

## Journal of Integrated Information Management

Vol 9, No 2 (2024)

Jul-Dec 2024



### Graph Databases and Graph Neural Networks

*Stratos Tsolakidis, Anastasios Tsolakidis, Evangelia Triperina, Nikitas Karanikolas, Christos Skourlas*

doi: [10.26265/jiim.v9i2.39198](https://doi.org/10.26265/jiim.v9i2.39198)

Copyright © 2024, Stratos Tsolakidis, Anastasios Tsolakidis, Evangelia Triperina, Nikitas Karanikolas, Christos Skourlas



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

### To cite this article:

Tsolakidis, S., Tsolakidis, A., Triperina, E., Karanikolas N., & Skourlas, C. (2024). Graph Databases and Graph Neural Networks. *Journal of Integrated Information Management*, 9(2), 24–32. <https://doi.org/10.26265/jiim.v9i2.39198>



where D. Washington acted are presented in the graph, as well as the other actors that also participated in the film “The Equalizer (2014)”.

There are two major research approaches focused on developing frameworks for graph data: the representation of graph data with Graph Databases and the application of machine learning methods on graph data with Graph Neural Networks (GNNs).

The technology of big data and the spread of its applications offer an opportunity to re-evaluate the technology of relational databases. GNNs are machine learning models designed for graph data that utilize the graph topology. Further investigation is needed to understand the perspectives of the two above approaches whereas a review of the two research approaches at a theoretical and a practical level is important. More specifically, the current study addresses the following research questions:

RQ1. Is there in the literature any satisfactory theoretical and experimental verification of the benefits of applying graph neural networks and graph databases to managing large-scale data in social networks?

RQ2. Are there any comparative advantages of using graph databases over relational databases in managing large-scale data in social networks?

The paper is structured as follows: the first section is the introduction, whereas the second section corresponds to the literature review. Section 3 outlines a comparative presentation of handling graph data and relational data by examining the data modeling, the queries and the representation of the data. The fourth section demonstrates how graph neural network can be applied in social recommendation and finally, in section 5, there is the conclusion.

## II. LITERATURE REVIEW

Although Relational Database Management Systems (RDBMS) are considered the most established technique for data management, graph databases seem to attract the interest of researchers, because of their ability to manage big data. According to Kumar Kaliyar, R. (2015) [2], “most of the real-world applications can be modeled as a graph and one of the best real-world

examples is social network”. According to Bhattacharyya & Chakravarty [3], the evolution of relational DBMS will be the Graph Databases with NoSql methodologies, “which is emerging as beyond of relational model”, while Tian [4] referred to the growth of big network data in industry that demands graph technologies, and denoted that the research projections showed a significant growth of the global market for graph databases in the subsequent years. Xirogiannopoulos & Deshpande [5] stated that the analysis of the graph structure among the underlying entities (or objects) in a dataset deliver meaningful information and value in various application domains and they mentioned the development of various graph databases, e.g., Neo4j, which address these needs.

Both Graph Databases and GNNs for graph data have been the focus of research and according to literature many benefits derive from their application. The benefits that may occur from their combination and from the proposal and the use of a unified framework are yet to be examined. According to the literature there is a research gap in the before mentioned research area. Among the efforts made to this direction, Besta et al. [6] combined GNN models with graph databases but, limited work has been done in this research area.

GNNs [7], [8] are deep learning-based methods that operate on graphs, in which the edges connecting the nodes express the underlying topology. GNNs can exploit this topology and combine it with features on the nodes, in order to provide predictions. According to Zhou et al., there are many variants of GNNs include graph convolutional network (GCN), graph attention network (GAT) and graph recurrent network (GRN). Wu et al. [9] provide an overview of GNNs, a new taxonomy, as well as a set of evaluation techniques for GNNs, including open-source codes and benchmark data sets. Xu et al. [10] highlight that there is limited understanding of the representational properties and the limitations of GNNs, describing an aggregation scheme, in which the representation vector of a node is calculated by accumulating recursively and transforming the representation vectors of the adjacent nodes. Eventually, they present a theoretical framework

for analyzing the ability of GNNs to capture different graph structures.

Spectral approaches of GNNs have their origin in signal processing, while working with a spectral representation of graphs. A graph signal is initially transformed by the graph Fourier transform and then the convolution operation is carried out, leading to the resulting signal, which is transformed once again using the inverse graph Fourier transform. According to Zhu & Koniusz [11], despite the great significance of Graph Convolutional Networks (GCNs) for learning, there is also the need for special architectures. As a direct consequence, a Simple Spectral Graph Convolution (S2GC) was proposed, so as to achieve the target performance. In general, a GNN-based technique could be applied on data extracted from popular social networks applications. For example, Behún [12] conducts an overview and a comparison of such GNN-based techniques, which are used to address problems of learning on graph data from the Tripadvisor website.

Based on the literature, Graph Databases are considered the most suitable technique for graph data, because of their ability to manage large-scale data from social networks. In addition, GNN can effectively be used for predictions by analyzing the topology of the graph. Nevertheless, there is a gap in the literature about the interoperation of GNN models to graph databases, which will be discussed at the sections below.

### III. A COMPARATIVE PRESENTATION OF HANDLING GRAPH DATA AND RELATIONAL DATA

Choosing and developing performant database, which will support operational functionalities and decision-making activities, seems to be a challenging task, especially when big datasets and complex data structures are involved. Conventional databases, such as mysql and oracle are designed to store relational data. The data are retrieved using joins among the tables. The performance of a query is directly associated with the complexity of the data model and the number of tables, used for the execution of the query. Those databases support codifying tabula structures and paper forms. Robinson et al. [13]

pointed out the significance of implementing and utilizing database infrastructures for large datasets with complex relationships. Nevertheless, issues occurred when the relationships among the entities increase, or / and the complexity of the queries augments. When the data are stored on graph databases, users manage data more naturally, because they are represented as network of nodes - entities. In such cases, Neo4j significantly outperforms relational databases, since it models the data as a graph network, while they are retrieved by querying the graph. One of the main advantages of relational databases is that they can eliminate the redundancy, and ensure the integrity of data, which originated from the normalization process. On the other hand, the relational databases are not so flexible on changes that may occur to the model. Sandell et al. [14] conducted a performance comparison between state-of-the-art graphs and relational databases, evaluating their efficiency and capabilities. They conducted Performance Comparison Analysis of ArangoDB, MySQL, and Neo4j and the results indicate that Neo4j performs faster in querying connected data than MySQL and ArangoDB.

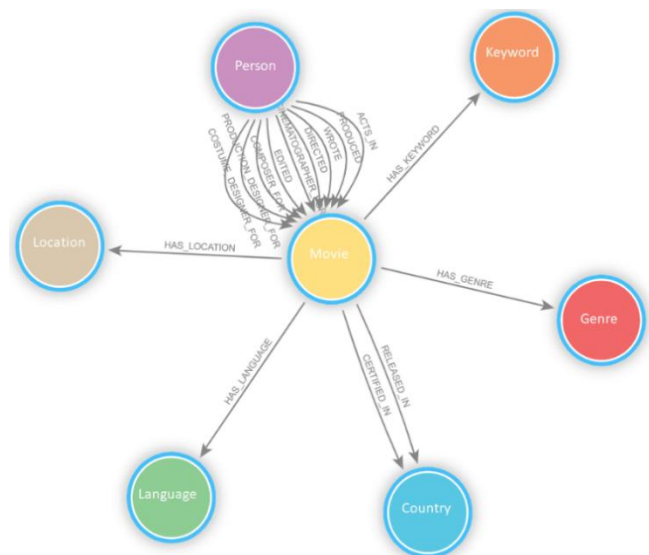


Figure 2 Graph Database Model

In this section, mysql and neo4j are compared, using an IMDB dataset of movies and people (actors, directors, etc.). In this dataset, there are seven entities (Country, Genre, Movie, Person, Language, Location and Keyword) and 15 relationships (Acts\_In, Has\_Keyword, Has\_Genre,

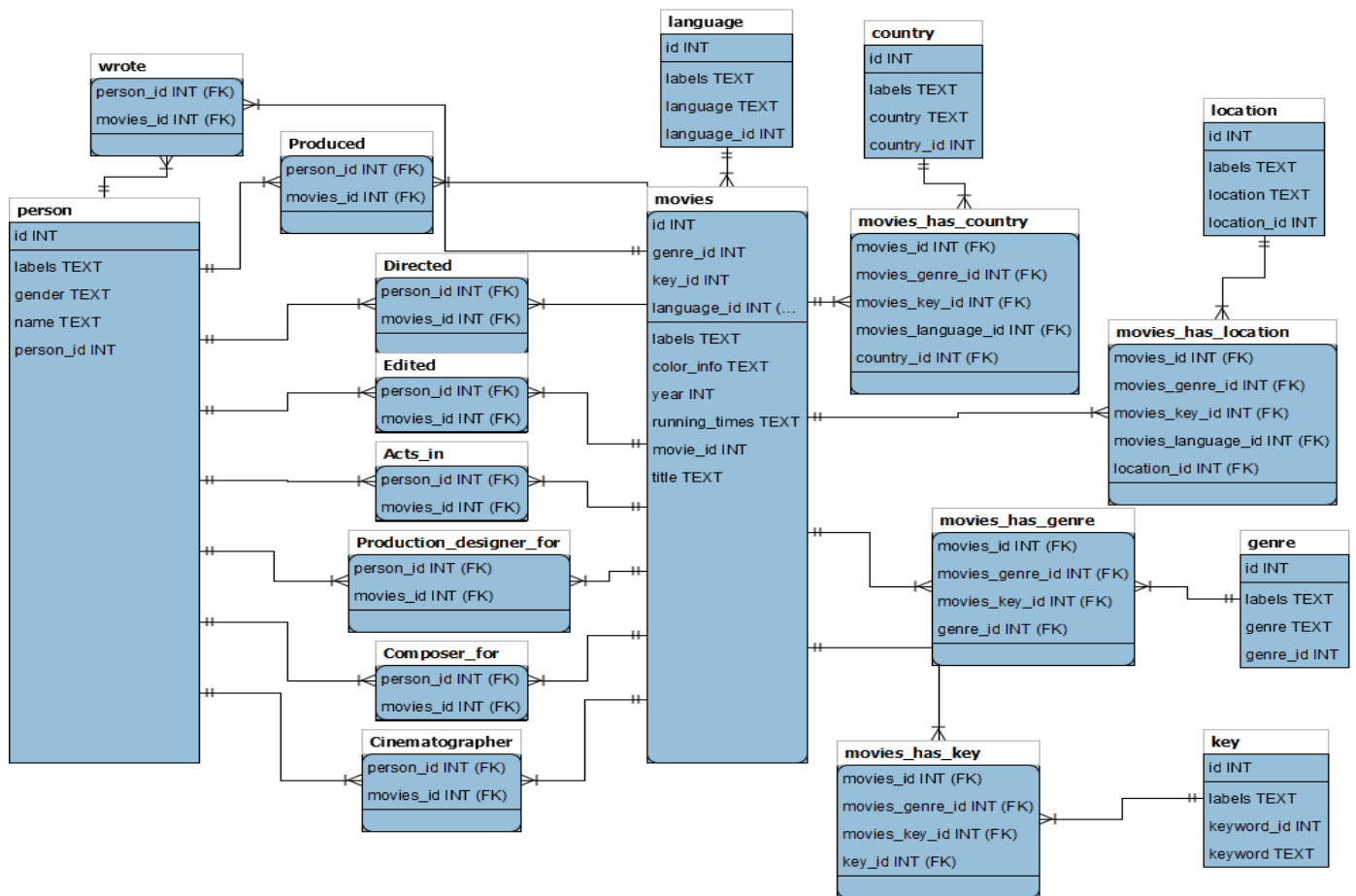


Figure 3 Relational Data Model

Released\_In, Produced, Wrote, Directed, Has\_Language, Cinematographer\_For, Edited, Has\_Location, Composer\_For, Certified\_In, Production\_Designer\_For, Costume\_Designer\_For). The comparison will be held on the following categories:

- Data Modeling
- Querying of Data
- Data Representation

#### A. Data Modeling

In figures 2 and 3, the models of graph databases (neo4j) and relational databases (mysql) are presented respectively. As it is obvious from the figures, the relational schema is more complex. Despite the fact that relational databases provide structured and high-quality data without inconsistencies, due to the normalization process, the relationships and the constraints that exist among the entities have to be considered. The schema of a relational database cannot efficiently incorporate changes on the relationships on the database, as the schema is inflexible. On the other hand, because

graph database is scalable and flexible it is easy to adopt any change on the relationships among the entities

#### B. Querying the Data

The performance of a query's execution is of vital importance and is associated with the data complexity and diversity. In relational databases, the performance is getting worst in case JOIN operations applied to large tables in contrast to graph databases. When the data are generated using two to three hops across tables, the RDBMS provide results in satisfactory response time. However, when more hops are required, then the response time is amplified, while in several cases, some tables will be locked, as they will wait the completion of the execution of the submitted queries.

#### C. Data Representation

The data representation is associated with the decision-making process, which depends on the selection of the database selected for data retrieval. The two main criteria that are used in

order to decide the appropriate visualizations are:

- the cost to produce the results and
- the way the representations lead to informative decisions.

Figures 4 and 5 present the data in Neo4j and in MySQL. The graph-based representation lead to decisions that are more accurate and provide more functionalities so as to identify in detail the associations among the entities. For instance, in figure 6 the movie with title “Mystery of Maya” is selected and the entities associated with that movie are displayed. Nevertheless, in order to examine all the entities related to the selected movie in a relational database, too many queries, have to be executed that cause high demands on resources and increase the required time.

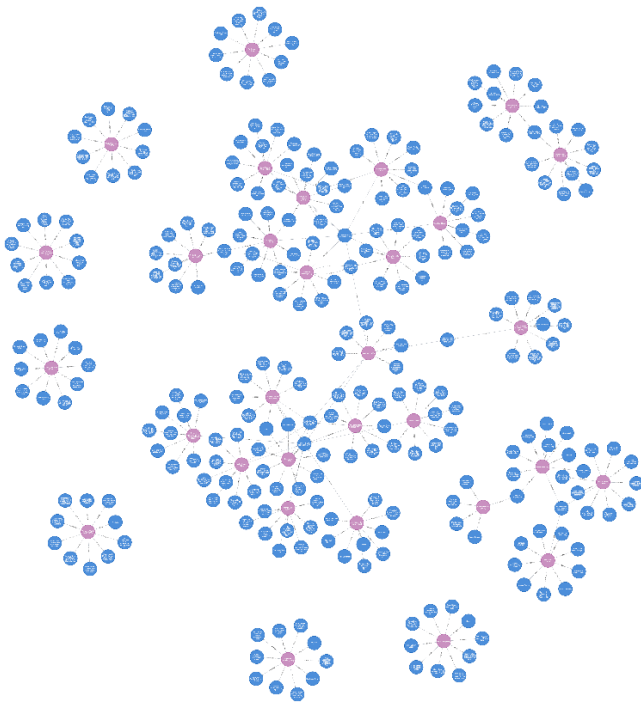


Figure 4 . Neo4j Graph Representation

title	location
American Hustle (2013)	Union Station, Worcester, Massachusetts, USA
American Hustle (2013)	Wang Center - 270 Tremont Street, Boston, Ma..
American Hustle (2013)	Worcester Art Museum, Worcester, Massachus..
American Hustle (2013)	Worcester, Massachusetts, USA
An Education (2009)	Bloomsbury Service Station - 6 Store Street, Blo..
An Education (2009)	Caf? Rosetta, Mattock Lane, Ealing, London, En.
An Education (2009)	Caf? de Paris, Coventry Street, Soho, London, ..

Figure 5 MySQL Tabular Representation

Table 1 illustrates some IMDB data used in the implementation process for both DBMSs.

Rank	1.0	2.0
Title	Guardians of the Galaxy	Prometheus
Genre	Action,Adventure,Sci-Fi	Adventure,Mystery,Sci-Fi
Description	A group of intergalactic criminals are forced to work together to stop a fanatical warrior from taking control of the universe.	Following clues to the origin of mankind, a team finds a structure on a distant moon, but they soon realize they are not alone.
Director	James Gunn	Ridley Scott
Actors	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe Saldana	Noomi Rapace, Logan Marshall-Green, Michael Fassbender, Charlize Theron
Year	2014.0	2012.0
Runtime (minutes)	121.0	124.0
Rating	8.1	7.0
Votes	757074.0	485820.0
Revenue (Millions)	333.13	126.46
Metascore	76.0	65.0

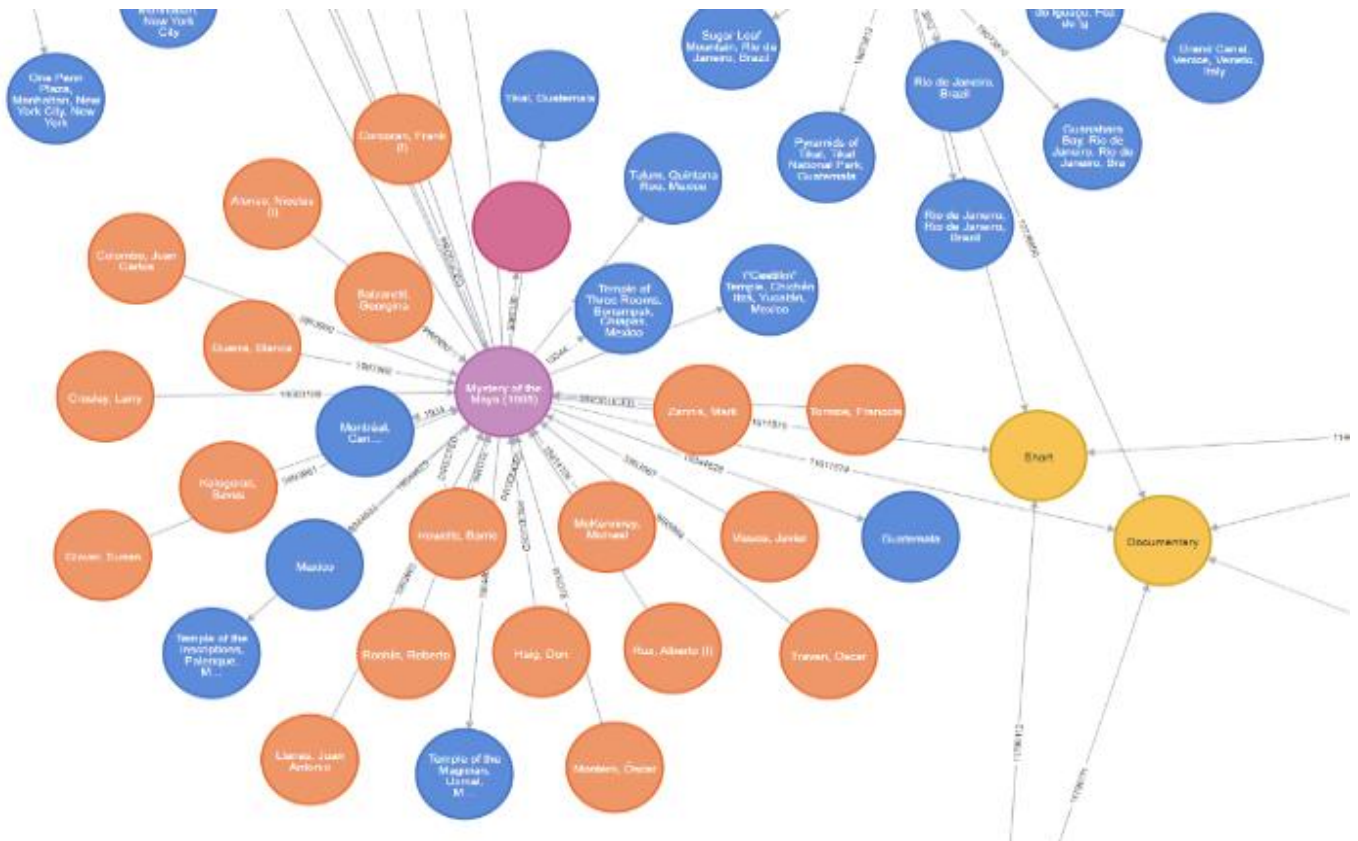
TABLE 1 IMDB DATASET

#### IV. GRAPH NEURAL NETWORKS IN SOCIAL RECOMMENDATION

In recent years, there is a vast development in social network applications (such as Facebook, X, Amazon, etc.) that cover diverse areas of consumer interests and needs. All these applications keep the users' data stored. In most cases, they also record their social connections, as well as their interactions over time. These data is also known as Big Data, due to their massive volume. In industry, there is an enormous interest in exploiting valuable information from the before mentioned data, especially for their application in Social Recommendation. Social Recommender Systems are used in order to predict new connections or actions for a person using their social activities, connections and interactions with the other users in their social network. Those systems are valuable for the industry and especially in e-commerce, to offer more personalized recommendations to the users, based on their previous behavior or their interactions in their social network.

In Social Recommendation, ignoring social relations would lead to loss of valuable information. Social relations that are apparent in a social network have an effect on the preferences of an individual. As McPherson M. et al. [15] suggest, social homophily indicates that a user is more likely to connect to another user with similar





attributes and preferences. Likewise, Marsden P., et al. [16] talk about social influence, which indicates that users which are connected to each other, tend to have similar preferences. It is evident that social relations among people are of significant interest and provide valuable insights to analysts. Many different models have been developed and as Gao C., et al. [17] and by Sharma K., et al. [18] mentioned in their article, the first efforts to build Social Recommender Systems was the Matrix Factorization (e.g. [19] and [20]), followed by the deep learning models (e.g. [21], [22]). Among the deep learning models, various GNNs models are proposed. A plethora of such models are proposed in literature, with an ever-increasing rate over the last recent years [18], due to the fact that GNNs utilize more effectively the high order connectivity in social networks. Often GNNs models for Social Recommendation outperform the previous proposed models on public benchmark datasets [23].

According to literature, there are surveys that cover different aspects of GNN based Social Recommender Systems and referencing many of the proposed models (e.g. [17], [18]). The two main aspects covered, are the input types and the

architectures of these systems. The input types refer to the graph types that model the data input and can be homogeneous graphs (i.e. edges connect only two nodes and there is one type of nodes and edges), heterogeneous graphs (i.e. edges connect only two nodes and there may be more than one type of nodes and edges), and hypergraphs (i.e. edges may connect more than two nodes). The input data are used by the Social Recommender System to be trained. As was suggested by Sharma, et al. [18], the architectures of these Social Recommender Systems have three key components: the encoder, the decoder and the loss function. The encoder is a GNN model that produces node embeddings (users and items). The GNN encoder must be suited for the corresponding graph of input data (e.g. homogenous, heterogenous). Decoders base their predictions on the embeddings, which are given by the encoders. Ultimately, the loss function trains the model so as to achieve accurate predictions.

Various GNN models are proposed as the encoder in the previous mentioned architecture, such as Graph Convolutional Networks [24], GraphSAGE [25], Graph Attention Networks [26], Gated Graph Neural Networks [27] and

Hypergraph Neural Networks [28]. Graph Convolutional Networks operate on spatial, or on spectral domain. On spatial domain, a neighbor of nodes is selected for each node and a convolution is applied in its neighbor. On spectral domain, the signal is transformed from spatial to spectral domain, a convolution is applied on the signal, and it is transformed back to the spatial domain. Either way, Graph Convolutional Networks leverage information from their neighboring nodes.

Social Recommender Systems have attracted research interest, since they have many possible application domains in industry and cover needs with high demand. Many GNNs based architectures have been proposed over the years. But GNNs have been also proven their efficiency for graph data in domains other than social networks, such as physical systems [29] and protein structure [30]. GNNs have been increasingly applied and evolve within machine learning for graph data, as they effectively leverage the complex, higher-order connectivity inherent in graph structures.

## V. CONCLUSION

In this study, the research questions RQ1, RQ2 were affirmatively addressed through a comprehensive review and analysis of relevant literature findings. Concerning the first research question whether there is any satisfactory theoretical and experimental verification in the literature of the benefits of applying graph neural networks and graph databases to managing large-scale data in social networks, a large increase in publications of research results in the topics of GNN and Graph database is identified and mentioned in the paper. In particular, a remarkable rise of research outcomes has been observed since 2020. There is also a continuous reference to big data management and applications. Graph databases are widely used to manage Big Data and social network data. GNNs are increasingly being proposed as machine learning techniques for graph data, particularly in the aforementioned context. Consequently, GNNs and Graph Databases can be incorporated to represent, query and make predictions in big network data. Regarding the second research

question if there are any comparative advantages of using graph databases over relational databases in managing large-scale data in social networks, the answer was experimentally validated, as the technology behind relational databases has reached maturity and is extensively utilized to meet business requirements. Relational databases are particularly well-suited for applications involving structured data, such as transaction processing and customer data management, while ensuring data integrity. Although the schema of a relational database is fixed, it allows for easy modifications. These databases can be scaled vertically with relative ease; however, their performance may decline as the size of the datasets increases. In contrast, graph database technology is relatively new, yet it excels at managing large datasets. These databases offer the advantage of intuitive usage. Additionally, they easily scale horizontally through partitioning. Moreover, graph databases give emphasis on supporting the relationships (edges) between objects (nodes), and, therefore, are well-suited for semantic search and recommendation engines.

## VI. REFERENCES

- [1] Almabdy, S. (2018). Comparative analysis of relational and graph databases for social networks. In 2018 1st International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-4). IEEE. <https://doi.org/10.1109/CAIS.2018.8441982>
- [2] Kumar Kaliyar, R. (2015, May). Graph databases: A survey. In International Conference on Computing, Communication & Automation (pp. 785-790). IEEE. <https://doi.org/10.1109/CCAA.2015.7148480>
- [3] Bhattacharyya, A., & Chakravarty, D. (2020, January). Graph database: A survey. In 2020 International Conference on Computer, Electrical & Communication Engineering (ICCECE) (pp. 1-8). IEEE. <https://doi.org/10.1109/ICCECE48148.2020.9223105>
- [4] Tian, Y. (2023). The world of graph databases from an industry perspective. ACM SIGMOD Record, 51(4), 60-67. <https://doi.org/10.1145/3582302.3582320>
- [5] Xirogiannopoulos, K., & Deshpande, A. (2017, May). Extracting and analyzing hidden graphs from relational databases. In Proceedings of the 2017 ACM International Conference on Management of Data (pp. 897-912). <https://doi.org/10.1145/3035918.3035949>
- [6] Besta, M., Iff, P., Scheidl, F., Osawa, K., Dryden, N., Podstawski, M., Chen, T., & Hoeffler, T. (2022, December). Neural graph databases. In Learning on Graphs Conference (pp. 31-1). PMLR.



- [7] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, 20(1), 61-80. <https://doi.org/10.1109/TNN.2008.2005605>
- [8] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI open*, 1, 57-81. <https://doi.org/10.1016/j.aiopen.2021.01.001>
- [9] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S.. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1), 4-24. <https://doi.org/10.1109/TNNLS.2020.2978386>
- [10] Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2018). How Powerful are Graph Neural Networks?. In *International Conference on Learning Representations*.
- [11] Zhu, H., & Koniusz, P. (2021, May). Simple spectral graph convolution. In *International conference on learning representations*.
- [12] Behún, M. (2023). Graph neural networks and their application to social network analysis [Master's thesis, Charles University]. Charles University Digital Depository.
- [13] Robinson, I., Webber, J., and Eifrem, E. (2015). *Graph Databases: New Opportunities for Connected Data* (2nd ed.). O'Reilly Media.
- [14] Sandell, J., Asplund, E., Ayele, W. Y., & Duneld, M. (2024). Performance Comparison Analysis of ArangoDB, MySQL, and Neo4j: An Experimental Study of Querying Connected Data. *Hawaii International Conference on System Sciences 2024*
- [15] McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1), 415-444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- [16] Marsden, P. V., & Friedkin, N. E. (1993). Network studies of social influence. *Sociological Methods & Research*, 22(1), 127-151. <https://doi.org/10.1177/0049124193022001006>
- [17] Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., Quan, Y., Chang, J., Jin, D., He, X. & Li, Y. (2023). A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems*, 1(1), 1-51. <https://doi.org/10.1145/3568022>
- [18] Sharma, K., Lee, Y. C., Nambi, S., Salian, A., Shah, S., Kim, S. W., & Kumar, S. (2024). A survey of graph neural networks for social recommender systems. *ACM Computing Surveys*, 56(10), 1-34. <https://doi.org/10.1145/3661821>
- [19] Jamali, M., & Ester, M. (2010). A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 135-142). <https://doi.org/10.1145/1864708.1864736>
- [20] Yang, B., Lei, Y., Liu, J., & Li, W. (2016). Social collaborative filtering by trust. *IEEE transactions on pattern analysis and machine intelligence*, 39(8), 1633-1647. <https://doi.org/10.1109/TPAMI.2016.2605085>
- [21] Deng, S., Huang, L., Xu, G., Wu, X., & Wu, Z. (2017). On deep learning for trust-aware recommendations in social networks. *IEEE transactions on neural networks and learning systems*, 28(5), 1164-1177. <https://doi.org/10.1109/TNNLS.2016.2514368>
- [22] Krishnan, A., Cheruvu, H., Tao, C., & Sundaram, H. (2019, November). A modular adversarial approach to social recommendation. In *Proceedings of the 28th ACM international conference on information and knowledge management* (pp. 1753-1762). <https://doi.org/10.1145/3357384.3357898>
- [23] Wu, S., Sun, F., Zhang, W., Xie, X., & Cui, B. (2022). Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5), 1-37. <https://doi.org/10.1145/3535101>
- [24] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*.
- [25] Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.
- [26] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. In *International Conference on Learning Representations*.
- [27] Li, Y., Tarlow, D., Brockschmidt, M., & Zemel, R. (2016). Gated graph sequence neural networks. In *International Conference on Learning Representations*.
- [28] Feng, Y., You, H., Zhang, Z., Ji, R., & Gao, Y. (2019, July). Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 3558-3565). <https://doi.org/10.1609/aaai.v33i01.33013558>.
- [29] Battaglia, P., Pascanu, R., Lai, M., & Jimenez Rezende, D. (2016). Interaction networks for learning about objects, relations and physics. *Advances in neural information processing systems*, 29.
- [30] Fout, A., Byrd, J., Shariat, B., & Ben-Hur, A. (2017). Protein interface prediction using graph convolutional networks. *Advances in neural information processing systems*, 30.



Stratos Tsolakidis is a PhD candidate in the Department of Informatics and Computer Engineering of the University of West Attica. His topic is Graph Neural Networks and applications on Social Networks. His previous studies include a graduate degree in Mathematics from the National and Kapodistrian University of Athens and an MSc in Computer Science from the Athens University of Economics and Business. Currently, he works at the Ministry of Education, Religious Affairs and Sports in Greece.



Dr. Anastasios Tsolakidis received his PhD degree in computer science from the University of Limoges, France, in 2015 and is currently an Assistant Professor at the Department of Archival, Library and Information Studies at the University of West Attica. His research interests lie in the fields of Visual Analytics, Decision Support Systems, Business Intelligence and E-Government. He has previously worked as a research associate in many European and Greek projects and as a Business Intelligent Analyst at "e-Government Center for Social Security (IDIKA SA)" at the sector of E-Health.



Evangelia Triperina is a PostDoc researcher at the Department of Archival, Library and Information Studies of University of West Attica. She holds a PhD in Computer Science from the University of Limoges (France), with a thesis entitled "Visual interactive knowledge management for multicriteria decision making and ranking in linked open data environments". She holds an MSc in Information Technology, Image Synthesis and Computer Graphics from the University of Limoges (France). She worked in European research projects at GRNET, Agro-Know Technologies and the University of West Attica.



Nikitas N. Karanikolas is a professor at the Department of Informatics and Computer Science of the University of West Attica. He was Head of the Technological Educational (1996) Institute's Library (TEI of Athens Library). One year later, November 1997, he got the position of chief in the Department of Informatics and Organization of the Aretaieio University Hospital until November 2004. He participates as a coordinator and/or key researcher in European and nationally funded research and development projects. His research work has been published in international journals and conference proceedings.



Christos Skourlas is an emeritus professor at the Department of Informatics and Computer Science of the University of West Attica. He was an analyst-programmer and head of the systems with the National Documentation Centre of Greece (1983- 89) and a research assistant with the Nuclear Research Centre "Demokritos" (1977-82). He was head of the research lab "Data, Information and Knowledge Management (InfoDat\_KM)". He participates as a coordinator and/or key researcher in European and nationally funded research and development projects. His research work has been published in international journals and conference proceedings.