University of Sindh Journal of Information and Communication Technology (USJICT)
Volume 8, Issue 1, 2024
ISSN-E: 2523-1235, ISSN-P: 2521-5582
Website: http://sujo.usindh.edu.pk/index.php/USJICT/

Exploring Inter-connectivity Link Prediction: An Insight from Social Network Science

Jawad Hussain Awan[1], Syed Raza Hussain Shah[2], Shazma Tahseen[3], and Shabreena[4]

[1] Faculty of Engineering, Sciences and Technology, Iqra University, Karachi, Pakistan

[2,4] Faculty of Engineering and Technology, University of Sindh, Jamshoro, Pakistan

[3] Department of Information Technology, Liaquat University of Medical and Health Sciences, Jamshoro, Pakistan

jawad.hussain@iqra.edu.pk[1], and raza.shah@usindh.edu.pk[2]

Abstract: Physical and computational science communities are becoming interested in link prediction for complicated networks. Thus, various algorithms can be used to extract missing data, detect erroneous interactions, assess network evolution mechanisms, and so on. The contributions from computational science views and approaches, such as arbitrary methods and maximum likelihood methods are highlighted in this paper, which summarizes current advances in link prediction algorithms. One of the frequently discussed research topics in the area of social network analysis is link prediction. Numerous practical applications of this issue may arise in the future, including modelling recommender systems, fraud detection, stock prediction, and more. One can comprehend the dynamics of network evolution if the absent links or the links that will appear in the future are foreseen. One may undoubtedly develop superior choice models for the business strategy based on the observed patterns of structural changes, which can have high performance value and lower market risks. In this paper, the author has presented overview of background of link prediction issue in Social Network Science and also presented the link prediction methods for modern research in network science.

Keywords: Link, Prediction, Inter-connectivity problem, and Social Networks;

I. INTRODUCTION

Generally, collaborative networks play major role in social network theory. In these networks, scientists are connected with each other forming authorships or co-authorships [1]. Two scientists are treated as connected or having tie whenever both have published a paper mutually. In this regard, various researchers have used databases [2] like (Medline, Los Ala-MOS, Stanford Public Information Retrieval System (SPIRES), and Networked Computer Science Technical Reference Library (NCSTRL)). These databases archive provide support to the researchers in analysis and constructing a collaborative network. These provide the archives as Medline (Medical and Biology), Los Ala-mas and SPIRES (Physics) and NCSTRL (Computer Science). Those refer, Number of authors per paper, Number of papers per author, Number of ties between authors or scientists, and degree of network clustering as basic statistical parameters, and evaluation parameters for this network in datasets. Although, shortest path, betweeness and funneling, average distances are measures essential for these networks [3] to understand the communication and pattern in which collaborative network is constructed. However, Geodesic is primary concept in collaborative network, although it is not unique for each connected node having two vertices. But, helpful for association. Then, shortest path is chosen to connect the vertices in network. Though, Standard breadth-first search algorithm is used in this network to calculate geodesic. For example, scientist A and B have no any tie or unconnected either they belong to different department/ institute/ university/ campus or they have not published a paper yet. Hence, referral chain concept is adopted to make new ties among unknown scientists. Further, parameters can be measured using following equations.

$$O(\mu v2) \quad (1)$$

$$O(v2) \quad (2)$$

$$O(\mu) \quad (3)$$

Equation (1) is applicable to calculate time of order (betweeness) of all vertices existing in network, while Eq. (2) used to find the shortest path and lastly, Eq. (3) evaluates the time taken in shortest path. Hence, Eq. (1) is applied to calculate shortest path in n vertices. Though, researchers have added weighted in collaborative network to compute the shortest path via mean distance with straightforward manner. In this way, generalization of betweeness is computed and the valuable information is available in bibliographic databases. Hence, it may be future direction for new researchers. Addition to theses, degree is also major component and refers as "the number of connected connections with focal point of

given network structure". Additionally, degree is only used in small level network because it is not efficient to reach at nodes quickly and access the resources, and information or knowledge. Hence, closeness is better choice for this purpose and it is defined as "the sum of shortest distance from focal point or node". It is also helpful for medium level networks, due to unstable nature. If two node's connection is broken then it becomes failure. Hence, it is restricted for large level of components. That is the reason, that Betweenness is appropriate in larger components of networks and centrality is suggested, and it is defined as Centrality is an approach by which closer and central nodes are identified. A greater number of ties along with well connectivity, reachability and good location in control flow are the advanced features of centrality.

II. LITERATURE REVIEW

In [7], the researchers investigate the challenge of identifying sentiment relationships in online social networks in the lack of sentiment-related information. Due to the absence of clear sentiment linkages, we first create a labelled, heterogeneous, and entity-level sentiment dataset. We offer Signed Heterogeneous Information Network Embedding (SHINE), a deep-learning-based network embedding framework to extract users' highly nonlinear descriptions while keeping the shape of actual networks, to know a lot from these heterogeneous networks. We carry out in-depth tests to gauge SHINE's effectiveness. Results from the experiments support SHINE's ability to compete against a number of strong base lines and show how social relationships and profile data are particularly useful in cold start scenarios [8].

In [9] study, the authors presented a unique PME model to incorporate heterogeneous data networks, which effectively resolves the difficult problem of modelling node and link heterogeneities in intricately built relation-specific spaces. Additionally, we use Euclidian Distance as a metric to embed nodes' distances, which fulfills the important constraint and simultaneously maintains first- and second-order closeness. In order to further accelerate model adaptation and address the significant skewness of the heterogenous link distribution w.r.t. relations, we additionally introduce a new loss-aware adaptable optimistic sampling technique. Our concept is adaptable and may be used to any network, with no restrictions on its use. A sizable Yelp heterogenous network was used for extensive testing, and our PME model significantly out performs the state-of-art heterogenous network embedding methods. They provide a brief overview of the two-stage approach for the partnership prediction challenge in the current publication. Statistical tests show the potential for using citation data to expand the co-authorship structure. The vectorized descriptions of the network component and other aspects, which are represented as features of the associated graph elements, are digital components in the dataset produced by GNNs. In the current work, our primary contributions are as follows [10]:

1. From the matching HEP-TH citation network, we obtain the co-authorship network.

2. The presence of important metadata enables us to interpret the papers' scientific significance metrics (e.g., impact factor).

3. To assess its impact on the defect prediction reliability, we combine structural data from the citation graph and apply it to the co-authorship network. The authors noticed that the region under the receiver operating characteristics assess the value of all classifiers increased after combining SIN and KMC, showing that SIN and KMC can significantly increase classification accuracy. Finally, based on the suggested model, we offer suggestions for studies in a cooperative way [11].

A significant issue with network-structured information is link predictions. Some scoring functions, like similar neighbors and the Katz index, are used by clustering heuristics to determine the likelihood of linkages. Due to their simplicity, interpretability, and for some of them, adaptability, they have a wide range of practical applications. Each heuristic, though, makes a firm prediction about when two nodes are likely to link, which restricts their use in channels where these expectations are incorrect. In this case, developing a suitable heuristic from an existing network is a more sensible approach than employing predetermined ones. In order to learn a "heuristic" that works for the existing service, we want to learn a function mapping the subgraph patterns to link existence by identifying a local subset on each targeted link.

In this research, we investigate the link prediction heuristic learning paradigm. We start by creating a brand-new fading heuristic theory. The theory demonstrates that all of these heuristics may be accurately approximated from local subgraphs and combines a variety of heuristics into a single framework. Our findings indicate that local subgraphs store a wealth of information about link presence. Second, using a graph neural network and the -decaying theory, we suggest a fresh approach for learning heuristics from local subgraphs (GNN). It routinely performs admirably on a broad range of issues, according to experimental data [12].

Since there are trillions of nodes and edges in social networks, modelling is still a difficult issue. In this study, we introduced a novel node-based similarity measure based on the width and degrees of the path. We looked to see how the direction affected the algorithm's reliability and found that in all datasets, the lengths of 2 and 3 produced good accuracy. Based on both approaches, we evaluated popular link prediction systems and uses the AUC to compare similarity nodes with the suggested technique. Results demonstrated the suggested method's exceptional performance for several networks based. Additionally, as a binary job, we solved the link prediction problem using machine learning algorithms. In fact, using the suggested algorithm as an extra feature boosts accuracy. The parameter-free proposed method based on path depth and nodes degree is effective and has great promise in solving the link prediction model for large social networks, according to the experimental tests [13].

In online social networks, link prediction is applied to identify potential interactions between its users. One of the key areas of focus for clustering research up to this point has been the coauthor ship networks. In order to make effective

recommendations for authors who can collaborate on a scientific project, studies have concentrated on understanding the situation and making responses. It is ideal to create a similarity measure among two omnipresent writers in a co-authorship network and use it to determine the most likely co-author (s). The assimilation of the paper's content in the metric itself was not considered by the pertinent studies, nevertheless. This is significant while thinking about scientific collaboration because it's probable that writers with similar research interests are more likely to publish a work together than those with distinct research preferences. In this research [14], the authors introduce a novel metric called LDA cosin based on information resemblance for link prediction in the coauthor ship network. We present a link prediction technique based on language models and algebraic ideas of the link prediction in the co-authorship network. On the publicly accessible bibliographical dataset, the new metric is experimentally confirmed.

III. THE LINK-PREDICTION PROBLEM

In [15], suppose two scientists are colleagues, working together in same environment and are closely attached, may write or publish articles in future under authorships. But both have no publication right now. In another scenario, two scientists have published and written an article but they apart from each other in few years and will not publish articles in co-authorships. In both scenarios, future situation becomes vice versa and predictions become unstable. Hence, link prediction is difficult in social networks as shown in Figure no. 01.

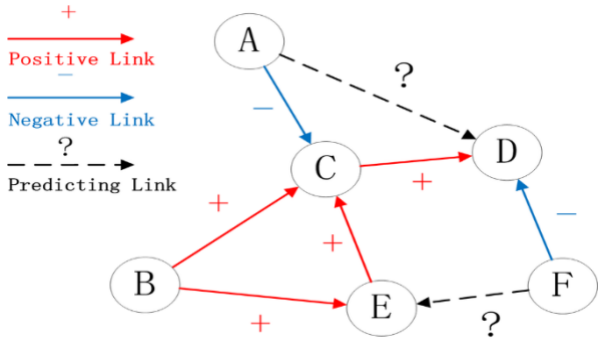

Figure 1: Problem of Link Prediction [16]

The authors mentioned a range about 40 to 50, which may lead to outperform. The link prediction is also inferring missing links from an experiential network. In this research, authors have chosen a collaboration network of five graphs and compared them by using two parameters (K training and K test) and Set core (predicted links in K training and K test). K training includes G [t0, t0'] (three years publications from 1994 to 1996) and K test includes G [t1, t1'] (three years publications from 1997 to 1999).
Following equation is applied for evaluation.

$$E^{*}new := Enew \cap (CoreCore) \text{ and } n := |E\nu\varepsilon\omega| \quad (5)$$

In addition, the paper also discusses two methods of measurement. First method is based on node neighborhoods while second one is based on ensemble of all paths. Node Neighborhoods, Common neighbors, Preferential attachment are methods based on node neighborhoods, while Katz, hitting time, PageRank, and variants and SimRank are methods based on the ensemble of all paths.

A. Link Predictions Methods

These all methods are applicable for link predictions between or among nodes in network as shown in figure no. 02.

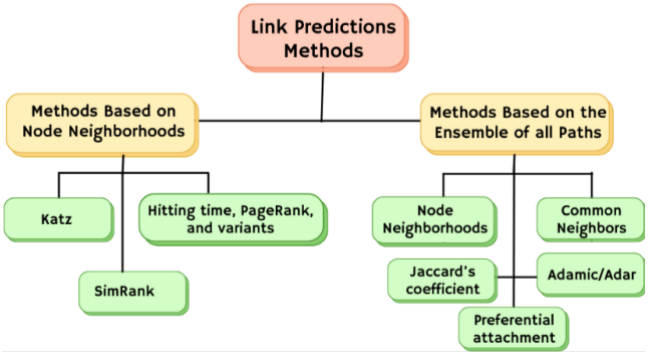

Figure 2: Link Prediction methods

1) Methods Based on Node Neighborhoods

a) Node Neighborhoods

Suppose Gcollab is network where x is a node and represents set of neighbors [17]. According to method, y node is neighbor node and more central to node x, may form a link in future. For example, two colleagues are working in same organization and may be close to each other in future and will submit or publish articles. Link may be formed between them.

b) Common Neighbors

In this approach, two nodes x and y are connected to each other and their neighbor common nodes may be linked or connected to focal nodes (x,y) in future [18]. Following equation is defined

$$\text{Score } (x,y) := |\neg (x) \cap \neg (y)| \quad (6)$$

c) Jaccard's coefficient

This method is mostly used as similarity metric and computes the probability of features resembles in nodes when random feature is selected existing in any node, x and y are taken as node and f as feature [19]. Following equation is applicable for computation when f is taken for neighbor nodes in Gcollab

$$\text{Score } (x,y) := |\neg (x) \cap \neg (y)| / |\neg (x) \cap \neg (y)| \quad (7)$$

d) Adamic/Adar

This approach is also similar to Jaccard's coefficient [20]. In this approach, decision is taken when two nodes (like web home pages) relate to each other strongly. Following equation is defined to compute similar features of nodes (web pages).

$$\Sigma\ 1/(\log f_0(\text{frequency}(z))) \quad (8)$$

Where z is shared feature of x and y nodes

e) Preferential attachment

This technique is proposed late due to experiential evidence [21]. Thus, the probability of co-authorship of nodes (x and y) is associated with "the product of the number of collaborators of x and y" and measures by using following equation.

$$\text{Score}(x,y) := |\neg(x)| . |\neg(y)| \quad (9)$$

B. Methods Based on the Ensemble of All Paths

This technique finds shortest path from set of nodes connected in network while above discussed methods are applicable or related to neighbor nodes.

1) Katz

This measurement approach directly combines the shortest paths from a set of paths and count shortest path based on length of linked node paths [22]. Following equation is also defined for this purpose.

$$\text{Score}(x,y) := \Sigma\_(l=1)^{\alpha}\beta^{l} * |\text{paths}(l)x,y| \quad (10)$$

2) Hitting time, PageRank, and variants

The hitting time (H) is expressed as the predictable number of steps necessary which starts from focal node x to reach at destination node y [23]. Because the hitting time is not in general symmetric, H is natural and is not considered in general symmetric. Hence, used as 'natural proximity measures.

$$Cx,y := Hx,y * Hy,x \quad (11)$$

When negated used and equation becomes as

$$\text{Score}(x, y) := Hx,y * \pi y + Hy,x * \pi x \quad (12)$$

PageRank is used to measure web pages for link prediction [24]. It uses two parameters either o or 1 where it moves to random neighbor nodes with (1-∞) probability. It also defines score(x, y) under the rooted.

3) SimRank

SimRank [25] is follows recursion approach where nodes are linked on similarity feature. If two nodes are similar to the extent, and both be linked to similar neighbors. Besides these, the authors have also discussed three high-level approaches such as Low-rank approximation, Unseen bigrams and Clustering. The researchers investigated 03 applications such as Ranking the Katz measure and common neighbors, and defining score (x, y) in low-rank approximation. Clustering is an advanced approach applied to improve the quality of a predictor. With respect to this, weak edges are removed from Gcollab. After this, predicator is executed on cleaned-up or new subgraph.

IV. RESEARCH METHODOLOGY

This section discusses about the completion of procedure to identify the missing link from the dataset. In this research three datasets (i-e: facebook, Google+ and Twitter) have been utilized. These datasets have been taken from Stanford University, USA. The detailed discussion and experimental results are illustrated in Section VI. Refer Figure No.3, at initial stage, Network Data (Vertices and Edges) is used to apply similarity feature extraction approach using any one method from the available i-e: Common, Jaccard, Preferential, Neighbor and Coefficient. After that, Feature vector approach is used along with training classifier using trained data. After the execution of these processes, link is predicted and optimized links are shown for future use.

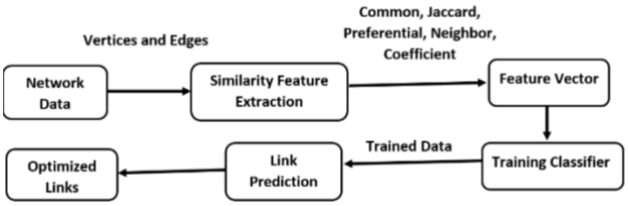

Figure 3: Research Methodology

V. RESULTS AND DISCUSSIONS

This section consists of three datasets i-e: Facebook, Google+ and Twitter. The datasets are received from the databases available at Stanford University.

A. Facebook

The Facebook dataset, consisted of friend list from Facebook. The data is also collected through a survey. The Facebook dataset consists of profiles as node features, their circles, and surrounded networks. The details of the dataset are discussed below.

TABLE I: DATASET STATISTICS OF FACEBOOK

Parameters	Values
Nodes	4039
Edges	88234
Nodes in largest WCC	4039 (1.000)
Edges in largest WCC	88234 (1.000)
Nodes in largest SCC	4039 (1.000)
Edges in largest SCC	88234 (1.000)
Average clustering coefficient	0.6055
Number of triangles	1612010
Fraction of closed triangles	0.2647
Diameter (longest shortest path)	8
90-percentile effective diameter	4.7

Table I illustrates the results of Facebook for nodes (4039), edges (88234), nodes in largest (4039), edges in largest (88234), average clustering coefficient (0.6055), number of triangles (1612010), fraction of closed triangles (0.2647), diameter (8) and 90-percentile effective diameter (4.7).

B. Google+

The Google+ dataset, consisted of circle list from Google+. The data is also collected through a survey. The Google+ dataset consists of profiles as node features, their circles, and surrounded networks. The details of the dataset are discussed below.

TABLE II: DATASET STATISTICS OF GOOGLE+

Parameters	Values
Nodes	107614
Edges	13673453
Nodes in largest WCC	107614 (1.000)
Edges in largest WCC	13673453 (1.000)
Nodes in largest SCC	69501 (0.646)
Edges in largest SCC	9168660 (0.671)
Average clustering coefficient	0.4901
Number of triangles	1073677742
Fraction of closed triangles	0.6552
Diameter (longest shortest path)	6
90-percentile effective diameter	3

Table II illustrates the results of Google+ for nodes (107614), edges (13673453), nodes in largest (107614), edges in largest (9168660), average clustering coefficient (0.4901), number of triangles (1073677742), fraction of closed triangles (0.6552), diameter (6) and 90-percentile effective diameter (3).

C. Twitter

The Twitter dataset, consisted of circle list from Twitter. The data is also collected through a survey. The Twitter dataset consists of profiles as node features, their circles, and surrounded networks. The details of the dataset are discussed below.

TABLE III: DATASET STATISTICS OF TWITTER

Parameters	Values
Nodes	81306
Edges	1768149
Nodes in largest WCC	81306 (1.000)
Edges in largest WCC	1768149 (1.000)
Nodes in largest SCC	68413 (0.841)
Edges in largest SCC	1685163 (0.953)
Average clustering coefficient	0.5653
Number of triangles	13082506
Fraction of closed triangles	0.06415
Diameter (longest shortest path)	7
90-percentile effective diameter	4.5

Table III illustrates the results of Twitter for nodes (81306), edges (1768149), nodes in largest (81306), edges in largest (1685163), average clustering coefficient (0.5653), number of triangles (13082506), fraction of closed triangles (0.06415), diameter (7) and 90-percentile effective diameter (4.5).

D. Comparison of Facebook, Google+, Twitter

This section analyses the comparison of three datasets of Facebook, Google+ and Twitter. The detailed illustration can be seen in Figure No. 04.

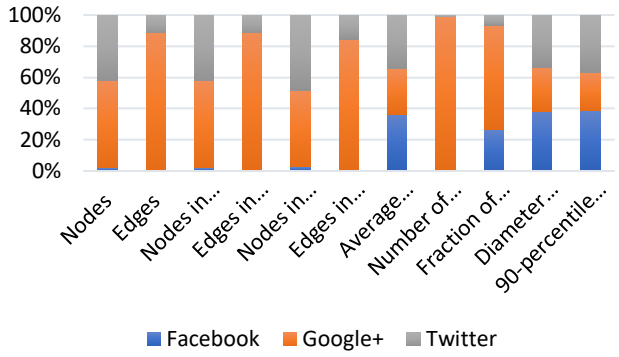

Figure 4: Comparative Analysis of Facebook, Twitter and Google+

Referring above figure, Google+ is most dominating in link prediction where most of contacts are linked easily as compared to Facebook and Twitter. While twitter is the second one as compared with Facebook.

TABLE IV: COMPARATIVE DATASET STATISTICS OF FACEBOOK, TWITTER AND GOOGLE+

Parameters	Facebook	Google+	Twitter
Nodes	4039	107614	81306
Edges	88234	13673453	1768149
Nodes in largest WCC	4039	107614	81306
Edges in largest WCC	88234	13673453	1768149
Nodes in largest SCC	4039	69501	68413
Edges in largest SCC	88234	9168660	1685163
Average clustering coefficient	0.6055	0.4901	0.5653
Number of triangles	1612010	107367774	13082506
Fraction of closed triangles	0.2647	0.6552	0.06415
Diameter (longest shortest path)	8	6	7
90-percentile effective diameter	4.7	3	4.5

Referring the Table IV, it is identified that Facebook has most 90 percentile effective diameter, diameter (longest shortest path) and Average clustering coefficient as compared to Twitter and Google+.

E. Evaluation Meterics

Precision, Recall, F1 score and Accuracy are evaluations metrics, which are used in this research to get experimental results. The equation of each evaluation metrics is given below.

For Precision,

Precision = TP / TP + FP' (13)

For Recall,
Recall = TP /TP + FN' (14)

For F1 score,
F1 = 2 × Precision × Recall / Precision + Recalls. (15)

For accuracy,
Accuracy = TP + TN / TP + TN + FP + FN' (16)

By applying the above-mentioned evaluation metrics, following results have been taken as shown in Figure no.5.

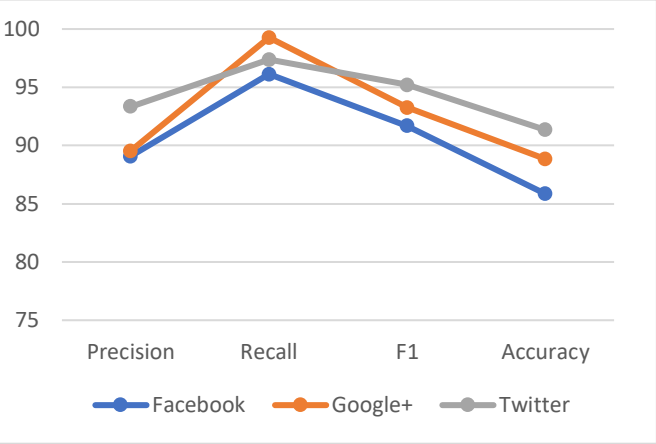

Figure 5: Results of Evaluation metrics

Additionally, Results were taken by preprocessing and training the dataset. Four metrics were set and Machine Learning approach is applied to achieve following results in Table V.

TABLE V: RESULTS OF EVALUATION METRICS

Metrics	Facebook	Google+	Twitter
Precision	89.10	89.55	93.36
Recall	96.13	99.27	97.38
F1	91.68	93.27	95.22
Accuracy	85.86	88.85	91.35

VI. APPLICATIONS

Link prediction plays vital role in modern fields ie cybersecurity, AI and emerging fields. Table VI shows various applications.

TABLE VI: APPLICATIONS OF LINK PREDICTION

Ref	Field	Application	Role of Link Prediction
[26]	Cybersecurity	Insider threat detection	Unusual or risky new links in device/user communication
[27]	Fraud Detection	Money laundering, identity fraud	Hidden or predicted connections between fraudsters
[28]	Social Media	Friend suggestion, bot detection	Predicting valid vs. fake user interactions

VII. CONCLUSION

We summarized the evolution of studies on link prediction in this paper, with a focus on recent work by statistical physicists. Whereas link prediction is a not a new topic in applied science, conventional methods have never really kept up with the new developments in network science, particularly the new viewpoints and tools that have emerged from complex network research. The research of link prediction and complex networks will benefit each other in future because a detailed analysis of network structure could be used to design sophisticated link prediction algorithms (e.g., using details relating the centralized structure and modular structure of real - world networks is used to effectively forecast missing links.), and the effectiveness of a link prediction algorithm can provide indication about design characteristics, and the schemes themselves would be used to enhance assumptions of actual network properties and to analysis.

REFERENCES

[1] H. Li, Q. Shang, and Y. Deng, "A generalized gravity model for influential spreaders identification in complex networks," Chaos, Solitons \& Fractals, vol. 143, p. 110456, 2021.

[2] R. Song, G. Ling, Q. Fan, M.-F. Ge, and F. Wang, "Link prediction based on heterogeneous degree penalization with extending neighbors and clustering coefficient," Int. J. Mod. Phys. C, p. 2250033, 2021.

[3] M. Klopschitz, A. Irschara, and D. Schmalstieg, "Robust Incremental Structure from Motion," Proc. Int. Syposium 3D Data Process. Vis. Transm. (3DPVT 2010), no. January, 2010.

[4] Y. Wang, S. Wang, and Y. Deng, "A modified efficiency centrality to identify influential nodes in weighted networks," Pramana, vol. 92, no. 4, pp. 1–11, 2019.

[5] K. Yang et al., 'Complex systems and network science: a survey', Journal of systems engineering and electronics, vol. 34, no. 3, pp. 543–573, 2023.

[6] H. Bashir, S. Hamdan, U. Ojiako, S. Haridy, M. Shamsuzzaman, and H. A. Al Zarooni, 'A weighted fuzzy social network analysis-based approach for modeling and analyzing relationships among risk factors affecting project delays', Eng. Manag. J., vol. 36, no. 1, pp. 3–13, Jan. 2024.

[7] R. Van Belle and J. De Weerdt, 'SHINE: A scalable heterogeneous inductive graph neural network for large imbalanced datasets', IEEE Trans. Knowl. Data Eng., vol. 36, no. 9, pp. 4904–4915, Sep. 2024.

[8] Q. Wang et al., 'Graph confident learning for software vulnerability detection', Eng. Appl. Artif. Intell., vol. 133, no. 108296, p. 108296, Jul. 2024.

[9] S. Yu, A. Li, Y. Chen, D. Wang, and X. Tang, 'Heterogeneous network-based algorithms in the biomedical data mining: A review

from technical perspective', Informatics and Health, vol. 1, no. 2, pp. 111–122, Sep. 2024.

[10] V. Navarro, S. del Rio, M. del Mar Millán, A. Messina, and J. Ventura-Traveset, 'GSSC Now: ESA Thematic exploitation platform for navigation digital transformation. Enhancing GNSS scientific research', Adv. Space Res., vol. 74, no. 6, pp. 2728–2751, Sep. 2024.

[11] Shanmuga, R. U., & Tamilpavai, G., 'DEEP Learning-based air quality monitoring model via BM-KMC using seasonal images", Environment, Development and Sustainability, pp. 1-31, 2024

[12] 'Relating structural connectivity to brain function using deep learning, graph theory, complexity, and disease', J. Neudorf and others, 2022.

[13] Z. Zhao et al., Mining node attributes for link prediction with a non-negative matrix factorization-based approach. Knowledge-Based Systems. 2024.

[14] W. Yanju, 'Augmenting college English teaching with advanced semantic representation and intelligent evaluation in E-learning', Comput. Aided Des. Appl., pp. 127–134, Mar. 2024.

[15] L. L. Gilson, 'Why be creative: A review of the practical outcomes associated with creativity at the individual, group, and organizational levels. Handbook of organizational creativity', pp. 303–322, 2024.

[16] F. Liu, B. Liu, C. Sun, M. Liu, and X. Wang, "Deep belief network-based approaches for link prediction in signed social networks," Entropy, vol. 17, no. 4, pp. 2140–2169, 2015, doi: 10.3390/e17042140.

[17] B. Bebensee, N. Nazarov, and B.-T. Zhang, "Leveraging node neighborhoods and egograph topology for better bot detection in social graphs," Soc. Netw. Anal. Min., vol. 11, no. 1, pp. 1–14, 2021.

[18] D. Malhotra and R. Goyal, "Supervised-learning link prediction in single layer and multiplex networks," Mach. Learn. with Appl., vol. 6, p. 100086, 2021.

[19] F. Sarhangnia, N. Ali Asgharzadeholiaee, and M. Boshkani Zadeh, "A Novel Multilayer Model for Link Prediction in Online Social Networks Based on Reliable Paths," J. Inf. \& Knowl. Manag., vol. 21, no. 02, p. 2250025, 2022.

[20] A. Havolli, A. Maraj, and L. Fetahu, "Building a content-based recommendation engine model using Adamic Adar Measure; A Netflix case study," in 2022 11th Mediterranean Conference on Embedded Computing (MECO), 2022, pp. 1–8.

[21] N. Fountoulakis and T. Iyer, "Condensation phenomena in preferential attachment trees with neighbourhood influence," Electron. J. Probab., vol. 27, pp. 1–49, 2022.

[22] D. U. Wulff, S. De Deyne, S. Aeschbach, and R. Mata, "Using network science to understand the aging lexicon: Linking individuals' experience, semantic networks, and cognitive performance," Top. Cogn. Sci., vol. 14, no. 1, pp. 93–110, 2022.

[23] B. Jhun, "Topological analysis of the latent geometry of a complex network," Chaos An Interdiscip. J. Nonlinear Sci., vol. 32, no. 1, p. 13116, 2022.

[24] M. Joodaki, M. B. Dowlatshahi, and N. Z. Joodaki, "An ensemble feature selection algorithm based on PageRank centrality and fuzzy logic," Knowledge-Based Syst., vol. 233, p. 107538, 2021.

[25] M. Zhang, L. Yang, H. Hu, T. Liu, and J. Wang, "Efficient index-free SimRank similarity search in large graphs by discounting path lengths," Expert Syst. Appl., vol. 206, p. 117746, 2022.

[26] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," in Proc. AAAI Workshop on AI for Cyber Security, 2025.

[27] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," Artif. Intell. Rev., vol. 34, no. 1, pp. 1–14, 2025.

[28] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," J. Am. Soc. Inf. Sci. Technol., vol. 58, no. 7, pp. 1019–1031, 2025.