JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Abu Bakar, J., Min, C. K., Mohd Radzi, M. Z., Baharom, F., Yusof, Y., Saip, M. A., Muhamad Rasli, R., Zulkifli, M. A. A., Jamaludin, N., & Abuzaraida, M. A. (2025). AI-driven influencer and market analysis: A social network approach to measure e-commerce relationships. *Journal of Information and Communication Technology*, *24*(3), 1-21. https://doi.org/10.32890/jict2025.24.3.1

AI-driven Influencer and Market Analysis: A Social Network Approach to Measure E-Commerce Relationships

[1]Juhaida Abu Bakar, [2]Chong Kar Min, [3]Mohd Zulhisham Mohd Radzi, [4]Fauziah Baharom, [5]Yuhanis Yusof, [6]Mohamed Ali Saip, [7]Ruziana Muhamad Rasli, [8]Muhammad Amirul Ariff Zulkifli, [9]Nuratikah Jamaludin & [10]Mustafa Ali Abuzaraida
[1,2,4,5,6,8&9]Data Management & Software Solution Research Lab, School of Computing, Universiti Utara Malaysia, Malaysia
[3]Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Malaysia
[7]School of Multimedia Technology & Communication, Universiti Utara Malaysia, Malaysia
[10]Department of Computer Science, Faculty of Information Technology, Misurata University, Libya

[*1]juhaida.ab@uum.edu.my
[2]karminchong88@gmail.com
[3]mzulhisham@unimap.edu.my
[4]fauziah@uum.edu.my
[5]yuhanis@uum.edu.my
[6]mdali@uum.edu.my
[7]ruziana@uum.edu.my
[8]m.amirul.ariff@uum.edu.my
[9]nuratikah.jamaludin@uum.edu.my
[10]abuzaraida@it.misuratau.edu.ly
[*]Corresponding author

Received: 15/1/2025 Revised: 1/4/2025 Accepted: 1/7/2025 Published: 31/7/2025

ABSTRACT

Social network analysis is a process of studying social structures and relationships using graph theory and data analysis techniques. It involves mapping and measuring connections and entities in a network. However, on online selling platforms, identifying influential entities such as individuals and high-value products remains a challenge due to the complexity of customer and seller interactions. This study aims to assess seller performance and product lifetime value using AI-driven network analysis involving a measure of centrality. AI-driven network analysis utilises artificial intelligence (AI) to identify influential individuals and predict emerging trends in consumer engagement. It uses weighted degree and betweenness centrality to assess their effectiveness in identifying influential entities, including sellers, products, or organisations in a commercial network. Weighted degree centrality measures the strength and frequency of direct connections, while betweenness centrality identifies entities that act as intermediaries across different network segments. The analysis reveals that weighted degree centrality, with a value of 3190 for annual seller performance, is more closely aligned with actual sales performance and stakeholder assessments, making it a more suitable metric for supporting business decisions in this context. The findings demonstrate that AI-driven analytics enable businesses to consistently identify high-performing sellers and products based on their structural positions within the network. It contributes to the development of more targeted marketing strategies, systematic recognition of top performers, and enhanced customer engagement through data-informed decision-making. Future research may explore the integration of dynamic network modelling with multi-layered e-commerce networks, thereby increasing the depth of analysis across various platforms and industries.

Keywords: Social network analysis, AI-driven market analysis, influencer identification, weighted degree centrality, business decision-making.

INTRODUCTION

The COVID-19 pandemic has accelerated global digital transformation, rendering online platforms indispensable for communication, e-commerce, and digital marketing. In Malaysia, 83.1% of the population, or approximately 28.68 million people, actively use social media, underscoring its integral role in daily life (Howe, 2024). Social media advertising has become the primary method of brand discovery, preferred by 40.1% of consumers over traditional search engines (Wan, 2024). As a result, many businesses have shifted to online platforms to expand their reach and increase engagement. However, this transition has intensified competition in the digital marketplace, making it difficult for small and medium-sized enterprises (SME) to attract and retain customers.

Despite the vast amount of customer interaction data available, many businesses struggle to analyse and utilise it effectively. Limited or inadequate datasets hinder businesses’ ability to monitor trends and make informed decisions (Gómez et al., 2023). Traditional data analysis methods exacerbate this issue, as they struggle to manage large volumes of data or capture the complex relationships among sellers, customers, and products (Camacho et al., 2020; Kusuma et al., 2023). The complexity of social networks necessitates effective visualisation to facilitate an understanding of online community trends. Therefore, visualisation tools such as network graphs play a vital role in simplifying relationships, revealing hidden patterns, and presenting data in an accessible format (Singh et al., 2025; Mondal et al., 2024; Recuero et al., 2019; Smith et al., 2021). Without effective visualisation, businesses face challenges in extracting actionable insights and risk missing opportunities to understand customer behaviour, identify popular products, and refine their strategies.

To address these challenges, this study employs Social Network Analysis (SNA) to examine transaction data from Malaysian e-commerce businesses. The comparison between weighted degree and betweenness centrality measures is used as an experiment-based approach to investigate key relationships between sellers, products, and customers. The findings enhance the understanding of network structures and provide valuable insights for marketing and business optimisation. This research establishes a foundation for data-driven decision-making, enabling businesses to refine strategies, enhance customer retention, and improve product management. It offers a practical and cost-effective method for small businesses to analyse customer behaviour and market trends, supporting sustainable growth in the competitive digital marketplace.

RELATED WORKS

This section provides an overview of SNA, including its application in e-commerce, marketing, education, healthcare, social media, and energy management. It concludes by highlighting the related literature and identifying the gaps addressed by the proposed research.

Overview of SNA

SNA has seen significant growth due to its ability to extract valuable insights from web-based connections (Camacho et al., 2020). SNA examines relationships between entities within networks to identify influential individuals based on direct and indirect interactions (Laghridat & Essalih, 2023). It is structured into shell, mantle, and core layers, enabling the identification of key factors influencing the core layer (Chai & Ye, 2024). SNA has been applied in various domains, including project management, risk management, tourism, and supply chain management (Anugerah et al., 2022). Additionally, it has been used to analyse the Irish bio economy network and the role of brokers in involving marginalised groups (Harrahill et al., 2023). The integration of big data, social media, and sentiment analysis further enhances its potential applications.

SNA in E-Commerce, Marketing, and Education

SNA plays a crucial role in understanding consumer behaviour and optimising marketing strategies on platforms like Facebook, Instagram, and TikTok (Almashaleh et al., 2025; Arab, 2021). It has also contributed to market expansion in niche industries, such as the cuttlefish market (Gómez et al., 2023). E-commerce platforms use collaborative filtering recommender systems alongside SNA to identify influential customers and enhance product recommendations, though they remain vulnerable to shilling attacks (Rahardjo, 2019). Efforts to address these challenges include leveraging user trust, ratings, and contextual factors for improved recommendation systems (Hamidi & Moradi, 2024). Beyond marketing, SNA is also valuable in education, helping educators analyse social presence and learning behaviours in online learning communities (Norz et al., 2023). The technique has been applied to measure student engagement and academic performance in virtual classrooms (Chai & Ye, 2024).

SNA in Healthcare, Social Media, and Energy Management

SNA provides valuable insights into healthcare access within rural communities, revealing key network structures that influence service availability (Mondal et al., 2024). It also aids in analysing patient behaviour and healthcare decision-making (Strozzi et al., 2019). In social media studies, SNA is widely used to track sentiment trends, misinformation spread, and network roles such as activists and opinion leaders (Recuero et al., 2019). For instance, it has been instrumental in understanding public reactions to COVID-19 on Twitter (Kusuma et al., 2023). Beyond social media, SNA is applied in energy resource management, facilitating efficient decision-making in water distribution networks and carbon emission analysis (Latifi et al., 2023; Zhu et al., 2024). It has also been used to assess energy trade networks within the EU, identifying key stakeholders and vulnerabilities (Zehir et al., 2023). These applications demonstrate SNA's versatility across various sectors, extending beyond traditional social networks. Table 1 presents a review of related literature and identifies gaps in the proposed research.

Table 1

Review of Related Literature and Gaps in the Proposed Research

Study	Focus Area	Key Findings	Gap with Proposed Works
Camacho et al. (2020)	SNA Growth	SNA extracts insights from web-based networks	General application of SNA, not specific to e-commerce
Laghridat and Essalih (2023)	Network Influence	Identification of key influencers in networks	Lacks emphasis on influencer marketing and e-commerce
Harrahill et al. (2023)	Bio Economy Networks	Role of brokers in marginalised group involvement	Focuses on economic structures, not consumer relationships
Gómez et al. (2023)	Marketing	SNA facilitates niche market expansion (cuttlefish)	Limited to a specific market, not broader e-commerce
Hamidi and Moradi (2024)	Recommender Systems	Enhancing recommendations using user trust and ratings	Focuses on system security, not market influence
Norz et al. (2023)	Education	Measures social presence in online learning	Different domain; does not analyse social commerce
Kusuma et al. (2023)	Social Media	Analyses COVID-19 interactions on Twitter	Public sentiment analysis, not influencer impact

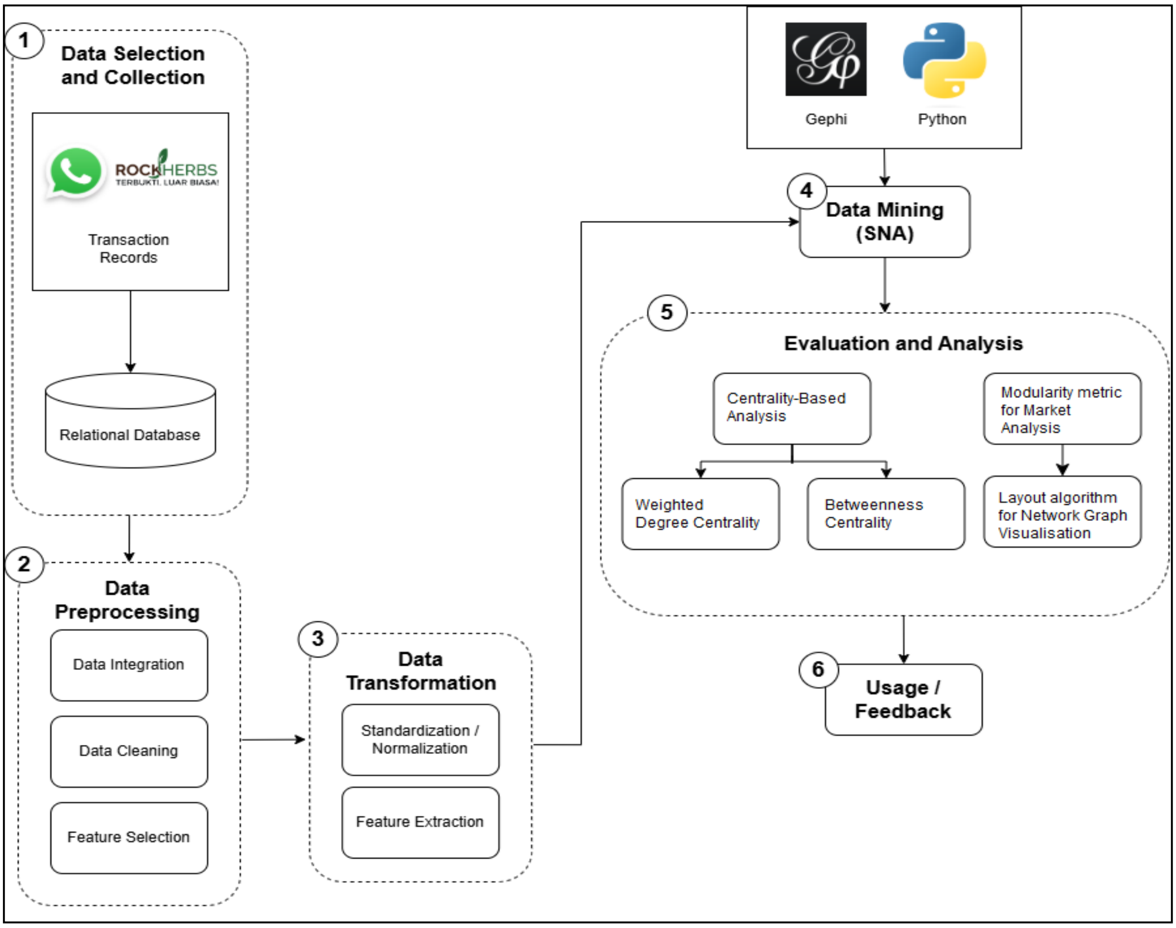
Referring to Table 1**,** although SNA has been applied in various fields, a theoretical gap remains in its application to e-commerce, particularly in understanding the role of influencers. The studies summarised provide insights into network structures and influencer identification (e.g., Camacho et al., 2020; Laghridat & Essalih, 2023) but do not adapt these concepts to the unique interactions between consumers and influencers in digital marketplaces. Research on bio economy and niche marketing (Harrahill et al., 2023; Gómez et al., 2023) explores network roles but lacks a focus on trust and influence mechanisms that are central to commercial contexts. Similarly, work on recommender systems and social media (Hamidi & Moradi, 2024; Kusuma et al., 2023) emphasises technical or sentiment-related aspects without developing theoretical frameworks for understanding how influencers drive consumer engagement and purchasing decisions in social commerce. This study aims to address this gap by advancing SNA theory in the context of e-commerce influencer networks, thereby enhancing the understanding of their structural dynamics, influence mechanisms, and implications for consumer behaviour in digital marketplaces. Identifying influencers is crucial in e-commerce because they can have a substantial impact on consumer behaviour, brand visibility, and sales (Singh et al., 2025; Park & Ahn, 2025). Influencers often have a trusted relationship with their followers; the recommendations they provide feel personal and authentic, and followers are more likely to try or purchase products based on word-of-mouth (Rizgar & Zeebaree, 2025). Influencer marketing represents an emerging marketing strategy compared to traditional digital advertising. Compared to traditional digital advertising, such as paid advertising, influencer marketing offers a better return on investment (ROI) when matched with the right audience. It can also provide lower costs (especially for micro-influencers) for niche markets, such as small to medium-sized enterprises (SMEs).

METHODOLOGY

In this study, the Knowledge Discovery in Databases (KDD) model was adapted to support the development of SNA. KDD helps extract hidden patterns from large datasets, identifying insights that are valid, novel, useful, and understandable (Behera et al., 2019). This iterative process requires refinement at different stages (Ljupco Markuseski et al., 2019). To achieve this, the study integrates the KDD and SNA models into a unified architecture (Figure 1), with each phase explained in detail to clarify their interaction.

Figure 1

The Unified Architecture Diagram



Data Selection and Collection Phase

The first step in the KDD model is to define the analysis objectives. This study aimed to investigate the relationships between sellers, products, and customers within an e-commerce system. The goal was to identify functional patterns to support business decisions. Before selecting data, researchers reviewed available data sources and their potential applications. The dataset, obtained with company approval, included transaction records from April 2023 to March 2024, sourced from WhatsApp and the official website (https://rockherbs.com.my/). It contained 65,235 rows (about 5,000 per month) and 38 attributes covering orders, customers, and sellers. The data was analysed to determine attributes suitable for the SNA process. Table 2 presents the total number of records per month, while Table 3 lists the attributes included in the dataset.

Table 2

Total Number of Records Per Month

Month	Total Data	Attributes
April 2023	4906	38
May 2023	4687	38
June 2023	4737	38
July 2023	8132	38
August 2023	7084	38
September 2023	4678	38
October 2023	6417	38
November 2023	5326	38
December 2023	5909	38
January 2024	5270	38
February 2024	3990	38
March 2024	4099	38
Total (One Year)	65235	38

Table 3

List of Attributes in the Dataset

Category	Attributes
Order details	Order ID, Order Status, Quantity X Product, Total Quantity, Instructions, Payment Method, Payment ID, Payment Status, Payment Date, Bank Code, Shipping Provider, Rider ID, Weight (g), Tracking Number, Cost, Amount, Shipping Charge, COD Charge, FPX Charge, GST, Discount, Grand Total, Order Date, Channel URL
Seller information	Seller Role, Seller Name, Seller ID, Sales Commission, Points
Customer information	Customer Name, Address, Street Address, Postcode, City, State, Country, Phone, Email

Data Preprocessing Phase

The data preprocessing phase ensured accuracy, consistency, and proper formatting for analysis, maintaining data integrity and reliability. Conducted in Python using Google Colaboratory, this phase leveraged its integration with Google Drive, cloud resources, and collaborative features. Key libraries, such as Pandas, NumPy, and Matplotlib, facilitated data cleaning, transformation, and visualisation. To simplify trend analysis, monthly datasets were aggregated into quarterly and annual datasets, reducing computational complexity while preserving detail. The cleaning process removed irrelevant and incomplete data, such as non-completed orders, and excluded redundant attributes to enhance efficiency. Feature selection retained only the most relevant attributes, ensuring the dataset was optimised for SNA. This refined dataset provided a strong foundation for advanced analysis.

This step aimed to reduce computational complexity while preserving the granularity needed for meaningful insights. Each quarterly dataset consisted of approximately 15,000 records (Table 4), representing a manageable yet comprehensive subset of the data for detailed analysis. The data cleaning process addressed several common issues, such as irrelevant, noisy, and missing data. For example, order statuses like "Rejected", "Returned", "Pending", and "Pending COD" were removed because they did not contribute to the study's objectives, which only focused on the completed orders. Attributes irrelevant to the construction of nodes and edges for the SNA, or those deemed redundant, were excluded to optimise the dataset and enhance computational efficiency. Feature selection techniques were rigorously applied to identify and retain only the most relevant attributes, ensuring the final dataset was both concise and rich in information. By the end of this phase, the dataset was fully refined and aligned with the requirements for advanced SNA, laying a robust foundation for the subsequent analysis.

Table 4

Total Number of Records Per Quartile

Quartile	Month	Total data
1	April 2023 May 2023 June 2023	14330
2	July 2023 August 2023 September 2023	19894
3	October 2023 November 2023 December 2023	17652
4	January 2024 February 2024 March 2024	13359
	Annual	65235

Data Transformation Phase

The data transformation phase standardised attributes for SNA by structuring the nodes and edges files based on the relationships under study. Seller performance analysis involved standardising customer phone numbers and structuring seller interactions, while product lifetime value analysis refined product labels, separated multi-product records, and excluded gift items to enhance accuracy. Customer loyalty analysis focused on repeat orders, using customer IDs and product labels to track engagement. Nodes and edges files were generated for quarterly and annual datasets, ensuring consistency and analytical precision. After transformation, all files were verified for accuracy and stored for further analysis.

During this phase, attributes were selected and extracted based on the relationships requiring analysis and transformed into a standardised format suitable for SNA. The primary objective of this phase was to ensure that the data was accurately formatted for a practical relationship analysis and visualisation. Nodes files included unique IDs and labels, while edges files comprised source, target, type, and weight information. For seller performance analysis, customer phone numbers were standardised into the “60” format, facilitating accurate edge file generation by linking seller performance to customer interactions. The node files were structured with “SellerID” as the ID and “SellerName” as the label. The edges files utilised “SellerID” for both source and target, marked as “Undirected”, with weights calculated based on connections between “SellerID” and customer “Phone”. This transformation ensured a clear representation of seller interactions within the dataset. Separate nodes and edges files were created for each quarterly and yearly dataset, maintaining consistency and granularity across timeframes.

For product lifetime value analysis, the “QuantityXProduct” column was split into two separate columns: “Quantity” and “ProductLabel”. The product label was further refined to ensure consistency, using only abbreviations of product names for simplicity and accuracy. Records containing multiple products were duplicated and separated into individual rows, with each row representing one product to enhance analytical precision. Gift items, such as “AROMATHERAPY”, “PERFUME”, and “KYJELLY”, were excluded since they did not represent actual sales. Product labels were standardised by removing suffixes like “RO”, “PROMO”, “WEB”, “FREE”, “PWP”, and “TOPGUN”, as these represented variations of the same product (“RO” = repeat orders or “PROMO” = promotions). Labels like “AEJ” were consolidated, removing size distinctions (“L”, “XL”, and “2XL”) to simplify analysis. Nodes files featured product IDs (numbered 1–67) as the ID and the refined “ProductLabel” as the label. Edges files utilised product IDs for source and target, with weights calculated from interactions between product IDs and “OrderDate”. Nodes and edges files were generated for each quarterly and yearly dataset to facilitate a detailed understanding of product trends over time.

For customer loyalty analysis, node files were created using customer IDs (generated sequentially from 1 - 40,030 for distinct customers) as the ID and “Phone” as the label. Edges files used customer IDs as both source and target, marked as “Undirected”, with weights calculated from the frequency of interactions between “Phone” and “ProductLabel.” Only records with product labels ending in “RO” were included in the analysis. The inclusion of these records was intentional because the “RO” label indicated repeat orders made by customers. Therefore, these records were deemed valuable for analysing customer loyalty, as they reflected the frequency and consistency of customer engagement with the company’s products. The nodes and edges files were generated for yearly data only, as stakeholders requested an annual analysis of customer loyalty. After the transformation process, all node and edge files were rigorously checked for accuracy and completeness. The finalised nodes and edges files were stored for use in subsequent analysis phases.

Data Mining (Social Network Analysis) Phase

The data mining phase used SNA to uncover hidden patterns and relationships within the dataset. Nodes and edges files were imported into Gephi, an open-source tool chosen for its powerful visualisation and analysis features. All relationships were set as “Undirected,” and key metrics like centrality measures and community detection were calculated. Modularity determined node colours, while the weighted degree metric set node sizes to highlight direct interactions. The Fruchterman-Reingold layout ensured that network graphs were clear and readable. Final visualisations were generated for seller performance, product lifetime value, and customer loyalty across quarterly and annual datasets.

The objective of this phase was to uncover and analyse hidden patterns, relationships, and community structures within the dataset using SNA. After transforming the dataset into nodes and edges files, these files were imported into Gephi for analysis. Gephi, an open-source software tool, was chosen for its robust capabilities in network visualisation and analysis. It provided essential features for calculating centrality measures, detecting communities, and visualising network structures. Gephi’s user-friendly interface allowed for the manipulation and customisation of visualisations, making it easier to interpret patterns and relationships in the data. Its ability to process large networks efficiently made it an indispensable tool for this study.

During the import process, all relationships were set to “Undirected”, as the connections for seller performance, product lifetime value, and customer loyalty were indirectly linked through hidden attributes. Key metrics, including network overview statistics, community detection, and node and edge overviews, were calculated. These metrics informed the selection of visualisation parameters. For node colours, modularity was used to identify and partition communities within the network. For node size, the weighted degree metric was selected, emphasising direct interactions and connections over other measures, such as betweenness centrality. The Fruchterman-Reingold layout algorithm was applied to ensure visually appealing, well-spaced graphs. This algorithm was chosen for its ability to enhance readability and clarity, which was critical for stakeholders’ understanding. Additional adjustments, including repositioning nodes, resizing labels, and fine-tuning edges, were performed to improve the visual presentation. The final network graphs were generated for each relationship type (seller performance and product lifetime value) and both quarterly and annual datasets.

Evaluation and Analysis Phase

In SNA, the evaluation and analysis phase compares different network metrics to understand relationships and influence. The weighted degree metric measures the total strength of a node’s connections by considering both the number and weight of edges. It helps identify nodes with the highest direct influence, making it useful for analysing seller performance, customer loyalty, and product lifetime value. A highly weighted degree indicates a highly connected and influential node in its immediate surroundings. See Equation 1 for the weighted degree centrality of a node in a network.

$$C_w(v) = \sum_{u \in N(v)} W_{vu} \tag{1}$$

where:

$C_w(v)$ is the weight degree centrality of node v,
$N(v)$ is the set of neighbors of node v,
W_{vu} is the weight of the edge between node v and node u.

The weighted degree centrality takes into account the strength (weight) of each connection, unlike degree centrality, which only counts the number of connections (edges) a node has. The higher the weighted degree centrality, the more influential the node is in terms of direct interactions.

In contrast, betweenness centrality measures how often a node acts as a bridge between other nodes. It identifies key connectors or intermediaries that influence the flow of information or transactions. Nodes with high betweenness centrality play a crucial role in connecting different parts of the network, making them essential for understanding critical points in the structure. See Equation 2 for the betweenness centrality of a node v is calculated as:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

where:

$C_B(v)$ is the betweenness centrality of node v,
σ_{st} is the total number of shortest paths between node s and t,
$\sigma_{st}(v)$ is the number of those shortest paths that pass through node v,
The sum is taken over all pairs of nodes s and t in the network, excluding v itself.

This metric measures the frequency at which a node lies on the shortest paths between other nodes. A node with higher betweenness centrality indicates that it acts as a bridge, playing a key role in network flow and connectivity. It is beneficial for identifying influential intermediaries in a network.

Usage or Feedback Phase

After the evaluation and analysis phases, including usability testing, the findings were documented and shared with stakeholders at the e-commerce company to transform insights into actionable strategies. The documented findings serve as a foundation for future business strategies and performance improvements.

EVALUATION AND ANALYSIS

Three significant findings in this section are the centrality-based analysis for influencer identification, the modularity metric for market analysis, and the layout algorithm for network graph visualisation.

Centrality-Based Analysis for Influencer Identification

This section analysed findings from the experiments to determine node size in network graphs, considering various metrics, including weighted degree centrality and betweenness centrality. The weighted degree metric was identified as the most suitable, as it measures both the number and strength of a node's connections, making it ideal for assessing seller performance, customer loyalty, and product lifetime value. This metric effectively emphasises the number and strength of a node's connections, making it particularly ideal for analysing direct influence within a network. Nodes with high weighted degrees represent highly connected or influential points in their immediate surroundings, providing valuable insights into seller performance, customer loyalty, and product lifetime value.

Betweenness centrality was also analysed for its role in identifying key connectors in the network. Betweenness centrality was considered relevant due to its ability to highlight nodes that serve as bridges or connectors within a network. This metric identifies critical nodes for network flow, which often act as key influencers or gateways. Betweenness centrality measures the extent to which other nodes depend on a specific node, providing insights into the structural importance of entities within the network. To verify and validate the appropriateness of these metrics for deriving insights aligned with e-commerce business sales records and real-world conditions, a comparative analysis was conducted during an interview session with a stakeholder. The weighted degree metric proved to be more effective in identifying direct influence, while betweenness centrality offered insights into bridging roles. However, the stakeholder emphasised that the weighted degree metric aligned more closely with the company's operational reality and objectives.

For instance, an analysis of seller performance from April 2023 to March 2024 revealed that Nor Fatin was the best-performing seller based on the weighted degree metric, followed by Muhammad Firdaus and Muhammad Naseem. In contrast, betweenness centrality highlighted Nurul Najwa as the top seller, with Muhammad Firdaus and Muhammad Naseem in subsequent positions. During the interview, the stakeholder confirmed that Nor Fatin, Muhammad Firdaus, and Muhammad Naseem consistently achieved the highest commissions and bonuses, validating the accuracy and relevance of the weighted degree metric. Tables 5 and 6 show the list of sellers' performance (annual) according to the weighted degree metric and betweenness centrality metric, in descending order (from highest to lowest).

Table 5

List of Sellers' Performance (Annual) According to the Weighted Degree Metric in Descending Order (Highest to Lowest)

Id	Label	Weighted Degree Centrality
37	NOR FATIN FARHANIM BINTI MOHD DEPI	3170.0
65	MUHAMMAD FIRDAUS AHMAD	2614.0
44	MUHAMMAD NASEEM BIN ABDUL MASHOOD NASEEM	1958.0
38	NURUL NAJWA BINTI AHMAD	1820.0
40	NURFAHMI ABDUL RAHIM	1507.0
30	KHAIRULNISSA SHARMILA	1266.0
91	MUHAMMAD SUKRI NGATIMAN	1187.0
39	NUR ILIANA FATIHAH BINTI ABDUL RASHID	1084.0
138	NIZAM TOP GUN	756.0

Table 6

List of Sellers' Performance (Annual) According to the Betweenness Centrality Metric in Descending Order (Highest to Lowest)

Id	Label	Betweenness Centrality
38	NURUL NAJWA BINTI AHMAD	0.09312
65	MUHAMMAD FIRDAUS AHMAD	0.08967
44	MUHAMMAD NASEEM BIN ABDUL MASHOOD NASEEM	0.080312
38	NOR FATIN FARHANIM BINTI MOHD DEPI	0.06996
30	KHAIRULNISSA SHARMILA	0.05454
170	ATIKAH SUHAIMI	0.040391
138	NIZAM TOP GUN	0.040391
40	NURFAHMI ABDUL RAHIM	0.0322074
39	NUR ILIANA FATIHAH BINTI ABDUL RASHID	0.028284

Similarly, for product lifetime value in Q4 (January–March 2024), the weighted degree metric identified Black Seed Honey (MHS), Honey Kids (MK), and Black Tumeric Coffee (KKHQ) as the top products, which the stakeholder confirmed as the most purchased. In contrast, betweenness centrality highlighted Black Seed Honey (MHS), *Sacha Inchi* Coffee (KSI), and Oh My Essay Book Basic Package (OMKMENENGAHBASIC), which did not align with actual customer behaviour. A similar trend appeared in the annual customer loyalty graph, where the weighted degree metric provided clearer insights into loyal customers and their connections. The stakeholder concluded that the weighted degree metric is more reliable for decision-making as it accurately reflects sales patterns and relationships. Refer to Tables 7 and 8 for detailed information on weighted degree centrality and betweenness centrality concerning product lifetime value.

Theoretically, the weighted degree metric is more effective in e-commerce settings because it captures both the strength and frequency of direct interactions, which are key indicators of sustained commercial activity, rather than focusing solely on structural positioning within the network. In contrast, betweenness centrality highlights the intermediary roles of nodes by identifying those that serve as bridges between otherwise disconnected parts of the network. Although this perspective is valuable in contexts where the flow of information or influence is critical, it may fail to recognise individuals or products that demonstrate consistently high levels of direct engagement. Consequently, weighted degree centrality is more closely aligned with performance metrics such as sales volume, customer loyalty, and product turnover, making it a more relevant and practical indicator for decision-making in transactional and digitally mediated marketplaces.

Table 7

List of Product's Lifetime Value (Q4) According to the Weighted Degree Metric in Descending Order (Highest to Lowest)

Id	Label	Weighted Degree Centrality
33	MHS	788.0
51	MK	772.0
2	KKHQ	320.0
9	KSI	253.0
4	KMRQ	212.0
11	KBG	174.0

Table 8

List of Product's Lifetime Value (Q4) According to the Betweenness Centrality Metric in Descending Order (Highest to Lowest)

Id	Label	Betweenness Centrality
33	MHS	0.157417
9	KSI	0.107854
50	OMKMENENGAHBASIC	0.073991
13	KMRR	0.07167
2	KKHQ	0.064649
19	OMKMENENGAHUPGRADE	0.053545
3	KKHB	0.043293
8	KC	0.040277
1	OMK	0.035749

The outcome of this proposed work is then visualised using SNA—figures 2 and 3 display network graphs analysing seller performance based on weighted degree centrality. The network graph can also be viewed across quarterly (Q1, Q2, Q3, Q4) and annual timeframes.

Figure 2

Social Network Analysis of Seller Performance (Annual)

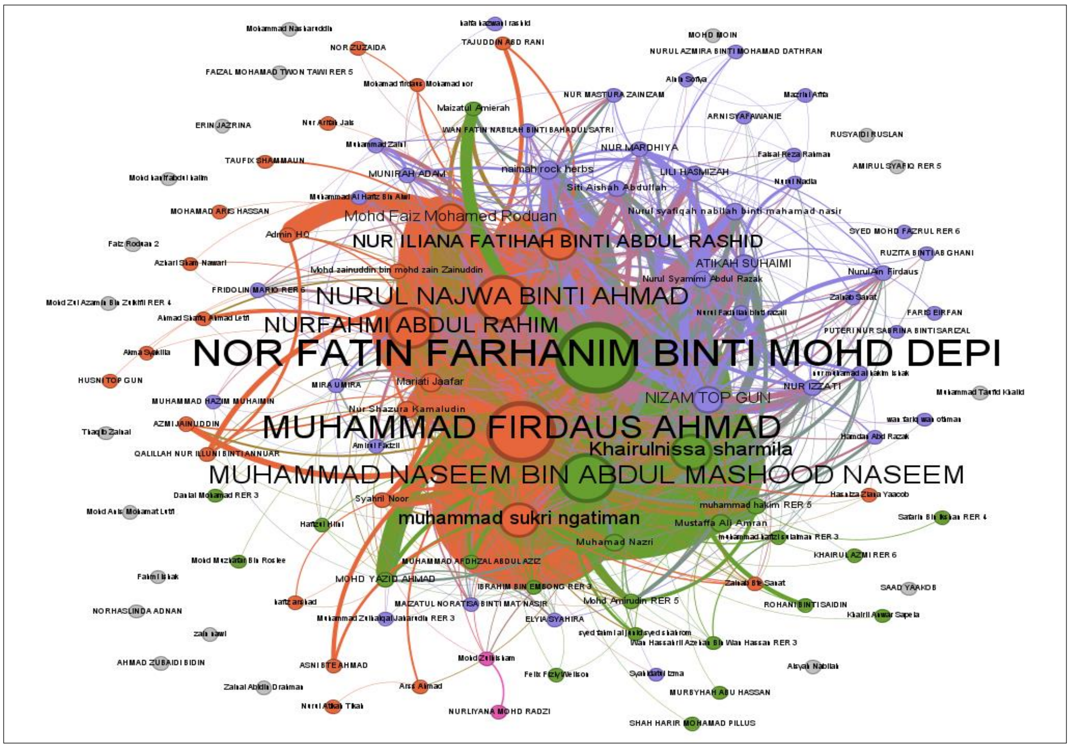
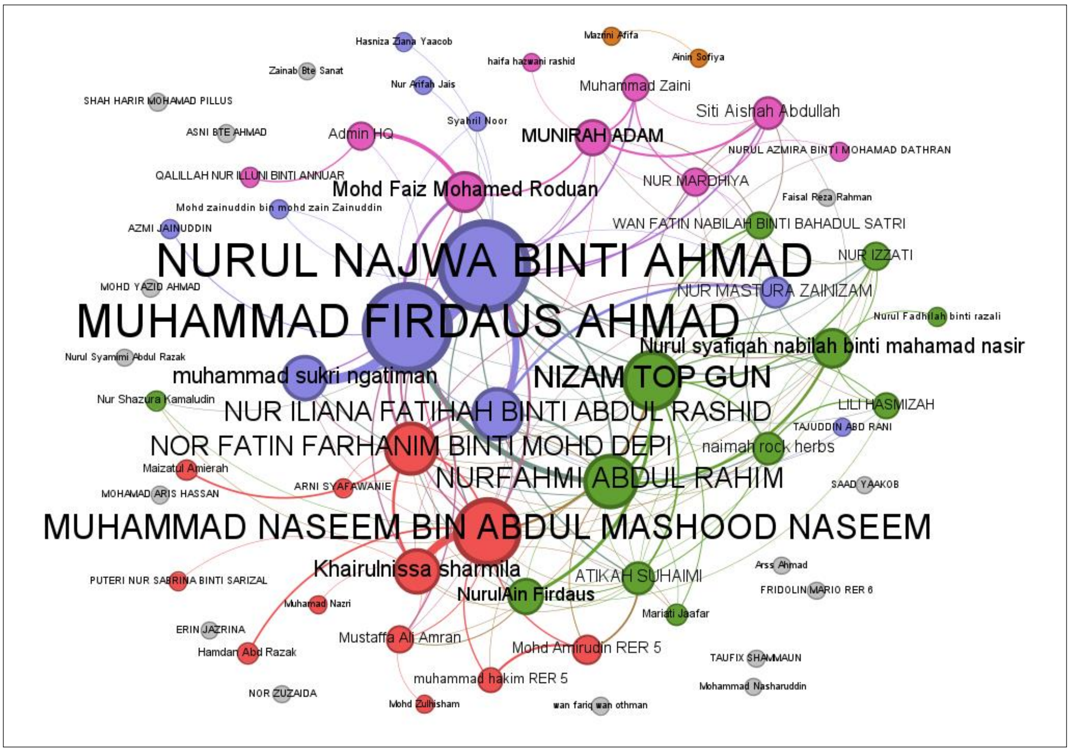


Figure 3

Social Network Analysis of Seller Performance (Quartile 4)



Similarly, Figures 4 and 5 display a social network graph focused on analysing product lifetime value, following the same interactive and time-based structure as the seller performance analysis based on weighted degree centrality.

Figure 4

Social Network Analysis of Product Lifetime Value (Quartile 1)

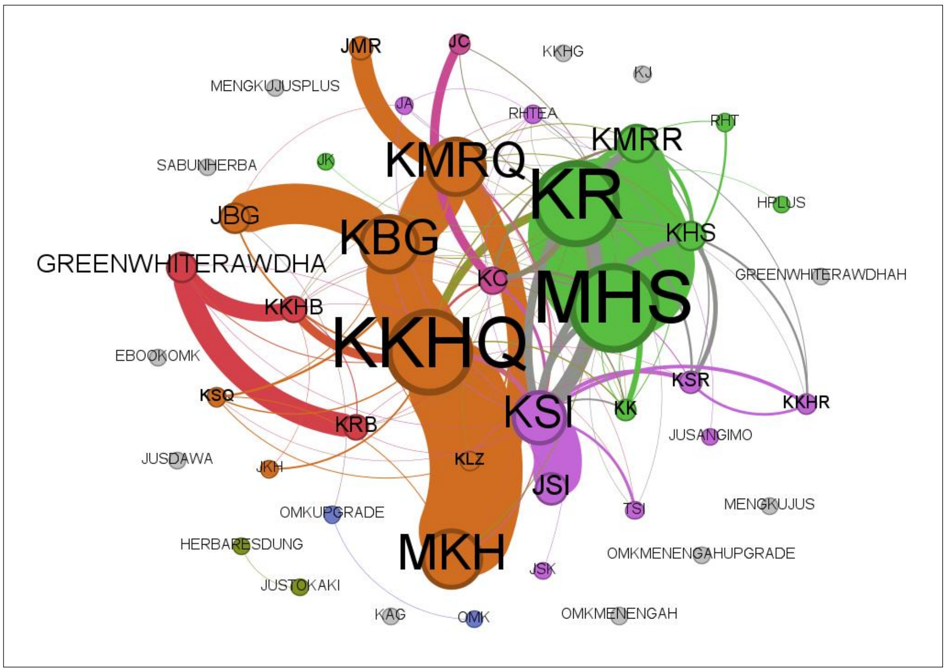
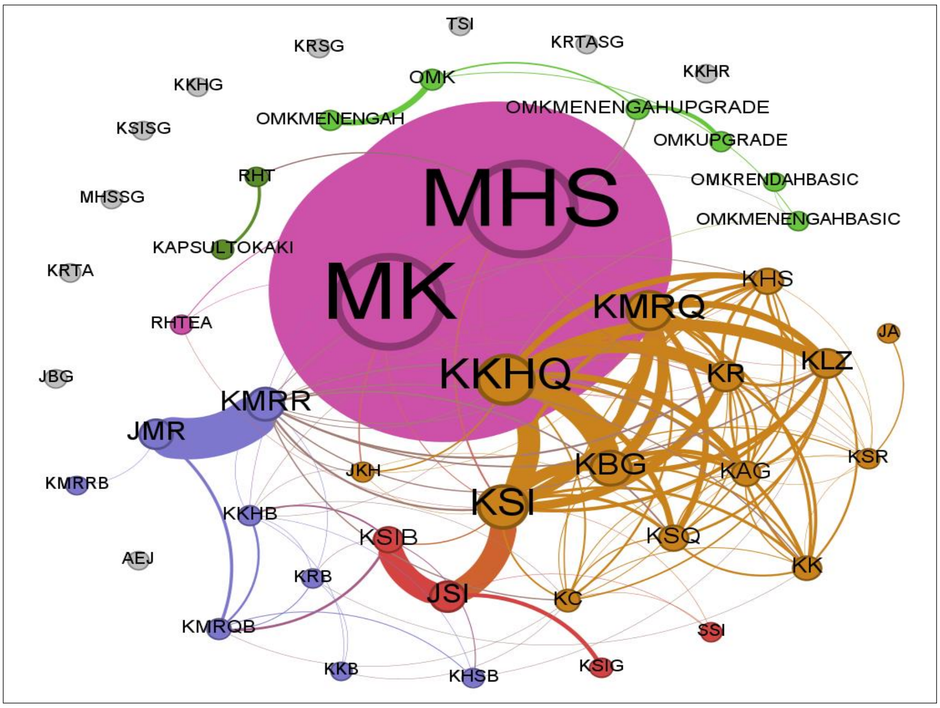


Figure 5

Social Network Analysis of Product Lifetime Value (Quartile 4)



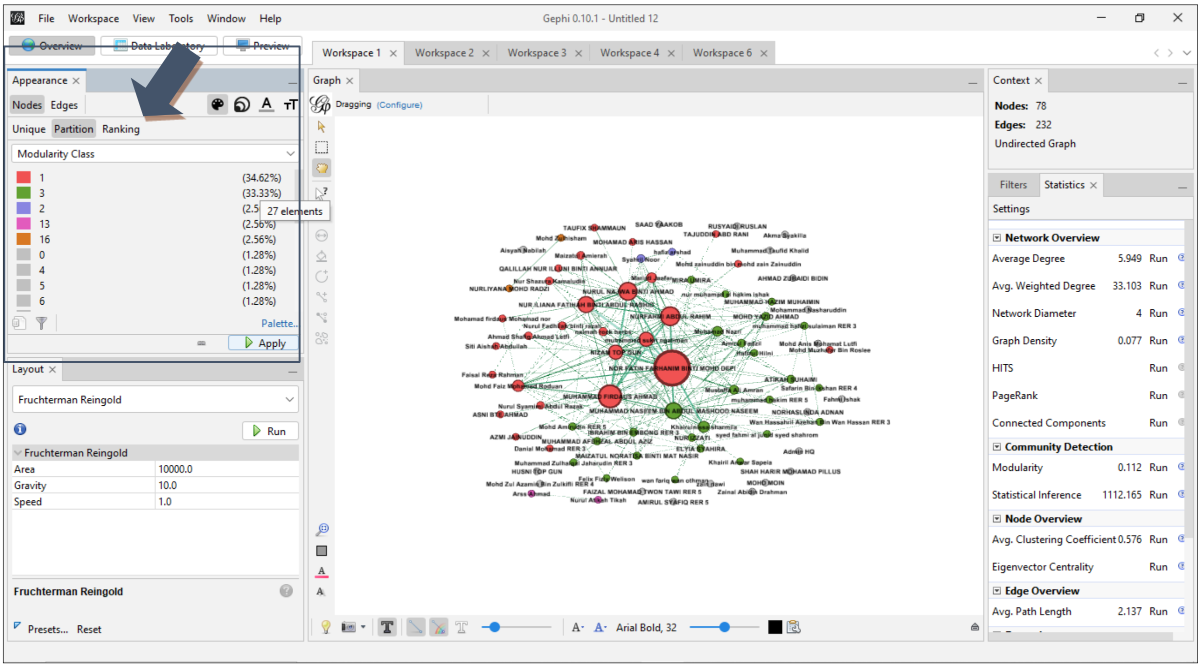
This process allows companies to understand how SNA improves marketing efforts, optimises product offerings, and improves seller performance. A network graph based on weighted degree centrality identifies high-performing sellers who can be incentivised, while product lifetime value highlights the total revenue a product is expected to generate over its entire life. This insight strengthens seller-customer relationships and enables informed strategic decisions.

Modularity Metric for Market Analysis

The modularity metric was used as a critical tool for community detection. This metric identifies clusters or groups within a dataset, with each detected community represented by a distinct colour, a process known as node colour partitioning in network graphs. By visually differentiating relationships between nodes, it enhances the interpretability of the network graph. For example, in the seller performance network graph for the annual period (Figure 6), four main communities were detected and distinguished by colour. The stakeholder confirmed that these communities accurately reflected the real-world structure of the company. For instance, sellers whose names ended in “RER 5” and “RER 6” were correctly grouped as part of the same team, which collaborates to sell products. This validation demonstrated the effectiveness of the modularity metric in capturing team dynamics.

Figure 6

Network Graph for the Seller Performance (Annual) using Modularity Metric



Similarly, in the product lifetime value network graph for the annual period, seven main groups were identified using the modularity metric. The stakeholder validated these groupings, noting that products such as KKHQ, KMRQ, and KSQ, which belong to the same category (“Qash”), were accurately grouped together. Likewise, products such as OMK, OMKMENENGAHBASIC, and OMKRENDAHBASIC (all OMK products) were placed in the same community, as were KSI and JSI (coffee and juice products, respectively). The stakeholder confirmed that the modularity metric effectively captured the relationships between sellers, products, and customers. These groupings reflected real-world categorisations, reinforcing the modularity metric's value in supporting the company's decision-making processes for marketing strategies and resource allocation.

Layout Algorithm for Network Graph Visualisation

The Fruchterman-Reingold layout algorithm was selected for its effectiveness in organising and structuring network graphs compared to other layout algorithms, such as ForceAtlas 2 and YifanHu. This force-directed algorithm simulates a physical system in which nodes repel each other, while edges act as springs that pull connected nodes closer together. This process minimises edge crossings and optimises the overall graph structure, making it particularly effective for large and complex networks. The algorithm offers a balance between computational efficiency and visual clarity, ensuring the network graphs are both functional and aesthetically appealing.

When compared to network graphs without any layout algorithm, the Fruchterman-Reingold layout significantly improved the organisation and readability of the graphs (Figures 7 and 8). Without the algorithm, the graphs appeared cluttered and failed to display communities and relationships between nodes effectively. Applying the Fruchterman-Reingold algorithm yielded structured and visually clear graphs, facilitating the identification of key relationships and the extraction of actionable insights. During the interview, the stakeholder expressed satisfaction with the algorithm's clarity and efficiency, reinforcing its suitability for this study. As a result, the Fruchterman-Reingold layout algorithm was retained and seamlessly integrated into the dashboard for consistent use across all network graph visualisations.

Figure 7

Network Graph for the Seller Performance (Quartile 1) without Layout Algorithm

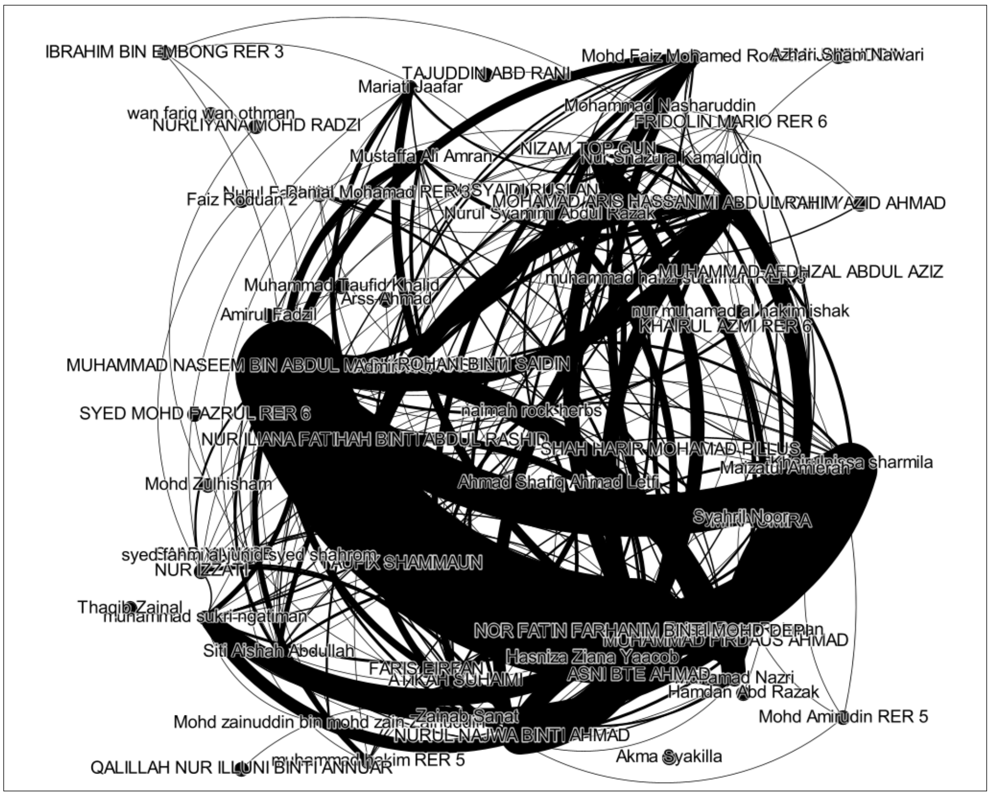
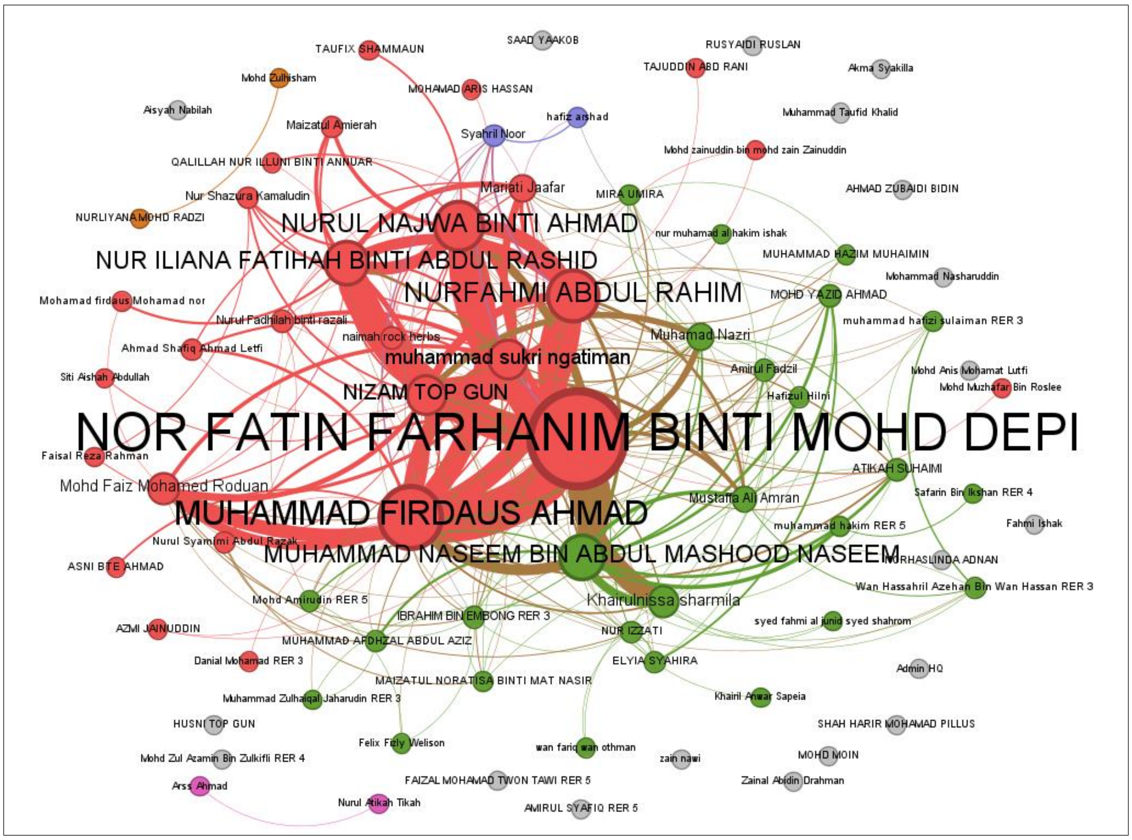


Figure 8

Network Graph for the Seller Performance (Q1) with Fruchterman-Reingold Layout Algorithm



CONCLUSION

Researchers have extensively analysed network relationships to identify key influencers. While SNA has been applied in healthcare, education, social media, and marketing, its use in business and management remains limited (Anugerah et al., 2022). Most business-related studies focus on brand awareness and community detection via hashtag analysis and social media comments.

This study advances SNA methodology by analysing real-world transaction records from the e-commerce company's social media platform and website database, shifting from digital engagement metrics to structured transaction data. This approach enables a more precise examination of network structures, identifying key influencers within the company's customer base and operations. By employing transaction-based network modelling, the study refines SNA techniques to extract actionable business insights. The integration of weighted degree and betweenness centrality measures enhances the analysis of relational dynamics, improving network structure interpretation in a commercial context. This methodological refinement broadens SNA's conventional scope, reinforcing its relevance for data-driven decision-making in business management.

This study successfully integrated centrality measures into Social Network Analysis to uncover hidden relationships among sellers, products, and customers, improving decision-making. Centrality measures, such as weighted degree and betweenness centrality, helped identify key influencers and connections within the network, allowing businesses to optimise marketing strategies, product offerings, and customer engagement. Key strengths included interactive visualisations and downloadable reports, which enhanced data-driven decision-making by providing clear insights into sales trends and customer behaviour. While limitations such as its static nature and limited sample size remain, the findings demonstrate its potential as a valuable business tool for strategic planning and operational improvement. Further research may explore the integration of dynamic network modelling with multi-layer e-commerce networks to capture complex commercial dynamics and improve the depth of analysis across various platforms and industries.

Additionally, future work will focus on expanding this SNA-based framework to other industries beyond e-commerce, including education, public health, and supply chain management. In education, for instance, transaction-based network modelling can be applied to analyse student engagement, peer collaboration, and learning pathways within digital learning environments. Similarly, in healthcare, this approach may reveal referral patterns, patient-provider interactions, or treatment outcomes across institutional networks. By adapting the methodology to different sectors, future studies can further validate the model’s flexibility and scalability, contributing to more targeted interventions, resource optimisation, and policy formulation across diverse organisational contexts.

ACKNOWLEDGMENT

The authors would like to extend their gratitude to the Data Management & Software Solution Research Lab, School of Computing, Universiti Utara Malaysia, for their generous support in sponsoring the publication of this article.

REFERENCES

Almashaleh, O., Wicaksono, H., & Valilai, O. F. (2025). A framework for social media analytics in textile business circularity for effective digital marketing. *Journal of Open Innovation: Technology, Market, and Complexity*, 100544. https://doi.org/10.1016/j.joitmc.2025.100544

Anugerah, A. R., Muttaqin, P. S., & Trinarningsih, W. (2022). Social network analysis in business and management research: A bibliometric analysis of the research trend and performance from 2001 to 2020. *Heliyon*, *8*(4). https://doi.org/10.1016/j.heliyon.2022.e09270

Arab, M. (2021). Providing an approach to predicting customer quality in E-commerce social networks based on big data and unsupervised learning method. *arXiv preprint arXiv:2109.02080*.

Behera, P. C., Dash, C., & Mohapatra, S. (2019). Data mining and knowledge discovery (KDD). *International Journal of Research and Analytical Reviews (IJRAR)*, *6*(1), 101-106.

Camacho, D., Panizo-LLedot, A., Bello-Orgaz, G., Gonzalez-Pardo, A., & Cambria, E. (2020). The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion, 63,* 88-120. https://doi.org/10.1016/j.inffus.2020.05.00

Chai, J., & Ye, J. H. (2024). A social network analysis of college students’ online learning during the epidemic era: A triadic reciprocal determinism perspective. *Heliyon*, *10*(6), e28107. https://doi.org/10.1016/j.heliyon.2024.e28107

Gómez, S., Patraca, B., & Molina, J. L. (2023). Improving seafood systems with social network analysis: The case of cuttlefish marketing in Catalonia. *Marine Policy*, *150*, 105517. https://doi.org/10.1016/j.marpol.2023.105517

Hamidi, H., & Moradi, R. (2024). Design of a dynamic and robust recommender system based on item context, trust, rating matrix and rating time using social networks analysis. *Journal of King Saud University-Computer and Information Sciences*, *36*(2), 101964. https://doi.org/10.1016/j.jksuci.2024.101964

Harrahill, K., Macken-Walsh, Á., & O'Neill, E. (2023). Identifying primary producers' positioning in the Irish bioeconomy using Social Network Analysis. *Cleaner and Circular Bioeconomy*, *5*, 100042. https://doi.org/10.1016/j.clcb.2023.100042

Howe, S. (2024). Social Media statistics for Malaysia. *Meltwater*. https://www.meltwater.com/en/blog/social-media-statistics-malaysia.

Kusuma, I. Y., Pratiwi, H., Khairunnisa, S. F., Pitaloka, D. A. E., & Kurnianto, A. A. (2023). The assessment of Twitter discourse on the new COVID-19 variant, XBB. 1.5, through social network analysis. *Vaccine: X*, *14*, 100322. https://doi.org/10.1016/j.jvacx.2023.100322

Laghridat, C., & Essalih, M. (2023). A set of measures of centrality by level for social network analysis. *Procedia computer science*, *219*, 751-758. https://doi.org/10.1016/j.procs.2023.01.348

Latifi, M., Kerachian, R., & Zali, R. B. (2023). Evaluating energy harvesting from water distribution networks using combined stakeholder and social network analysis. *Energy Strategy Reviews*, *49*, 101158. https://doi.org/10.1016/j.esr.2023.101158

Ljupco Markuseski, L., Zdravkoski Igor, I. Z., Andonovski Miroslav, M. A., & Jovanoska Aleksandra, A. J. (2019). Knowledge discovery databases (KDD) process in data mining. In *International Conference Proceeding* (pp. 529-539). Faculty of Economic Prilep.

Mondal, T., Sen, J., Goswami, R., & Nag, P. K. (2024). Community adaptation to heat stress: Social network analysis. *Climate Risk Management*, *44*, 100606. https://doi.org/10.1016/j.crm.2024.100606

Norz, L. M., Dornauer, V., Hackl, W. O., & Ammenwerth, E. (2023). Measuring social presence in online-based learning: An exploratory path analysis using log data and social network analysis. *The Internet and Higher Education*, *56*, 100894. https://doi.org/10.1016/j.iheduc.2022.100894

Park, J., & Ahn, S. (2025). Unpacking masstige brand perception: A multimethod study on value dimensions and brand extension. *Journal of Retailing and Consumer Services*, *87*, 104376. https://doi.org/10.1016/j.jretconser.2025.104376

Rahardjo, D. A. S. (2019). Use of social network analysis as a strategy for e-commerce recommendations. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, *10*(2), 617-622.

Recuero, R., Zago, G., & Soares, F. (2019). Using social network analysis and social capital to identify user roles on polarised political conversations on Twitter. *Social Media+ Society*, *5*(2), 2056305119848745. https://doi.org.10.1177/2056305119848745

Rizgar, F., & Zeebaree, S. R. (2025). The rise of influence marketing in e-commerce: A review of effectiveness and best practices. *East Journal of Applied Science*, *1*(1), 18-34. https://doi.org/10.63496/ejas.Vol1.Iss1.38

Singh, S. S., Singh, S., Singh, K., Srivastava, V., & Shakya, H. K. (2025). Big data meets social networks: A survey of analytical strategies and research challenges. *IEEE Access (13),* 98668 - 98698. https://doi.org/10.1109/ACCESS.2025.3576143

Smith, M. B., Blakemore, J. K., Ho, J. R., & Grifo, J. A. (2021). Making it (net) work: A social network analysis of "fertility" in Twitter before and during the COVID-19 pandemic. *F&S Reports*, *2*(4), 472-478. https://doi.org/10.1016/j.xfre.2021.08.005

Strozzi, F., Garagiola, E., & Trucco, P. (2019). Analysing the attractiveness, availability and accessibility of healthcare providers via social network analysis (SNA). *Decision Support Systems*, *120*, 25-37. https://doi.org/10.1016/j.dss.2019.03.007

Wan, B. (2024). 2024 Malaysia's Latest Digital Trends: 9 Key Highlights Unveiled. *cacaFly Malaysia*. https://cacaflymalaysia.com/en/archives/15309

Zehir, C., Yücel, M., Borodin, A., Yücel, S., & Zehir, S. (2023). Strategies in energy supply: A social network analysis on the energy trade of the European Union. *Energies*, *16*(21), 7345. https://doi.org/10.3390/en16217345

Zhu, C., Su, Y., Fan, R., Qin, M., & Fu, H. (2024). Exploring provincial carbon-pollutant emission efficiency in China: An integrated approach with social network analysis and spatial econometrics. *Ecological Indicators*, *159*, 111662. https://doi.org/10.1016/j.ecolind.2024.111662