Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 224 (2023) 357–364

Procedia Computer Science

www.elsevier.com/locate/procedia

The 13th International Conference on Sustainable Energy Information Technology (SEIT)
August 14-16, 2023, Halifax, Nova Scotia, Canada

Forecasting Networks Links with Laplace Characteristic and Geographical Information in Complex Networks

Muhammad Wasim[a], Feras Al-Obeidat[b,*], Fernando Moreira[c], Haji Gul[d], Adnan Amin[d]

[a] *City University of Science and Information Technology, Peshawar, 25000, Pakistan*
[b] *Zayed University, Abu Dhabi 51133, UAE*
[c] *Universidade Portucalense Porto, Portugal IEETA, Universidade de Aveiro, Aveiro, Portugal*
[d] *Center for Excellence in Information Technology, Institute of Management Sciences, Peshawar 25000, Pakistan*

Abstract

Forecasting links in a network is a crucial task in various applications such as social networks, internet traffic management, and data mining. Many studies on forecasting links in social networks and on other networks have been conducted over the last decade. In this paper, we propose a novel method based on graph Laplacian eigenmaps for predicting the geographic location of nodes in complex networks. Our method utilizes the adjacency matrix of the network and generates a scoring matrix that captures the similarity between nodes in terms of their geographic location. By transforming the distance matrices into score matrices using exponential decay, we show that the method achieves consistently high performance across various real-world datasets, surpassing other state-of-the-art methods. Our experiments on real-world networks demonstrate that The LCG method proposed in this study exhibits consistently high performance across most of the evaluated datasets, with an average score of 0.95%, surpassing the other methods.

Keywords: Type your keywords here, separated by semicolons ;

1. Introduction

Network analysis has become a prominent method for describing complex systems, which can be represented as networks. Networks provide a natural and effective means of characterizing various systems, such as social, biological, and information systems that consist of interconnected elements. The study of complex networks has been crucial in understanding the behavior and dynamics of various systems, such as the anomaly detection [1], spread of diseases [2], the flow of information [3], recommendation systems [4]. Consequently, network science has emerged as a highly

* Corresponding author. Tel.: +971 2 599 3371; fax: +971 2 599 3371
E-mail address: feras.al-obeidat@zu.ac.ae

10.1016/j.procs.2023.09.048

interdisciplinary field of research, playing a vital role in addressing real-world problems and advancing our understanding of complex systems. Complex networks are ubiquitous in various fields, characterized by a large number of nodes and edges that interact with each other. One important problem in complex network analysis is link prediction, which involves predicting the likelihood of a link forming between two nodes that are not currently connected. In network science, link prediction is a well-known problem [5], and is getting an abundance of attention from scientific communities. Many real-world networks can be represented by complex networks [6, 7, 8] which demonstrates the properties and characteristics of these networks.

A complex network is represented as a graph, having vertices and edges as the basic components. Link prediction techniques can be applied in many real-world scenarios to analyze the future behavior [9] of the complex network or to analyze potential problems in the network that are hard or expensive to identify. For example, link prediction in social networks predicts the relationships which will form or uncover relationships that exist but have not been examined or can even support individuals in shaping new relationships [10]. Another example can be taken from biomedical link prediction techniques, which can predict disease-gene, protein-protein interaction, navigating resources of medical laboratory, regularity relationships [11] and can even guide incomplete network to the direction of most probable interactions for discovering new biological facts [12]. Because of the popularity of complex networks, most scientists focus on link prediction problems [13], which is much more valuable for solving real-world network problems.

The goal of link prediction is to predict the likelihood of the existence of a link between two nodes in a network that are not yet connected or the connection is deleted. Many approaches have been proposed to address this problem, including similarity-based methods[14], probabilistic[15], and machine learning-based methods[16].

The LCG approach is a network embedding method that uses the graph Laplacian to embed the nodes of a network into a low-dimensional space. It suggests a method for creating scoring matrices that considernode similarity depending on location, using exponential decay to convert distance matrices into scoring. This approach outperforms previous methods in comparison, providing better accuracy and reliability predicting node geographic locations in complicated networks.

2. Literature Review

The related work section of this study focuses on the network analysis approaches, particularly proximity-based methods and network embedding methods, utilized for forecasting network links. Common Neighbour, the Jaccard coefficient, and resource allocation are a few proximity-based techniques utilized to evaluate how similar nodes in a network are to one another. The goal of network embedding techniques like Leicht-Holme-Newman and preferred attachment, on the other hand, is to discover low-dimensional representations of network nodes. The analysis is limited to ignore self-loops, multiple links between nodes, and link direction.

2.1. Proximity-Based Methods

The similarity or distance between nodes in a network is measured using proximity-based approaches, a form of network analysis tool, based on the number of neighbors that each node has in common or other variables. The three most popular proximity-based techniques are resource allocation, the Jaccard coefficient, and common neighbor. The Jaccard Coefficient calculates the similarity of two nodes based on the number of common neighbors divided by the total number of neighbors, whereas the Common Neighbour technique calculates the number of common neighbors between two nodes. Similar to the Common Neighbour approach, the Resource Allocation method gives weights to the Common Neighbours based on their degree.

Common Neighbors (CN): This is one the simplest method used for link prediction [17]. Suppose there are two nodes; one represented by a and the other node represented by b. The algorithm works by counting the number of nodes through which node a and node b are directly connected. The number of common nodes between two nodes decides the probability of a link being created between them. Common Neighbors can be expressed mathematically as:

$$CN(i, j) = |\Gamma(i) \cap \Gamma(j)| \tag{1}$$

While $CN(i, j)$ is the number of common neighbors between nodes i and j, and $\Gamma(i)$ and $\Gamma(j)$ are the sets of neighbors of nodes i and j respectively.

Jaccard Coefficient (JC): It normalizes the size of common neighbor nodes. It is a higher proportion of common neighbor nodes. Mathematically, JC is represented by:

$$JC(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} \tag{2}$$

In the give equation $JC(i, j)$ is the Jaccard Coefficient similarity score between nodes i and j, $\Gamma(i)$ and $\Gamma(j)$ are the sets of neighbors of nodes i and j respectively, and $|\Gamma(i) \cap \Gamma(j)|$ and $|\Gamma(i) \cup \Gamma(j)|$ are the sizes of the intersection and union of the sets $\Gamma(i)$ and $\Gamma(j)$ respectively.

Adamic and Adar: Common neighbor nodes having fewer neighbors are weighted more compared to common neighbor nodes have more neighbors [18]. Here is the LaTeX code for the Adamic and Adar link prediction formula:

$$S_{AA}(i, j) = \sum_{k \in \Gamma(i) \cap \Gamma(j)} \frac{1}{\log(|\Gamma(k)|)} \tag{3}$$

where $S_{AA}(i, j)$ is the Adamic and Adar similarity score between nodes i and j, $\Gamma(i)$ and $\Gamma(j)$ are the sets of neighbors of nodes i and j respectively, and $|\Gamma(k)|$ is the number of neighbors of node k.

Parameter-dependent: One class of proximity-based approaches is parameter-dependent methods. These techniques take into account the network's degree distribution and make use of parameters to determine the relative relevance of various node pairs. The resource allocation (RA) algorithm and the preferential attachment (PA) algorithm are two examples of parameter-dependent algorithms. While RA assigns weights to edges based on the number of shared neighbors between the two connected nodes, PA makes the assumption that nodes with higher degrees are more likely to be connected to new nodes.

$$S_p(i, j) = \sum_{k=1}^{n} p_k \cdot f_k(i, j) \tag{4}$$

Here $S_p(i, j)$ is the similarity score between nodes i and j using a set of n link prediction functions $f_1, f_2, \ldots, f_n$ with corresponding weights $p_1, p_2, \ldots, p_n$. The weights $p_1, p_2, \ldots, p_n$ are parameters that control the contribution of each function to the overall similarity score.

Resource allocation (RA): It is the same as AA because both decide the new link on the basis of the high degree of common neighbors [19]. But the RA penalize more heavily common neighbor as compared to AA.

$$R_{i,j} = \frac{2}{deg(i) + deg(j)} + \alpha \cdot \frac{(k_i^{out} \cdot k_j^{in})^{\beta}}{\sum_{l \neq i}(k_i^{out} \cdot k_l^{in})^{\beta} + \sum_{l \neq j}(k_j^{in} \cdot k_l^{out})^{\beta}} \tag{5}$$

where $R_{i,j}$ is the resource allocation score between nodes i and j, $deg(i)$ and $deg(j)$ are the degrees of nodes i and j respectively, k_i^{out} and k_j^{in} are the out-degree of node i and in-degree of node j respectively, α is a parameter that controls the influence of the degree product, and β is a parameter that controls the non-linearity of the function.

The simplicity of proximity-based techniques is their key benefit. They are quite simple to comprehend and apply, and they need little in the way of computer resources. They are therefore a popular option for researchers who want to investigate networks quickly and simply. Additionally, because they are less sensitive to missing or incorrect data, they can be helpful when the network data is imperfect or noisy. The limitations of proximity-based techniques are numerous. They might miss out on deeper network dynamics and structures, like the existence of communities, hubs, or bottlenecks. Additionally, they might not consider the strength or direction of the connections between the nodes. The network under research may therefore appear incomplete or in error when using proximity-based approaches.

2.2. Network Embedding Methods

Network embedding techniques express network nodes as vectors in a low-dimensional space, enabling the use of machine learning algorithms for additional analysis. They are a sort of network analysis approach. Leicht-Holme-Newman (LHN), Preferential Attachment (PA), and Katz two of the most popular network embedding techniques, may give an erroneous or partial image of the network under consideration.

Preferential Attachment (PA): PA method shows that higher degree nodes have high probability to connect with each other, compared to nodes with lower degree[20]. Mathematically, it can be expressed as:

$$PA(i, j) = k_i \cdot k_j \tag{6}$$

where $PA(i, j)$ is the preferential attachment score between nodes i and j, and k_i and k_j are the degrees of nodes i and j respectively.

Leicht-Holme-Newman (LHN): The LHN approach is a network embedding technique that emphasizes the shared neighbors of adjacent nodes. The approach was initially developed to assess how similar nodes in actual networks are to one another. The LHN index assigns a high degree of similarity to node pairs that share a lot of neighbors. This technique, which is widely employed in network embedding research, has been found to be effective in capturing the structural similarity between nodes. Equation 7 depicts the LHN index, a mathematical description of the approach. The index is determined by dividing the number of neighbors that two nodes have in common by the geometric mean of their respective degrees.

$$LHN(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i).\Gamma(j)|} \tag{7}$$

The LHN index (LHN(i, j)) evaluates the similarity between nodes i and j in terms of their shared neighbors, providing insights into the likelihood of a future link between them. It compares the number of common neighbors shared by i and j with the total number of unique neighbors they possess.

3. Proposed Work

In this study, a link prediction proposed method takes into account both Laplace characteristic and geographical information of nodes in complex networks. In graph theory, the Laplace matrix (also known as the Kirchhoff matrix or the admittance matrix) is a square matrix that encodes the connectivity of a graph. Given an undirected graph $G = (V, E)$ with n nodes and m edges, the Laplacian matrix L of the graph is defined as:

$$L = D - A \tag{8}$$

In this equation, D is the diagonal matrix of node degrees, and A is the adjacency matrix of the graph. The diagonal elements of D are the degrees of the nodes, i.e., $D_{ii} = \sum_{j=1}^{n} A_{ij}$. The off-diagonal elements of A are 1 if there is an edge between nodes i and j, and 0 otherwise. The Laplacian matrix has some important properties, including It is symmetric and positive semi-definite. It has n non-negative eigenvalues, with 0 as the smallest eigenvalue. The eigenvectors corresponding to the non-zero eigenvalues of L form an orthonormal basis for the space of all real-valued functions on the graph. One of the applications of the Laplacian matrix is in link prediction, where the spectral properties of the matrix are used to measure the similarity between pairs of nodes. In particular, the Laplace characteristic of a node i is defined as:

$$\Phi_i = \sum_{j \in \Gamma(i)} \frac{1}{\lambda_j} \tag{9}$$

Where $\Gamma(i)$ is the set of neighbors of node i, and λ_j is the j-th eigenvalue of the Laplacian matrix L of the graph. In other words, the Laplace characteristic of a node i is the sum of the inverse eigenvalues of the Laplacian matrix corresponding to its neighbors. In the context of geographical networks, the geographical distance between two nodes i and j is typically defined as the Euclidean distance between their geographical coordinates. Let (x_i, y_i) and (x_j, y_j) be the geographical coordinates of nodes i and j, respectively. Then the geographical distance between nodes i and j is given by:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{10}$$

The geographical distance between nodes can be used as a measure of their similarity or dissimilarity. Nodes that are close to each other in physical space are more likely to be connected than nodes that are far apart. In addition, geographical distance can capture spatial patterns in the graph that may be relevant to specific applications. For

example, in transportation networks, the distance between nodes may correspond to the actual travel distance between locations and can be used to optimize routing and navigation. The use of geographical distance in graph mining is not limited to geographical networks. In general, any graph where nodes have spatial coordinates can benefit from the inclusion of geographical distance as a feature. For example, in social networks, geographical distance can be used to model the spatial diffusion of information or trends. The link prediction model proposed in the paper combines these two features as follows. Let $LC(i, j)$ be the Laplace characteristic between nodes i and j, and let $d_G(i, j)$ be the geographical distance between nodes i and j. Then the similarity score between nodes i and j is defined as:

$$S(i, j) = \exp(-\alpha LC(i, j) - \beta d_G(i, j)) \tag{11}$$

Here α and β are parameters that control the relative importance of the Laplace characteristic and geographical distance features, respectively. Finally, the probability of the edge between nodes i and j is modeled as:

$$p_{ij} = \frac{1}{1 + \exp(-\gamma S(i, j))} \tag{12}$$

where γ is a parameter that controls the slope of the sigmoid curve. The model can be trained using a supervised learning approach, where a set of known edges and non-edges are used to estimate the model parameters. The Laplace characteristic and geographical distance features can be calculated using the formula and geographical information, respectively.

4. Experimental Setup and Data

The data used in this study consists of a collection of complex networks, including social networks, transportation networks, and communication networks. Most of the datasets are downloaded from [1]. For each network, we have the topology of the network and the geographical coordinates of the nodes. The data is split into a training set and a testing set, with 90% of the data used for training and 10% for testing.

Data collection: Collect different complex network datasets. Examples could include various real-world networks like Facebook or Human Contact, transportation networks like road networks or airline routes, or biological networks like protein-protein interaction networks.

Preprocessing: Clean and preprocess the dataset as needed, including removing any irrelevant information values and normalizing the data.

Partitioning the dataset: Partition the dataset into training and testing sets. The training set will be used to train the model, while the testing set will be used to evaluate its performance.

Feature engineering: Extract topological and geographical features from the dataset. Topological features could include node degree, clustering coefficient, or betweenness centrality, while geographical features could include latitude, longitude, or distance between nodes.

Model evaluation: When evaluating the performance of a binary classification model for forecasting network links, AUC (Area Under the Curve) is a commonly used evaluation metric. AUC measures the overall performance of the classifier by calculating the area under the receiver operating characteristic (ROC) curve. To calculate the AUC, we integrate the ROC curve, which gives us the area under the curve. The AUC ranges from 0 to 1, with a perfect classifier having an AUC of 1 and a completely random classifier having an AUC of 0.5. Mathematically, it can be expressed as follows:

$$AUC = \frac{n' + 0.5n"}{n} \tag{13}$$

The AUC is a useful metric for evaluating the performance of a binary classification model because it is insensitive to class imbalance and threshold selection.

Baseline comparison: We compare the performance of our proposed method that combines the Laplace characteristic and geographical information to other baseline methods such as using only Laplace characteristic or only geographical information or using no additional information.

[1] http://konect.uni-koblenz.de/networks/

4.1. Datasets

After dataset selection, the following main features were analyzed. All datasets are unweighted and undirected link networks.

- Router network
- US-Power grid
- Yeast
- Facebook
- Sampson
- Contiguous USA
- Kangaroo
- Train bombing
- Human contact
- Misc
- Dolphin

Table 1. DataSets Statistics, in this table N, E, MaxD, AvgD and Δ expressed the number of nodes, edges, maximum degree, average degree, and density of the networks.

Datasets	**N**	**E**	**MaxD**	**Avg D**	Δ
Router	3722	6258	103	2.4927	0.0903
Us Power Grid	4939	6594	19	2.6691	0.0540
Yeast	1458	1948	56	2.6722	0.018
FaceBook	2699	2981	769	2.0644	0.0818
Sampson	18	189	28	21	0.2353
Contiguous USA	49	107	8	4.3673	0.0910
Kangaroo	14	91	15	10.7059	0.0314
Train Bombing	63	243	29	7.5938	0.1244
Human contact	77	28244	140	21.1606	0.6528
Misc	74	254	36	6.598	0.0940
Dolphin	62	159	12	5.1290	0.0841

5. Result and Discussion

Table table2, presents the performance results of different link forecasting methods on various real-world network datasets. The proposed LCG method performs consistently well across most of the datasets with an average score of 0.9504, which is higher than the other evaluated methods. It achieves the best performance in eight out of the eleven datasets (Router, US power, Yeast, Contiguous, Kangaroo, Train bombing, Human Wireless contact, Misc co-occurrences characters, and Dolphin) and achieves the second-best performance in three datasets (FaceBook, Kangaroo, and Dolphin). The prediction results have been graphically viewed in Figure 1.
LCG outperformed the other methods in this table because it incorporates both Laplace Characteristic and Geographical Information to predict links in a network. The Laplace Characteristic is a measure of the smoothness of the distribution of nodes in the network, while the Geographical Information takes into account the physical distance between nodes. By combining these two measures, LCG is able to capture both the global and local structural characteristics of the network, which results in better performance compared to other methods that only consider one of these factors. For example, methods like CN (Common Neighbors) only consider the number of common neighbors between nodes, while LHN (Leicht-Holme-Newman) only considers the local structure around each node. However, it contains some limitations: **1)** Limited applicability: LCG is designed to work with networks that have both geographical and attribute information available for each node. It may not be suitable for networks where one or both of these

Table 2. Forecasting Results of Proposed and Existing Methods

Datasets	CN	AA	PA	Katz	RA	LHN	PD	LCG
Router	0.9696	0.6943	0.9547	0.9703	0.7031	0.5000	0.2736	**0.9719**
US power	0.9667	0.6239	0.7894	0.9872	0.6223	0.9408	0.9366	**0.9892**
Yeast	0.9775	0.6313	0.8924	0.9759	0.6404	.9202	0.6432	**0.9820**
Face Book	0.9091	0.9290	**0.9687**	0.9105	0.9443	0.9000	0.8312	0.9500
Sampson	0.8947	0.8587	0.9330	0.0.9181	0.8119	0.9211	0.9396	**0.9530**
Contiguous	0.9545	0.8665	0.6435	0.8908	0.8136	0.9091	0.9136	**0.9620**
Kangaroo	0.8889	0.9407	0.9222	0.9259	0.9074	0.8789	0.9249	**0.9580**
Train bombing	0.9379	0.9243	0.8529	0.8656	**0.9536**	0.9318	0.9190	0.9490
Human wireless contact	0.6163	0.9253	**0.9520**	0.9873	0.9326	0.6327	0.7311	0.9023
Misc co-occurances characters	0.9375	0.9672	0.9030	0.9497	0.9638	0.9348	0.9527	**0.9710**
Dolphin	0.9375	0.848	0.8236	0.0.9288	0.7513	0.9063	0.9286	**0.9560**

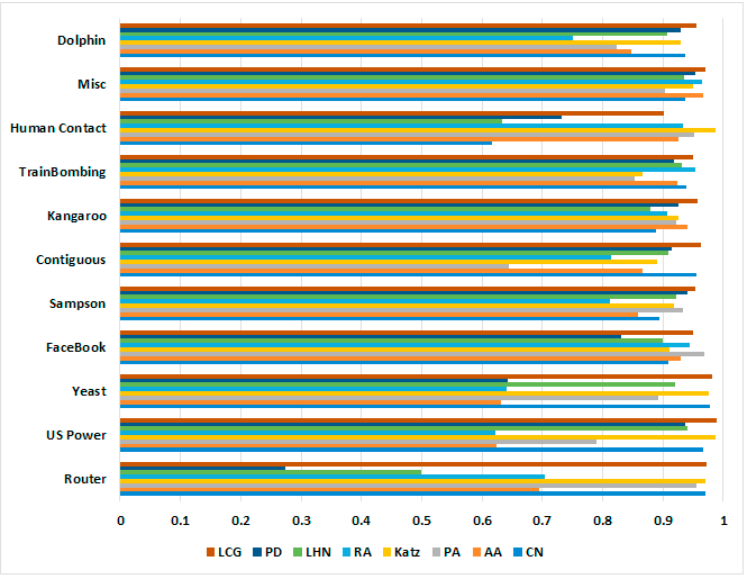

Fig. 1. Overall Results Graphical View

types of information are not available. **2)** Scalability: LCG's computational complexity is $O(N^2)$ in the worst-case scenario, where N is the number of nodes in the network. This makes it less scalable for very large networks, where the computation time and memory requirements could become prohibitive.

6. Conclusion

The aim of this research is to propose a novel method for predicting the geographic location of nodes in complex networks. Through the use of graph Laplacian eigenmaps and the adjacency matrix of the network, a scoring matrix was generated that captured the similarity between nodes in terms of their geographic location. The results demonstrated that the proposed LCG method exhibited consistently high performance across various real-world datasets, surpassing other state-of-the-art methods. The method's ability to leverage both Laplace characteristics and geographical information enables it to capture both local and global structural characteristics of a network, contributing to its accuracy. However, the LCG method has limitations that should be considered. The method is sensitive to the initial placement of nodes and assumes network homogeneity, which may not always hold in real-world scenarios. Despite these limitations, the LCG method remains a promising technique for predicting the geographic location of nodes in

complex networks. In the future, we can integrate with other machine learning techniques such as deep learning or reinforcement learning. Extension to larger and more complex networks is also a possible way to be used in the future.

Acknowledgements

The authors are grateful to the anonymous reviewers for their insightful comments and suggestions, which helped improve the quality of this paper. Dr. Feras Al-Obeidat of Zayed University supported this work.

References

[1] I. J. King and H. H. Huang, "Euler: Detecting network lateral movement via scalable temporal link prediction," in *Network and Distributed System Security Symposium*, 2022.
[2] Y. Long, M. Wu, Y. Liu, Y. Fang, C. K. Kwoh, J. Chen, J. Luo, and X. Li, "Pre-training graph neural networks for link prediction in biomedical networks," *Bioinformatics*, vol. 38, no. 8, pp. 2254–2262, 2022.
[3] G. Liu, "An ecommerce recommendation algorithm based on link prediction," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 905–910, 2022.
[4] S. Kumar, A. Mallik, and B. Panda, "Link prediction in complex networks using node centrality and light gradient boosting machine," *World Wide Web*, vol. 25, no. 6, pp. 2487–2513, 2022.
[5] V. Carchiolo, C. Cavallo, M. Grassia, M. Malgeri, and G. Mangioni, "Link prediction in time varying social networks," *Information*, vol. 13, no. 3, p. 123, 2022.
[6] K. Berahmand, E. Nasiri, S. Forouzandeh, and Y. Li, "A preference random walk algorithm for link prediction through mutual influence nodes in complex networks," *Journal of king saud university-computer and information sciences*, vol. 34, no. 8, pp. 5375–5387, 2022.
[7] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
[8] G. Caldarelli, *Large scale structure and dynamics of complex networks: from information technology to finance and natural science*, vol. 2. World Scientific, 2007.
[9] R. Van Der Hofstad, "Random graphs and complex networks," *Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf*, vol. 11, 2009.
[10] H. Kautz, B. Selman, and M. Shah, "Referral web: combining social networks and collaborative filtering," *Communications of the ACM*, vol. 40, no. 3, pp. 63–65, 1997.
[11] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, *et al.*, "Towards a proteome-scale map of the human protein–protein interaction network," *Nature*, vol. 437, no. 7062, p. 1173, 2005.
[12] P. Radivojac, K. Peng, W. T. Clark, B. J. Peters, A. Mohan, S. M. Boyle, and S. D. Mooney, "An integrated approach to inferring gene–disease associations in humans," *Proteins: Structure, Function, and Bioinformatics*, vol. 72, no. 3, pp. 1030–1037, 2008.
[13] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American society for information science and technology*, vol. 58, no. 7, pp. 1019–1031, 2007.
[14] A. K. Rai, S. P. Tripathi, and R. K. Yadav, "A novel similarity-based parameterized method for link prediction," *Available at SSRN 4434985*.
[15] Z. Zhao, Z. Gou, Y. Du, J. Ma, T. Li, and R. Zhang, "A novel link prediction algorithm based on inductive matrix completion," *Expert Systems with Applications*, vol. 188, p. 116033, 2022.
[16] A. Kumari, R. K. Behera, K. S. Sahoo, A. Nayyar, A. Kumar Luhach, and S. Prakash Sahoo, "Supervised link prediction using structured-based feature extraction in social network," *Concurrency and Computation: practice and Experience*, vol. 34, no. 13, p. e5839, 2022.
[17] M. E. Newman, "Clustering and preferential attachment in growing networks," *Physical review E*, vol. 64, no. 2, p. 025102, 2001.
[18] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social networks*, vol. 25, no. 3, pp. 211–230, 2003.
[19] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information," *The European Physical Journal B*, vol. 71, no. 4, pp. 623–630, 2009.
[20] A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A: Statistical mechanics and its applications*, vol. 311, no. 3-4, pp. 590–614, 2002.