Link Prediction Algorithm Based on Local Generalized Gravity Model

Yongqing Wu, Xinxin Li

***Abstract*—As an extension and further development of the network evolution mechanism at the link level, link prediction is experiencing significant growth in both physics and computer science domains. In this study, we introduce a novel similarity metric that employs the Local Generalized Gravity Model, which computes similarity scores based on the local clustering coefficient and the spreading ability of nodes within the network. We evaluate its performance on eight real-world networks, and the results demonstrate that our proposed algorithm achieves higher prediction accuracy than other benchmarks based on the AUC evaluation criterion.**

***Index Terms*—Complex network, Generalized gravity model, Spreading ability, Link prediction**

I. Introduction

COMPLEX networks have emerged as indispensable tools for analyzing complex systems, providing a comprehensive framework for representation across diverse real-world domains, including financial networks[1], social networks[2], and air transportation networks[3], among others. The formalized graphical descriptions offered by complex networks have significantly enhanced their efficacy in depicting these systems. One of the primary objectives in modeling complex systems as networks is to unravel the intricate interplay between network structure and functionality. This focus has spurred the development of network science, establishing it as a paramount interdisciplinary research field. In the context of modeling, complex networks utilize nodes to represent individuals or entities within complex systems such as bacteria, neurons, roads, airlines, and computers. These nodes are interconnected by edges that signify interaction relationships governed by specific rules or connections. For instance, in scientific collaboration networks, nodes represent authors and edges denote collaborations between them; in neural networks, nodes represent neurons and edges indicate interactions among them; in road transportation networks, nodes correspond to cities and edges represent the connections between those cities. Research in network science encompasses various aspects, including community detection[4], identifying influential spreaders[5][6], synchronization[7] and link prediction[8]. Among these, link prediction stands out as one of the paradigmatic challenges[9]. It involves estimating the probability of existing but undiscovered connections or potential future connections within a given network based on its structure and existing link information. Investigations within this domain are essential for deciphering the evolutionary trends of networks and for forecasting their prospective actions. To clarify this, a network diagram is depicted in Fig.1. The left graph depicts the actual linkages within the current network, consisting of five connecting edges formed between five nodes, thereby constituting a non-complete graph. We hypothesize that certain links may be missing from this graph. Given this premise, we implement a link prediction technique. Through comprehensive analysis of the network and subsequent deductions, we derive the predictive outcomes illustrated in the right graph. The red dashed lines signify the connections present in the network but yet to be identified. To forecast such potential connections, a similarity measure must be given to every pair of nodes that lack a direct link in the network. These similarity values are then ranked in descending order, with the assumption assumed that the pairs at the top of the ranking have the highest probability of forming a connection. The prediction of connections plays a crucial role both theoretically and practically. Theoretically, it serves as an robust analytical framework for dissecting the intricacies of complex network in greater detail. Practically, link prediction can be employed to forecast potential connections that may arise in evolving networks over time. For instance, within social networks, the ability to predict links can be utilized to enhance privacy management[10]. In the context of spam email detection, analyzing the relationships between email senders and recipients can aid in determining the likelihood of an email being classified as spam[11]. Additionally, in recommendation systems, the quality of recommendations can be improved by suggesting new friends to users based on their historical browsing behavior or by recommending new products on shopping platforms[12].

In recent years, numerous algorithms have been proposed for predicting network connections. However, achieving precise predictions of node connections remains a challenging task. Among these algorithms, similarity indices based on network structure have attracted considerable interest because of their reliable performance and efficient computation. These indices are generally categorized into three types: local, global, and random walk-based similarities. Local similarity indices primarily focus on the immediate neighborhood structure of a node, resulting in lower computational complexity compared to their global counterparts. This characteristic makes them advantageous for handling large networks, facilitating swift and efficient link predictions. In contrast,

Manuscript received April 18, 2024; revised September 23, 2024.

This work was supported by the National Natural Science Foundation of China (Grant No. 52174184).

Y. Wu is an associate professor of College of Science, Liaoning Technical University, Fuxin 123000, China (corresponding author to provide e-mail: yqwuyywu@163.com).

X. Li is a postgraduate student of College of Science, Liaoning Technical University, Fuxin 123000, China (e-mail: 2217742712@qq.com).

global similarity indices necessitate consideration of the entire network structure, thereby demanding higher computational resources. Moreover, random walk algorithms require multiple iterations[13], which result in increased computational costs, especially in large-scale networks. Local similarity indices demonstrate exhibit heightened sensitivity to local network changes, enabling them to can accurately reflect similarity relationships between nodes and maintain robust performance in the presence of local structural changes or noise within the network. Given on these advantages, this paper focuses on methodologies grounded in local similarity principles. The fundamental premise of local similarity is straightforward: nodes with more common neighbors are considered more similar and are more likely to be connected. The most basic measure of similarity, which relies on common neighbors, is known as the Common Neighbors Index (CNI)[14]. Building on this concept, several metrics take into account the influence of node degrees at both ends, including the Preference Attachment Index (PAI)[15], Hub Depressed Index (HDI)[16], Hub Promoted Index (HPI)[16], Sorenson Index (SI)[17] and Jaccard Index (JI)[18]. Recently, Xu et al. introduced a novel algorithm known as the Local Neighbor Gravity Model (LNGM), which employs the principles of the gravity model[19]. Drawing inspiration from the law of universal gravitation, this algorithm integrates neighborhood and path information. It considers node's degree as its mass and the distance between nodes, based on the shortest path, to identify influential spreaders. However, these methods predominantly depend on common neighbor information to extract local topological characteristics of nodes, which results in an underutilization of node-specific information extraction and ultimately limits predictive accuracy. In contrast, Li et al. recently introduced a generalized gravity centrality, positing that the spreading ability provides a more precise measure of local information than node degree alone[20]. Furthermore, research indicates that local clustering coefficients significantly influence node spreading ability[21].

Building on the preceding discourse, this study employs the spreading ability of nodes as an analogy for to mass in the gravity model. The capacity of a node to propagate information is influenced by two key factors: its degree and its local clustering coefficient. Nodes with identical degrees but higher local clustering coefficients exhibit denser connections among their neighbors, thereby enhancing the likelihood of information dissemination within these local networks rather than externally. Consequently, the impact of that node is reduced. Within the scope of link prediction challenges, this implies a reduced similarity between two nodes. Therefore, leveraging the spreading ability of nodes to predict potential links between them is underpinned by robust theoretical foundations.

The layout of this paper is arranged in the subsequent order: Section II presents several classical link prediction indices utilized for comparison, accompanied by a description of the generalized gravity model and our proposed algorithm. Section III offers an experimental evaluation of the generalized gravity model. Finally, Section IV summarizes and discusses the results.

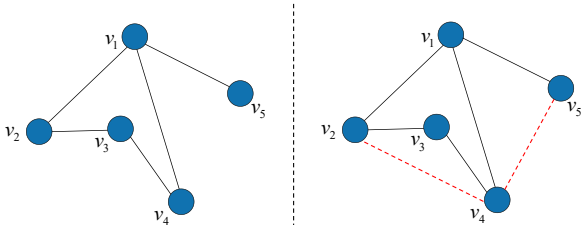

Fig. 1. Link prediction schematic diagram. The red dashed lines represent predicted future edges.

II. Method

The gravity model is widely utilized to describe interactions among objects with mass. In this study, we consider a node's ability to spread information as its mass within the gravity model framework, and the length of the shortest path connecting two nodes signifies the distance as per the model's parameters. The spreading ability of nodes is quantified using local clustering coefficients and node degrees. This paper aims to demonstrate that assessing mass in terms of node spreading ability yields more accurate prediction outcomes compared to solely using node degree as a measure of mass.

The network's local clustering coefficient for node i, which has a degree of k_i, is specified as:

$$C_i = \frac{2E_i}{k_i\left(k_i-1\right)}, \tag{1}$$

where E_i and $k_i\left(k_i-1\right)/2$ denote the current count of edges and the theoretical maximum of potential edges among the neighbors of node i, respectively.

The spreading ability of the node i is defined as

$$sp(i) = \exp(-qC(i))k(i), \tag{2}$$

where q is an adjustable parameter with $q \geq 0$, allowing for flexible modification in various applications. In this work, we set $q = 2$ for calculation[20].

The influence between nodes in the network is defined as follows:

$$G_{ij} = \frac{sp(i) \times sp(j)}{d_{ij}^2}, \tag{3}$$

where d_{ij} is the shortest distance between the nodes i and j. The similarity calculation score is defined as

$$s_{ij} = \alpha \sum_{d_{im\leq 2}} G_{mj}/k_i + \beta \sum_{d_{jm\leq 2}} G_{im}/k_j, \tag{4}$$

where $\alpha, \beta \in (0,1)$ and $\alpha+\beta = 1$. α is a variable parameter that can be adjusted according to the requirements of practical scenarios.

A. Compared algorithm

We perform a comparative analysis of our suggested algorithm against the subsequent algorithms.

1) Common Neighbors Index(CNI): Regarding two nodes within the network, their neighbor sets are defined as $\Gamma(i)$ and $\Gamma(j)$. The similarity score is calculated by tallying the common neighbors among the two distinct nodes. The formula for computation is presented below:

$$s_{ij} = \left|\Gamma(i) \cap \Gamma(j)\right|, \tag{5}$$

2) Preference Attachment Index(PAI): The application of the preferential attachment index can result in a scale-free network. The core concept behind the preferential attachment index is that the likelihood of a new connection forming with a node is directly proportional to the node's degree k. Thus, the chance of a new link forming between nodes i and j is relative to the multiplication of their respective degrees. The computation of the similarity score between a pair of nodes is specified as follows:

$$s_{ij} = k_i k_j, \tag{6}$$

3) Hub Depredded Index(HDI): This metric is employed to portray the topological resemblance between each set of reactants within the metabolic network. It is determined by the proportion of the shared neighbors of nodes i and j to the highest degree among the two nodes. The formula for calculating the similarity score between a pair of nodes is outlined below:

$$s_{ij} = \frac{|\Gamma(i) \cap \Gamma(j)|}{\max\{k_i, k_j\}}, \tag{7}$$

4) Hub Promoted Index(HPI): HPI is defined similarly to the HDI, the difference is that the denominator takes the minimum of the two node degrees, i.e.

$$s_{ij} = \frac{|\Gamma(i) \cap \Gamma(j)|}{\min\{k_i, k_j\}}, \tag{8}$$

5) Sorenson Index(SI): In the Sorenson Index, the similarity score for two nodes is derived by doubling the count of their mutual neighbors and then dividing by the aggregate of their individual degrees. The expression is delineated below:

$$s_{ij} = \frac{2 \times |\Gamma(i) \cap \Gamma(j)|}{k_i + k_j}, \tag{9}$$

6) Jaccard Index(JI): The Jaccard coefficient measures the similarity score between two nodes i and j by taking the ratio of common neighbors to the union of all neighbors of the nodes. The calculation formula is as follows:

$$s_{ij} = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|}, \tag{10}$$

7) Local Neighbors Gravity Model(LNGM): The local neighbors gravity model consider the combined influence of 1-hop and 2-hop neighbor nodes. The influence G_{ij} of nodes i and j in the network is outlined below:

$$G_{ij} = \frac{k(i) \times k(j)}{d_{ij}^2}, \tag{11}$$

the potential for future connections between nodes i and j is characterized as

$$s_{ij} = \alpha \sum_{d_{ib} \leq 2} G_{bj}/k_i + (1-\alpha) \sum_{d_{jb} \leq 2} G_{ib}/k_j. \tag{12}$$

B. Datasets

In order to assess the algorithm's forecasting capabilities, eight classic real-world network datasets have been selected. The data descriptions are as follows:

1) USAir: The USAir dataset represents the U.S. air transportation network, where each node corresponds to an airport, and edges between nodes indicates the existence of direct flights between the corresponding airports[22].

2) NS: The NS dataset depicts a collaborative network of scientists, emphasizing two review articles and incorporating additional significant literature. Nodes in this network represent scientists, and edges signify collaboration relationships among them[23].

3) Email: The Email dataset illustrates a network of email interactions, with nodes signifying users and edges representing email exchanges among them[24].

4) PB: The PB network represents a network of U.S. political blogs, with nodes representing distinct blog pages and edges indicating hyperlinks that interconnect these pages[25].

5) Facebook: The Facebook network dataset captures a social networking landscape, with nodes denoting users and edges signifying connections due to mutual following[26].

6) Power: The Power Network represents the Western U.S. Power Grid, with nodes representing substations or converter stations, and edges depicting the high-voltage lines connecting them[27].

7) Jazz: The Jazz network depicts collaborations among jazz musicians, where nodes representing individual musicians and edges represent the collaborative relationships between them[28].

8) Wikivote: The Wikivote dataset represents Wikipedia, an open-source encyclopedia edited by a global community of volunteers. The network focus on active users who participate in the nomination and voting process for administrative roles. During elections, users express support, opposition, or neutrality towards candidates through public votes. The individual with the highest number of votes becomes the appointed administrator. This dataset delineates a social networking platform, where nodes stand for users, and edges signify the voting interactions among them[29].

Table I summarizes the key characteristics of the selected datasets, characterized by: N for the node count, M for the edge count, C for the clustering coefficient, k for the average degree, d for the average network distance, and r for the assortative factor.

TABLE I
CHARACTERIZATION OF EIGHT REAL NETWORKS.

Datasets	N	M	C	k	d	r
USAir	332	2126	0.7494	12.8072	2.7381	-0.2079
NS	379	914	0.7981	4.8232	6.0419	-0.0817
Email	1133	5451	0.2540	9.6222	3.6060	-0.0782
PB	1222	16741	0.3600	27.3552	2.7375	-0.2213
Facebook	4039	88234	0.6170	43.6910	3.6925	0.0636
Power	4941	6549	0.1065	2.6691	18.9892	0.0035
Jazz	198	2742	0.6334	27.6970	2.2350	0.0202
Wikivote	7066	100736	0.2090	28.5129	0.2475	-0.0833

C. Divided training sets

Consider an undirected network $G = (V, E)$, consist of a set of nodes V and the set of edges E. let U be the set of all possible edges between nodes in G, consisting of $N(N-1)/2$ pair of nodes. In link prediction, known edges in the network are typically split into a training set E^{train} and a test set E^{test}. Here, $E = E^{train} \cup E^{test}$ and $E^{train} \cap E^{test} = \emptyset$. Fig.2 illustrates this process visually. Various methods exist for partitioning datasets, including random sampling[30], item-by-item traversal, and k-fold cross-validation. In this study, we adopt the widely accepted method of random sampling. In a network denoted as G, which is comprised of N distinct nodes and M connecting edges, we set the splitting ratio $p \in (0, 1)$, meaning the test set comprises pM randomly chosen edges. Through experimentation, we establish $p = 0.2$ as the optimal ratio.

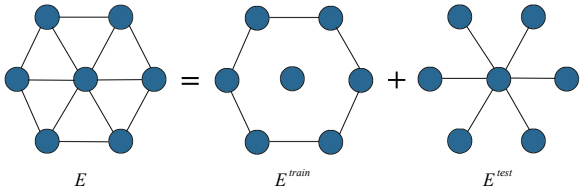

Fig. 2. Divide the original graph into a training set and a test set.

D. Evaluation criterion

To assess the effectiveness of our proposed algorithm, we utilize the AUC value to evaluate its performance on a specific dataset and compare it with other algorithms. The AUC value indicates the probability of a model assigning a higher similarity score to an actual edge from the test set compared to a non-existent edge. For every one of the n independent experiments, the model receives 1 point if the score for a test set edge is higher than that for a non-existent edge, and 0.5 points if the scores match. The AUC is defined in the following way[31]:

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

n' instances where the similarity values of edges in the test set exceed those of non-existent edges are denoted, while n'' indicates cases where the similarity values are equal. Higher AUC values approaching 1 signify greater accuracy of the algorithm.

The construction details of the LGGM are outlined in Algorithm 1.

Precision is another commonly used evaluation metric for link prediction, characterized by the proportion of accurately predicted positive samples among the top-L ranked predicted links. Specifically, if m correctly predicted links are found among the top-L links ranked by the likelihood of link occurrence and these m links are indeed present in the test set, the calculation for precision can then be carried out:

$$precision = \frac{m}{L}, \tag{14}$$

clearly, the prediction accuracy depends on achieving a higher precision.

The Root Mean Square Error (RMSE) metric is an assessment tool that quantifies the deviation between estimated and actual values. Its calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{15}$$

in this context, y_i corresponds to the actual observed value of the node pair within the test set, whereas $\hat{y}_i$ signifies the predicted value obtained from the training set. n denotes the number of observations. A larger value of RMSE, indicates a greater deviation between the predictive model and the actual observations; conversely, a smaller value of RMSE suggests that the model exhibits higher prediction accuracy.

Algorithm 1: LGGM

Input: Network G=(V, E), with N nodes and M edges;
Output: AUC;

```
for i = 1 → N do
    Calculate the degree k_i and the clustering coefficient C(i);
end
for i, j ∈ N, i ≠ j do
    Calculate the shortest distance d_ij;
end
for d_ij ≤ 2 do
    Calculate the influence G_ij between i and j by Eq.(3);
    for d_ij > 2 do
        G_ij = 0;
    end
end
Compute the similarity score S_ij by Eq.(4);
Arrange the list of all S_ij in descending order;
Calculate AUC by Eq.(12);
return AUC;
```

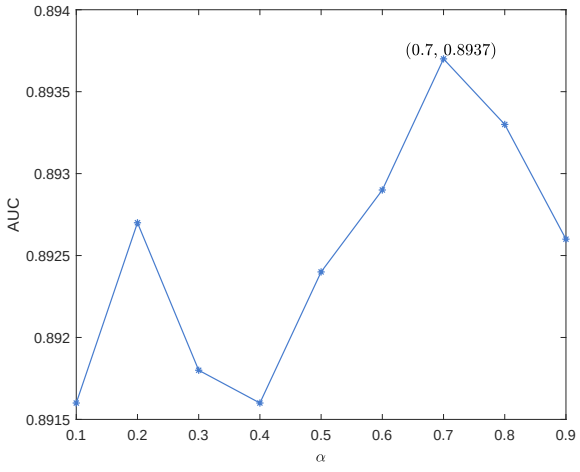

Fig. 3. The AUC for LGGM at different values of α.

TABLE II
BEST AUC VALUES AND THE CORRESPONDING VALUES OF α.

	USAir	NS	Email	PB	Facebook	Power	Jazz	Wikivote
AUC	0.9501	0.9603	0.8343	0.9187	0.9918	0.6159	0.9537	0.9317
α	0.4	0.7	0.3	0.6	0.4	0.6	0.6	0.7

TABLE III
AUC OF DIFFERENT METHODS IN DIFFERENT NETWORKS.

Datasets	CNI	PAI	JI	SI	HPI	HDI	LNGM	LGGM
USAir	0.933	0.8838	0.897	0.8966	0.8198	0.8902	0.9333	0.9501
NS	0.9493	0.6279	0.947	0.9469	0.9094	0.9458	0.9495	0.9588
Email	0.8481	0.7827	0.8465	0.8466	0.8077	0.8466	0.8486	0.8349
PB	0.917	0.8991	0.8711	0.871	0.8231	0.8673	0.9171	0.9175
Facebook	0.9924	0.8314	0.9905	0.9906	0.9823	0.9896	0.9924	0.9904
Power	0.5591	0.4715	0.5496	0.5496	0.2545	0.5497	0.5595	0.6097
Jazz	0.9541	0.7839	0.8742	0.9253	0.9332	0.9525	0.9511	0.9567
Wikivote	0.908	0.9073	0.8737	0.8738	0.5934	0.8728	0.9082	0.9315

TABLE IV
COMPARISON OF LNGM'S AND LGGM'S ACCURACY QUANTIFIED BY PRECISION.

Algorithms	USAir	NS	Email	PB	Facebook	Power	Jazz	Wikivote
LNGM	0.5874	0.2218	0.264	0.4332	0.9521	0.0963	0.5817	0.342
LGGM	0.5961	0.2382	0.2679	0.47	0.9569	0.1226	0.5863	0.339

Fig. 4. Average AUC for each link prediction model on all datasets.

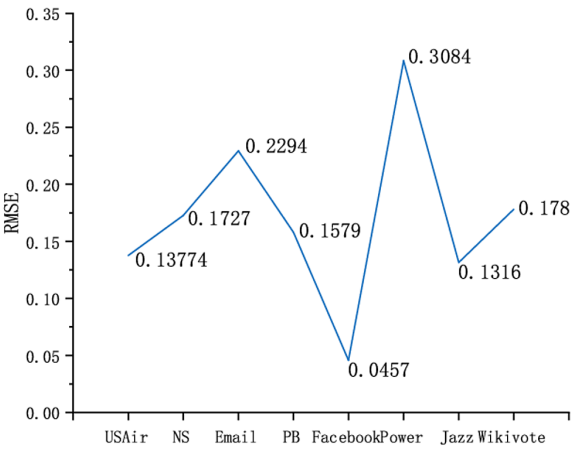

Fig. 5. RMSE analysis of LGGM on different datasets.

III. RESULT AND DISCUSSIONS

To investigate the impact of different α values on AUC, we conducted experiments using the LGGM with α values spanning from 0.1 to 0.9.

The average AUC values across all datasets are presented in Fig.3. As illustrated in the line graph in Fig. 3, the AUC of LGGM exhibits minimal variation with changing α values. The highest AUC value of 0.8937 is attained when α is set to 0.7, while the lowest AUC of 0.8916 occurs at $\alpha = 0.4$. The difference between these two values is merely 0.0021, which is not statistically significant. Although we attempted to identify the optimal α value using various datasets presented in Table II, the results did not reveal indicate a clear optimal choice. However, it is noteworthy that α values of 0.1, 0.2, 0.8, and 0.9 are not considered optimal. Furthermore, the optimal α value varies depending on the dataset. For example, the NS and Wikivote datasets exhibit optimal AUC values at $\alpha = 0.7$ (0.9603 and 0.9317, respectively), while the Email dataset achieves an optimal AUC value of 0.8343 at $\alpha = 0.3$. The optimal AUC values for the remaining datasets are detailed in Table II. Based on the information in Table II and Fig.3, there is a considerable degree of randomness in selecting the optimal α value. To mitigate prediction errors associated with this randomness, we have chosen $\alpha = 0.7$ as the benchmark for performance

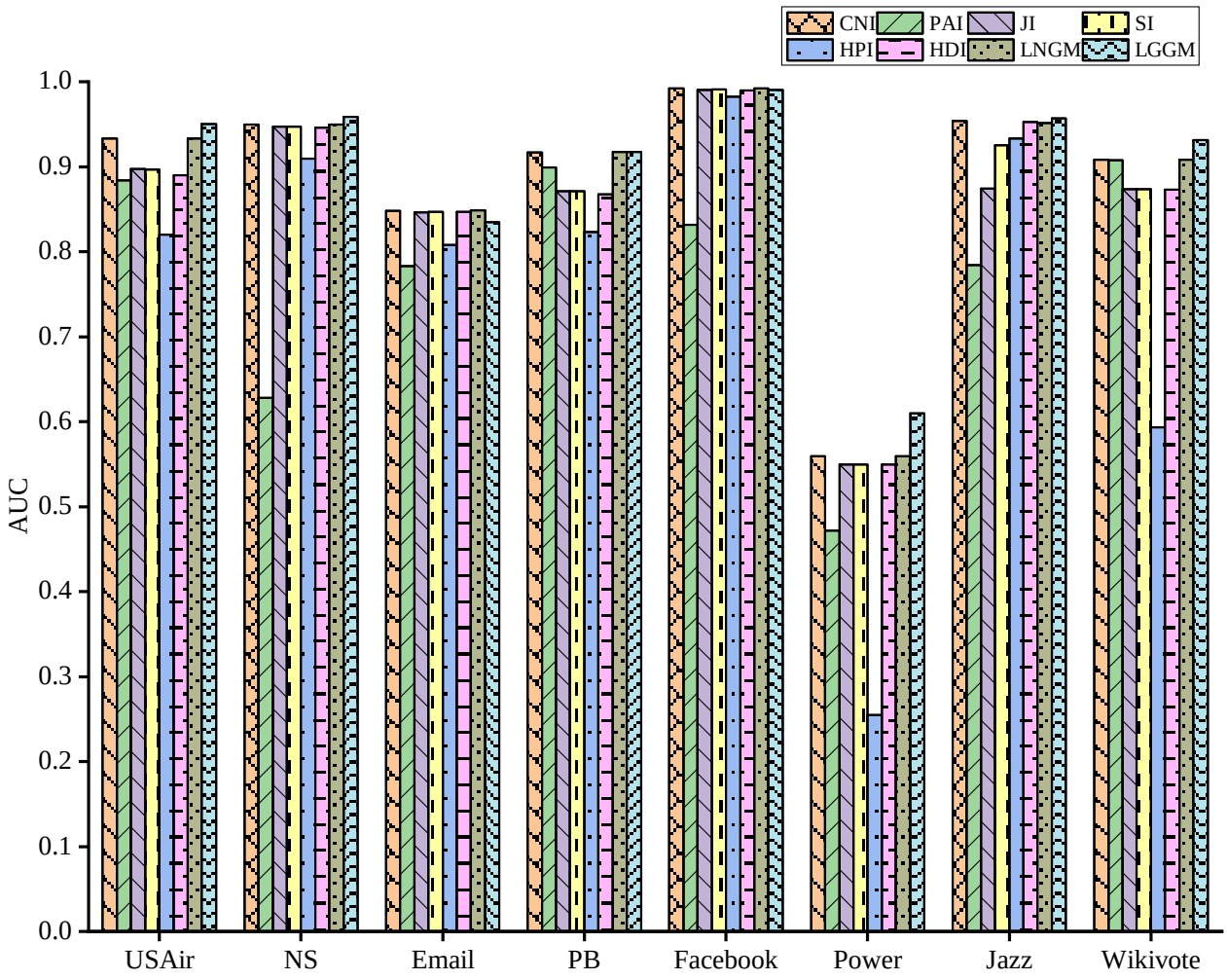

Fig. 6. The AUC for each link prediction model on each dataset.

comparison in this study.

We compared the proposed algorithm with several classical algorithms. Table III summarizes the AUC values of each algorithm across different datasets. To validate the stability of the outcomes, we conducted 20 separate experiments and calculated the average AUC values. Fig. 4 presents the average AUC values of the considered algorithms across all datasets. Our findings indicate that among this group of algorithms, LGGM has the highest average AUC value thereby outperforming all the compared algorithms. Specifically, LGGM attains an average AUC value of 0.8935, representing a 16% increase compared to HPI's lowest average AUC value of 0.7654, and 1.6% higher than LNGM's average AUC value. This demonstrates that the accuracy of LGGM has improved and surpasses that of the other algorithms. Notably, we observed that the AUC of LNGM was higher than that of LGGM for the Email and Facebook datasets. Therefore, we further compared the precision of both algorithms. Additionally, we evaluated the precision of LNGM and LGGM from the perspective of accuracy. The experimental results indicate that LGGM consistently exhibits significantly higher precision than LNGM across nearly all datasets. This observation underscores LGGM's robust generalization capability and stability, enabling effective identification of genuine links across diverse datasets of varying types and scales while minimizing false positives. Specific results are presented in Table IV, illustrating LGGM's superior performance. Overall, LGGM provides competitive and robust predictive outcomes when stacked against alternative methods. Fig.5 illustrates the RMSE values of the LGGM across various datasets. The figure indicates that the RMSE value for the Power dataset is the highest, which indicates that the prediction results of LGGM on this dataset are somewhat different from the observation results. Conversely, the RMSE for the Facebook dataset is the lowest, indicating a better model fit in this case. In addition, the RMSE values for the algorithm across the eight datasets fall within the range of (0, 0.3), with the maximum value being less than 0.5. This performance is generally regarded as indicative of good overall results.

In Fig.6, we illustrate the AUC value for each algorithm on eight datasets, this is to determine which datasets poses more challenges for link prediction compared to others. Two key factors explain this phenomenon: Firstly, in social networks like Facebook, user behavior patterns are crucial for the accuracy of link prediction. Subsequent relationships between users often reflect specific behavior patterns including interests, social circles, and activity levels. Link prediction algorithms leverage these patterns to accurately predict future connections. The rich and diverse interactions among Facebook users create a complex, in-

formative network structure, offering valuable insights into user behavior dynamics. Secondly, the scale and density of data within the Facebook datasets significantly enhance the effectiveness of link prediction algorithms. With millions of users and abundant interaction data, algorithms have access to a vast amount of information. This extensive data allows the algorithms to learn and adapt to intricate network dynamics, resulting in superior predictive performance on the Facebook dataset as compared to others with smaller data sizes or lower interaction densities. In contrast, LGGM shows the lowest AUC value among the Power datasets. The Power datasets represents a power network, a type of technological network within artificial network systems. The formation of this network is not instantaneous but involves long-term evolutionary development. The lower AUC value suggests that predicting links in technological networks is particularly challenging.

IV. Conclusion

To tackle the challenges associated with link prediction within complex networks, we propose an innovative algorithm based on the Local Generalized Gravity Model. This approach replaces mass in the gravity equation with node spreading ability, utilizing local clustering coefficients and degree metrics to extract topological insights from nodes. We compare the proposed method with other benchmark algorithms. The findings indicate that our method consistently surpasses benchmark algorithms in predictive accuracy across various datasets. Additionally, we assess the efficiency of the suggested algorithm under different parameter configurations.

References

[1] Q. A. Huang, J. C. Zhao, and X. Q. Wu, "Financial risk propagation between chinese and american stock markets based on multilayer networks," *Physica A: Statistical Mechanics and its Applications*, vol. 586, p. 126445, 2022.

[2] N. N. Daud, S. H. Ab Hamid, M. Saadoon, F. Sahran, and N. B. Anuar, "Applications of link prediction in social networks: A review," *Journal of Network and Computer Applications*, vol. 166, p. 102716, 2020.

[3] S. Wandelt, Y. f. Xu, and X. Sun, "Measuring node importance in air transportation systems: On the quality of complex network estimations," *Reliability Engineering & System Safety*, vol. 240, p. 109596, 2023.

[4] T. Ma, Q. Liu, J. Cao, Y. Tian, A. Al-Dhelaan, and M. Al-Rodhaan, "LGIEM: Global and local node influence based community detection," *Future Generation Computer Systems*, vol. 105, pp. 533–546, 2020.

[5] T. Zhang, K. Zhang, L. Lv, and D. Bardou, "Co-ranking for nodes, layers and timestamps in multilayer temporal networks," *Chaos, Solitons & Fractals*, vol. 125, pp. 88–96, 2019.

[6] K. Berahmand, A. Bouyer, and N. Samadi, "A new centrality measure based on the negative and positive effects of clustering coefficient for identifying influential spreaders in complex networks," *Chaos, Solitons & Fractals*, vol. 110, pp. 41–54, 2018.

[7] J. Zhu, T. Xue, P. Pei, X. Chen, H. Gao, and Q. Liu, "The synchronization ability of a class of multi-layer coupled networks," *IAENG International Journal of Applied Mathematics*, vol. 53, no. 1, pp. 209–216, 2023.

[8] A. Kumar, S. S. Singh, K. Singh, and B. Biswas, "Link prediction techniques, applications, and performance: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 553, p. 124289, 2020.

[9] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 6, pp. 1150–1170, 2011.

[10] S. Al-Oufi, H. N. Kim, and A. El Saddik, "Controlling privacy with trust-aware link prediction in online social networks," in *Proceedings of the Third International Conference on Internet Multimedia Computing and Service*, 2011, pp. 86–89.

[11] Z. Huang and D. D. Zeng, "A link prediction approach to anomalous email detection," in *2006 IEEE International Conference on Systems, Man and Cybernetics*, vol. 2. IEEE, 2006, pp. 1131–1136.

[12] Y. Sevim, G. K. Orman, and M. T. Yöndem, "Machine learning-based link prediction for hotel network," *IAENG International Journal of Computer Science*, vol. 49, no. 4, pp. 1268–1275, 2022.

[13] 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.

[14] M. E. Newman, "Clustering and preferential attachment in growing networks," *Physical Review E*, vol. 64, no. 2, p. 025102, 2001.

[15] 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.

[16] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–1555, 2002.

[17] T. Sorensen, "A method of establishing group of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. i kommission hos e," *Biologiske Skrifter*, vol. 7, pp. 1–34, 1948.

[18] P. Jaccard, "Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 241–272, 1901.

[19] Y. Xu, T. Ren, and S. Sun, "LNGM: A link prediction algorithm based on local neighbor gravity model," *International Journal of Modern Physics C*, vol. 33, no. 10, p. 2250134, 2022.

[20] 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.

[21] Z. Liu, J. L. He, K. Kapoor, and J. Srivastava, "Correlations between community structure and link formation in complex networks," *PloS One*, vol. 8, no. 9, p. e72908, 2013.

[22] I. M. Diop, C. Cherifi, C. Diallo, and H. Cherifi, "Revealing the component structure of the world air transportation network," *Applied Network Science*, vol. 6, pp. 1–50, 2021.

[23] J. Wang, "Knowledge creation in collaboration networks: Effects of tie configuration," *Research Policy*, vol. 45, no. 1, pp. 68–80, 2016.

[24] M. R. Barnes, M. Karan, S. McQuistin, C. Perkins, G. Tyson, M. Purver, I. Castro, and R. G. Clegg, "Temporal network analysis of email communication patterns in a long standing hierarchy," *arXiv preprint arXiv:2311.13442*, 2023.

[25] P. Chunaev, "Community detection in node-attributed social networks: a survey," *Computer Science Review*, vol. 37, p. 100286, 2020.

[26] T. S. Tata Sutabri, "Design of a web-based social network information system," *International Journal of Artificial Intelligence Research*, vol. 6, no. 1, pp. 310–316, 2023.

[27] C. Moya and G. Lin, "Dae-pinn: a physics-informed neural network model for simulating differential algebraic equations with application to power networks," *Neural Computing and Applications*, vol. 35, no. 5, pp. 3789–3804, 2023.

[28] Z. Wang, D. Hou, C. Gao, J. Huang, and Q. Xuan, "A rapid source localization method in the early stage of large-scale network propagation," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 1372–1380.

[29] E. Rijshouwer, J. Uitermark, and W. de Koster, "Wikipedia: a self-organizing bureaucracy," *Information, Communication & Society*, vol. 26, no. 7, pp. 1285–1302, 2023.

[30] A. E. Berndt, "Sampling methods," *Journal of Human Lactation*, vol. 36, no. 2, pp. 224–226, 2020.

[31] Y. Yang, R. N. Lichtenwalter, and N. V. Chawla, "Evaluating link prediction methods," *Knowledge and Information Systems*, vol. 45, pp. 751–782, 2015.

Yongqing Wu received the PhD degree in computational mathematics from Lanzhou University, China, in 2011. He worked as a visiting scholar with the University of Texas at Arlington, Texas, USA, from 2019 to 2020. Currently, he is an associate professor of College of Science, Liaoning Technical University, Fuxin, China. His current research interests include recommendation systems, machine learning algorithms, complex systems and complex networks.