

Otimização de Vendas no E-commerce de Supermercados por Meio de Recomendação Baseada em Grafos

João Francisco Teles da Silva

Centro Federal de Educação Tecnológica de Minas Gerais

Divinópolis, Brasil

joaoteles0505@gmail.com

Abstract—This paper investigates the optimization of sales in e-commerce for supermarkets through a recommendation system based on graph theory. The study uses transaction data from Instacart to build graphs that represent product co-occurrence, facilitating the identification of associations and patterns in consumer behavior. By modeling product co-occurrences in purchase baskets, the study explores how graph-based recommendations can enhance cross-selling strategies and maximize sales. The research compares two departments: Produce (high-density) and Personal Care (low-density), showing how product interconnectivity influences recommendation systems. The Maximum Spanning Tree (MST) algorithm is employed to prioritize the most significant connections in the graph, reducing complexity and improving the accuracy of recommendations. The results demonstrate the potential of graph-based analysis in optimizing sales strategies and consumer experience in supermarket e-commerce.

Index Terms—Basket analysis, graph theory, e-commerce, recommendation system, co-occurrence, sales maximization.

I. INTRODUÇÃO

A evolução tecnológica tem transformado profundamente a forma como os consumidores realizam suas compras, especialmente no setor de supermercados. Para as empresas, essa mudança representa não apenas uma ampliação do alcance de mercado, mas também oportunidades para reduzir custos operacionais e aprimorar estratégias de venda.

Entretanto, apesar do acesso crescente a dados de consumo dos clientes, muitos supermercados ainda não exploram plenamente essas informações para otimizar suas vendas.

Dessa maneira, a análise das cestas dos usuários surge como uma ferramenta valiosa para compreender os hábitos de compra desses consumidores e identificar associações relevantes entre produtos. A descoberta de padrões de consumo permite a implementação de recomendações, estratégias de precificação dinâmica e ações promocionais mais eficazes, impulsionando a maximização das vendas no comércio digital.

Neste estudo, a análise de cestas de compras é abordada por meio do uso de grafos, que modelaram a coocorrência de itens nas compras. Com base nessa estrutura, foi desenvolvido um simulador de sistema de recomendação, o qual possibilita explorar, por meio dos dados de coocorrência, oportunidades para aprimorar recomendações e otimizar estratégias de venda.

Para demonstrar o impacto da análise de cestas de compras na maximização das vendas, este trabalho está estruturado da

seguinte forma: a Seção 2 apresenta o referencial teórico sobre teoria dos grafos, sistemas de recomendação e maximização de vendas. A Seção 3 detalha os métodos, incluindo pré-processamento, modelagem dos grafos e análise de conexões entre produtos. A Seção 4 expõe os resultados e discussões, como padrões de compra e associações entre itens. Por fim, a Seção 5 discute as conclusões e aplicações da abordagem proposta no e-commerce.

II. REFERENCIAL TEÓRICO

O crescimento do comércio eletrônico exige soluções eficientes para otimizar vendas e melhorar a experiência do consumidor. Nesse contexto, sistemas de recomendação baseados em grafos têm se mostrado promissores, com estudos evidenciando o impacto da modelagem de coocorrência de produtos e estratégias de maximização de vendas [1], [3].

Ohashi et al. (2021) destacam a importância dos sistemas de recomendação na automação da força de vendas (SFA), permitindo que vendedores sugiram produtos com base nas compras do cliente [1]. Essa abordagem é essencial para grandes portfólios, onde identificar combinações de itens ideais pode ser desafiador. Seguindo essa linha, este estudo utiliza a análise de cestas de compras para modelar a coocorrência de produtos e aprimorar recomendações.

A teoria dos grafos tem sido amplamente aplicada na modelagem de relações entre produtos, possibilitando a criação de redes que representam conexões entre itens frequentemente adquiridos juntos. Maskar et al. (2020) empregaram a Árvore Geradora Máxima (MST) para reduzir a complexidade da rede e destacar relações mais relevantes [2]. Neste trabalho, essa técnica foi utilizada para filtrar as conexões mais significativas, garantindo recomendações baseadas em padrões sólidos de compra.

Além disso, as estratégias de cross-selling são fundamentais para impulsionar vendas. Campelo et al. (2022) ressaltam como a recomendação de produtos complementares pode aumentar o número de compras dos clientes [3]. Neste estudo, a análise da coocorrência permitiu sugerir itens complementares, otimizando estratégias comerciais.

A gestão de grandes volumes de dados no e-commerce exige ferramentas automatizadas para extrair padrões de compra com eficiência [1]. Este estudo investiga como a modelagem

de grafos pode otimizar recomendações e maximizar vendas, avançando na aplicação dessa abordagem para estratégias comerciais mais eficazes.

III. METODOLOGIA

A. Dados

Para o estudo, foi utilizado a base de dados disponibilizada pelo Instacart, um serviço de compras e entregas de supermercado. A base contém mais de 3 milhões de pedidos realizados por mais de 200.000 usuários, abrangendo um histórico de compras organizado em múltiplos arquivos relacionais, permitindo uma análise mais valiosa para o estudo.

Os dados obtidos pelo Instacart é composta por um conjunto de arquivos que descrevem os pedidos dos clientes ao longo do tempo. Cada usuário possui entre 4 e 100 pedidos registrados, possibilitando a análise de padrões de recompra e hábitos de consumo.

TABLE I
PRINCIPAIS ATRIBUTOS DA BASE DE DADOS

Atributo	Descrição
product_id	Identificador único do produto
product_name	Identificador do nome do produto
user_id	Identificador do usuário
order_id	Identificador do pedido
department_id	Identificador do departamento
department_name	Identificador do nome do departamento
aisle_id	Identificador do corredor
aisle_name	Identificador do nome do corredor

Além dos atributos, é importante mencionar a distribuição de tais dados em 6 diferentes arquivos relacionais, os quais foram divididos da seguinte maneira:

TABLE II
PRINCIPAIS ARQUIVOS DE DADOS

Arquivo	Descrição
products.csv	Lista de produtos com seus IDs, categorias e departamentos.
aisle.csv	Identificadores e nomes das categorias de produtos.
department.csv	Identificadores e nomes dos departamentos.
order_products_prior.csv	Histórico de compras com produtos e reordenações.
order_products_train.csv	Dados de pedidos usados para treinamento.
orders.csv	Metadados dos pedidos, incluindo usuário, horário e frequência.

B. Pré-processamento

Para modelar a coocorrência de produtos em cestas de compras, foi necessário um pré-processamento dos dados a fim de selecionar as informações mais relevantes e garantir que estivessem organizadas de maneira apropriada para a construção do grafo.

Visto que, a base de dados da InstaCart contém múltiplos arquivos e dados, foram selecionados somente aqueles que possibilitariam uma melhor análise dos dados. Sendo assim, os dados escolhidos para a modelagem incluíram os arquivos

order_products_prior.csv, o qual contém o histórico de produtos comprados em cada pedido, o arquivo products.csv, o qual contém o nome e departamento correspondente de cada produto e o arquivo departments.csv, que contém a categorização de todos os produtos.

O pré-processamento foi realizado através da leitura dos arquivos pela biblioteca pandas, da linguagem python, assim foi utilizada a função merge para fusão dos dados, os quais por fim foram salvos em um novo arquivo de dados, nomeado de processed_data_with_department.csv.

Dessa maneira, com essa fusão os dados foram agrupados de forma que cada produto estivesse relacionado a seu nome e departamento, respectivamente.

Esse pré-processamento foi essencial para preparação dos dados, garantindo uma melhor organização e estruturação dos dados.

C. Estruturação do Grafo

A estruturação foi realizada por meio de um grafo não direcionado e ponderado, criado com o objetivo de analisar as coocorrências entre os produtos, ou seja, a quantidade de vezes que dois produtos foram comprados juntos, dessa maneira, os nós, arestas e pesos foram modelados da seguinte forma:

- 1) **Nós** representam os produtos;
- 2) **Arestas** indicam que dois produtos foram comprados juntos em um mesmo pedido;
- 3) **Pesos das arestas** correspondem ao número de vezes que essa coocorrência ocorreu.

Para tornar a análise mais eficiente, os produtos foram agrupados por departamento, resultando na geração de 21 subgrafos, um para cada departamento presente na base de dados. Essa segmentação reduziu a complexidade do grafo e permitiu uma análise mais precisa das relações entre produtos dentro de cada categoria.

Devido ao grande número de linhas do arquivo, no código foi realizado a divisão do arquivo por chunks, visando não sobrecarregar a memória ram da máquina. Criado os chunks o programa agrupou os pedidos por order_id, dividiu os produtos dentro de cada pedido por departamento e gerou combinações para cada par de produtos que foram comprados juntos, criando posteriormente o grafo de cada departamento através da biblioteca networkX.

Cada grafo foi salvo em um arquivo .pickle com o objetivo de não refazer todo o processamento toda vez que esse fosse utilizado, sendo necessário somente carregá-lo no programa.

A visualização dos grafos foi feita utilizando matplotlib para impressão de imagens estáticas e Plotly, permitindo uma navegação interativa pelos nós e conexões. A Figura 1 apresenta um exemplo de um dos grafos gerados, correspondente ao departamento "Bulk". Essa imagem ilustra a estrutura de coocorrência dos produtos dentro de uma categoria específica.

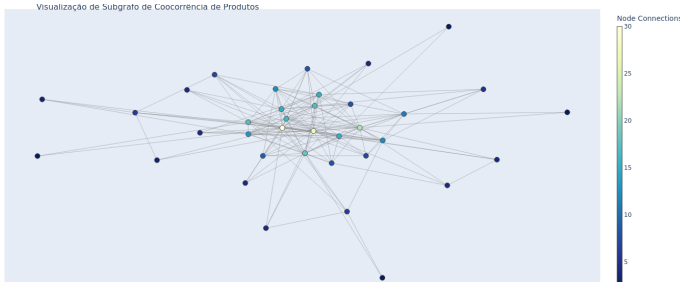


Fig. 1. Exemplo de grafo gerado para um departamento (Bulk).

D. Densidade do Grafo

Visando fazer uma análise mais aprofundada dentro de um determinado departamento, o estudo em questão foi realizado encima de dois grafos determinados pelas suas densidades, sendo um deles com alta densidade e outro com baixa densidade.

A densidade de um grafo é uma métrica utilizada para avaliar o nível de conectividade entre os nós, sendo definida como a razão entre o número de arestas $|E|$ e o número máximo de conexões possíveis $\binom{|V|}{2}$:

$$D = \frac{|E|}{\binom{|V|}{2}} = \frac{2|E|}{|V|(|V| - 1)} \quad (1)$$

onde $|E|$ representa o número total de arestas e $|V|$ é o número de nós no grafo.

A escolha do critério de densidade foi utilizada para determinar quais dos departamentos seriam analisados, visto que, com um grafo de alta densidade e um de baixa densidade seria possível comparar diferentes padrões de conexão entre produtos e suas influências nas vendas.

No grafo mais denso, onde há muitas conexões entre os produtos, é possível entender como produtos fortemente conectados influenciam na venda uns dos outros, além disso, esse grafo indica que os clientes costumam comprar vários produtos juntos com frequência, o que pode indicar padrões de consumo mais previsíveis e estratégias de recomendação diretas.

No grafo mais esparsa, cuja densidade é baixa, a dinâmica de coocorrência pode ser diferente, permitindo avaliar como produtos menos conectados afetam as compras. Nesse sentido, as compras desse departamento são mais ocasionais e menos correlacionadas.

A fórmula utilizada para o cálculo da densidade foi a 1, visto que essa diz respeito a uma equação de densidade para um grafo não direcionado, como o do estudo. Analisando as densidades dos 21 grafos, obteve-se os resultados apresentados na tabela abaixo:

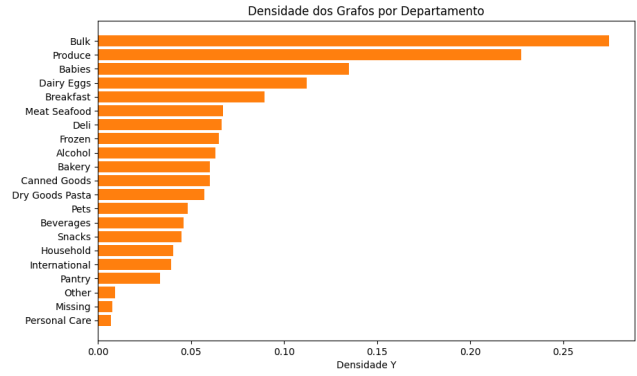


Fig. 2. Densidade dos Grafos de cada Departamento.

Embora o grafo mais denso obtido tenha sido o do departamento Bulk, o grafo escolhido para a análise como representante de maior densidade foi o segundo colocado, Produce. Essa escolha justifica-se pelo fato de que o departamento Bulk possui um número reduzido de produtos, resultando em uma conectividade quase completa entre os itens. Esse alto grau de interconectividade, apesar de gerar uma densidade elevada, não reflete necessariamente um padrão realista de compras generalizadas, mas sim um efeito da baixa cardinalidade do conjunto de produtos.

Por outro lado, o departamento Produce contém um número significativamente maior de produtos (1.684 itens) e conexões entre eles (322.514 conexões), possibilitando uma análise mais robusta e representativa da relação entre coocorrências de compra. A maior quantidade de produtos permite identificar padrões de associação mais variados, tornando a análise mais relevante para compreender como a densidade impacta as interações entre os itens dentro de um ambiente de compras real.

Além do grafo com alta densidade, Produce, também foi escolhido um grafo de baixa densidade para o estudo, o qual entre os resultados que foram obtidos resultou em ser o grafo do departamento Personal Care, com alto número de produtos (6493 itens), mas baixa conexão entre eles (147.171 conexões).

A geração dos grafos evidenciou claramente a diferença na densidade entre os departamentos analisados. O grafo do departamento Produce apresenta uma alta densidade, com um grande número de conexões, resultando em uma estrutura visualmente mais compacta e interligada. Em contraste, o grafo do departamento Personal Care mostrou-se significativamente mais esparsa, evidenciando uma menor conectividade entre os produtos. A visualização de ambos pode ser observada nas Figuras 3 e 4.

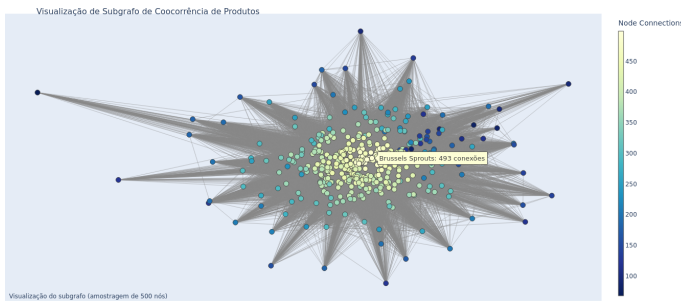


Fig. 3. Grafo com Densidade Alta (Produce).

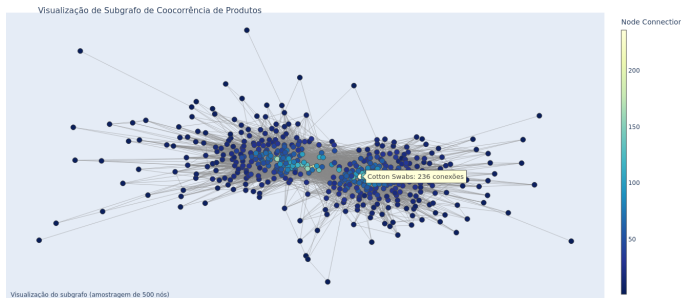


Fig. 4. Grafo com Densidade Baixa (Personal Care).

E. MST Máxima no Grafo

Para otimizar a análise das relações entre produtos e reduzir a complexidade do grafo original, aplicou-se a Árvore Geradora Máxima (MST Máxima) a cada um dos 21 grafos de departamentos gerados. A MST Máxima preservou apenas as conexões mais fortes entre os produtos, eliminando arestas de menor relevância e garantindo que o sistema de recomendação utilizasse apenas as relações de maior impacto.

A construção da MST Máxima baseou-se no conceito da Árvore Geradora Mínima (MST Mínima), que normalmente encontra a estrutura de menor soma de pesos. No entanto, para priorizar as conexões mais relevantes, foi invertido os pesos das arestas, aplicando o algoritmo de Kruskal para extrair a MST e, em seguida, restaurando os valores originais.

A implementação foi realizada em Python, utilizando a biblioteca NetworkX, que oferece suporte para operações em grafos, incluindo a função `minimum_spanning_tree()`, responsável pela extração da MST. Para garantir que apenas as conexões mais fortes fossem preservadas, os pesos das arestas foram multiplicados por -1 antes da aplicação do algoritmo, revertendo os valores posteriormente para recuperar os pesos originais.

Para avaliar o impacto da MST Máxima na estrutura dos grafos, foi analisado a variação no número de arestas e no peso médio das conexões antes e depois da aplicação do método.

Os resultados indicam uma redução expressiva no número de arestas, uma vez que a MST Máxima mantém apenas $|V| - 1$ conexões, eliminando associações de menor impacto. Esse efeito pode ser visualizado na Figura 5, que compara o número de arestas antes e depois da aplicação do método.

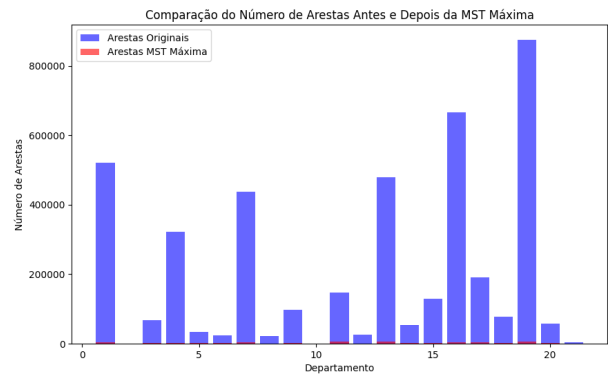


Fig. 5. Comparação do número de arestas antes e depois da MST Máxima.

Além disso, a média dos pesos das arestas aumentou substancialmente em todos os departamentos, conforme apresentado na Figura 6. Isso demonstra que a MST Máxima priorizou as conexões mais frequentes, garantindo que o sistema de recomendação utilizasse apenas as relações de maior relevância para sugerir produtos.

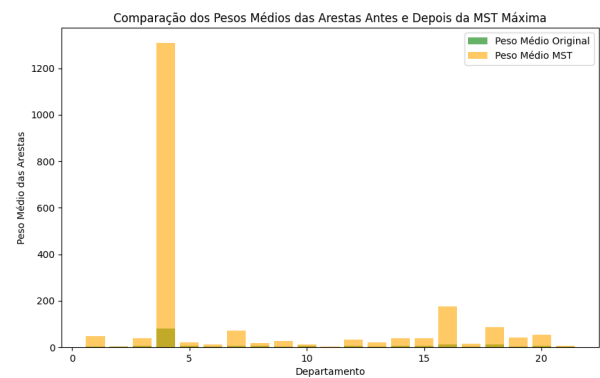


Fig. 6. Comparação do peso médio das arestas antes e depois da MST Máxima.

Esses fatores contribuíram para otimizar a recomendação de produtos no e-commerce, tornando o sistema mais eficiente e centrado nas relações de compra mais significativas.

F. Sistema de Recomendação

O sistema desenvolvido neste estudo se baseia na Filtragem Colaborativa Baseada em Itens (Item-Based Collaborative Filtering), onde a recomendação é feita a partir da coocorrência de produtos em compras anteriores. Para modelar essas relações, foi utilizado grafo de coocorrência estruturado.

A implementação foi desenvolvida em Python, utilizando as bibliotecas NetworkX para manipulação de grafos, Pandas para processamento de dados e Tkinter para a interface gráfica.

O processo de recomendação ocorre em três etapas:

- 1) O usuário seleciona um departamento do supermercado virtual;

- 2) O sistema carrega o grafo de coocorrência correspondente ao departamento;
- 3) O usuário insere produtos, e o sistema identifica os itens mais conectados, ordenando-os pelo peso das conexões.

Para otimizar o grafo, foi aplicada a Árvore Geradora Máxima (MST Máxima), garantindo que apenas as conexões mais relevantes sejam consideradas.

A recomendação segue a seguinte fórmula:

$$\text{Score}(q) = \sum_{p \in P} w(p, q) \quad (2)$$

onde:

- P é o conjunto de produtos informados pelo usuário;
- q representa os produtos candidatos à recomendação;
- $w(p, q)$ é o peso da conexão entre os produtos p e q .

Os produtos com maior score são os retornados como as melhores recomendações.

A interface gráfica, desenvolvida com Tkinter, permite que o usuário:

- Selecione um departamento e carregue seu grafo correspondente;
- Insira produtos manualmente para buscar recomendações;
- Visualize sugestões personalizadas com base na coocorrência dos itens.

A Figura 7 apresenta a interface gráfica do sistema.

Fig. 7. Interface do Sistema de Recomendação

O sistema também inclui um mecanismo de correção para sugestões de produtos caso um nome seja digitado incorretamente.

IV. RESULTADOS E DISCUSSÕES

Após a aplicação dos processos metodológicos descritos na seção anterior, foram obtidos resultados promissores que revelam informações relevantes sobre a correlação entre produtos, a centralidade de certos itens como pontos estratégicos para vendas e possíveis estratégias para impulsionar as compras dentro de cada departamento. Além disso, a estrutura dos grafos permitiu uma análise aprofundada da conectividade entre os produtos, auxiliando na identificação de padrões de compra e oportunidades para otimização das recomendações.

Como parte dessa análise, foi desenvolvido um sistema de recomendação baseado na coocorrência de produtos, que permitiu explorar a relação entre os itens e avaliar estratégias para sugerir-lhes de forma mais eficaz. Esse sistema possibilitou

uma melhor compreensão dos produtos mais influentes no grafo e das conexões que podem ser exploradas para incentivar compras combinadas, tornando-se uma ferramenta valiosa para maximizar as vendas no e-commerce.

Nas subseções seguintes, os resultados serão apresentados separadamente para os dois grafos estudados: o Produce, que possui maior densidade de conexões, e o Personal Care, que apresenta uma estrutura mais esparsa. Essa abordagem permite uma análise comparativa mais clara e detalhada sobre como a modelagem por grafos e o sistema de recomendação podem contribuir para o aumento da conectividade entre produtos e a otimização das estratégias de venda.

A. Produce

O departamento Produce corresponde ao setor do supermercado destinado à venda de frutas e vegetais frescos para consumo humano, podendo incluir também laticínios e nozes. Na base de dados analisada para o estudo, esse departamento contém 1684 produtos diferentes. Dentre todos os departamentos, o Produce é o que registra o maior volume de vendas, ultrapassando 9 milhões de vendas computadas.

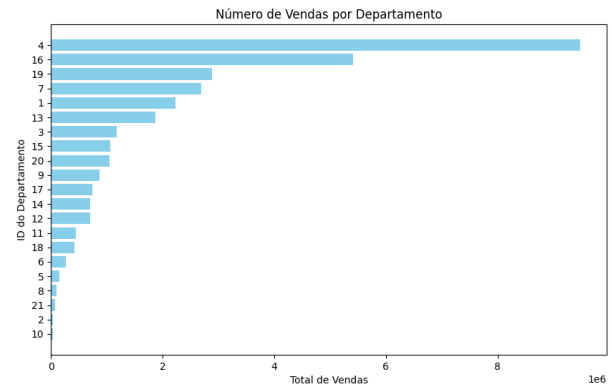


Fig. 8. Total de vendas por Departamento

Além do alto volume de vendas, a análise da densidade do grafo revelou que este departamento possui uma estrutura altamente conectada, indicando que os produtos dentro dele são frequentemente comprados em conjunto. Essa característica o torna um ponto estratégico para o supermercado virtual, não apenas pelo seu alto desempenho de vendas, mas pela possibilidade de fidelizar, atrair e direcionar clientes para outros setores do e-commerce.

Os produtos desse departamento são em sua maioria perecíveis e, por consequência, levam os consumidores a realizarem compras frequentes. Isso sugere que o departamento Produce pode funcionar como um ponto de ancoragem para a recorrência de compras no e-commerce. Ou seja, ao retornar para adquirir os itens desse departamento, o consumidor pode ser levado a comprar mais produtos, por meio do sistema de recomendações.

Dentro do grafo, é possível identificar produtos que se destacam por estarem altamente conectados a diversos outros itens. Esses produtos atuam como hubs dentro da rede de

compras e podem ser usados como gatilhos para impulsionar vendas complementares.

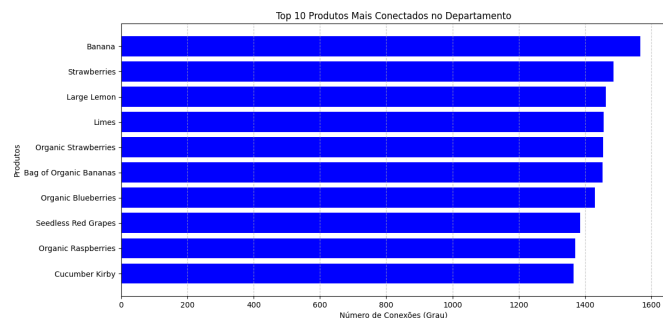


Fig. 9. Top 10 Produtos com mais conexões dentro de Produce

Ao analisar os produtos mais conectados, percebe-se que banana, morango e limão ocupam as primeiras posições, o que indica que eles desempenham um papel essencial na rotina alimentar dos consumidores. Além disso, há uma forte presença de produtos orgânicos na lista, reforçando a tendência de consumo saudável.

Esses produtos podem ser utilizados para campanhas de cross-selling e promoções direcionadas. Por exemplo, um desconto na compra de bananas e morangos pode incentivar os clientes a adquirir mais dentro do departamento.

Outra análise realizada partiu da extração da Maximum Spanning Tree (MST), a qual revelou as conexões mais relevantes dentro do grafo, indicando os pares de produtos mais frequentemente comprados juntos. Essas informações adquiridas podem ser aplicadas para estratégias de recomendação e disposição dos produtos na interface do e-commerce.

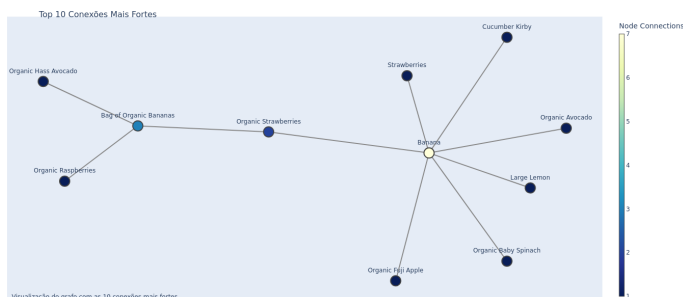


Fig. 10. Top 10 Produtos com conexões mais fortes dentro de Produce

A tabela abaixo mostra o peso das conexões entre esses produtos com conexão mais forte:

TABLE III
TOP 10 CONEXÕES MAIS FORTES

#	Conexão	Peso
1	Bag of Organic Bananas - Organic Hass Avocado	62341
2	Bag of Organic Bananas - Organic Strawberries	61626
3	Banana - Organic Strawberries	56156
4	Banana - Organic Avocado	53395
5	Organic Baby Spinach - Banana	51394
6	Banana - Strawberries	41231
7	Banana - Large Lemon	40880
8	Bag of Organic Bananas - Organic Raspberries	40501
9	Banana - Organic Fuji Apple	33942
10	Banana - Cucumber Kirby	32097

Foi observado que a banana se destacou como o item com o maior número de conexões fortes. Esse dado revelou que a banana pode ser utilizada para a criação de combos promocionais, combinando-a com outros produtos populares. Por exemplo, a forte associação entre Bag of Organic Bananas e Organic Hass Avocado sugere que esses itens fazem parte da rotina alimentar dos clientes. Oferecer um desconto na compra conjunta pode incentivar consumidores que normalmente comprariam apenas um deles a adicionarem o outro ao carrinho.

Além de impulsionar vendas diretas, essa análise também pode otimizar a interface do e-commerce. Produtos frequentemente comprados juntos podem ser exibidos lado a lado na plataforma, facilitando a navegação do usuário e incentivando compras adicionais. Essa estratégia, comum em lojas físicas por meio da organização das prateleiras, pode ser aplicada digitalmente para tornar a experiência do cliente mais fluida e intuitiva.

Além da análise estrutural do grafo e da identificação dos produtos mais conectados, o sistema de recomendação baseado em grafos implementado foi usado para avaliar como a conectividade entre os produtos pode ser utilizada para sugerir itens complementares e maximizar as vendas no e-commerce. Para validar o funcionamento do sistema, foram testadas diferentes cestas de compras e analisadas as recomendações geradas.

A primeira cesta de compras foi composta por produtos com alta popularidade dentro do departamento Produce, representando uma seleção voltada para consumidores que buscam conveniência e alimentação saudável. Os produtos selecionados foram:

- Mini Watermelon
- Santa Fe Style Salad Cafe Bowl
- Organic Whole Peeled Baby Carrots
- Organic Broccoli Crowns
- 3 lb Clementines

Ao processar essa cesta no sistema de recomendação, os seguintes produtos foram sugeridos como complementares:

Produto Recomendado
Bag of Organic Bananas
Banana
Cranberry Walnut Salad Cafe Bowl
Organic Pacific Rose Apples
Cherries Rainier Organic

TABLE IV

RECOMENDAÇÕES GERADAS PARA A PRIMEIRA CESTA DE COMPRAS.

Os resultados mostram que o sistema identificou corretamente a forte conectividade entre bananas e outros produtos de alta demanda, reforçando o papel desse item como um hub central no grafo. Além disso, a recomendação de saladas prontas e frutas orgânicas sugere que consumidores que compram essa cesta podem estar interessados em produtos naturais e prontos para consumo. Essa informação pode ser explorada comercialmente por meio de estratégias como a oferta de descontos progressivos em frutas frescas ou combos promocionais envolvendo saladas e snacks saudáveis.

Para avaliar o desempenho do sistema em um contexto diferente, uma segunda cesta de compras foi testada, incluindo produtos com forte relação no grafo, mas com menor volume de vendas individuais. Os itens selecionados foram:

- Yellow Bell Pepper
- Yukon Gold Potatoes Bag
- Organic Broccolini
- Ginger Root
- Small Hass Avocado

As recomendações geradas pelo sistema podem ser observadas na tabela abaixo:

Produto Recomendado
Banana
Orange Bell Pepper
Bunched Cilantro
Portabella Mushroom Caps
Large Lemon

TABLE V

RECOMENDAÇÕES GERADAS PARA A SEGUNDA CESTA DE COMPRAS.

Os produtos sugeridos pelo sistema revelam padrões interessantes de coocorrência, indicando que os consumidores que compram ingredientes para refeições caseiras tendem a adicionar temperos frescos, cogumelos e frutas cítricas ao carrinho. Isso sugere que o e-commerce pode explorar essa conexão para incentivar compras adicionais, oferecendo sugestões automáticas de itens complementares ou organizando a interface de forma a destacar produtos frequentemente adquiridos juntos.

B. Personal Care

O departamento Personal Care abrange produtos voltados para higiene e cuidados pessoais, sendo um setor essencial para o dia a dia dos consumidores. No entanto, apesar de contar com mais de 6.000 itens diferentes, esse departamento apresenta um volume de vendas relativamente baixo em comparação a outros setores do supermercado. Conforme ilustrado na Figura 8, ele ocupa a 14ª posição em vendas, ficando bem atrás de departamentos mais populares como Produce.

Além do menor volume de vendas, o grafo de coocorrência construído para este departamento revelou uma baixa densidade, ou seja, os produtos são comprados juntos com menor frequência. Isso resulta em um grafo menos conectado, onde apenas alguns produtos possuem um alto número de conexões, enquanto a maioria apresenta poucas ou nenhuma coocorrência

significativa, conforme ilustrado pelo grafo criado para esse departamento pela figura 4.

Apesar da baixa densidade geral do grafo, alguns produtos se destacam por apresentarem fortes conexões com outros itens. A Figura 11 mostra os 10 produtos mais conectados dentro do departamento Personal Care, classificados de acordo com seu grau no grafo.

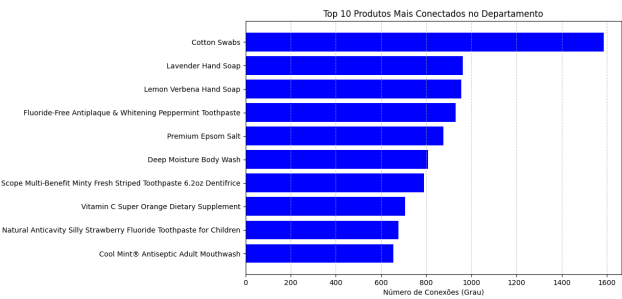


Fig. 11. Top 10 Produtos com mais conexões dentro de Personal Care

Essa análise revela que produtos como hastes flexíveis (Cotton Swabs), sabonetes líquidos (Lavender Hand Soap, Lemon Verbena Hand Soap), pastas de dente e enxaguantes bucais possuem um número significativo de conexões. Esses produtos funcionam como hubs dentro do grafo, ou seja, atuam como pontos centrais de compra, frequentemente adquiridos com outros itens.

Para ilustrar melhor essas relações, a Figura 12 destaca os pares de produtos que possuem as conexões mais fortes dentro do departamento Personal Care. Essas conexões representam os produtos que são comprados juntos com maior frequência, tornando-se essenciais para a construção de recomendações eficazes.

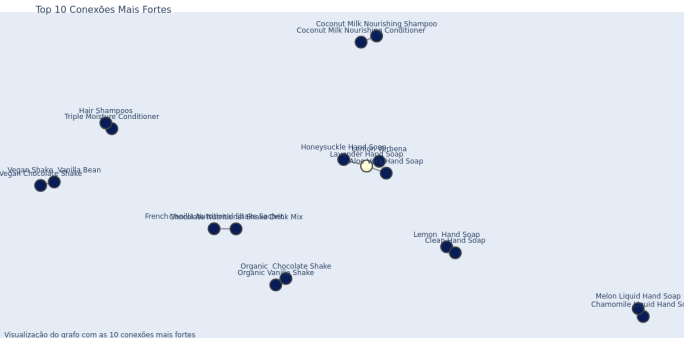


Fig. 12. Grafo destacando as conexões mais fortes entre produtos de Personal Care

Um exemplo dessa forte relação ocorre entre os produtos Lavender Hand Soap e Lemon Verbena Hand Soap, que apresentam a maior coocorrência no grafo (450 conexões). Esse padrão sugere que os clientes frequentemente compram mais de um tipo de sabonete líquido na mesma transação, possivelmente para uso em diferentes ambientes da casa ou para experimentar novas fragrâncias.

Esse comportamento também é observado em outras categorias, como os shakes nutricionais e os cuidados capilares, onde produtos de sabores ou fórmulas complementares são frequentemente adquiridos juntos. Esses padrões são fundamentais para o sistema de recomendação, pois ajudam a sugerir itens relevantes e incentivam compras combinadas.

Dessa forma, o sistema de recomendação pode explorar essas relações para oferecer sugestões mais precisas, aumentando a conectividade entre os produtos e impulsionando as vendas dentro do departamento Personal Care.

A partir da análise do grafo de coocorrência, foi desenvolvido um sistema de recomendação que utiliza a conectividade entre os produtos para sugerir itens complementares aos clientes. Esse sistema foi projetado para identificar padrões de compra e gerar recomendações, proporcionando uma experiência de compra mais eficiente e aumentando as vendas por meio de compras combinadas.

Para validar o desempenho do sistema, foram simuladas diferentes cestas de compras e foram analisadas as sugestões geradas. A primeira cesta de compras testada incluía produtos de categorias variadas dentro do departamento Personal Care, sendo composta por Cotton Swabs, Vanilla Whey Protein Powder, Vitamin C Super Orange Dietary Supplement e Fluoride-Free Antiplaque & Whitening Peppermint Toothpaste. Com base nesses produtos, o sistema gerou as seguintes recomendações:

Produto Recomendado
Vitamin C 1,000 mg Raspberry Dietary Supplement
Natural Chocolate Flavor Whey Protein Powder
Lavender Hand Soap
Tangerine 1,000 mg Vitamin C Dietary Supplement
Wicked Fresh! Mouthwash - Cool Mountain Mint

TABLE VI

RECOMENDAÇÕES GERADAS PARA A PRIMEIRA CESTA DE COMPRAS.

Os resultados mostram que o sistema identificou corretamente associações entre produtos que costumam ser adquiridos juntos. A recomendação de diferentes variações de Vitamin C e Whey Protein reflete o comportamento de clientes que preferem experimentar sabores alternativos ou manter uma variedade de suplementos em estoque. Esse tipo de recomendação pode ser explorado comercialmente por meio de ofertas de produtos combinados, promovendo descontos progressivos para compras múltiplas e aumentando o ticket médio por transação.

Além disso, a recomendação de um produto essencial de higiene pessoal, como o Lavender Hand Soap, reforça a presença de itens com alta conectividade dentro do departamento. A sugestão do Wicked Fresh! Mouthwash - Cool Mountain Mint segue a mesma lógica, indicando que produtos de higiene bucal frequentemente aparecem juntos nas compras. Isso demonstra que o sistema não apenas recomenda produtos similares, mas também identifica padrões de complementaridade que podem ser aproveitados para estratégias de cross-selling no e-commerce.

Para avaliar como o sistema responde a cestas mais homogêneas, uma segunda simulação foi realizada com produtos

voltados especificamente para cuidados pessoais, incluindo Dry Nose Oil, Lasting Color Shampoo, Daily Moisturizing Lotion Sheer Hydration e Sunscreen SPF 100+. O sistema gerou as seguintes recomendações:

Produto Recomendado
Lasting Color Conditioner
Wet Skin Kids Sunscreen Spray SPF 70+
Dry Touch Sunscreen SPF 100+
Instant Hand Sanitizer With Aloe
Geranium Scent Liquid Hand Soap Refill

TABLE VII

RECOMENDAÇÕES GERADAS PARA A SEGUNDA CESTA DE COMPRAS.

Os resultados confirmam que o sistema compreende relações diretas entre produtos de uma mesma linha, como a recomendação de Lasting Color Conditioner para clientes que compraram o shampoo correspondente. Essa estratégia pode ser usada para sugerir kits de produtos complementares, reforçando a conectividade entre os itens do departamento. Além disso, a recomendação de diferentes variações de protetores solares, como Wet Skin Kids Sunscreen Spray SPF 70+ e Dry Touch Sunscreen SPF 100+, mostra que os clientes frequentemente compram múltiplas opções desse tipo de produto, possivelmente para diferentes finalidades ou para membros da família.

A presença do Instant Hand Sanitizer With Aloe e do Geranium Scent Liquid Hand Soap Refill entre as recomendações sugere que o sistema também reconhece padrões mais amplos de consumo, incluindo produtos de higiene geral dentro de cestas voltadas para cuidados pessoais. Isso indica que o sistema pode ser utilizado para aumentar a conectividade do grafo, associando itens menos populares a produtos altamente conectados, o que contribui para um aumento na densidade da rede ao longo do tempo.

Os resultados obtidos evidenciam que o sistema de recomendação baseado em grafos não apenas melhora a experiência do consumidor ao oferecer sugestões mais precisas, mas também serve como uma ferramenta estratégica para maximizar as vendas no e-commerce. Ao sugerir produtos complementares e fortalecer as relações entre os itens do departamento, o sistema permite aumentar a conectividade entre produtos, otimizar vendas cruzadas, criar novas oportunidades comerciais e gerar um ciclo contínuo de crescimento de rede.

V. CONCLUSÃO

Este estudo investigou a aplicação de análise de cestas de compras baseada em grafos para otimizar recomendações e maximizar vendas no e-commerce de supermercados.

Utilizando dados da plataforma Instacart, foram construídos grafos representando a coocorrência de produtos, permitindo uma modelagem detalhada dos padrões de compra dos consumidores. A abordagem proposta demonstrou que a teoria dos grafos é uma ferramenta eficaz para estruturar as relações entre itens e explorar conexões que podem ser usadas estrategicamente para impulsionar vendas.

A análise comparativa entre dois departamentos — Produce (alta densidade) e Personal Care (baixa densidade) — permitiu compreender diferentes dinâmicas de consumo e suas implicações para estratégias comerciais.

No caso de Produce, a forte conectividade entre os produtos possibilitou a identificação de hubs de compra, como a banana, que pode ser utilizada para promoções e ações de cross-selling. Já no departamento Personal Care, onde a conectividade entre os itens é mais esparsa, foi possível mapear produtos-chave que atuam como pontos de ancoragem para recomendações, como sabonetes líquidos e suplementos vitamínicos.

A implementação de um sistema de recomendação baseado na Árvore Geradora Máxima (Maximum Spanning Tree - MST) mostrou-se eficiente ao priorizar as conexões mais relevantes, reduzindo a complexidade do grafo e tornando as sugestões mais precisas. Os testes realizados com diferentes cestas de compras validaram a eficácia do sistema, que conseguiu identificar corretamente associações entre produtos frequentemente adquiridos juntos e sugerir itens complementares com base na estrutura da rede de coocorrência.

Os resultados obtidos destacam o potencial da modelagem por grafos para transformar estratégias de recomendação no e-commerce de supermercados, proporcionando recomendações mais precisas e alinhadas aos hábitos de consumo dos clientes. Além de otimizar e impulsionar o faturamento, incentivando compras complementares e fortalecendo a fidelização dos consumidores.

[14] Disponível em: <https://github.com/joaofranciscoteles/Trabalho-Final.git>

REFERENCES

- [1] Ohashi et al., ‘Sistemas de Recomendação em Automação de Vendas’, 2021.
- [2] Maskar et al., ‘Árvores Geradoras Máximas e Análise de Dados’, 2020.
- [3] Campelo et al., ‘Maximização de Vendas no Comércio Digital’, 2022.
- [4] Medium, ‘Pré-processamento de dados com Python’, 2022. Disponível em: <https://medium.com/data-hackers/pr%C3%A9-processamento-de-dados-com-python-53b95bcf5ff4>
- [5] Medium, ‘Market Basket Analysis do Instacart’, 2022. Disponível em: <https://medium.com/@ndosanjos/market-basket-analysis-do-instacart-bbf6b4e0403a>
- [6] Medium, ‘Network Models for Recommender Systems’, 2022. Disponível em: <https://medium.com/dunhumby-data-science-engineering/network-models-for-recommender-systems-7f0d6d210ccf>
- [7] Medium, ‘Entendendo Sistemas de Recomendação’, 2022. Disponível em: <https://arturlunardi.medium.com/entendendo-sistemas-de-recomenda%C3%A7%C3%A3o-c50a20856394>
- [8] Medium, ‘Instacart Market Basket Analysis’, 2022. Disponível em: <https://medium.com/kaggle-blog/instacart-market-basket-analysis-feda2700cded>
- [9] Medium, ‘Recommendation System for Retail Customer’, 2022. Disponível em: <https://taufik-azri.medium.com/recommendation-system-for-retail-customer-3f0f80b84221>
- [10] Wikipedia, ‘Dense graph’, 2022. Disponível em: https://en.wikipedia.org/wiki/Dense_graph
- [11] USP, ‘Algoritmos para Grafos: MST’, 2022. Disponível em: https://www.ime.usp.br/~pf/algoritmos_para_grafos/aulas/mst.html
- [12] Medium, ‘Grafos e Recomendação de Produtos: Comparando Spark GraphFrames com Neo4j’, 2022. Disponível em: <https://medium.com/neurolake/grafos-e-recomenda%C3%A7%C3%A3o-de-produtos-comparando-spark-graphframes-com-neo4j>
- [13] Medium, ‘Modelando e Analisando Relacionamentos em Grafos com Docker, Python, Neo4j’, 2022. Disponível em: <https://medium.com/xp-inc/modelando-e-analisando-relacionamentos-grafos-com-docker-python-neo4j>