenta-se o Objetivo Geral e os Objetivos Específicos que nortearam este trabalho. O Objetivo Geral é recomendar produtos aos consumidores do *marketplace* com base no histórico de carrinhos. Os Objetivos Específicos são: (1) definir se o sistema de recomendação será elaborado no nível de produto ou em algum dos níveis de categoria que o *marketplace* adota, (2) definir o algoritmo de recomendação que será utilizado, (3) selecionar as regras válidas de associação dos produtos, (4) gerar a rede que apresenta as regras de associação, e (5) calcular as métricas de centralidade da rede (*degree*, *closeness* e *betweenness*).

2 REFERENCIAL TEÓRICO

A crescente migração do consumo do varejo de alimentos do ambiente físico para plataformas *online* já é percebida no Brasil. Por exemplo, diversas redes de supermercado já comercializam seus produtos pela internet, como as bandeiras do Grupo Pão de Açucar (Pão de Açucar e Extra) e a rede Distribuidora Internacional de Alimentação (DIA %). A comodidade de consumir produtos alimentícios sem sair de casa também fomenta as plataformas do iFood e do Uber Eats. O consumo *online* de alimentos é definitivamente um fenômeno sem retrocesso.

Hoje, consumir *online* parece algo simples, mas a uma década atrás era algo pouco comum. Os *smartphones* estavam chegando ao mercado e os consumidores dificilmente confiavam em realizar compras pela internet. Segundo Kotler, Kartajaya e Setiawan (2017), a conectividade dos consumidores na internet, embora não possa mais ser considerada uma novidade, vem mudando muitas facetas do mercado e não mostra sinais de desaceleração.

Com o aumento da mobilidade e da conectividade, conforme sugerem Kotler, Kartajaya e Setiawan (2017), os consumidores terão cada vez mais um tempo limitado para examinar e avaliar as marcas e é com base nessa limitação que os sistemas de recomendação podem auxiliar o consumo *online* no varejo de alimentos.

Conforme Ricci *et al.* (2011), os Sistemas de Recomendação são ferramentas e técnicas de software que fornecem sugestões para itens que podem ser usados por um usuário. As sugestões fornecidas visam apoiar os usuários em vários processos de tomada de decisão, como quais itens comprar, que música ouvir ou quais notícias ler. Segundo os autores, os Sistemas de Recomendação provaram ser ferramentas valiosas para que usuários *online* lidem com a sobrecarga de informação e se tornaram ferramentas poderosas e populares no comércio eletrônico de diversos tipos de produtos.

Diversas técnicas para geração de recomendação foram propostas na última década, muitas delas também foram implantadas com sucesso em ambientes de negócios. Dentre essas técnicas, conforme apontam Ricci *et al.* (2011), está a mineração de regras de associação, que se concentra em encontrar regras que prevejam a ocorrência de um item com base nas ocorrências de outros itens em uma transação. Os autores ressaltam que o fato de dois itens estarem relacionados significa co-ocorrência, mas não causalidade.

Na identificação de regras de associação, conforme destacam Ricci *et al.* (2011), define-se um *itemset* como uma coleção de um ou mais itens, como, por exemplo, leite, cerveja e fralda. Um *k-itemset* é um conjunto de itens que contém k itens. A frequência de um determinado conjunto de itens é conhecida como contagem de suporte, por exemplo, leite, cerveja e fralda aparecerem em 100 carrinhos diferentes. Já o suporte do conjunto de itens é a fração (contagem de suporte / total de carrinhos) de transações que o contém, por exemplo, se a quantidade total for 1.000 carrinhos, o suporte do conjunto leite, cerveja e fralda é de 10% (100/1.000). Um conjunto de itens frequente é um conjunto de itens com um suporte maior ou igual a um limite mínimo de suporte definido pelo pesquisador, por exemplo, de 2%.

Ricci *et al.* (2011) apontam que uma regra de associação é uma expressão da forma $X \Rightarrow Y$ (lê-se "X para Y"), onde X e Y são conjuntos de itens, por exemplo, "leite, fralda $\Rightarrow$ cerveja". Nesse caso, o suporte da regra de associação é a fração (%) de transações que têm X e Y. Por outro lado, a confiança da regra é a frequência com que os itens em Y aparecem nas transações que contêm X.

Segundo Ricci *et al.* (2011), dado um conjunto de transações T, o objetivo da mineração de regras de associação é encontrar todas as regras que possuam suporte $\geq$ limite mínimo de suporte (definido pelo pesquisador) e confiança $\geq$ limite mínimo de confiança (definido pelo pesquisador). Um abordagem de força bruta, conforme sugerem os autores, seria listar todas as regras de associação possíveis, computar o suporte e a confiança para cada regra e, em seguida, remover as regras que não satisfaçam ambas as condições. Isto é, no entanto, computacionalmente muito caro. Por esta razão, adota-se uma abordagem em duas etapas: (1) Identificação dos conjunto de itens frequentes, que significa gerar todos os conjunto de itens

cujo suporte ≥ limite mínimo de suporte (definido pelo pesquisador); (2) Identificação das regras de associação, que é gerar as regras que possuam alta confiança de cada conjunto de itens frequentes.

Existem várias técnicas para otimizar a geração de conjuntos de itens frequentes, conforme destacam Ricci *et al.* (2011). Em um sentido amplo, elas podem ser classificadas em técnicas que tentam minimizar o número de candidatos (M), que tentam reduzir o número de transações (N) e que tentam reduzir o número de comparações (NM). Segundo os autores, a abordagem mais comum, porém, é reduzir o número de candidatos usando o princípio *apriori*. Este princípio afirma que, se um conjunto de itens é frequente, todos os seus subconjuntos também devem ser frequentes. Isso é verificado, conforme afirma Ricci *et al.* (2011), usando a medida de suporte, pois o suporte de um conjunto de itens nunca excede o de seus subconjuntos. O algoritmo *apriori*, que foi utilizado para construir o Sistema de Recomendação para o *marketplace* desta pesquisa, é uma implementação prática do princípio.

3 METODOLOGIA

A pesquisa teve corte transversal, conforme proposto por Malhotra (2006) e Marconi e Lakatos (1996), e a unidade de análise considerada foi o histórico dos carrinhos de compra dos consumidores do *marketplace*. O histórico dos carrinhos utilizados compreendeu o período de 01/01/2017 e 17/05/2019. É válido destacar que nem todos os carrinhos desse período foram utilizados, pois, segundo os gestores do *marketplace*, algumas parcerias com os estabelecimentos físicos foram descontinuadas, o que poderia prejudicar a construção do Sistema de Recomendação.

Para a construção do Sistema de Recomendação de Produtos foram estudados aproximadamente 30 mil carrinhos, de 9 mil clientes exclusivos, de 16 estabelecimentos parceiros diferentes e que continham 750 mil produtos comprados. O Sistema de Recomendação de Produtos focou nos produtos comprados em cada carrinho, buscando encontrar padrões de compra entre esses produtos. As variáveis utilizadas para construção do Sistema de Recomendação foram: (1) Client_id: código do cliente no banco de dados do *marketplace*, (2) Partnet_id: código da loja parceira onde foi realizada a compra, (3) Order_id: código do pedido, (4) Product_id: código do produto no banco de dados do *marketplace*, (5) Product_name: nome do produto, (6) Product_brand_name: nome da marca do produto, (7) Category_name: categoria do produto atribuída pelo *marketplace* e (8) Categorias_SR: categoria do produto construída especificamente para o Sistema de Recomendação.

É válido destacar que o algoritmo utilizado para gerar as regras de associação dos produtos considerou a variável Categorias_SR. Esta variável foi construída especificamente para a elaboração do Sistema de Recomendação de Produtos, pois, sua granularidade permitiu identificar regras de associação mais gerencialmente implementáveis pelo *marketplace*.

4 ANÁLISE DOS DADOS E RESULTADOS ENCONTRADOS

A exposição do Sistema de Recomendação é feita neste trabalho em duas partes. A primeira versa sobre a seleção das melhores regras de recomendação e a segunda sobre a análise da rede de conexão entre os produtos gerada a partir das regras selecionadas.

4.1. Seleção das melhores regras de recomendação

Para elaborar o Sistema de Recomendação, inicialmente, realizou-se a classificação dos produtos em categorias com granularidade intermediária, de forma que produtos com marca, utilidade ou característica semelhante (cor, sabor, textura) ficassem agrupados na mesma categoria para o Sistema de Recomendação. Esse agrupamento mostrou-se necessário, pois a base disponibilizada pelo *marketplace* agrupava aproximadamente 48 mil produtos em 250 categorias, o que não permitia a elaboração robusta de um Sistema de Recomendação. As categorias foram criadas e os produtos nelas alocados em conjunto com os gestores do *marketplace* e, portanto, por eles validadas. Ao final da classificação obteve-se 4.996

categorias, doravante referenciadas como “Categorias_SR”. O histórico de compras analisado possui 29.357 carrinhos e 18.484 produtos diferentes. Ao combinar carrinho e produto exclusivos o número de combinações é de 602.723. Uma vez que o objetivo deste trabalho foi avaliar as associações entre produtos, produtos iguais no mesmo carrinho foram contados como apenas um. Uma vez criadas as Categorias_SR, os produtos comprados pelos clientes na base analisada foram classificados conforme as Categorias_SR. A Tabela 1 compara o número de produtos com os as quantidades de Categorias_SR obtidas e a combinação delas com os carrinhos.

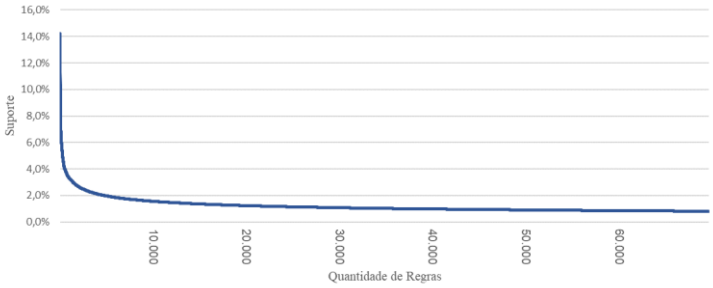
Tabela 1 – Quantidade de produtos, carrinhos e categorias: totais e exclusivos

Quantidade de Carrinhos Exclusivos	Quantidade de Produtos Exclusivos	Combinação Carrinho e Produto	Combinação Carrinho e Produto Exclusivos	Quantidade de Categorias_SR exclusivas	Combinação Carrinho e Categorias_SR Exclusivos
29.355	18.484	740.040	602.723	2.899	554.142

Fonte: Elaborado a partir da base disponibilizada pelo *marketplace*.

É válido destacar que se obteve 4.996 diferentes Categorias_SR, mas apenas 2.899 foram associadas aos produtos contidos no histórico dos carrinhos de compra analisado. As 554.142 combinações exclusivas entre Carrinhos e Categorias_SR foram submetidas ao algoritmo *apriori* a fim de se obter as regras de associação entre elas. Segundo Cavique (2007), o algoritmo “*apriori*” utilizado neste estudo para construção do sistema de recomendação, utiliza as compras (no algoritmo denominadas “transações”) e fornece as regras de associação, representadas no formato “Esquerdo ➔ Direito”. Por exemplo, dado um conjunto de itens {*A,B,C*}, a regra {*B,C*} ➔ {*A*} deve ser lida como “se o consumidor compra em conjunto {B,C}, ele provavelmente também comprará {A}”. Para calibragem das regras de associação dois parâmetros devem ser estabelecidos: suporte e confiança mínimos. Para as análises iniciais desta pesquisa, foi estabelecido um suporte mínimo de 0,3%, pois os valores abaixo de 0,3% de suporte forneciam uma quantidade exorbitante de regras (acima de 10 milhões). Para auxílio na tomada de decisão de quanto usar como suporte mínimo na versão final do Sistema de Recomendação, foi realizada a análise relacionando diferentes níveis de suporte com a quantidade de regras obtidas em cada um deles.

Gráfico 1 – Relação entre diferentes níveis de suporte e quantidade de regras



Fonte: Elaborado a partir da saída do algoritmo *apriori.*

Conforme pode ser observado no Gráfico 1 acima, a partir de 10.000 regras, cada regra gerada pouco contribui para o suporte global do modelo, pois o suporte global é a soma de todos os suportes individuais das regras encontradas pelo algoritmo para o Sistema de Recomendação. Baseado na análise acima, e após validação com os gestores do *marketplace*, definiu-se um ponto de corte com aproximadamente 10.000 regras, o que exigiu um suporte mínimo de 0,0155339 (~1,55%). A partir da aplicação desses parâmetros obteve-se pelo algoritmo 10.020 regras. Além das métricas de suporte e confiança, o algoritmo apriori fornece também o *lift* de cada regra de associação. Conforme apresentado por Yanamoto (2009), o *lift* é uma medida de interesse em sistemas de recomendação, pode assumir

intervalos [0...1...∞] e tem a representação matemática para a regra $A \rightarrow B$ conforme equação abaixo.

$$lift = \frac{P(AB)}{P(A)P(B)}$$

Como resultado do algoritmo *apriori*, foram obtidas as regras e métricas específicas de cada uma delas: suporte, confiança e *lift*. A Tabela 2 apresenta as 3 regras com maior *lift*.

Tabela 2 – Regras e métricas específicas

Regras	Suporte	Confiança	*lift*
{Açúcar Refinado,Feijão Carioca} => {Óleo de Soja}	0,016931	0,529286	7,620012
{Brócolis,Tomate Italiano} => {Couve Flor}	0,015977	0,352367	6,299466
{Brócolis,Cenoura} => {Couve Flor}	0,017919	0,347884	6,21932

Fonte: Elaborado a partir da saída do algoritmo *apriori.*

Yanamoto (2009) propõe a seleção da regra de maior *lift* quando se tem várias regras com o mesmo lado esquerdo, porém, a regra com maior *lift* não necessariamente possui maior confiança. Com base nesta constatação, ou seja, regras com maiores *lifts* e outras com maiores confianças, conforme exemplificado na Tabela 3, surge a necessidade de definir uma forma composta de priorização das regras encontradas. É válido ressaltar que diversos artigos científicos foram pesquisados com o intuito de avaliar se a proposição de priorização de regras considerando *lift* e confiança conjuntamente foi adotada, porém, nenhum texto científico pesquisado considerou a conjunção das duas métricas. Sendo assim, acredita-se que a presente pesquisa contribui cientificamente ao propor esta forma de priorização das regras para um Sistema de Recomendação.

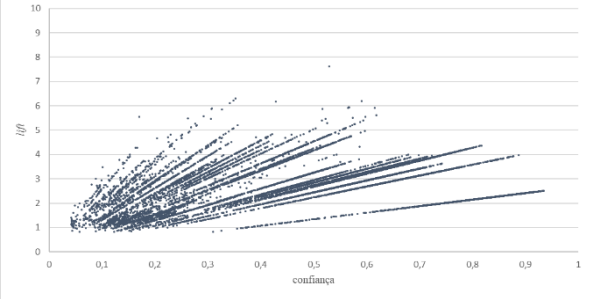
Tabela 3 – Exemplos de regras e suas métricas

Regras	Suporte	Confiança	*lift*
{Abacaxi,Mamão} => {Cenoura}	0,015772	0,54152	2,896034
{Abacaxi,Mamão} => {Limão}	0,016794	0,576608	2,816362
{Abacaxi,Mamão} => {Cebola}	0,015909	0,546199	2,444157
{Abacaxi,Mamão} => {Banana}	0,024664	0,846784	2,26904

Fonte: Saída do algoritmo *apriori*.

O Gráfico 2 apresenta o diagrama de dispersão das 10.020 regras conforme a confiança e alavancagem de cada uma.

Gráfico 2 – Dispersão entre confiança (X) e *lift* (Y) das 10.020 regras

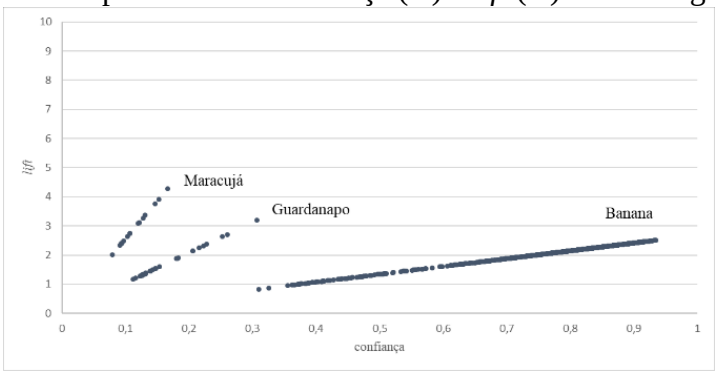


Fonte: Elaborado a partir da saída do algoritmo *apriori*.

Pelo Gráfico 2 é possível inferir que existe uma relação linear entre confiança e alavancagem, inclusive, cada uma das "retas" observadas no gráfico correspondem a uma mesma Categoria_SR recomendada (lado direito da regra de recomendação). Como

ilustração, no Gráfico 3, apresenta-se o diagrama de dispersão das regras cujo lado direito possuem as Categorias_SR “Banana”, “Maracujá” e “Guardanapo”.

Gráfico 3 – Dispersão entre confiança (X) e *lift* (Y) de 3 Categorias_SR



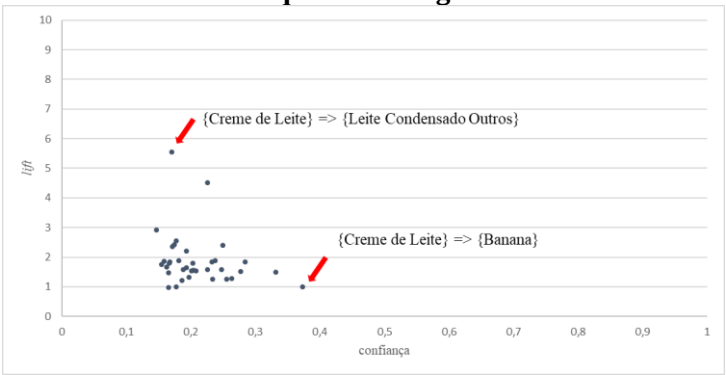
Fonte: Elaborado a partir da saída do algoritmo *apriori*.

Combinando as equações para cálculo do *lift* e da confiança (*C*) podemos encontrar a relação linear entre estas duas grandezas (fórmula abaixo a esquerda). Considera-se o coeficiente angular (C_{ang}) da relação linear entre *lift* e *C* a relação abaixo a direita.

$$lift = C.\frac{P(AB)}{P(B).P(A \cap B)} \qquad C_{ang} = \frac{P(AB)}{P(B).P(A \cap \mathrm{B})}$$

O coeficiente angular representa a inclinação das retas do Gráfico 3 acima e pode ser interpretado como a taxa de variação da alavancagem em relação à confiança. Avaliando-se as regras a partir do lado esquerdo da relação A → B, pode-se plotar a dispersão de todos os produtos recomendados por determinada regra. O Gráfico 4, abaixo, apresenta 59 regras onde a Categoria_SR “Creme de Leite” aparece no lado esquerdo da regra.

Gráfico 4 – Dispersão entre confiança (X) e *lift* (Y) da Categoria_SR “Creme de Leite” no lado esquerdo da regra



Fonte: Elaborado a partir da saída do algoritmo *apriori*.

Pode-se observar que a regra {Creme de Leite} => {Leite Condensado Outros} é a que possui maior *lift* (5,55), entretanto, possui apenas 17,05% de confiança, enquanto a regra {Creme de Leite} => {Banana} é a que possui maior Confiança (37,27%), mas o seu *lift* é de apenas 0,99. Nesse sentido, surge então a seguinte pergunta de negócio: Deve ser priorizada a regra com maior *lift* ou a regra com maior confiança?

Para responder essa pergunta, propõe-se nesta pesquisa a ponderação do *lift* pela confiança, selecionando assim, a regra com maior *lift* levando em conta a probabilidade de ser efetiva. Desta forma, define-se como *lift ponderado* ($lift_p$) a multiplicação do *lift* pela confiança:

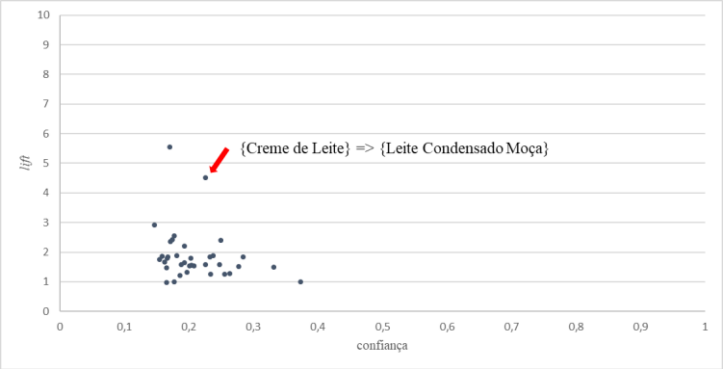
$$lift_p = C \,.\, lift$$

Pode-se expressar o *lift ponderado* como função das probabilidades e da confiança:

$$lift_p = C^2 . \frac{P(AB)}{P(B).P(A \cap B)}$$

Analisando o $lift_p$ das 59 regras cujo lado esquerdo possui a Categoria_SR "Creme de Leite" conclui-se que não necessariamente a regra com maior *lift ponderado* é a regra com maior *lift* e/ou confiança. A proposta deste trabalho é que a priorização das regras que possuem a mesma Categoria_SR do lado esquerdo se dê pelo *lift ponderado*, uma vez que faz a composição entre o *lift* de cada regra e a probabilidade de a regra ser concretizada no próximo produto escolhido pelo cliente. Aplicando-se tal proposta de priorização, a regra que deve ser aplicada passa a ser "{Creme de Leite} => {Leite Condensado Moça}" que tem o *lift ponderado de* 1,01831346, diferente das duas candidatas iniciais: {Creme de Leite} => {Leite Condensado Outros} e {Creme de Leite} => {Banana}. O Gráfico 5 abaixo apresenta a dispersão das regras com a Categoria_SR do lado esquerdo e a posição da regra proposta, escolhida a partir do *lift ponderado.*

Gráfico 5 – Dispersão entre confiança (X) e *lift* (Y) da Categorias_SR "Creme de Leite" no lado esquerdo da regra com a regra proposta para o Sistema de Recomendação



Fonte: Elaborado a partir da saída do algoritmo *apriori*.

Como tratamento final no conjunto de regras, foram eliminadas aquelas que possuem *lift* abaixo de 1, pois nestas regras as Categorias_SR do lado esquerdo possuem um efeito de desalavancagem nas Categorias_SR do lado direito. Removendo as regras que se encontram nesta condição, a quantidade final de regras foi de 9.965, sendo 2.348 regras no formato 1 → 1, 3.764 regras no formato 2 → 1, 2.904 regras no formato 3 → 1, 895 regras no formato 4 → 1 e 54 regras no formato 5 → 1. Desta forma, acredita-se que o *marketplace* pode utilizar todas as 9.965 regras finais concebidas e adotar o *lift ponderado* como um critério sólido de decisão para a escolha da ordem das recomendações que devem ser feitas para o consumidor.

4.2 Avaliação das métricas da rede sem ponderação das arestas

Para se avaliar as métricas da rede de recomendações obtida foi realizado ajuste para que cada nó da rede fosse representado por uma Categoria_SR, pois assim poderiam ser avaliados aspectos de Categoria_SR com maior centralidade, maior impacto e maior proximidade com todas as demais Categorias_SR.

As 9.965 regras foram convertidas em vértices e arestas, de forma que apenas uma Categoria_SR representasse um vértice e a relação Categoria_SR A→Categoria_SR B representasse uma aresta. Regras que possuíssem mais de uma Categoria_SR do lado esquerdo foram desmembradas em duas ou mais relações Categoria_SR A→Categoria_SR B.

Aplicando-se tal regra para conversão de regras em arestas, obteve-se 135 Categorias_SR exclusivas, portanto, 135 vértices e 2.362 relações diferentes entre estas

Categorias_SR, portanto 2.362 arestas. Conforme observado na rede, Categorias_SR com muitas conexões apresentam alta densidade e difícil visualização. A densidade total da rede é de 13,05694%. O diâmetro, conforme Albert e Barabasi (2002), é a distância máxima entre qualquer par de vértices, e no caso da rede de recomendações, é apenas 3, o que sugere que existem muitas conexões em poucos vértices. O número de cliques, conforme Albert e Barabasi (2002) é o subconjunto de vértices tais que cada dois vértices do subconjunto são conectados por uma aresta. Na rede de recomendação existem 8 cliques com 22 elementos (maior número de elementos em um único clique da rede de recomendações).

4.2.1 Centralidade de grau

Centralidade de grau da rede, conforme Newman (2006) é definida como a quantidade de vértices conectados em um vértice específico. A centralidade de grau para redes direcionadas pode ser observada avaliando-se arestas que chegam (centralidade de grau de entrada) ou partem (centralidade de grau de saída) do vértice em análise. Os 3 nós com maiores centralidades de grau de entrada e saída foram: Banana (124), Cebola (93) e Tomate Italiano (84).

4.2.2 Centralidade de proximidade

Centralidade de proximidade da rede, conforme Newman (2006) é definida como a quantidade de caminhos que deve ser percorrido de cada vértice a fim de se visitar todos os vértices da rede. A centralidade de proximidade para redes direcionadas pode ser observada avaliando-se arestas que chegam (centralidade de proximidade de entrada) ou partem (centralidade de proximidade de saída) do vértice em análise. Os 3 vértices com maiores centralidades de proximidade de saída foram: Refrigerante Guaraná Antarctica (367), Leite Codensado Outros (366) e Leite Condensado Moça (366). Os 3 vértices com menores centralidades de proximidade de saída foram: Banana (144), Cebola (175) e Tomate Italiano (184). Conclui-se que quanto menor a centralidade de proximidade de saída, mais próximo está a Categoria_SR de recomendar todos os produtos da rede, enquanto quanto maior a centralidade de proximidade de saída, mais lentamente a Categoria_SR recomendará todas as demais Categorias_SR.

4.2.3 Centralidade de intermediação

Centralidade de intermediação da rede, conforme Newman (2006) é a fração de caminhos entre os outros vértices caso o vértice em análise seja removido. No grafo de recomendação, uma Categoria_SR com alta centralidade de intermediação representa uma Categoria_SR que, se removida, deixará de promover a relação entre as Categorias_SR que não estão relaciondadas entre sí. Produtos pertencentes a Categorias_SR com alto grau de intermediação são produtos que, analogamente ao trabalho de Newman (2006), promovem a conexão entre grupos de Categorias_SR que não estão diretamente conectadas. Os 10 vértices com maiores centralidades de intermediação são: Banana (8682,0), Cebola (1590,3), Totame Italiano (899,9), Leite Líquido Integral (814,4), Limão (668,7), Batata (605,9), Creme de Leite (585,2), Cenoura (570,3), Mamão (568,9) e Água Sanitária (322,6).

4.2.4 Interpretação dos indicadores

A partir das observações neste capítulo, verifica-se que a rede é pouco densa (13,06%), entretanto, bastante contectada em alguns poucos produtos. Uma vez que “Banana”, “Cebola” e “Tomate Italiano” foram os *rankings* superiores nos indicadores de centralidade calculados, tais produto devem sempre ser fortemente ofertados pois são ao mesmo tempo ponto de partida para recomendação dos demais (conforme centalidade de grau e proximidade de saída), resultado do sistema de recomendação de parte significativa das Categorias_SR (conforma centralidade de grau e proximidade de entrada) e são Categorias_SR que conectam grupos de outras Categorias_SR desconectadas entre sí (centralidade de intermediação).

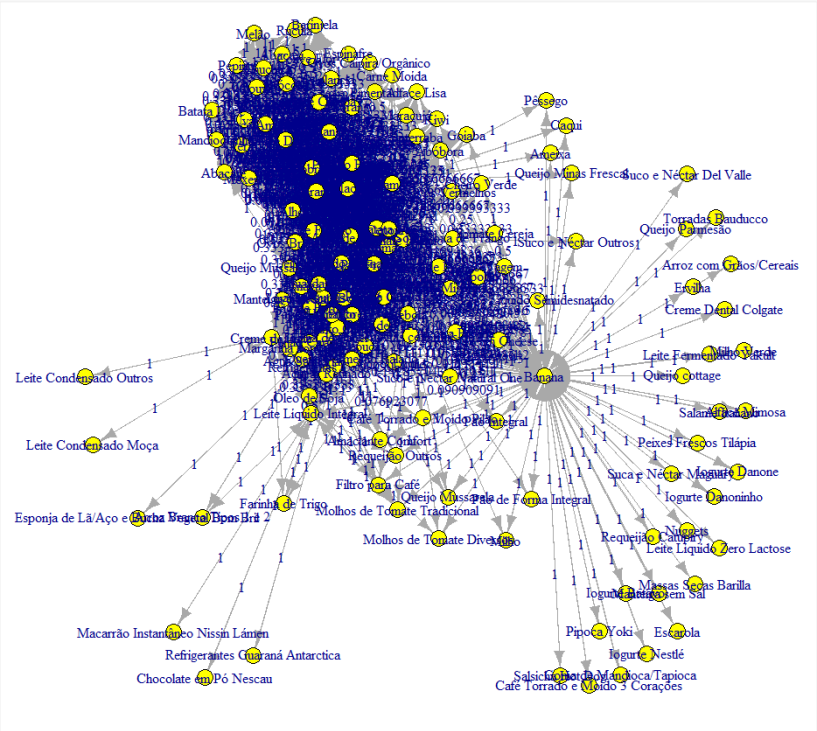
4.3 Avaliação das métricas da rede com ponderação das arestas

De acordo com o critério estabelecido apresentado na sessão 4.2 (Avaliação das métricas da rede sem ponderação das arestas), o lado A das regras de recomendação foi desmembrado de forma que cada Categoria_SR presente no lado A ficasse separadamente associada à Categoria_SR do lado B. Uma vez que se tem a mesma Categoria_SR em diversas regras do lado A cuja Categoria_SR do lado B também é a mesma, existirá repetição na composição das arestas Categoria_SR X → Categoria_SR Y. Conforme Biggs, Lloyd, e Wilson (1999), o passeio *p(X,Y)* que interliga dois vértices X e Y, sendo X e Y interligados por uma aresta X → Y pode ser considerado seu "custo". Fazendo-se analogia com este trabalho, a proposta é a utilização do número de repetições dos vértices como insumo para atribuição dos pesos. Como a repetição das arestas sugerem maior proximidade entre as Categorias_SR X e Y é natural concluir que quanto maior o número de repetições menor o custo de se recomendar um Catagoria_SR Y a partir de uma Categoria_SR X. Dado tal cenário a proposta deste trabalho é considerar o peso das arestas como a proporção inversa da quantidade de repetições.

$$p(X,Y) = \frac{1}{Quantidade\ de\ repetições}$$

Obteve-se, então, pesos que variam de 0,00308642 até 1. A Figura 1 abaixo apresenta a rede obtida com os devidos pesos representados em cada aresta.

Figura 1 – Rede do Sistema de Recomendação com ponderação das arestas



Fonte: Elaborado pelos autores.

Calculou-se, então, as métricas inerentes a toda a rede: densidade, diâmetro e número de cliques e comparou-se com os valores obtidos com a rede sem definição de pesos para arestas.

Tabela 4 – Comparação das redes sem e com pesos

Métrica	Rede sem pesos	Rede com pesos
Densidade	13,05694%	13,05694%
Diâmetro	3	2,416667
Cliques	8 com 22 elementos	8 com 22 elementos

Fonte: Elaborado pelos autores.

Conforme pode ser observado na Tabela 4 acima, apenas o diâmetro da rede é influenciado pelos pesos das arestas, já a densidade e configuração de cliques são afetados pela configuração de vértices e arestas, independente dos pesos.

4.3.1 Centralidade de grau

A Centralidade de grau de entrada e de saída para a rede com pesos, para todos os vértices, foi a mesma para a rede sem pesos, uma vez que a grandeza centralidade independe dos pesos das arestas, mas sim da quantidade de arestas interligadas a cada vértice.

4.3.2 Centralidade de proximidade

Uma vez que a centralidade de proximidade é uma função da distância entre vértices, é esperado que existam vértices com centralidade de proximidade diferentes nas redes com e sem pesos. Segundo Dangalchev (2006), matematicamente a centralidade de proximidade pode ser representada conforme fórmula abaixo.

$$C_c(v) = \sum_{t \in V} 2^{-d_G(v,t)}$$

Onde *v* é o vértice em questão, *t* é o vértice para o qual se calculará a centralidade, *V* é o número de vértices da rede *G* e $d_G(v,t)$ é a distância ponderada entre *v* e *t*. A Tabela 5 abaixo ilustra os 3 vértices com maior diferença entre centralidade de proximidade de saída e entrada entre as redes com e sem pesos.

Tabela 5 – Categorias_SR com maiores diferenças entre centralidade de proximidade de saída

Categoria_SR	Proximidade de saída sem pesos	Proximidade de saída com pesos	Proximidade de entrada sem pesos	Proximidade de entrada com pesos
Couve Flor	260	92,24796558	260	91,93733
Rúcula	259	92,16513766	259	91,8545
Chuchu	259	92,27808192	259	91,96745

Fonte: Elaborado pelos autores.

4.3.3 Centralidade de intermediação

Conforme observado na centralidade de proximidade, a centralidade de intermediação também é uma função dos pesos das arestas ligadas nos vértices em questão. Conforme Freeman (1977), a intermediação de um vértice *v* é obtida com a fórmula abaixo.

$$g(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Onde σ_{st} é o número total de caminhos do vértice *s* para o vértice *t* e $\sigma_{st}(v)$ é o número de menores caminhos que passam por *v*. A Tabela 6 abaixo apresenta os 3 vértices com maiores diferenças entre centralidade de intermediação da rede com e sem pesos.

Tabela 6 – Categorias_SR com maiores diferenças entre centralidade de intermediação

Categoria_SR	Intermediação sem pesos	Intermediação com pesos
Banana	8682,004	16451,97
Tomate Italiano	899,9033	99,22051
Limão	668,71	48,59194

Fonte: Elaborado pelos autores.

4.3.4 Interpretação dos indicadores

Uma vez que a quantidade de regras que associam dois vértices ou duas Categorias_SR influenciam na probabilidade de recomendação do produto do lado direito da regra, entende-se que é razoável utilizar a rede ponderada para estabelecer os produtos prioritários para estoque. As Tabelas 7, 8 e 9 a seguir apresenta os 6 vértices com menores centralidades de proximidade de entrada e saída e maiores centralidade de intermediação obtidas através da rede ponderada.

Tabela 7 – Categorias_SR com menores centralidades de proximidade de entrada

Categoria_SR	Centralidade de proximidade de entrada
Banana	76,44267
Cebola	83,64172
Batata Doce	88,17690
Tomate Italiano	88,35108
Pêra	89,09079
Mandioquinha	89,66258

Fonte: Elaborado pelos autores.

Tabela 8 – Categorias_SR com menores centralidades de proximidade de saída

Categoria_SR	Centralidade de proximidade de saída
Banana	76,75330
Cebola	83,64172
Tomate Italiano	88,35108
Batata Doce	88,48754
Pêra	89,40143
Mandioquinha	89,97321

Fonte: Elaborado pelos autores.

Tabela 9 – Categorias_SR com maiores centralidades de intermediação

Categoria_SR	Centralidade de intermediação
Banana	16451,97
Cebola	1719,492
Leite Líquido Integral	807,3538
Creme de Leite	531,1538
Água Sanitária	270,4062
Refrigerantes Coca-Cola	267,8857

Fonte: Elaborado pelos autores.

Conforme pode ser observado “Banana”, “Cebola” e “Tomate Italiano” deixaram de ser sempre os três primeiros colocados nas três métricas acima. “Tomate Italiano” passou de terceiro lugar no *ranking* dos três indicadores da rede não ponderada para quarto (no caso da centralidade de proximidade de entrada) e oitavo (no caso da centralidade de intermediação).

A Tabela 10 consolida, as métricas dos primeiros 35 vértices classificados pelo maior Grau de Saída da rede de recomendações.

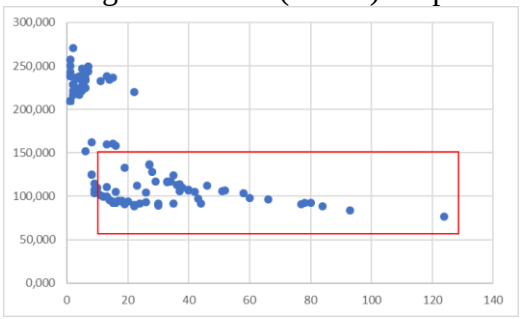
Tomando como ponto de partida as métricas de saída, pois são as Categorias_SR “geradoras” de recomendação, foram analisadas a centralidade de grau e a centralidade de proximidade. O Gráfico 6 apresenta a dispersão das Categorias_SR por estas duas métricas.

Tabela 10 – Métricas consolidadas dos primeiros 35 vértices pelo maior Grau de Saída

Categoria_SR	Grau Saída	Grau Entrada	Proximidade Saída	Proximidade Entrada	Intermediação	Proximidade Saída Ponderada	Proximidade Entrada Ponderada	Intermediação Ponderada
Banana	124	124	144	144	8682,004	76,753	76,443	16451,967
Cebola	93	93	175	175	1590,341	83,642	83,642	1719,492
Tomate Italiano	84	84	184	184	899,903	88,351	88,351	99,221
Limão	80	80	188	188	668,71	92,093	92,093	48,592
Batata	78	78	190	190	605,906	92,591	92,591	27,687
Cenoura	77	78	192	190	570,326	90,654	92,147	21,467
Mamão	66	69	207	199	568,973	96,131	102,512	14,667
Maçã	60	62	214	208	236,826	98,005	101,611	1,667
Laranja	58	58	211	211	187,72	103,222	103,008	0
Ovos Brancos	52	52	216	216	143,859	106,615	106,615	1,611
Alho	51	51	217	217	119,631	105,594	105,594	0
Filé de Frango	46	46	222	222	117,195	112,538	112,538	2,057
Manga	44	44	231	231	79,124	91,786	91,475	0
Abobrinha	43	43	228	228	79,98	97,032	96,817	0
Leite Líquido Integral	42	40	226	228	814,425	104,983	104,421	807,354
Bisnaga	40	40	228	228	66,192	107,065	107,065	2,04
Pão de Forma	40	40	228	228	68,185	106,532	106,532	2,04
Água Sanitária	38	37	230	231	322,614	110,138	111,572	270,406
Queijo Mussarela Fatiada	37	37	231	231	40,017	105,991	105,991	0
Toalha de Papel	37	37	231	231	51,994	114,036	114,036	3,411
Lava Louças Bombril	36	36	232	232	52,142	112,763	112,763	3,071
Brócolis	35	35	240	240	29,67	91,74	91,429	0
Creme de Leite	35	35	233	233	585,179	123,814	123,814	531,154
Refrigerantes Coca-Cola	34	31	234	237	294,463	117,162	132,499	267,886
Açúcar Refinado	33	33	235	235	81,549	117,232	117,232	255,121
Papel Higiênico Neve	33	33	235	235	30,221	116,113	116,113	1,971
Pepino	30	30	245	245	11,543	91,958	91,647	0
Pêra	30	30	245	245	11,42	89,401	89,091	0
Uva	30	30	245	245	11,607	91,143	90,833	0
Guardanapo	29	29	239	239	6,148	117,304	117,304	0
Lava Roupas em Pó Omo	28	28	240	240	14,843	128,338	128,338	0,733
Feijão Carioca	27	27	242	242	124,928	136,676	136,676	3,154
Limpadores Veja	27	27	241	241	14,093	136,291	136,291	0,733
Couve	26	26	249	249	5,915	93,294	92,983	0
Presunto Cozido	26	26	242	242	26,321	104,482	104,482	0

Fonte: Elaborado pelos autores.

Gráfico 6 – Dispersão entre o grau de saída (eixo x) e a proximidade de saída (eixo y)



Fonte: Elaborado pelos autores.

Como se deseja alta centralidade de grau e baixa centralidade de proximidade, fica evidente as Categorias_SR mais relevantes para a navegação da recomendação por toda a rede. A Tabela 11 abaixo apresenta as 59 Categorias_SR contidas na área destacada no Gráfico 6 acima, sugestão de pontos de corte deste trabalho. A centralidade de grau proposta é maior ou igual a 10 e centralidade de proximidade menor ou igual a 150,00.

Tabela 11 – Categorias_SR mais relevantes

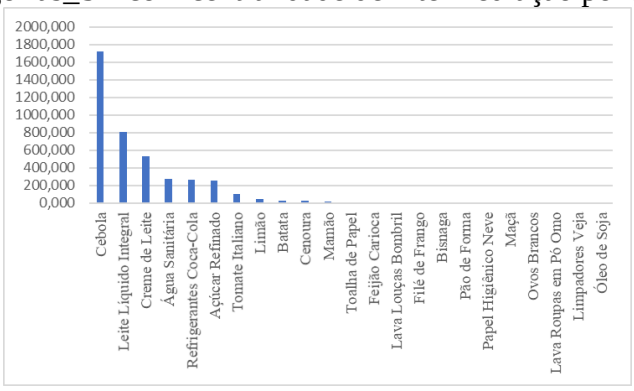
Categoria_SR	Grau Saída	Proximidade Saída
Banana	124	76,753
Cebola	93	83,642
Tomate Italiano	84	88,351
Limão	80	92,093
Batata	78	92,591
Cenoura	77	90,654
Mamão	66	96,131
Maçã	60	98,005
Laranja	58	103,222
Ovos Brancos	52	106,615
Alho	51	105,594
Filé de Frango	46	112,538
Manga	44	91,786
Abobrinha	43	97,032
Leite Líquido Integral	42	104,983
Pão de Forma	40	106,532
Bisnaga	40	107,065
Água Sanitária	38	110,138
Queijo Mussarela Fatiada	37	105,991
Toalha de Papel	37	114,036
Lava Louças Bombril	36	112,763
Brócolis	35	91,740
Creme de Leite	35	123,814
Refrigerantes Coca-Cola	34	117,162
Papel Higiênico Neve	33	116,113
Açúcar Refinado	33	117,232
Pêra	30	89,401
Uva	30	91,143
Pepino	30	91,958
Guardanapo	29	117,304

Categoria_SR	Grau Saída	Proximidade Saída
Lava Roupas em Pó Omo	28	128,338
Limpadores Veja	27	136,291
Feijão Carioca	27	136,676
Couve	26	93,294
Presunto Cozido	26	104,482
Abacate	24	91,727
Manteiga com Sal	23	111,881
Batata Doce	22	88,488
Mandioquinha	22	89,973
Alface Americana	20	93,733
Melão	19	90,751
Leite Líquido Desnatado	19	132,659
Tomate Diversos	18	94,057
Mexerica	18	94,901
Abacaxi	17	94,692
Rúcula	16	92,165
Chuchu	16	92,278
Água de Coco	16	104,823
Couve Flor	15	92,248
Berinjela	15	94,064
Melancia	14	95,650
Espinafre	13	99,435
Alface Crespa	13	99,476
Patinho Bovino	13	110,976
Ovos Caipira/Orgânico	12	99,227
Morango	12	99,552
Peito de Peru	11	101,284
Maracujá	10	104,001
Pimentão	10	109,636

Fonte: Elaborado pelos autores.

Para análise da intermediação, são 24 as Categorias_SR com centralidade de intermediação ponderada acima de 0, e sua priorização para reposição de estoque ou negociações estratégicas pode ser apoiada pelo Gráfico 7 abaixo. Do gráfico foi retirada a Categoria_SR "Banana" por conta da distorção na escala devido ao alto valor de centralidade de intermediação, entretanto ela deve ser considerada para fins gerenciais.

Gráfico 7 – Categorias_SR com centralidade de intermediação ponderada acima de 0.



Fonte: Elaborado pelos autores.

A Tabela 12 apresenta as 24 Categorias_SR com centralidade de intermediação ponderada acima de 0.

Tabela 12 – Categorias_SR com centralidade de intermediação ponderada acima de 0

Categoria_SR	Intermediação Ponderada
Banana	16451,967
Cebola	1719,492
Leite Líquido Integral	807,354
Creme de Leite	531,154
Água Sanitária	270,406
Refrigerantes Coca-Cola	267,886
Açúcar Refinado	255,121
Tomate Italiano	99,221
Limão	48,592
Batata	27,687
Cenoura	21,467
Mamão	14,667

Categoria_SR	Intermediação Ponderada
Toalha de Papel	3,411
Feijão Carioca	3,154
Lava Louças Bombril	3,071
Filé de Frango	2,057
Bisnaga	2,040
Pão de Forma	2,040
Papel Higiênico Neve	1,971
Maçã	1,667
Ovos Brancos	1,611
Lava Roupas em Pó Omo	0,733
Limpadores Veja	0,733
Óleo de Soja	0,500

Fonte: Elaborado pelos autores.

5 IMPLICAÇÕES GERENCIAIS PARA O *MARKETPLACE*

A partir das regras encontradas para o Sistema de Recomendação, o *marketplace* adquire o potencial de desenvolver estratégias mais direcionadas ao comportamento de compra de cada consumidor e de conceber ações de marketing de forma mais efetiva e customizada. Uma outra implicação gerencial para o *marketplace* é que o Sistema de Recomendação fornece informação relevante para a tomada de decisão e priorização das negociações de promoções e descontos com os estabelecimentos parceiros. É válido destacar também que o Sistema de Recomendação desenvolvido nesta pesquisa, para ser efetivo, deverá ser implementado na plataforma do *marketplace* no momento da compra *online* do consumidor. Porém, como já anteriormente exposto, as regras de recomendação consideram categorias de produtos e não os produtos em si, sendo assim, deverão ser estabelecidos pela equipe de *analytics* do *marketplace* os critérios para escolha de qual(is) produto(s) da Categoria_SR recomendada pelo modelo será efetivamente apresentado na tela do cliente. Vale a pena ressaltar que este trabalho não pesquisou qual é o momento mais oportuno para a recomendação dos produtos durante a jornada de navegação, seleção e composição do carrinho de compras pelo consumidor. No entanto, sugere-se dois principais momentos durante a jornada do consumidor onde a recomendação pode ser adequada: (1) imediatamente após a inclusão de um produto no carrinho que está contemplado em uma das 9.965 regras finais concebidas ou (2) apenas no momento do fechamento do carrinho, antes da finalização da compra. Cabe aos gestores do *marketplace* definir o momento mais propício de fazer a recomendação com o objetivo de maximizar o valor da compra pelo consumidor.

Por fim, um segundo ponto também não abordado nesta pesquisa foi o dimensionamento da infraestrutura tecnológica (*software* e *hardware*) necessária para seleção dos produtos no carrinho, identificação da Categoria_SR de cada um dos produtos, seleção da(s) regra(s) aplicáveis às Categorias_SR identificadas, definição do(s) produto(s) a serem recomendados e a recomendação propriamente dita na tela do cliente.

6 CONCLUSÃO

Esta pesquisa objetivou desenvolver um Sistema de Recomendação para um *marketplace online* do varejo de alimentos. Aproximadamente 30 mil carrinhos foram analisados, de aproximadamente 9 mil clientes exclusivos, de 16 estabelecimentos parceiros diferentes e que continham aproximadamente 750 mil produtos comprados. Foi necessário categorizar os produtos na variável Categorias_SR, pois, sua granularidade permitiu identificar regras de associação mais gerencialmente implementáveis pelo *marketplace.* A variável construída Categorias_SR ficou com uma granularidade de aproximadamente 5 mil grupos exclusivos, mas apenas 2.899 foram associados aos produtos contidos no histórico dos carrinhos de compra analisado. Após a definição das medidas de confiança, suporte e *lift* no algoritmo *apriori*, foram encontradas 9.965 regras válidas de recomendação, sendo 2.348

regras no formato 1 → 1, 3.764 regras no formato 2 → 1, 2.904 regras no formato 3 → 1, 895 regras no formato 4 → 1 e 54 regras no formato 5 → 1. Esta pesquisa propôs a medida de *lift* ponderado para priorização das regras encontradas. As 9.965 regras válidas contemplaram apenas 135 Categorias_SR, o que denotou uma alta concentração destes agrupamentos no Sistema de Recomendação elaborado. A análise da rede das 135 Categorias_SR apresentou 2.362 relações diferentes (arestas) entre estas Categorias_SR. As centralidades de grau (*degree*), de proximidade (*closeness*) e de intermediação (*betweenness*) foram calculadas para a rede com arestas não-ponderadas e ponderadas, o que permitiu classificar as Categorias_SR que devem ser priorizadas pelo *marketplace* para a melhoria da *performance* gerencial da empresa através do crescimento da receita, do crescimento do lucro e da concepção de ações de marketing mais eficientes e eficazes.

Uma limitação desta pesquisa foi a não realização de testes das regras encontradas diretamente na plataforma *online* de vendas do *marketplace*. Acredita-se que a utilização das regras encontradas durante o momento de compra do consumidor será capaz de aumentar, por exemplo, a quantidade de produtos comprados e o valor total da compra dos consumidores, mas tal inferência não foi validada através de uma verificação empírica. Sendo assim, recomenda-se para pesquisas futuras o teste das regras encontradas no momento da compra dos consumidores, assim como a construção de algoritmos de *reinforcement learning* para refinar e aprimorar o Sistema de Recomendação em tempo real.

7 REFERÊNCIAS

ALBERT, R., BARBASI, A-L. Statistical Mechanics of Complex Networks. **Reviews of Modern Physics**, V. 74, 2002.

BIGGS, N. L., LLOYD, E. K., WILSON, R. J. **Graph Theory 1736-1936**. Clarendon Press, 1999.

BRADLOW, E. T., GANGWAR, M., KOPALLE, P., VOLETI, S. The Role of Big Data and Predictive Analytics in Retailing. **Journal of Retailing**. V. 93, I. 1, p. 79-95, 2017. DOI doi.org/10.1016/j.jretai.2016.12.004.

CAVIQUE, L. A Scalable Algorithm for the Market Basket Analysis. **Journal of Retailing and Consumer Services**. V. 14, I. 6, p. 400-407, 2007. DOI doi.org/10.1016/j.jretconser.2007.02.003.

DANGALCHEV, C. Residual Closeness in Networks. **Physica A: Statistical Mechanics and its Applications**. V. 365, I. 2, p. 556-564, 2006. DOI doi.org/10.1016/j.physa.2005.12.020.

FREEMAN, L. C. A Set of Measures of Centrality Based on Betweenness. **Sociometry**. V. 40, I. 1, p. 35-41, 1977. DOI doi.org/10.2307/3033543.

GREWAL, D., ROGGEVEEN, A. L., NORDFÄLT, J. The Future of Retailing. **Journal of Retailing**. V. 93, I. 1, p. 1-6, 2017. DOI doi.org/10.1016/j.jretai.2016.12.008.

KOTLER, P., KARTAJAYA, H., SETIAWAN, I. **Marketing 4.0**: do tradicional ao digital. Rio de Janeiro: Sextante, 2017.

MALHOTRA, N. K. **Pesquisa de Marketing**: uma orientação aplicada. 4. ed. Porto Alegre: Bookman, 2006.

MARCONI, M. A., LAKATOS, E. M. **Técnicas de Pesquisa**: planejamento e execução de pesquisas, amostragens e técnicas de pesquisas, elaboração, análise e interpretação de dados. 3. ed. São Paulo: Atlas, 1996.

NEWMAN, M. E. J. **The Mathematics of Networks**. Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI, 2006.

RICCI, F., ROKACH, L., SHAPIRA, B., KANTOR, P. B. (eds) **Recommender Systems Handbook**. Springer Science. Springer, Boston, MA, 2011. DOI doi.org/10.1007/978-0-387-85820-3.

YANAMOTO, C., H. **Visualização como Suporte à Extração e Exploração de Regras de Associação**. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação ICMC/USP São Carlos. São Paulo, 2009.