The Comprehensive Contributions of Endpoint Degree and Coreness in Link Prediction

Yang Tian, Yanan Wang, Hui Tian, and Qimei Cui

1 State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
2 School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Hui Tian; tianhui@bupt.edu.cn

Received 12 June 2021; Revised 14 July 2021; Accepted 27 July 2021; Published 11 August 2021

1. Introduction

The research of link prediction aims to find the lost, false, or possible links through the observed network structure and information [1–5]. Therefore, link prediction algorithms have been applied to many fields. For example, link prediction algorithms can remove noise of the networks [6]. Furthermore, link prediction algorithms also can be applied to friends’ influence on online social networks [7–11] and products’ recommendation on e-commerce websites [12–16]. Moreover, link prediction algorithms provide references for biological experiments, which can reduce the cost of experiments [17–19]. In addition, link prediction algorithms can reveal network evolution mechanism and organization pattern [20–22].

To reveal the structure of complex networks, researchers propose a large number of link prediction models. Specifically, the models based on local information have been gained more attentions. For example, Kossinets [23] finds that two strangers become friends if they have more common friends in social networks. Newman [24] finds that two scientists have more likely to establish cooperation relationship in the future if they have more common cooperators. Based on this phenomenon, researchers propose the common neighbors’ model (CN). Based on CN, some researchers propose improved models, such as Salton [25] and LHN-I [26]. Furthermore, according to the different similarity contributions of common neighbors, Adamic and Adar propose the AA model [27]. Zhou et al. [28] propose the resource-allocation model (RA). Moreover, Cannistraci et al. [29] propose that CN, AA, RA, and other algorithms can be weighted by local community information, which can further improve the performances of these models. However, the models based on common neighbors only consider the influence of endpoints on one-step paths. Though further research studies, Lü et al. [30] propose local path model (LP) through considering the influence of endpoints on three-
step paths. In addition, some models consider global information, such as Katz [31] and hierarchical structure model [32]. Besides, some models consider quasi-local information, which can compensate the defect of low accuracy in local information and high-computational complexity in global information. For example, local random walk model (LRW) [33] considers a random walker within a quasi-local range, and superposed random walk model (SRW) [33] considers the effects of LRW with different path lengths. Based on SRW, HSRW [34] and CSRW [34] models consider the roles of H-index [35] and coreness [36] with different path lengths, respectively. Simple hybrid influence model (SHI) [37] synthetically considers the role of endpoint degree and H-index as hybrid influence with different path lengths.

At present, many link prediction models only consider the degree [33] of endpoints, such as Sørensen [38], LHN [26], LRW [33], and SRW [33]. These models illustrate that the source endpoint can effectively spread its influence to the target endpoint if the source endpoint has more neighbors to connect the target endpoint. Through abundant study, Lü et al. [39] find that H-index shows a better performance to quantify the influence of endpoint than degree and coreness. Zhu et al. [37] find that an endpoint possessing large syntethetical degree and H-index can acquire a more extensive maximal connected subgraph, which can help the endpoint to attract other nodes. Through further investigations, we find that the endpoint influence can be expressed by the aggregation degree of neighbors. The large aggregation degree of neighbors illustrates that the endpoint has the extensive maximal connected subgraph, leading to attract more nodes. The aggregation degree of neighbors can only be quantified by coreness of endpoints. Thus, we synthesize the endpoint degree and coreness (or H-index and coreness) to quantify the endpoint influence and build the new link prediction models. Although the SHI model based on the syntethetical degree and H-index has been explored, the syntethetical endpoint degree and coreness (or the syntethetical H-index and coreness) has not been fully verified.

Figure 1 shows a clear illustration. In Figure 1, endpoint \( b \) possesses degree \( d = 6 \), H-index \( l = 3 \), and coreness \( \gamma = 3 \), respectively. The influence intensity of endpoint \( b \) is size 6 in consideration of only degree. However, only the degree cannot express the depth and scope of the influence of endpoints accurately. Due to the role of coreness in influence of endpoints, synthezizing degree, and coreness or synthetical H-index and coreness can better quantify the maximal connected subgraph of endpoints and the aggregation degree of neighbors. For endpoint \( b \), the product of degree (H-index) and coreness is 18 (9). Obviously, degree and H-index indicate the different sizes of maximal connected subgraph belonging to endpoint \( b \) with the same coreness, leading to different influences of endpoints. Therefore, the prediction performance on the influence of endpoints based on the different quantification index needs to be further explored.

In real world, we find many phenomena to confirm our idea. For example, in Weibo, an ordinary individual possesses the limited influence because he/she only has many individual followers from colleagues, classmates, relatives, or friends, indicating that he/she only has large degree. However, public figures possess extensive and strong influence because they have large number of fan club, indicating that they have large coreness to strengthen their influence. In scientists' collaboration network, if a scientist only cooperates with many scholars, meaning he/she has large degree but small coreness, the scientist cannot be known by more researchers and can hardly further attract them to cooperate. In e-commerce network, the applicability of products depends on purchase groups with similar identities, such as male/female group, student group, and teacher group, which shows the importance of aggregation degree. In paper-citation network, the value of a paper depends on the citation of researchers in the same field, not the citation of researchers in the different fields.

In summary, in this paper, we define the hybrid influence of syntethetical degree and coreness (syntethetical H-index and coreness) to redefine the SRW and propose two improved models DCHI and HCHI to further explore the accuracy of link prediction. Experimental results on twelve real networks show that DCHI exhibits better performances of link prediction.

The rest of this paper is organized as follows. In Section 2, we build two models based on the syntethetical degree and coreness and the syntethetical H-index and coreness, respectively. In Section 3, the thirteen benchmark experimental datasets are introduced. In Section 4, a link prediction metric and eight mainstream baselines are described, respectively. In Section 5, the experimental results are discussed. In Section 6, the conclusion is described.

2. Models Based on Hybrid Influence of Endpoints

Firstly, we study link prediction models in an undirected simple network \( G(V, E) \), where \( E \) is the set of links (\(|E|\) denotes the number of all edges.) and \( V \) refers to the set of nodes. Multiple links and self-connections are eliminated. For every pair of nodes, \( x, y \in V \), a score, \( s_{xy} \), is given to calculate the probability of their future connection. In this paper, we set the similarity value as a score directly, and a larger score illustrates that the potential link has more possibility to be found.

Secondly, we show two models based on the degree (SRW [33]) and the syntethetical degree and H-index (SHI [37]) separately as follows.

2.1. SRW Model. Liu et al. [33] build the similarity model using random walk, which finds all intermediate nodes sequentially between two endpoints according to a Markov chain with one-step transmission probability \( p_{xy} = \frac{a_{xy}}{d_{x}} \), where \( k_{x} \) represents the degree of node \( x \) and \( a_{xy} = 1 \) if node \( x \) successfully connects \( y \) and \( a_{xy} = 0 \) if not. The sequence of node with t-step between \( x \) and \( y \) is expressed as \( \{x = x_{0} = y_{1} = x_{1} = y_{2}, \ldots, x_{r-1} = y_{r}, x_{r} = y_{0} = y\} \). Thus, the t-step transmission probability from \( x \) to \( y \) is denoted by \( \pi_{xy}(t) = \prod_{i=0}^{t-1} p_{x_{i},x_{i+1}} \) and \( \pi_{yx}(t) = \prod_{i=0}^{t-1} p_{y_{i},y_{i+1}} \). Importantly, Liu et al. consider the degree \( k_{x} \) and \( k_{y} \) to quantify the influence of endpoints and define the SRW as
where $k_x$ and $k_y$ denote the degree of endpoint $x$ and $y$, respectively, and $|E|$ indicates the number of links in the network. $(k_x/2|E|)$ and $(k_y/2|E|)$ describe the influence of endpoint $x$ and $y$, respectively.

2.2. SHI Model. Zhu et al. [37] find that the H-index can represent the maximal connected subgraph of endpoints and describe the influence intensity. Thus, Zhu et al. simply synthesize degree and H-index as the hybrid influence of node $(x)$ and endpoint $(y)$, respectively, and describe the influence intensity of endpoints. -% hus, we consider the role of neighbors and the maximal connected subgraph of endpoints to quantify the influence of endpoints. -% he indicates the number of links in the endpoint influence, they only represent the number of neighbors and the maximal connected subgraph of endpoints separately, ignoring the influence intensity of endpoints. The influence intensity of endpoints can be expressed by the coreness of endpoints because the coreness can quantify the aggregation degree of neighbors which represents the endpoint influence intensity. Thus, we consider the role of coreness for endpoint influence. Finally, we build two models based on synthetical degree and coreness (DCHI) and synthetical H-index and coreness (HCHI) separately as follows.

2.2. SHI Model. Zhu et al. [37] find that the H-index can represent the maximal connected subgraph of endpoints and describe the influence intensity. Thus, Zhu et al. simply synthesize degree and H-index as the hybrid influence of node $(x)$ and endpoint $(y)$, respectively, and describe the influence intensity of endpoints. -% hus, we consider the role of neighbors and the maximal connected subgraph of endpoints to quantify the influence of endpoints. -% he indicates the number of links in the endpoint influence, they only represent the number of neighbors and the maximal connected subgraph of endpoints separately, ignoring the influence intensity of endpoints. The influence intensity of endpoints can be expressed by the coreness of endpoints because the coreness can quantify the aggregation degree of neighbors which represents the endpoint influence intensity. Thus, we consider the role of coreness for endpoint influence. Finally, we build two models based on synthetical degree and coreness (DCHI) and synthetical H-index and coreness (HCHI) separately as follows.

$$s_{xy}^{SHI}(t) = \sum_{l=1}^{t} \left( \sqrt{\frac{k_x \times h_x}{2|E|}} \times \pi_{xy}(l) + \sqrt{\frac{k_y \times h_y}{2|E|}} \times \pi_{yx}(l) \right),$$

where $(\sqrt{\frac{k_x \times h_x}{2|E|}})$ and $(\sqrt{\frac{k_y \times h_y}{2|E|}})$ denote the hybrid influence of node $x$ and $y$ based on synthetical degree and H-index, respectively.

Although the endpoint degree and H-index can quantify the endpoint influence, they only represent the number of neighbors and the maximal connected subgraph of endpoints separately, ignoring the influence intensity of endpoints. The influence intensity of endpoints can be expressed by the coreness of endpoints because the coreness can quantify the aggregation degree of neighbors which represents the endpoint influence intensity. Thus, we consider the role of coreness for endpoint influence. Finally, we build two models based on synthetical degree and coreness (DCHI) and synthetical H-index and coreness (HCHI) separately as follows.

2.3. DCHI Model. Through the explanation in Section 1 and the illustration in Figure 1, we synthesize degree and coreness to quantify the influence of endpoints and replace the degree in SRW to define a simple hybrid influence model (SHI) as

$$s_{xy}^{DCHI}(t) = \sum_{l=1}^{t} \left( \sqrt{\frac{k_x \times c_x}{2|E|}} \times \pi_{xy}(l) + \sqrt{\frac{k_y \times c_y}{2|E|}} \times \pi_{yx}(l) \right),$$

where $(\sqrt{\frac{k_x \times c_x}{2|E|}})$ and $(\sqrt{\frac{k_y \times c_y}{2|E|}})$ denote the hybrid influence of node $x$ and $y$ based on synthetical degree and coreness, respectively.

2.4. HCHI Model. Furthermore, we synthesize H-index and coreness to quantify the influence and replace the degree in SRW to build a new model HCHI as

$$s_{xy}^{HCHI}(t) = \sum_{l=1}^{t} \left( \sqrt{\frac{h_x \times c_x}{2|E|}} \times \pi_{xy}(l) + \sqrt{\frac{h_y \times c_y}{2|E|}} \times \pi_{yx}(l) \right),$$

where $(\sqrt{\frac{h_x \times c_x}{2|E|}})$ and $(\sqrt{\frac{h_y \times c_y}{2|E|}})$ denote the hybrid influence of node $x$ and $y$ based on synthetical H-index and coreness, respectively.

3. Experimental Data

In this section, we introduce 12 real network datasets to prepare the following experiments. (1) US Air97 (USAir) [40] represents the US airline network. (2) Yeast PPI (Yeast) [41] represents the yeast network of relationship between proteins. (3) Food Web (Food) [42] represents the relations of carbon exchanges in the cypress wetlands of Florida ecosystem. (4) Power Grid (Power) [43] represents the western US's electrical power transmission network. (5) NetScience (NS) [44] represents partnerships between scientists in publishing papers concerning the subject of networks. (6) Jazz [45] represents the networks of Jazz musicians. (7) e-mail network (e-mail) [46] represents e-mail communication network of University Rovira i Virgili (URV) in Spain. (8) Slavko [47] represents the friendship network of Slavko Zitnik on Facebook. (9) UC Irvine dealing with social network (UCsocial) [48] represents an online social network composed of students in the University of California, Irvine. (10) Infectious (Infect) [49] represents the offline contact network of visitors in the course of the exhibition named "Infectious: Stay Away" in the Science Gallery in Dublin, 2009. (11) EuroSiS web (EuroSiS) [50] represents network between Science in Society actors from twelve European countries. (12) C. elegans (CE) [43] represents the network of neurons in the C. elegans worm. Table 1 lists the mentioned networks fundamental topological features.

To achieve preprocess, arcs are changed as nondirectional links, and loops and multiedges are eliminated to ensure the network unwighted and undirected. Subsequently, the largest linked simplified network subgraph is extracted to make sure the connectivity.

In the beginning, the set of network links is divided into the training set $E'$ containing 90% links in a random manner, and the testing set $E''$ containing 10% links, while the connectivity of $E'$ is ensured [1]. Besides, 30 divisions are identically and separately conducted on the network. Next, experimental processes are performed over the 30 separated training and testing sets, the averaged accuracy is achieved in a statistical manner, and metrics is recalled more than 30 times realization.

4. Experimental Methods

4.1. Metric. AUC [36], a metric of accuracy, can be interpreted as the probability that a potential link (a link in $E'$) ranks a higher score than a nonexistent link (a link in $U-E$, where $U$ denotes the universal link set). In the specific implementation, among $n$ independent comparisons if the potential link ranks higher in $n'$ times and the same as the nonexistent link in $n''$ times, and the total score accumulates $n'$ and $0.5n''$. After that, AUC expresses the averaged score over $n$-time comparisons as

$$AUC = \frac{n' + 0.5n''}{n}.$$  

AUC evaluates the performance of a model globally. If all scores originate from an independent and identical
distribution, the value should equal to 0.5. Therefore, the extent to which the accuracy exceeds 0.5 suggests how much better a model performs than pure chance.

4.2. Baselines. Comparatively, we introduce eight fundamental models as follows:

1. Common neighbors (CN) [24] describe the similarity between endpoints by calculating the number of common neighbors, defined as

\[ s_{xy}^{CN} = |\Gamma(x) \cap \Gamma(y)|, \]

where \( \Gamma(X) \), \( X \in \{x, y\} \), represents the set of neighbors of endpoint \( X \) and \( |\Gamma(x) \cap \Gamma(y)| \) refers to the number of common neighbors of endpoints \( x \) and \( y \).

2. Adamic/Adar (AA) [27], based on CN, suppresses the contributions of common neighbors with big degree by applying the inverse logarithm, which is defined as

\[ s_{xy}^{AA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(k_z)}, \]

where \( k_z \) represents the degree of node \( z \).

3. Resource-Allocation (RA) [28], analogous to AA, suppresses the large degree of common neighbors by applying the reciprocal of the degrees of common neighbors, which is defined as

\[ s_{xy}^{RA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z}. \]

4. Local Path Index (LP) [30] considers the similarity on two-step and three-step paths between endpoints simultaneously, with the two-step paths preferred, which are defined as

\[ s_{xy}^{LP} = A^2 + \epsilon A^3, \]

where \( A \) represents the adjacency matrix and \( \epsilon \) is a punishment parameter.

5. Superposed Random Walk (SRW) [33] is introduced in Section 2.

6. CSRW [34] exploits the coreness to quantify the influence of endpoint and replace the degree influence in SRW, which is defined as

\[ s_{xy}^{CSRW}(t) = \sum_{l=2}^{t} \left[ \frac{c_x}{2|E|} \pi_{xy}(l) + \frac{c_y}{2|E|} \pi_{yx}(l) \right], \]

where \( c_x \) and \( c_y \) represent the coreness of node \( x \) and \( y \), respectively.

7. HSRW [34] exploits the H-index to quantify the influence of endpoint and replace the degree influence in SRW, defined as

\[ s_{xy}^{HSRW}(t) = \sum_{l=2}^{t} \left[ \frac{h_x}{2|E|} \pi_{xy}(l) + \frac{h_y}{2|E|} \pi_{yx}(l) \right], \]

where \( h_x \) and \( h_y \) represent the H-index of node \( x \) and \( y \), respectively.

8. Simple hybrid influence (SHI) [33] is introduced in Section 2.

5. Results and Discussion

To explore the prediction performances of the proposed models, extensive simulations are conducted on 12 real datasets. Through comparisons with several main baselines in terms of accuracy metric, we obtain the experimental results on the models and discuss the findings in the following.

SHI, HCHI, and DCHI models mainly consider two aspects: random walk on paths and hybrid influences of endpoints. Through simulations, the experimental results show that the number of steps in random walk between endpoints can affect the accuracy of link prediction. For illustrating the changes of prediction accuracy on the number of steps \( t \), we plot the relation curves in Figure 2.

In Figure 2, SHI (synthetical degree and H-index), HCHI (synthetical degree and coreness), and DCHI (synthetical H-index and coreness) models show their prediction performances on the random steps \( t \), and they exhibit different optimal accuracies at certain number of steps \( t \), respectively. Specifically, SHI shows optimal AUC values at \( t = 15 \) in food, power, NS, e-mail, UCsocial, and Eurosis, \( t = 5 \) in USAir and CE, \( t = 3 \) in yeast, \( t = 2 \) in Jazz, \( t = 6 \) in Slavko, and \( t = 9 \) in

![Figure 1: Illustration of the influences of endpoint b based on degree, H-index, and coreness. Endpoint b has degree = 6, H-index = 3, and coreness = 3. Node a and c represent two target endpoints, respectively. The blue nodes represent the medial nodes between endpoint b and two target endpoints.](image-url)
Table 1: The basic topological features of the twelve benchmark networks. There are the properties: \(|V|\) denoting the number of nodes, \(|E|\) denoting the number of links, \(\langle k \rangle\) denoting the average degree, \(\langle d \rangle\) representing the average distance, \(C\) representing the clustering coefficient, \(r\) indicating the assortativity coefficient, and \(H\) indicating the degree heterogeneity and defined as \(H = \langle k^2 \rangle / \langle k \rangle^2\).

| Nets     | \(|V|\) | \(|E|\) | \(\langle k \rangle\) | \(\langle d \rangle\) | \(C\)   | \(r\)   | \(H\)   |
|----------|--------|--------|----------------------|----------------------|--------|--------|--------|
| Usair    | 332    | 2128   | 12.81                | 2.74                 | 0.749  | -0.208 | 3.36   |
| Yeast    | 2370   | 10904  | 9.2                  | 5.16                 | 0.378  | 0.469  | 3.35   |
| Food     | 128    | 2075   | 32.42                | 1.78                 | 0.334  | -0.112 | 1.24   |
| Power    | 4941   | 6594   | 2.669                | 15.87                | 0.107  | 0.003  | 1.45   |
| NS       | 1461   | 2742   | 3.75                 | 5.82                 | 0.878  | 0.461  | 1.85   |
| Jazz     | 198    | 2742   | 27.7                 | 2.24                 | 0.633  | 0.02   | 1.4    |
| E-mail   | 1133   | 5451   | 9.62                 | 3.61                 | 0.254  | 0.078  | 1.94   |
| Slavko   | 334    | 2218   | 13.28                | 3.05                 | 0.488  | 0.247  | 1.62   |
| Ucsocial | 1893   | 13825  | 14.62                | 3.06                 | 0.138  | -0.188 | 3.81   |
| Infec    | 410    | 2765   | 13.49                | 3.63                 | 0.467  | 0.226  | 1.39   |
| Eurosis  | 1272   | 6454   | 10.15                | 3.86                 | 0.382  | -0.012 | 2.46   |
| CE       | 453    | 2025   | 8.94                 | 2.66                 | 0.655  | -0.225 | 4.49   |

Figure 2: Continued.
Infec. Obviously, the optimal number of steps on SHI mainly appears on the long path $t = 15$, illustrating that long paths can further facilitate the hybrid influence spreading based on the degree and the H-index. However, HCHI and DCHI all show optimal AUC values at $t = 5$ in USAir, Food, Slavko, Infec, Eurosis, and CE, illustrating that quasi-local paths can further facilitate the hybrid influence spreading based on H-index and coreness or degree and coreness. Importantly, we find that the influence concerning coreness can easily leak in the random-walk process on longer paths, which leads to weaken the intensity of influence spreading between endpoints. However, in power, the prediction performances of HCHI and DCHI reach the optimal value at $t = 15$ because power network includes large numbers of long paths with average distance $\langle d \rangle = 15.87$ much longer than other datasets (referring to Table 1). In addition, DCHI, compared with SHI and HCHI, has larger size of maximal connected subgraph and more paths to spread the hybrid influence of endpoints. Therefore, DCHI shows the best prediction performances in ten datasets (black mark on each dataset) except yeast and CE.

In addition, we compare HCHI and DCHI with eight link prediction models CN, AA, RA, LP, SRW, CSRW, HSRW, and SHI. To exhibit the experimental results, we show the averaged AUC values over 30 simulations in Table 2 for all models. The underlined bold fonts represent the best AUC values in each dataset and the numbers in parenthesis indicate the optimal random walk steps $t$, at which HCHI and DCHI obtain the optimal AUC values in eight datasets altogether.

As can be seen from Table 2, optimal values on seven datasets exist in DCHI with Power, NS, Jazz, Email, Slavko, Infec, and Eurosis. In contrast, local models CN, AA, and RA show worst prediction performances because they only consider the local paths and ignore the influence of endpoints. Then, optimal values on three datasets exist in LP with yeast, food, and UCsocial, illustrating that the quasi-local paths can limitedly promote the prediction performances. And then, SRW, CSRW, and HSRW also show worst performances because they only consider separately the contributions of degree, coreness, and H-index, meaning that degree, coreness, and H-index all cannot quantify the
influence of endpoints comprehensively. Finally, we focus on the performances of SHI, HCHI, and DCHI. In twelve datasets, there are seven optimal performances in DCHI. DCHI, compared with SHI and HCHI, shows the effective influence of endpoints (e.g., extensive maximal connected subgraph of endpoints and aggregation degree of neighbors) and finds sufficient paths between two unconnected endpoints. Therefore, because the synthetical degree and coreness as hybrid influence can be a good quantification index, DCHI can better enhance prediction accuracy than SHI and HCHI in many cases of link prediction.

Besides, the low computation complexity is a necessary condition in link prediction. The time complexity of the product of two $N \times N$ matrices is $O(N^3)$. According to the definitions of the baseline models, CN, AA, and RA possess the time complexity of $O(N^3)$ and LP, SRW, CSRW, HSRW, and SHI have $M \times O(N^3)$ with coefficient $M$. Although HCHI and DCHI have the same time complexity $M \times O(N^3)$, two proposed models, especially DCHI, show greater performance improvement. Therefore, the proposed models show a better performance with no increase in complexity.

6. Conclusions

At present, researchers pay more attention to the contributions of the influence of endpoints for link prediction based on local, quasi-local, or global similarity. To quantify the influence of endpoints, researchers consider the degree, H-index, or coreness separately, which all cannot evaluate the influence of endpoints comprehensively. Specifically, the endpoint degree only represents the number of neighbors of endpoints, but cannot describe the maximal connected subgraph. The H-index can express the maximal connected subgraph of endpoints to quantify the influence scope. However, the endpoint degree and H-index cannot quantify the influence intensity of endpoints and result in incomplete influence expression. We find that the coreness can represent the aggregation degree of endpoints, which can quantify the influence intensity of endpoints accurately.

Through abundant investigations, we find that the synthetical degree and coreness and the synthetical H-index and coreness can quantify the influence of endpoints accurately and comprehensively. Therefore, we synthesize degree (H-index) and coreness as the hybrid influence of endpoints and replace the degree in SRW to build two models DCHI and HCHI.

We explore the prediction performances of DCHI and HCHI by the comparisons among CN, AA, RA, LP, SRW, CSRW, HSRW, and SHI on twelve real datasets. As a result, we show that DCHI obviously outperform other models on the metric AUC and do not increase computational complexity. The outstanding improvement in accuracy illustrates the synthetical degree and coreness as hybrid influence of endpoints can describe the endpoint influence intensity accurately and can attract more nodes to produce links.

Although our models have been verified on the datasets, the models only make a simple synthesis between endpoint degree, H-index, and coreness. We find degrees differ in different networks, and so do H-indices and coreness. The network heterogeneity characterized by heterogeneous degrees, H-indices, and coreness directly results in heterogeneous influences. And, we find that endpoints in network with smaller heterogeneous influence can attract each other more likely. For such characteristic, we will further carry out research on the heterogeneous hybrid influence model based on DCHI and HCHI. In the future research studies, the impact of heterogeneous complex networks will become a crucial problem.

In addition, our study may provide new findings relating to link prediction based on similarity in future. Our research results can be applied to friends’ recommendation, products’ recommendation, scientists’ cooperation, biological experiments, and so on.

Data Availability

The data used to support the findings of the study are available at http://vlado.fmf.uni-lj.si/pub/networks/data/ and http://snap.stanford.edu/data/index.html.

Conflicts of Interest

The authors declare that they have no conflicts of interest.
Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61471060).

References

[1] L. Lu and T. Zhou, "Link prediction in complex networks: a survey," Physica A: Statistical Mechanics and Its Applications, vol. 390, 2010.
[2] B. Barzel and A. L. Barabási, "Network link prediction by global silencing of indirect correlations," Nature Biotechnology, vol. 31, no. 8, pp. 720–725, 2013.
[3] L. Lu, L. Pan, T. Zhou, Y. C. Zhang, and H. Stanley, "Toward link predictability of complex networks," Proceedings of the National Academy of Sciences of the United States of America, vol. 112, 2325 pages, 2015.
[4] Y. Wang, Y. Wang, X. Lin, and W. Wang, "The influence of network structural preference on link prediction," Discrete Dynamics in Nature and Society, vol. 2020, Article ID 6148273, 9 pages, 2020.
[5] K. Chi, G. Yin, Y. Dong, and H. Dong, "Link prediction in dynamic networks based on the attraction force between nodes," Knowledge-Based Systems, vol. 181, Article ID 104792, 2019.
[6] M. Newman, "Network structure from rich but noisy data," Nature Physics, vol. 14, 6 pages, 2018.
[7] L. M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines, and F. Menczer, "Friendship prediction and homophily in social media," ACM Transactions on the Web, vol. 6, no. 2, pp. 1–33, 2012.
[8] P. Wang, B. Xu, Y. Wu, and X. Zhou, "Link prediction in social networks: the state-of-the-art," Science China Information Sciences, vol. 58, no. 1, pp. 1–38, 2014.
[9] L. Pan, T. Zhou, L. Lu, and C. K. Hu, "Predicting missing links and identifying spurious links via likelihood analysis," Scientific Reports, vol. 6, no. 1, Article ID 22955, 2016.
[10] D. Gurini, F. Gasparetti, A. Micarelli, and G. Sansonetti, "Temporal people-to-people recommendation on social networks with sentiment-based matrix factorization," Future Generation Computer Systems, vol. 78, pp. 430–439, 2017.
[11] F. Xiong, X. Wang, S. Pan, H. Yang, H. Wang, and C. Zhang, "Social recommendation with evolutionary opinion dynamics," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 10, pp. 1–13, 2018.
[12] L. Lu, M. Medo, C. H. Yeung, Y. C. Zhang, and Z. T. Zhang, "Recommend system," Physics Reports-Review Section of Physics Letters, vol. 519, pp. 1–49, 2012.
[13] 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, Article ID 124289, 2020.
[14] W. Wang, Q. H. Liu, J. Liang, Y. Hu, and T. Zhou, "Co-evolution spreading in complex networks," Physics Reports, vol. 820, pp. 1–15, 2019.
[15] X. Chen, K. Gong, R. Wang, S. Cai, and W. Wang, "Effects of heterogeneous self-protection awareness on resource-epidemic coevolution dynamics," Applied Mathematics and Computation, vol. 385, Article ID 125428, 2020.
[16] X. Chen, Q. Liu, R. Wang, Q. Li, and W. Wang, "Self-awareness-based resource allocation strategy for containment of epidemic spreading," Complexity, vol. 2020, Article ID 3256415, 12 pages, 2020.
[17] M. Angeles Serrano and F. Sagues, "Network-based scoring system for genome-scale metabolic reconstructions," BMC Systems Biology, vol. 5, 76 pages, 2011.
[18] M. A. Serrano, M. Boguñá, and F. Sagués, "Uncovering the hidden geometry behind metabolic networks," Molecular bioSystems, vol. 8, no. 3, pp. 843–850, 2012.
[19] X. Chen, R. Wang, D. Yang, J. Xian, and Q. Li, "Effects of the awareness-driven individual resource allocation on the epidemic dynamics," Complexity, vol. 2020, Article ID 8661493, 12 pages, 2020.
[20] W. Wang, Q. Zhang, and T. Zhou, "Evaluating network models: a likelihood analysis," Europhysics Letters (Epl), vol. 98, pp. 28004–28009, 2011.
[21] F. Tan, Y. Xia, and B. Zhu, "Link prediction in complex networks: a mutual information perspective," PLoS One, vol. 9, no. 9, Article ID e107056, 2014.
[22] B. Zhu and Y. Xia, "Link prediction in weighted networks: a weighted mutual information model," PLoS One, vol. 11, no. 2, Article ID e0148265, 2016.
[23] G. Kossinets, "Effects of missing data in social networks," Social Networks, vol. 28, no. 3, pp. 247–268, 2006.
[24] M. Newman, "Clustering and preferential attachment in growing networks," Physical Review E, vol. 64, Article ID 025102, 2001.
[25] M. Davenport, "Introduction to modern information retrieval," Journal of the Medical Library Association: JMLA, vol. 100, no. 1, 75 pages, 3rd edition, 2012.
[26] A. L. Adamic and A. Eytan, "Friends and neighbors on the web," Social Networks, vol. 25, pp. 211–230, 2003.
[27] T. Zhou, L. Lu, and Y. C. Zhang, "Link prediction of missing links via local information," The European Physical Journal B, vol. 71, no. 4, pp. 623–630, 2009.
[28] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi, "From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks," Scientific Reports, vol. 3, no. 1, 1613 pages, 2013.
[29] L. Lu, C. H. Lü, and T. Zhou, "Similarity index based on local paths for link prediction of complex networks," Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, vol. 80, Article ID 046122, 2009.
[30] L. Katz, "A new status index derived from sociometric analysis," Psychometrika, vol. 18, pp. 39–43, 1953.
[31] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," Nature, vol. 453, no. 7191, pp. 98–101, 2008.
[32] W. Liu and L. Lu, "Link prediction based on local random walk," Europhysics Letter, vol. 89, Article ID 58007, 2010.
[33] X. Zhu, Y. Yang, L. Li, and S. Cai, "Roles of degree, h-index and coreness in link prediction of complex networks," International Journal of Modern Physics B, vol. 32, Article ID 1850097, 2018.
[34] J. Hirsch, "An index to quantify an individual’s scientific research output," Proceedings of the National Academy of Sciences of the United States of America (PNAS), vol. 102, pp. 16569–16572, 2005.
[35] M. Kitsak, L. K. Gallos, S. Havlin, and F. Liljeros, "Identification of influential spreaders in complex networks," Nature Physics, vol. 6, pp. 888–893, 2010.
[36] X. Zhu, W. Li, H. Tian, and S. Cai, "Hybrid influence of degree and h-index in the link prediction of complex networks," EPL, vol. 122, Article ID 68003, 2018.
[38] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons," Biologiske Skrifter/Kongelige Danske Videnskabernes Selskab, vol. 5, 1948.

[39] L. Liu, T. Zhou, Q. M. Zhang, and H. Stanley, "The h-index of a network node and its relation to degree and coreness," Nature Communications, vol. 7, Article ID 10168, 2016.

[40] V. Batagelj and A. Mrvar, Pajek-Program For Large Network Analysis, University of Ljubljana, Ljubljana, Slovenia, 1999.

[41] D. Bu, Y. Zhao, L. Cai, and H. Xue, "Topological structure analysis of the protein-protein interaction network in budding yeast," Nucleic Acids Research, vol. 31, pp. 2443–2450, 2003.

[42] C. Melian and J. Bascompte, "Food web cohesion," Ecology, vol. 85, pp. 352–358, 2004.

[43] G. Yan, T. Zhou, B. Hu, Z. Q. Fu, and B. Wang, "Efficient routing on complex networks," Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, vol. 73, Article ID 046108, 2006.

[44] P. Holme and M. Newman, "Nonequilibrium phase transition in the coevolution of networks and opinions," Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, vol. 74, Article ID 056108, 2006.

[45] P. Gleiser and L. Danon, "Community structure in jazz," Advances in Complex Systems, vol. 6, 2003.

[46] R. Guimerà, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, "Self-similar community structure in a network of human interactions," Physical Review E, vol. 68, 2003.

[47] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," Social Networks, vol. 31, 2009.

[48] N. Blagus, L. Subelj, and M. Bajec, "Self-similar scaling of density in complex real-world networks," Physica A: Statistical Mechanics and Its Applications, vol. 391, 2012.

[49] L. Isella, J. Stehle, A. Barrat, C. Cattuto, and J. F. Pinton, "What’s in a crowd? analysis of face-to-face behavioral networks," Journal of Theoretical Biology, vol. 271, 2011.

[50] B. Ermis, E. Acar, and A. Cemgil, "Link prediction in heterogeneous data via generalized coupled tensor factorization," Data Mining and Knowledge Discovery, vol. 29, pp. 203–236, 2013.