JURNAL EMACS
[Engineering, MAthematics and Computer Science] Vol.6 No.3 September 2024: 215-218 e-ISSN: 2686-2573 DOI: 10.21512/emacsjournal.v6i3.11693

Building Customer and Product Networks with Cosine Similarity in Graph Analytics for Deep Customer Insight

Aan Albone

Data Science Program, Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
aan.albone@binus.ac.id

Correspondence: aan.albone@binus.ac.id

Abstract – *The goal of social networks is to establish a link that facilitates information sharing and product recommendations between users. For the purpose of comprehending and evaluating links, relationships, and networks, graph theory is indispensable. Our customer network allows us to engage current customers with more products from similar customers and recommend products from one customer to another that are connected by a cosine similarity to the 70% above. If the price of a product is higher than that of comparable products, we can observe the demand for that product in the product network.*

Keywords: *Social Networking; Graph Analysis; Cosine Similarity; Deep Customer Insight*

I. INTRODUCTION

Normally, we predict customer's decisions based on their historical transactions and profile (e.g., age, gender, address, etc.). For instance, customers who want to buy a new handphone or electronic device will decide based on their experience and preference. They will also base their choices on the social context or external influence in which they find themselves; in this case, they get suggestions from their friends or other people who have similarities in terms of profile and preference.

Customers can also make product comparisons based on product characteristics and price. The product's price is an important factor in the customer's decision. We need to analyze the price gap between one product and similar products in order to know the demand for a product if the product's price is higher than similar products.

The idea of graph theory can be applied to the mining and utilization of this type of data about connections, relationships, or networks. At the same time, knowledge graph embeddings are gaining more and more popularity, explicitly considering various types of relations in the data. (Egorovaa et. al, 2022)

Social networks, such as those based on friendship or business, are another type of network in real life. People make up nodes in friendship networks, while social interactions make up edges. Over the past ten years, social networks like Facebook and LinkedIn have grown in popularity. Typically, these social networks aim to create a connection that helps their customers exchange information and provide product recommendations. (Koranyi et. al, 2022)

In the past, we found difficulties in deriving information from networks, evaluating a network's community structure, and building networks out of unstructured data. Nowadays, graph analytics has proven to be a useful resource for addressing these problems. (Campbell et. al, 2013)

The first step is always to find a relevant network and create it using the available data. In retail, we have four primary data sources: customer, supplier, product, and transaction. We will create customer and product networks with cosine similarity in graph analysis in order to get deep customer insight. (1) smart cross-selling and (2) price gap analysis

II. METHODS

Relational databases are great for creating lists but terrible for managing networks of diverse entities. A knowledge graph, based on graph database technology, is built to handle a diverse network of processes and entities. In a knowledge graph, we have nodes that represent people, events, places, resources, documents, etc. And we have relationships (edges) that represent links between the nodes.

In Graph theory, networks comprise independent components known as nodes (or vertices) connected by edges. (Thangavel et. al,, 2023)

A graph database is a way to store data that is based on a graph. Instead of storing the data in a table, with each user entry having references to the product entries in the product table (among other data), all of the data is stored in nodes on a graph, with edges connecting related data. We will use 4 nodes: customer, transaction, product, and supplier, which will be connected by an edge.

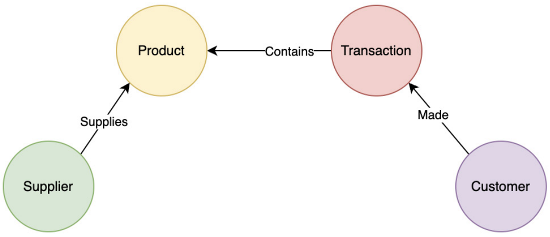

Figure 1. Four primary data sources

An emerging area of data analytics called graph technology enables many different kinds of organizations to comprehend their data more deeply. An intimate shared understanding of the customer's expressed and implicit present and future needs is known as deep customer insights, or DCI. A good insight goes beyond what consumers say and can influence future choices, attitudes, and behaviors. (Zarei et. al, 2022)

We will find deep customer insight from graph analysis with 2 use cases: (1) Smart cross selling and (2) Product's Price Gap Analysis.

2.1 Smart Cross Selling

Cross-selling is the practice of discovering existing customer's purchase preferences and engaging them with additional products or services, for increasing sales and customer loyalty. Products in transactions discover groups of customers with similar purchase preferences for cross-selling opportunities to increase sales and customer loyalty. (Lili Zhang et. al, 2021)

Data processing methods like feature integration and cosine similarity matrix generation are part of the cosine similarity process. Through careful analysis, the results show how effective the system is. The analysis demonstrates that cosine similarity can effectively identify and recommend similar customers, confirming the efficacy of the suggested recommendation system. (Smutek et. al, 2024)

The framework relies on a feature extraction algorithm to recommend communities or groups based on the cosine similarity measure and term frequency, which is combined with customer profiling. After the user provides data, the system follows their behavior, determines the relationships between them, and then suggests one or more communities based on their preferences. (Shaha et. al, 2021)

$$cos(A, B) = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_i A_i B_i}{\sqrt{\sum_i A_i^2}\sqrt{\sum_i B_i^2}}$$

This algorithm begins with two customer's purchases (Customer A & B), and the fact that at least one product item has been purchased by both of them determines the cosine similarity between those customers in the graph. The cosine similarity can be used to quantify the measurement of similarity between two customers.

A cosine similarity of 1 indicates that the two individuals are exactly alike, while a cosine similarity of 0 would indicate that the two documents have nothing in common. A value constrained by a range of 0 to 1 is called a cosine similarity. When the value approaches zero, it indicates that the two vectors are perpendicular or orthogonal to one another.

Customer's purchases:

- Rudy (Customer A) purchased product items: 1, 2, 3, 4, 5 and 6.
- Jhony (Customer B) purchased product items: 1, 2, 3, and 6.

Table I. Customer's Purchases

Product | Cust	item 1	item 2	item 3	item 4	item 5	item 6
Rudy (A)	1	1	1	1	1	1
Jhony (B)	1	1	1	0	0	1

Customer's purchases matrix:
Rudy (A) :[1, 1, 1, 1, 1, 0]
Jhony (B) :[1, 1, 1, 0, 0, 1]

Table II. Customer's Cosine Similarity

	item 1	item 2	item 3	item 4	item 5	item 6	Formula
A*B	1	1	1	0	0	1	4
Magnitude A	1	1	1	1	1	1	2.449489743
Magnitude B	1	1	1	0	0	1	2
Cosine Similarity							0.816496581

Calculate the cosine similarity: (4)/(2.2360679775*2) = 0.81 (81 percent similarity between Rudy and Jhony).

By using cosine similarity in graph analysis, we can determine how similar a vertex is to all other vertices. In Table II, the cosine similarity between A & B is 0.81 or 81%.

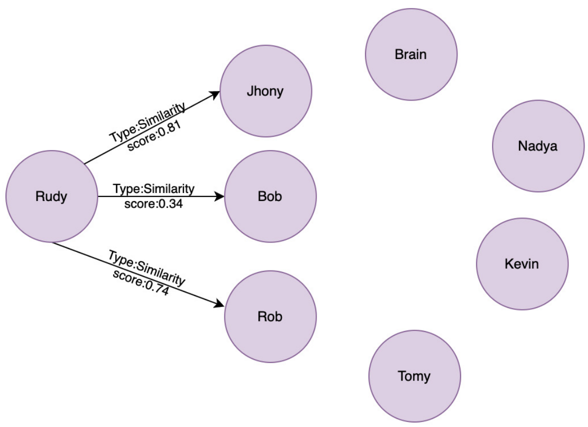

Figure 2. Customer Network with Cosine Similarity

As a result, we are able to recommend products from one customer to another that are connected by a cosine similarity to the 70% above. In this case, we can recommend product items 4 and 5 to Jhony.

2.2 Product's Price Gap Analysis

A fascinating feature of the financial markets is price gap analysis, which provides insightful information about the dynamics of asset prices. The study of these gaps is known as gap analysis, and it is essential to many fields, including technical analysis, trading, and competitive pricing strategies in business.

On comparatively transparent online marketplaces, one can compare the prices of the majority of goods. As a result, prices set by competitors are having an increasing impact on demand, so they shouldn't be disregarded. (Gerpott & Berends, 2022)

Price gaps indicate the variation in cost between one product and comparable products in this instance. Businesses are able to make more informed decisions about pricing and sales when they are aware of price differences for comparable goods.

We must first determine how similar the products are. A product's attributes are what define it. It implies that in order to compare two products, we are able to contrast their qualities. A collection of attributes, such as title, description, type, size, etc. Based on attributes, we develop a generic similarity function.

Our objective is to compute the degree of similarity between one product and other products that are described by any set of features. We'll examine how much the price differs between one product and similar products.

Table III. Product Attributes and Values

	Product A	Product B	Product C	Product D
Brand	HP	Acer	Sony	ASUS
Processor	Intel	Intel	Intel	Intel
Internal Memory	8GB	16GB	8 GB	8 GB
Internal Memory Type	DDR3-SDRAM	DDR3-SDRAM	DDR3-SDRAM	DDR3-SDRAM
Operating System	Windows 10 Pro	Windows 10 Pro	Windows 10 Pro	Windows 10 Pro
Hard Drive	256 GB	256 GB	256 GB	256 GB
Price	$1,049	$1,221	$1,236	$1,100

Table IV. Product's Cosine Similarity

	Intel	8GB	16 GB	DDR3-SDRAM	Windows 10 Pro	256GB	
Product A	1	1	0	1	1	1	
Product B	1	0	1	1	1	1	
							Formula
A*B	1	0	0	1	1	1	4
Magnitude A	1	1	0	1	1	1	2.236067977
Magnitude B	1	0	1	1	1	1	2.236067977
Cosine Similarity							0.8

This table demonstrates how similarity varies based on the selected product attributes. A product attributes is represented numerically by the averaged vector that was produced. After creating vector representations for every product, we can compare the degree of similarity between them using similarity metrics.

$$cos(A, B) = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_i A_i B_i}{\sqrt{\sum_i A_i^2}\sqrt{\sum_i B_i^2}}$$

Calculate the cosine similarity:

(4)/(2.2360679775*2.2360679775) = 0.80

(80 percent similarity between Product A and B).

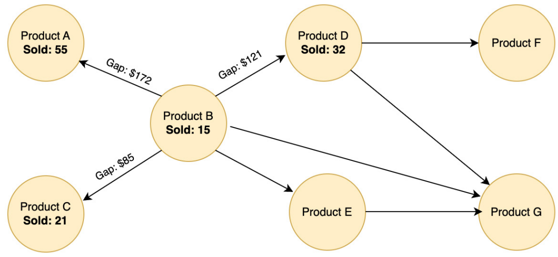

Figure 3. Product Network with Cosine Similarity

Product B has connections with products A, C, D, and G but not with F because the cosine similarity between product B and F is below 0.7, or 70%. The arrow's direction means that the price gap is positive from B to A, C, D, and G. We can see in this product network that if the products have similarities, then the product with a lower price will have more sales or demand.

A significant price difference between similar products can have several impacts on sales:

- If the price of the product is higher than similar products, consumer demand may decrease as consumers tend to seek the best value for their money.
- A product's ability to compete in the market may be impacted by a large price difference. If the product is more affordable than comparable products, this could make it more competitive and entice buyers to buy it. Higher prices, though, could cause the product to lose market share to less expensive options.

IV. CONCLUSION

Customer and product networks are developed from the Graph analysis. Graph theory is essential for understanding and assessing connections, relationships, and networks. The cosine similarity can be used to quantify the measurement of similarity between two objects in a graph.

Based on cosine similarity value, we are able to find deep customer insight through those 2 use cases: Smart Cross Selling, and Product's price gap analysis:

- According to the product's price gap analysis, a product's price difference from similar products can have a variety of effects on sales. For example, if a product costs more than similar products, consumer demand may decline because buyers typically look for the best deal.
- In product's price gap analysis, we can find a significant price difference between one product and similar products can have several impacts on sales, if the price of the product is higher than similar products, consumer demand may decrease as consumers tend to seek the best value for their money.

REFERENCES

Azim Zarei, Sima Alipour and Maryam Asgharinajib, (2022). Competitive pricing on online markets: a literature review, Journal of Revenue and Pricing Management (2022) 21:596–622

Azim Zarei, Sima Alipour and Maryam Asgharinajib, (2022). Discovering the Customer Insight using Netnography and Photography Methods, Asian Journal of Business Research Volume 12 Issue 3.

Elena Egorovaa, Gleb Glukhova, Egor Shikov. (2022). Customer transactional behaviour analysis through embedding Interpretation, 11th International Young Scientist Conference on Computational Science.

Rita Koranyi, Jose A. Mancera, & Michael Kaufmann, (2022). GDPR-Compliant Social Network Link Prediction in a Graph DBMS: The Case of Know-How Development at Beekeeper, International Journal of Molecular Sciences Switzerland Multidisciplinär Digital Publishing Institute (MDPI).

Shaha Al-Otaibi, Nourah Altwoijry, Alanoud Alqahtani, Latifah Aldheem, Mohrah Alqhatani, Nouf Alsuraiby, Sarah Alsaif, Shahla Albarrak, Cosine similarity-based algorithm for social networking recommendation, International Journal of Electrical and Computer Engineering (IJECE), Vol. 12, No. 2, April 2022, pp. 1881~1892

Thangavel, Shyamala Anto Mary, Kavitha, Deiwakumari. (2023). Graph Theory and Network Analysis: Exploring Connectivity in Computer Science, Journal of Namibian Studies, 35 S1.

Tomasz Smutek, Marcin Kowalski, Olena Ivashko, Robert Chmura, Justyna Sokolowska-Wozniak, (2022). A Graph-Based Recommendation System Leveraging Cosine Similarity for Enhanced Marketing Decisions, European Research Studies Journal, Volume XXVII, Special Issue 2, 83-93

William M. Campbell, Charlie K. Dagli, and Clifford J. Weinstein, (2013). Social Network Analysis with Content and Graphs, Lincoln Laboratory Journal, Volume 20, Number 1.