Graph-based methods for transaction databases: a comparative study

Wael Ahmad AlZoubi[1], Ibrahim Mahmoud Alturani[1], Roba Mahmoud Ali Aloglah[2]
[1]Department of Applied Sciences, Ajloun University College, Al-Balqa Applied University, Ajloun, Jordan
[2]Department of Management Information Science, Amman College for Financial and Managerial Sciences, Al-Balqa Applied University, Amman, Jordan

Article Info

Article history:

Received Jul 29, 2024
Revised Nov 3, 2024
Accepted Nov 14, 2024

Keywords:

Data mining
Graph
Rule mining
Structured data
Transaction database

ABSTRACT

There has been an increased demand for structured data mining. Graphs are among the most extensively researched data structures in discrete mathematics and computer science. Thus, it should come as no surprise that graph-based data mining has gained popularity in recent years. Graph-based methods for a transaction database are necessary to transform all the information into a graph form to conveniently extract more valuable information to improve the decision-making process. Graph-based data mining can reveal and measure process insights in a detailed structural comparison strategy that is ready for further analysis without the loss of significant details. This paper analyzes the similarities and differences among four of the most popular graph-based methods that is applied to mine rules from transaction databases by abstracting them out as a concrete high-level interface and connecting them into a common space.

Corresponding Author:

Wael Ahmad AlZoubi
Department of Applied Sciences, Ajloun University College, Al-Balqa Applied University
Ajloun 26816, Jordan
Email: wa2010@bau.edu.jo

1. INTRODUCTION

Graph-based methods for a transaction database are necessary to transform all the information into a graph form to conveniently extract more valuable information [1]–[3]. Graph-based data mining can reveal and measure process insights in a detailed structural comparison strategy that is ready for further analysis without the loss of significant details [4]. In addition, the graph-based methods process can be considered as a process mining method.

This research aims to systematically understand the trade-offs among graph-based methods for mining transaction datasets by comparing them. There are four main methods to mine transaction datasets using graphs, they are: clique percolation system [5], adjacency matrix [6], graph neural network (GNN) [7] and network-based visualization [8]. Each one of these methods follow the same general idea: constructing a graph that captures the relations between different parts of the structured data. Despite the diversity of methods and the variations in the exact form that the final task-related graph takes, some clear organizing principles emerge.

A transaction database is a collection of records; each record contains pieces of data. These records are also called transactions. A graph database is a database management system that uses graph structures to store, map and query relationships. Every element contains a direct pointer to its adjacent element and can also be used to perform search in constant time using hash index [9]. The transaction database management system supports transactions from multiple customers and does not contain any customer master data. A transaction database does not allow for the full capabilities of a transaction to be represented. It abstracts the transactions to a form that is compatible with the machinery of the transaction database. A graph database attempts to capture the full detail of a transaction [10].

We outlined a comparative study on the graph-based approaches for mining different useful patterns by growing algorithms in case of the transaction database [11]. Table 1 briefly explains some of the main characteristics of these methods. This table helps to focus the different features and applications of each method for network analysis and visualization.

Table 1. Graph-based mining methods' characteristics

Method	Description	Uses	Graph representation	Interactivity
1. Clique percolation system	System used to find and analyze complete sub graphs (cliques) in networks, focusing on identifying fully connected groups of nodes.	Identifying interconnected groups and communities within networks.	Focuses on identifying cliques, not a direct visual representation.	Minimal interaction: manually inspecting identified cliques is frequently necessary.
2. Adjacency matrix	This method represents the relationships among the nodes in 2D array (matrix) showing connections as binary values (presence or absence of edges).	Studying network construction accurately, computing network metrics like degrees and shortest routes.	Represents connections between nodes in a matrix form.	Static representation, needs manual adjustment for network changes.
3. GNN method	Neural network approach to learn node and edge features for prediction and classification tasks in networks.	Node classification, link prediction, and community detection in complex networks.	Learns node and edge features using deep learning techniques.	Interactive for network exploration and predictive tasks.
4. Network-based visualization	Visual representation technique for networks, showing nodes and links in a graphical and interactive manner.	Visual exploration of network structures, understanding relationships and identifying patterns.	Provides visual insights into network topology and dynamics.	Highly interactive, allows real-time exploration and analysis.

This study covers graph-based algorithms for data analysis of transaction databases and provides a comparative analysis regarding selected property descriptors. Retail datasets of 1000 transactions will be taken as a case study to clarify the role of each method in extracting the desired association rules, compare among them and so enhance the decision-making process. To the best of our knowledge, we introduce a comparative study of the graph-based methods used to discover rules from transaction datasets.

The overall structure of the research is organized as follows. Section 2 talks about the main graph-based methods for transaction datasets. Sction 3 explains briefly the research methodology. Section 4 discusses the comparative analysis of these methods. Section 5 the results of previous studies were comprehensively reviewed and analyzed using the criteria described there. Lastly, section 6 concludes this paper.

2. GRAPH BASED METHODS FOR TRANSACTION DATASETS

As we mentioned earlier in the introduction, a dataset of retail sales will be studied and analyzed since this type of datasets has been developed safely with the coming of president data science methods and tools [12]. Nowadays, retail enterprises create advanced techniques to derive meaningful conclusions from massive volumes of transactional data [13]. The most common among these techniques are: the clique percolation system, adjacency matrix analysis, GNNs, and network-based visualization. These algorithms offer powerful ways to uncover hidden patterns, complex relationships between products and customers will be discovered, and totally improve decision-making. We will examine how these techniques can be successfully used in retail sales environments to enhance consumer engagement, optimize strategies, and spur business growth. Retail companies can improve customer satisfaction, boost operational efficiency, and improve their marketing strategy by incorporating these tactics and analyzing the links and trends in their sales data. In the following sub-sections, we will describe briefly how these techniques are used in the context of retail sales dataset.

2.1. Clique percolation method

The clique percolation method is a common method for examining the overlying public construction of networks. The clique percolation system can be used in retail sales to find products or category clusters that are commonly purchased together, as well as significant correlations between them. For instance, it can reveal product groups that are frequently purchased together or close connections between categories.

2.2. Adjacency matrix

The adjacency matrix offers a matrix representation of nodes and their pairwise relationships based on transaction interactions showing connections as binary values (existence or nonexistence of edges). In retail sales data, links between items or product categories are represented by the adjacency matrix. A product or category is represented by each row and column, and the matrix shows whether there is a relationship between them or not. You can use this matrix to look at relationships and find fresh patterns in sales data.

2.3. Graph neural network

In discovery of complex associations from transaction data, the GNNs play an important role in finding hidden rules that represent the relations among products. GNNs signify the transactions as graphs to forecast conclusions such as customer comportment, product commendations, or deceitful activity. GNN algorithms are used to assess retail sales data and anticipate buyer behavior by means of product relationships and prior purchase patterns. GNNs are useful for understanding complicated linkages between goods and consumers as well as examining how marketing and promotions affect these connections.

2.4. Network-based visualization

This method gives graphical depiction for networks, displaying nodes and edges in a graphical and collaborative way. Visual representation and analysis of the outcomes of the GNN, adjacency matrix, and clique percolation system predictions in retail sales data are done by network-based visualization. It helps analysts and managers make based on data strategic decisions by offering an illustration of the complex relationships among products.

3. RESEARCH METHODOLOGY

The same set of data across all tested methods is used during the comparative study. This approach ensures fairness and consistency in evaluating the performance of different graph-based methods for mining transaction datasets [14]. The main graph-based methods to mine rules from transaction datasets, i.e., clique percolation, adjacency matrix, GNN and graph visualization are tested over the same set of transactions. An intuitive choice is to use a graph database as a new type of database and thus this technology has generated great attention. There are several surveys in the literature that summarize the existing graph databases and their applications [15].

A comparative study focusing on graph-based methods used for mining transaction datasets involves evaluating various techniques within this domain will be discussed. Figure 1 highlights the main steps to discover the find out the best choice by do an efficient comparison among graph-based methods from customer data. These steps improve the accuracy and truth of the comparative study's results, this will lead to worthy remarks into the best method(s) for extracting desired rules from transaction datasets. The following subsections talks briefly about each one of these steps.

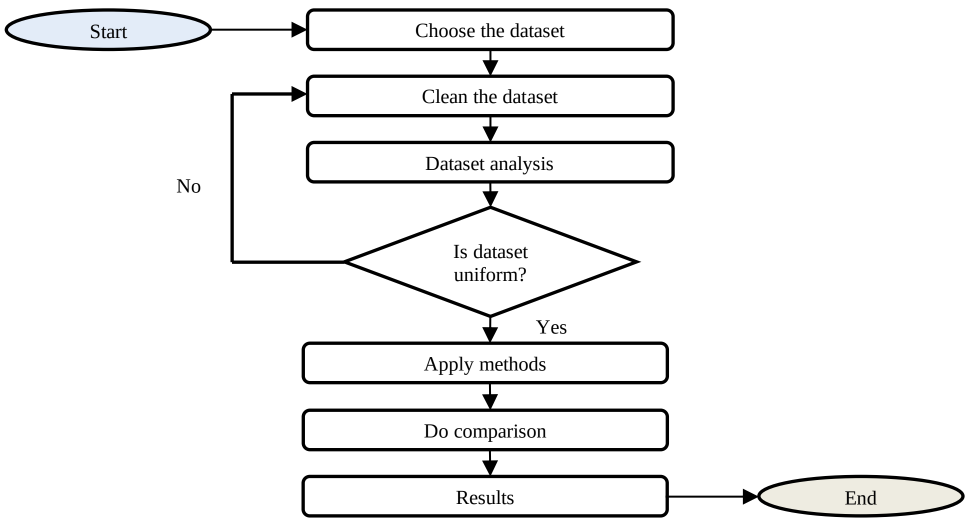

Figure 1. The flowchart of the experimental methods applied

3.1. Dataset selection

Choosing the right data set is not as simple as many people think, as there are criteria for choosing the appropriate data set, such as being compatible with the field of interest or study, and it must afford adequate transactional data. The chosen dataset should also be complete, accurate and free of outliers. The same set of data will be used for each method under investigation during the comparison analysis. This methodology guarantees impartiality and uniformity while assessing the efficacy of various graph-based techniques for transaction dataset mining.

3.2. Dataset cleaning and preprocessing

Data cleaning is an important step in improving the superiority of the data and confirm that we can infer eloquent rules. To guarantee consistency and quality of data, clean up and preprocess the dataset. Depending on the requirements of each approach, this stage may involve resolving missing values, normalizing data, and encoding categorical variables.

3.3. Apply methods on uniform dataset

When the selected dataset is ready to be used, i.e. it is cleaned from any outliers or missing values, the graph-based methods will be used directly to assist in making right decisions and the overall mining process will be improved. Utilize the standardized dataset with every graph-based technique, following the same guidelines. To ensure comparability and remove bias, all methods must use the same preprocessing procedures and settings.

3.4. Analysis and evaluation

It is very important to analyze and evaluate the results after applying the different graph-based methods on the selected transaction dataset. This phase aids us realize the efficiency of the chosen approach, measure the performance of each method, and find what must be improved. Gather and examine each method's output according to predetermined assessment criteria. These criteria might include outcomes interpretability, computational efficiency, scalability in managing big datasets, and accuracy of transaction pattern recognition.

3.5. Comparison

The performance of the chosen graph-based methods must be compared depending on five criteria, they are: scalability, accuracy, complexity, interpretability and versatility to be able to determine which one is the best in dealing with transaction dataset. Based on the evaluation metrics, compare how well each technique performs. Determine the advantages and disadvantages of each approach in comparison to the others, emphasizing any compromises that might affect how well-suited each is for a given kind of transactional data analysis.

3.6. Comparative analysis of graph-based methods

Graph-based methods have been used extensively with transaction databases. For this comparative study, we focus on the most widely used close n-vertices adjacency graph representation. This representation defines a graph where each node represents an item in the database and n-vertices are qualified as adjacent to each other if they appear together in a transaction. It is also referred to as the unique-itemset-content-compatible graph (UCC graph) [16], [17].

Retail dataset is one of the popular datasets used in data analysis and pattern mining studies in retail and sales. This group includes data on purchases that are typically recorded through point-of-sale (POS) systems in stores and shops. Data usually includes:

- Product information: such as name, description, and category.
- Customer information: such as age, gender, and location of residence.
- Purchase details: such as date, time, and amount paid.
- Store information: such as location, branches, and departments.
- Payment methods: such as cash, credit cards, and electronic payment.

Using a retail dataset can help analyze customer purchasing behaviors, discover common patterns in purchasing, forecast product demand, and improve inventory management and marketing strategies. This kit is ideal for research studies and business analysis in the retail industry [16]–[18]. It will be efficient to assess and select the best graph-based technique for generating rules from transactional datasets by applying this structured comparative study, considering the features of the dataset and the users' unique requirements. Table 2 is an expanded table that includes the evaluation for each method: clique percolation system, adjacency matrix, network-based visualization, and GNN. This table provides a comprehensive overview of how each method is evaluated in terms of analysis, visualization, and prediction capabilities based on the available data.

Table 2. The evaluation of the graph-based mining methods from transaction datasets

Method	Evaluation	Details of evaluation
1. Clique percolation system	Analysis of discovered cliques and comparison against expectations and requirements	Evaluation of clique size and frequency comparison across various clique percolation system settings (e.g., changing k if applicable). Effectiveness of cliques in predicting future network or data behavior.
2. Adjacency matrix	Analysis of relationships between categories and measuring relationship strengths	Analysis of existing relationships in the adjacency matrix. Measurement of relationship strengths between categories based on values in the matrix. Comparison of adjacency matrices under different bases (e.g., quantity or price).
3. Network-based visualization	Visual understanding of relationships and representation of developments over time	Visual understanding of relationships between different categories. Representation of developments over time if using temporal network visualization. Comparison of different network visualizations based on drawing techniques and emphasizing key relationships between categories.
4. GNN	Improvement in product categorization or sales prediction based on networks	Evaluation of GNN's ability to control network data for improving product categorization or sales prediction. Examination of GNN's performance in learning intricate relationships between categories based on available data. Comparison of GNN results with traditional methods.

4. RESULTS AND DISCUSSION

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily [19], [20]. In the literature [21]–[25], there are many studies about the different graph based methods for transaction datasets, we used the same set of data for each method under investigation during the comparison analysis. This methodology guarantees impartiality and uniformity while assessing the efficacy of various graph-based techniques for transaction dataset mining.

Five different criteria were used to offer a complete structure for allocating numbers to the tables that reflects an exhaustive evaluation of the effectiveness of each technique in relation to network data analysis and visualization [4], [26]. The criteria are:

– Scalability: assesses how well each technique can manage increasing amounts of data without sacrificing efficiency and concert.
– Complexity: evaluates each method's computational cost and resource usage (memory and CPU time).
– Accuracy: evaluates each method's capacity to produce accurate and dependable outcomes in tasks involving investigation and presentation.
– Interpretability: evaluates the ease of comprehension and interpretation of the outputs and outcomes produced by each method.
– Versatility: examines the adaptability of each method to a broad range of activities and applications.

Each of these criteria will be tested separately for each of these methods and then the results will be compared as in the following sections.

4.1. Scalability

Each method's scalability differs greatly depending on how it is designed and intended to be used. The modest scalability of the clique percolation system makes it appropriate for medium-sized networks, but it might be problematic for very large datasets [26], [27]. The adjacency matrix, on the other hand, shows good scalability and is effective for big, static networks, but it could need a lot of assets for networks that are dynamic [27]. When properly designed, the GNN exhibits significant scalability as well, making it a viable option for efficiently processing huge datasets [28], [29]. Depending on the amount of the dataset and the display capabilities, network-based visualization [30] provides strong scalability for visual exploration, making it easier for users to explore network structures easily. These findings aid in the suitable technique choosing, considering the scalability requirements for analysis or visualization chores.

Based on the allocated numerical values, this representation makes it easier for consumers or researchers to understand how the procedures differ from one another in a more structured way. It makes decision-making easier depending on certain analysis requirements or intended results. Figure 2 and Table 3 illustrate graphically the scalability of each one of these methods on the selected retail dataset.

4.2. Complexity

The complexity degree of each method is shown by the "complexity" results. The clique percolation system exhibits low complexity by using simple methods that are effective in terms of processing speed and memory utilization. The complexity of the adjacency matrix ranges from low to reasonable, depending on the extent of the entire network and memory needs [31]. Because they employ deep learning techniques, GNNs exhibit enormous complexity, requiring substantial processing resources and a lengthy training period [7], [32]. Network-based visualization is low to moderately complicated, with simple display operations at the base [33]. Large networks or interactive functionality may call for additional resources. The findings shed light on how each technique manages the complexity and processing demands of network data analysis and visualization. Figure 3 and Table 4 illustrate graphically the complexity of each one of these methods on the selected retail dataset.

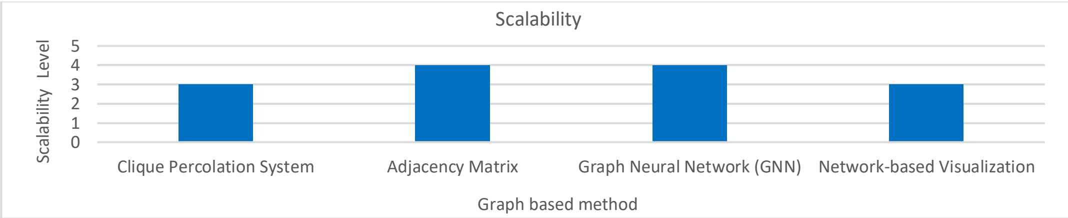

Figure 2. Graphical representation of the scalability among the graph-based methods for retail dataset

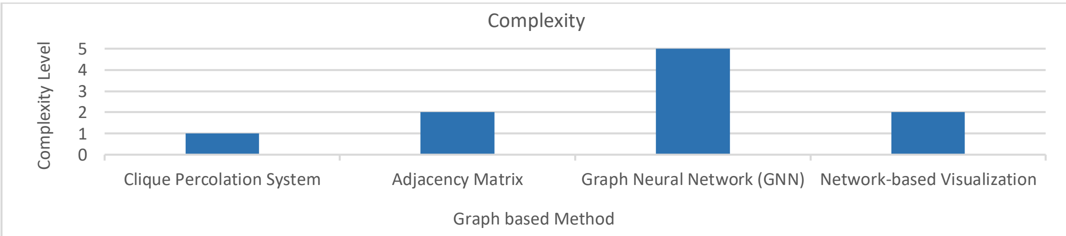

Figure 3. Graphical representation of the complexity among the graph-based methods for retail dataset

Table 3. Scalability of graph-based methods

Method	Scalability
Clique percolation system	3
Adjacency matrix	4
GNN	4
Network-based visualization	3

Explanation of values:
Scalability:
1: Low scalability
2: Moderate scalability
3: High scalability
4: Scalable for large datasets
5: Highly scalable with appropriate architecture

Table 4. Complexity of graph-based methods

Method	Complexity
Clique percolation system	1
Adjacency matrix	2
GNN	5
Network-based visualization	2

Explanation of values:
Complexity
1: Low complexity
2: Low to moderate complexity
3: Moderate complexity
4: High complexity due to deep learning techniques
5: Very high complexity

4.3. Accuracy

The "accuracy" results show how accurate each method is. The clique percolation system is a good tool for recognizing communities within networks since it shows good accuracy in identifying cohesive groups, or cliques. The adjacency matrix is a visual aid that makes node connections easier to understand while offering excellent accuracy in computing network metrics like node degrees and shortest paths [27]. When learning node and edge features, GNNs demonstrate exceptional accuracy, which makes them useful for intricate pattern recognition applications [7], [29]–[31]. Depending on the methods used and the level of user experience, network-based visualization exhibits medium to high accuracy in displaying network architecture and spotting patterns [33]. These points demonstrate how each technique complies with requirements for accuracy while examining and displaying network data. Figure 4 and Table 5 illustrate graphically the complexity of each one of these methods on the selected retail dataset.

4.4. Interpretability

The term "interpretability" describes how simple and intuitive it is to understand and examine the outcomes of any given method [4], [26]. Because the clique percolation system mainly finds cohesive groups (cliques) without offering a clear visual representation, it is difficult to intuitively grasp the results, which contributes to its low interpretability [27]. The adjacency matrix, on the other hand, provides excellent interpretability by graphically depicting node connections, making it possible to comprehend network interconnections and structure with clarity [28]. Given that they learn intricate node and edge properties, which may call for more in-depth research to properly interpret, GNNs exhibit intermediate interpretability [7], [29]–[34]. High interpretability is achieved using network-based visualization, which makes it simple to identify important network properties by providing a clear visual understanding of network topology and patterns [35]. These variations highlight how the interpretability of each approach meets various requirements for efficiently understanding and analyzing network data. Figure 5 and Table 6 illustrate graphically the interpretability of each one of these methods on the selected retail dataset.

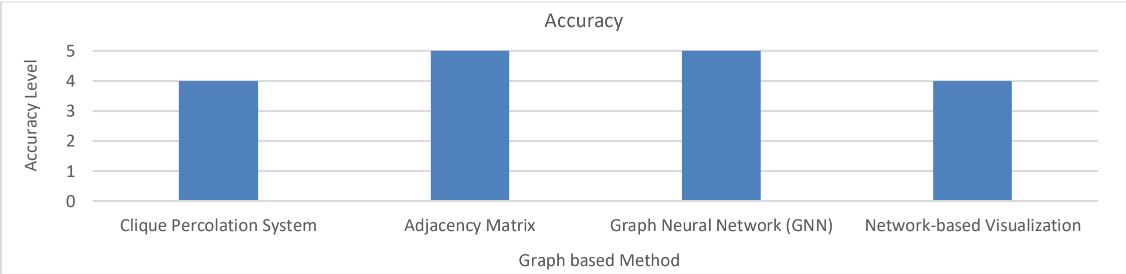

Figure 4. Graphical representation of the accuracy among the graph-based methods for retail dataset

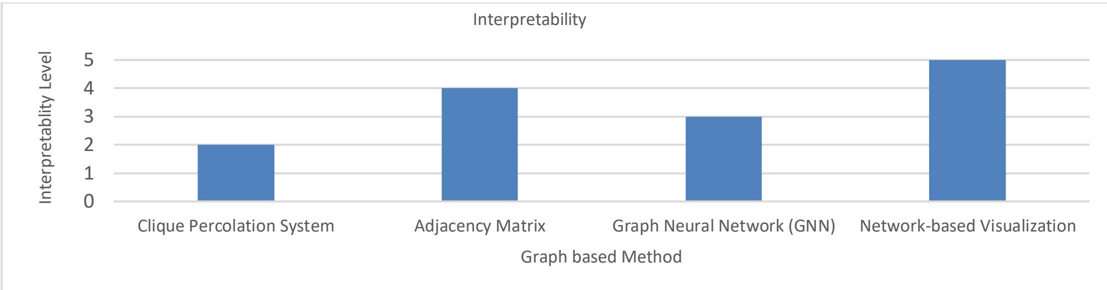

Figure 5. Graphical representation of the interpretability among the graph-based methods for retail dataset

Table 5. Accuracy of graph-based methods

Method	Accuracy
Clique percolation system	4
Adjacency matrix	5
GNN	5
Network-based visualization	4
Explanation of values: Accuracy: 1: Low accuracy 2: Low to medium accuracy 3: Medium accuracy 4: High accuracy 5: Very high accuracy	

Table 6. Interpretability of graph-based methods

Method	Interpretability
Clique percolation system	2
Adjacency matrix	4
GNN	3
Network-based visualization	5
Explanation of values: Interpretability: 1: Low interpretability 2: Moderate interpretability 3: High interpretability 4: High interpretability; matrix format visually represents node connections 5: Highly interpretable; provides basic visual insights	

4.5. Versatility

The degree to which a method can be tailored to a variety of activities and applications is referred to as its versatility. With its narrow scope of applicability, the clique percolation system is mainly useful for studying organized groups in networks. For a variety of analytical and mathematical activities requiring the structural representation of the network and the computation of different metrics, the adjacency matrix provides good adaptability [36]. GNNs are very versatile; they can handle a wide range of jobs because they can recognize intricate patterns and adjust to various kinds of network input [37], [38]. Additionally, network-based visualization offers great variety by enabling interactive and visual network exploration and analysis, which makes it easier to fully comprehend network patterns and structures [39]. These differences show how each approach fits requirements for network data analysis and visualization in various application contexts. Figure 6 and Table 7 illustrate graphically the versatility or adaptability of each one of these methods on the selected retail dataset. The retail dataset used in the literature contains 1000 transactions distributed over three main categories [25], i.e. clothes, electronics and cosmetics or beauty tools. Table 8 shows some data from the retail dataset chosen in the experiments. The schema or the description of the dataset is given in Table 9.

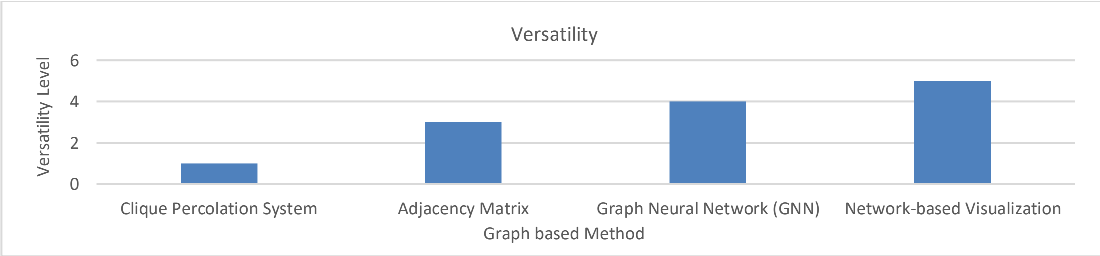

Figure 6. Graphical representation of the versatility among the graph-based methods for retail dataset

Table 7. Versatility of graph-based methods

Method	Versatility
Clique percolation system	1
Adjacency matrix	3
GNN	4
Network-based visualization	5

Explanation of values:
Versatility:
1: Limited versatility
2: Moderate versatility
3: Versatile for various tasks
4: Versatile for various tasks including node classification and link prediction detection
5: Highly versatile for exploratory analysis

Table 8. Retail dataset used in the comparison

#	Transaction ID	Date	Customer ID	Gender	Age	Product category
0	1	2023-11-24	CUST001	Male	34	Beauty
1	2	2023-02-27	CUST002	Female	26	Clothing
2	3	2023-01-13	CUST003	Male	50	Electronics
3	4	2023-05-21	CUST004	Male	37	Clothing
4	5	2023-05-06	CUST005	Male	30	Beauty
	Quantity	Price per unit ($)		Total amount		
0	3	50		150		
1	2	500		1,000		
2	1	30		30		
3	1	500		500		
4	2	50		100		

Table 9. Retail dataset schema

#	Attribute	Count	Null	Data type
0	Transaction ID	1,000	non-null	Int64
1	Date	1,000	non-null	object
2	Customer ID	1,000	non-null	object
3	Gender	1,000	non-null	object
4	Age	1,000	non-null	Int64
5	Product category	1,000	non-null	object
6	Quantity	1,000	non-null	Int64
7	Price per unit	1,000	non-null	Int64
8	Total amount	1,000	non-null	Int64

5. CONCLUSION

Since the development of sophisticated data science methods and tools, retail sales analytics has undergone substantial change. Retail businesses now have access to advanced techniques for deriving useful conclusions from massive volumes of transactional data. The clique percolation system, adjacency matrix analysis, GNNs, and network-based visualization are important methods among these. These approaches provide effective means of revealing latent patterns, comprehending intricate interactions between goods and consumers, and eventually improving decision-making. In this talk, we look at how these techniques can be used in retail sales scenarios to enhance customer engagement, optimize strategies, and spur corporate growth.

ACKNOWLEDGEMENTS

We thank the employees and programmers of the Computer and Information Center at our beloved university, Al-Balqa Applied University, for their cooperation and providing what is necessary to complete this research. We also thank the administration of Ajloun University College for the support it provided throughout the preparation of this scientific research. We can't forget our families for their patience and support.

REFERENCES

[1] M. Besta *et al.*, “Demystifying graph databases: analysis and taxonomy of data organization, system designs, and graph queries,” *ACM Computing Surveys*, vol. 56, no. 2, pp. 1–40, 2024, doi: 10.1145/3604932.
[2] Y. Shao and N. Nakashole, “On linearizing structured data in encoder-decoder language models: insights from text-to-SQL,” in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2024, pp. 131–156, doi: 10.18653/v1/2024.naacl-long.8.
[3] M. E. Coimbra, A. P. Francisco, and L. Veiga, “Study on resource efficiency of distributed graph processing,” *arXiv-Computer Science*, pp. 1–23, 2017.
[4] A. Baudin, M. Danisch, S. Kirgizov, C. Magnien, and M. Ghanem, “Clique percolation method: memory efficient almost exact communities,” in *Advanced Data Mining and Applications*, 2022, pp. 113–127.
[5] J. Kim, S. Lee, Y. Kim, S. Ahn, and S. Cho, “Graph learning-based blockchain phishing account detection with a heterogeneous transaction graph,” *Sensors*, vol. 23, no. 1, 2023, doi: 10.3390/s23010463.
[6] X. Ren, K. Zhao, P. J. Riddle, K. Taskova, Q. Pan, and L. Li, “DAMR: Dynamic adjacency matrix representation learning for multivariate time series imputation,” *Proceedings of the ACM on Management of Data*, vol. 1, no. 2, pp. 1–25, 2023, doi: 10.1145/3589333.
[7] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A comprehensive survey on graph neural networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021, doi: 10.1109/TNNLS.2020.2978386.
[8] H. Chen *et al.*, “G-tran,” *Proceedings of the VLDB Endowment*, vol. 15, no. 11, pp. 2545–2558, 2022, doi: 10.14778/3551793.3551813.
[9] D. Lin, J. Wu, Q. Yuan, and Z. Zheng, “Modeling and understanding ethereum transaction records via a complex network approach,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 11, pp. 2737–2741, 2020, doi: 10.1109/TCSII.2020.2968376.
[10] A. Pismerov and M. Pikalov, “Applying embedding methods to process mining,” in *ACM International Conference Proceeding Series*, 2022, pp. 1–5, doi: 10.1145/3579654.3579730.
[11] Z. Yang, Y. Bi, L. Wang, D. Cao, R. Li, and Q. Li, “Development and application of a field knowledge graph and search engine for pavement engineering,” *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-11604-y.
[12] M. Wu, X. Yi, H. Yu, Y. Liu, and Y. Wang, “Nebula graph: An open source distributed graph database,” *arXiv-Computer Science*, pp. 1–18, 2022.
[13] A. Ferhati, “Applying a label propagation algorithm to detect communities in graph databases,” *M.Sc. Thesis*, Department of Computer Science & Engineering, University of Bergamo, Bergamo, Italy, 2022.
[14] S. Biswas, M. Bhattacharyya, and S. Bandyopadhyay, “Topological analysis on multi-scenario graphs: Applications toward discerning variability in SARS-CoV-2 and topic similarity in research,” *Transactions of the Indian National Academy of Engineering*, vol. 7, no. 1, pp. 365–374, 2022, doi: 10.1007/s41403-021-00306-y.
[15] H. Seiti, A. Makui, A. Hafezalkotob, M. Khalaj, and I. A. Hameed, “R.Graph: A new risk-based causal reasoning and its application to COVID-19 risk analysis,” *Process Safety and Environmental Protection*, vol. 159, pp. 585–604, 2022, doi: 10.1016/j.psep.2022.01.010.
[16] A. B. Ammar, “Query optimization techniques in graph databases,” *International Journal of Database Management Systems*, vol. 8, no. 4, pp. 1–14, 2016, doi: 10.5121/ijdms.2016.8401.
[17] M. Mohajer, “A graph-based platform for customer behavior analysis using applications’ clickstream data,” *arXiv-Computer Science*, pp. 1–23, 2020, doi: 10.48550/arXiv.2002.10269.
[18] P. Mehrotra, V. Anand, D. Margo, M. R. Hajidehi, and M. Seltzer, “SoK: The faults in our graph benchmarks,” *arXiv-Computer Science*, pp. 1–26, 2024.
[19] P. Wills and F. G. Meyer, “Metrics for graph comparison: A practitioner’s guide,” *PLOS ONE*, vol. 15, no. 2, Feb. 2020, doi: 10.1371/journal.pone.0228728.
[20] C. Lezcano and M. Arias, “Characterizing transactional databases for frequent itemset mining,” *CEUR Workshop Proceedings*, vol. 2436, 2019.
[21] J. Sandell, E. Asplund, W. Y. Ayele, and M. Duneld, “Performance comparison analysis of ArangoDB, MySQL, and Neo4j: An experimental study of querying connected data,” in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2024, pp. 7760–7769.
[22] A. S. Reddy, P. K. Reddy, A. Mondal, and U. D. Priyakumar, “Mining subgraph coverage patterns from graph transactions,” *International Journal of Data Science and Analytics*, vol. 13, no. 2, pp. 105–121, 2022, doi: 10.1007/s41060-021-00292-y.
[23] M. Lei *et al.*, “Mining top-k sequential patterns in transaction database graphs: A new challenging problem and a sampling-based approach,” *World Wide Web*, vol. 23, no. 1, pp. 103–130, 2020, doi: 10.1007/s11280-019-00686-w.
[24] Z. Yao, “Visual customer segmentation and behavior analysis: A SOM-based approach,” *M.Sc. Thesis,* Department of Information Technologies, Åbo Akademi University, Turku, Finland, 2013.
[25] W. A. Alzoubi, “Dynamic graph based method for mining text data,” *WSEAS Transactions on Systems and Control*, vol. 15, pp. 453–458, 2020, doi: 10.37394/23203.2020.15.45.
[26] A. Bóta and M. Krész, “A high resolution clique-based overlapping community detection algorithm for small-world networks,” *Informatica*, vol. 39, no. 2, pp. 177–187, 2015.

[27] S. Tabassum, F. S. F. Pereira, S. Fernandes, and J. Gama, "Social network analysis: An overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 5, 2018, doi: 10.1002/widm.1256.

[28] Z. Huang, S. Zhang, C. Xi, T. Liu, and M. Zhou, "Scaling up graph neural networks via graph coarsening," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2021, pp. 675–684, doi: 10.1145/3447548.3467256.

[29] X. Liu *et al.*, "Survey on graph neural network acceleration: an algorithmic perspective," in *IJCAI International Joint Conference on Artificial Intelligence*, 2022, pp. 5521–5529, doi: 10.24963/ijcai.2022/772.

[30] V. Yoghourdjian, Y. Yang, T. Dwyer, L. Lawrence, M. Wybrow, and K. Marriott, "Scalability of network visualisation from a cognitive load perspective," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1677–1687, 2021, doi: 10.1109/TVCG.2020.3030459.

[31] M. Hlawatsch, M. Burch, and D. Weiskopf, "Visual adjacency lists for dynamic graphs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 11, pp. 1590–1603, 2014, doi: 10.1109/TVCG.2014.2322594.

[32] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, 2019, doi: 10.1186/s40649-019-0069-y.

[33] I. Amaral, "Complex networks," in *Encyclopedia of Big Data*, Cham: Springer International Publishing, 2022, pp. 198–201.

[34] H. Xuanyuan, P. Barbiero, D. Georgiev, L. C. Magister, and P. Liò, "Global concept-based interpretability for graph neural networks via neuron analysis," *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023*, vol. 37, no. 9, pp. 10675–10683, 2023, doi: 10.1609/aaai.v37i9.26267.

[35] H. Rawlani, "Visual interpretability for convolutional neural network," *Towards Data Science*, pp. 1–20, 2018.

[36] M. Li, Y. Deng, and B. H. Wang, "Clique percolation in random graphs," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 92, no. 4, 2015, doi: 10.1103/PhysRevE.92.042116.

[37] I. R. Ward, J. Joyner, C. Lickfold, Y. Guo, and M. Bennamoun, "A practical tutorial on graph neural networks," *ACM Computing Surveys*, vol. 54, no. 10, pp. 1–35, 2022, doi: 10.1145/3503043.

[38] B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, "A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions," *Journal of Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-023-00876-4.

[39] S. Dutta and S. Roy, "Complex network visualisation using JavaScript: a review," in *Intelligent Systems*, vol. 431, 2022, pp. 45–53.

BIOGRAPHIES OF AUTHORS

Wael Ahmad AlZoubi holds a doctor of computer sciences from National University of Malaysia in 2013. He also received his B.Sc. and M.Sc. (computer science) from Yarmouk University, Jordan in 2000 and 2004, respectively. He is currently an Assistant Professor at Department of Computer Science in Al-Balqa Applied University, Ajloun, Jordan. His research includes meta-heuristics, global optimization, machine learning, data mining, bioinformatics, graph theory and parallel programming. He has published over 20 papers in international journals and conferences. He can be contacted at email: wa2010@bau.edu.jo.

Dr. Ibrahim Mahmoud Alturani is an instructor in the Department of Computer Science at Ajloun College, Al-Balqa Applied University, Jordan. He earned his B.S. and M.S. degrees in computer science from Yarmouk University, Jordan, in 2004 and 2007, respectively, and completed his Ph.D. in computer science at the University Malaysia Terengganu, Malaysia, in 2021. He began his academic career as a part-time lecturer in the Department of Computer Science at Yarmouk University from 2007 to 2008 before joining Al-Balqa Applied University as an instructor, where he has been teaching since 2008. He has published several papers in international journals, with research interests encompassing knowledge representation through ontology and knowledge graphs, natural language processing, content-based retrieval, and artificial intelligence. He can be contacted at email: traini111@bau.edu.jo.

Roba Mahmoud Ali Aloglah received her bachelor's degree of information technology from Al-Balqa Applied University in 2004. She received the master's degree from the Arab academy Jordan, Amman in 2005. She is a lecturer of computer science and information technology at Department of Management Information Science, Amman College for Financial and Managerial Sciences, Al-Balqa Applied University, Amman, Jordan since 2008. Her research interests include algorithms, computer networks, artificial intelligence and computer security. She can be contacted at email: robajabali@bau.edu.jo.