

· 复杂性科学 ·

## Predicting Missing Links of Complex Network via Effective Common Neighbors

WANG Kai, LIU Shu-xin\*, YU Hong-tao, and LI Xing

(National Digital Switching System Engineering and Technological R&D Center Zhengzhou 450002)

**Abstract** Link prediction can predict the missing links of complex networks, which promotes a better understanding of evolution mechanisms in real networks. Many similarity indices have been proposed based on a topology structure for link prediction. Local topological information, especially common neighbors, plays an important role in calculating the similarity between two endpoints. However, plenty of similarity indices ignore the effectiveness of common neighbors under different topology structures. Considering the local topological information around common neighbors, an effective common neighbor index is proposed. Firstly, we analyze the effectiveness of all neighbor links of common neighbors. Then, based on the local topology on both sides of two endpoints around common neighbors, the effectiveness of two sides of common neighbors is quantified separately. Finally, the similarity between two endpoints is described through the effect of common neighbors' effectiveness on bilateral resource allocation process. Empirical study on 15 real networks shows that the index proposed can achieve higher prediction accuracy, compared with 9 mainstream baselines.

**Key words** complex network; effective common neighbors; link prediction; network topology; similarity

## 基于共同邻居有效性的复杂网络链路预测算法

王 凯, 刘树新\*, 于洪涛, 李 星

(国家数字交换系统工程技术研究中心 郑州 450002)

**【摘要】**链路预测旨在预测网络中的缺失连边, 对于实际网络演化机制的了解具有重要意义。虽然现有研究已经提出了很多相似性指标, 但它们都忽视了不同网络结构下共同邻居的有效性, 而局部拓扑结构信息尤其是共同邻居结构在计算节点间相似性中发挥重要作用。考虑到共同邻居周围局部拓扑信息, 该文提出了一种高效共同邻居指标。该指标首先分析了共同邻居所有连边的有效性, 分别从端点两侧量化了节点的有效性; 然后, 通过分析共同邻居节点拓扑有效性对两侧资源分配过程的影响刻画节点间相似性。15个实际网络数据实验表明, 相比现有经典的9种方法, 所提方法具有较高的预测精度。

**关键词** 复杂网络; 共同邻居有效性; 链路预测; 网络拓扑; 相似性

**中图分类号** TP391; N94

**文献标志码** A

doi:10.3969/j.issn.1001-0548.2019.03.020

With the development of complex network research, many more real complex systems have been described as complex network<sup>[1-5]</sup>. Link prediction aims to predict the likelihood that a link exists between two endpoints in complex networks<sup>[6]</sup>. Predicting missing links plays an important role in discovering the unknown connections in protein-protein interaction networks<sup>[7]</sup>, forecasting the future relationships in

online social networks<sup>[8]</sup>, and identifying spurious flights in air transportation networks<sup>[9]</sup>.

Based on evolution mechanisms of complex networks, a lot of link prediction methods have been proposed. Among these methods, topology-based similarity indices have got a great success due to their high performance and low complexity<sup>[10]</sup>. Similarity indices can be classified as local indices and global

收稿日期: 2018-03-21; 修回日期: 2018-11-15

基金项目: 国家自然科学基金创新研究群体项目(61521003)

作者简介: 王凯(1980-), 男, 博士生, 副研究员, 主要从事社会网络分析方面的研究。

通信作者: 刘树新, Email: liushuxin11@126.com

indices according to complexity and length of paths<sup>[11]</sup>. Considering the local information including common neighbors and short paths, local indices, such as common neighbor (CN) index<sup>[12]</sup> directly using the number of common neighbors, resource allocation (RA) index<sup>[13]</sup> and Adamic-Adar (AA) index<sup>[14]</sup> weighting the common neighbors by their node degree, local path (LP) index<sup>[15]</sup> and extended resource allocation (ERA)<sup>[16]</sup> index adding paths with length 3 to CN and RA respectively, perform very well in real networks with lower complexity. Ref. [17] proposed a local weighted method, which can improve the prediction accuracy of several similarity indices (including CN, AA, RA, Salton, Jaccard and LP). Similarly, Ref. [18] proposed a local weighted path (LWP) index to differentiate the contributions between paths. Considering the clustering coefficient of nodes, some clustering-based methods<sup>[19-20]</sup> have been proposed to improve the prediction accuracy. Ref. [21] believed that the regularity of a network was reflected in the consistency of structural features before and after a random removal of a small set of links and proposed a universal structural consistency index. To improve the prediction accuracy, many global indices have been proposed based on global paths, including Katz index<sup>[22]</sup>, LHN-II<sup>[23]</sup>, average commute time (ACT)<sup>[24]</sup> and cosine similarity time (Cos+)<sup>[25]</sup>. The global indices can achieve a good performance, but most of them are not suitable for large networks due to their higher complexity. Therefore, how to effectively use the local topological information may be a good way to predict the missing links of large-scale networks.

In the real world, such as human relationship network, if there are more indirect connections between two people and their common friends, they are more likely to be friend. Considering two pairs of endpoints  $x, y$  and  $x', y'$  with a common neighbor. There are two longer paths between  $x', y'$ , and two indirect connections between  $x, y$  and their common neighbor. According to CN, RA, AA and LP index, the similarity between two pairs of nodes are the same, though the links between endpoints  $x$  and  $y$  are more likely to exist. Because there are more indirect connections between  $x, y$  and their common neighbor,

common neighbor of  $x$  and  $y$  is more effective in similarity calculating. In a word, the topology information related to common neighbors plays an important role in link prediction.

Based on the above discussion, an effective common neighbor index (ECN) is proposed, which weights the common neighbors by the local topology information around them. With a parameter adjusting the strength of effective common neighbors, ECN can achieve the optimal prediction accuracy for different networks. Empirical study has shown that the ECN index proposed can significantly improve the prediction accuracy, compared with 9 similarity indices in 15 real networks.

## 1 Baselines

Several classical similarity indices are introduced here, including 5 local indices: CN, RA, AA, CAR and LP index, and 4 global indices: Katz, ACT, Cos+ and SPM index. A brief description is shown as follows:

1) CN index<sup>[12]</sup> measures the similarity of two endpoints by the number of their common neighbors:

$$s_{xy}^{CN} = |\Gamma(x) \cap \Gamma(y)| \quad (1)$$

where  $\Gamma(x)$  is the set of neighbors of node  $x$ , and  $\Gamma(x) \cap \Gamma(y)$  denotes the common neighbors between nodes  $x$  and  $y$ .

2) RA index<sup>[13]</sup> publishes the common neighbors with big degree:

$$s_{xy}^{RA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z} \quad (2)$$

where  $k_z$  is the node degree of node  $z$ .

3) AA index<sup>[14]</sup> weights the common neighbors by nodes degree:

$$s_{xy}^{AA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(k_z)} \quad (3)$$

4) CAR index<sup>[26]</sup> believes that two nodes are more likely to link together if their common-first-neighbors are members of a strongly inner-linked cohort:

$$s_{xy}^{CAR} = |\Gamma(x) \cap \Gamma(y)| \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{|\gamma(z)|}{2} \quad (4)$$

where  $|\gamma(z)|$  refers to the sub-set of neighbors of  $z$  that are also common neighbors of nodes  $x$  and  $y$ .

5) LP index<sup>[15]</sup> adds the path information with

length 3 to CN:

$$\mathbf{S} = \mathbf{A}^2 + \alpha \mathbf{A}^3 \quad (5)$$

where  $\alpha$  is the parameter and  $\mathbf{A}$  is the adjacency matrix.

6) Katz index<sup>[22]</sup> considers all the path information between two endpoints and gives more weights to the shorter paths:

$$s_{xy}^{\text{Katz}} = \sum_{l=1}^{\infty} \varepsilon^l |\text{path}_{xy}^l| \quad (6)$$

where  $\text{path}_{xy}^l$  is the set of paths with length  $l$  between  $x$  and  $y$ , and  $\varepsilon$  is the adjust parameter.

7) ACT<sup>[24]</sup> measures similarity by calculating the average steps of random walk between two endpoints:

$$s_{xy}^{\text{ACT}} = \frac{1}{l_{xx}^+ + l_{yy}^+ - 2l_{xy}^+} \quad (7)$$

where  $l_{xy}^+$  is the corresponding entry in  $\mathbf{L}^+$ , and  $\mathbf{L}^+$  denotes the pseudo-inverse of matrix  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ .

8) Cos<sup>[25]</sup> is based on  $\mathbf{L}^+$  by calculating similarity of two vectors:

$$s_{xy}^{\text{Cos}^+} = \frac{\mathbf{v}_x^T \mathbf{v}_y}{\|\mathbf{v}_x\| \|\mathbf{v}_y\|} = \frac{l_{xy}^+}{\sqrt{l_{xx}^+ l_{yy}^+}} \quad (8)$$

9) SPM<sup>[21]</sup> measures the structural consistency of a network by perturbing its adjacency matrix and observing the change of eigenvalues provided the fixed eigenvectors:

$$s_{xy}^{\text{SPM}} = \sum_{k=1}^N (\lambda_k + \Delta \lambda_k) \mathbf{x}_k \mathbf{x}_k^T \quad (9)$$

where  $\lambda_k$  and  $\mathbf{x}_k$  are the eigenvalue and the corresponding orthogonal and normalized eigenvector for  $\mathbf{A}^R$ , respectively.

## 2 Effective Common Neighbor Index

Considering an unweighted undirected network  $G(V, E)$ ,  $V$  and  $E$  are the sets of nodes and links respectively. A score  $s_{xy}$  is assigned to each pair of nodes  $x$  and  $y$  by prediction algorithms and it can be a measure of similarity between two endpoints. For each nonexistent link of network, this score represents the likelihood that the link exists.

**Definition 1** Considering a pair of nodes,  $x, y \in V$ , node  $z$  is one of their common neighbors. Among all the links connected to node  $z$ , if there are more links on the local paths (here, length  $l \leq 2$ ) between endpoints and node  $z$ , it will be more effective

for common neighbor  $z$  in calculating the similarity of two endpoints. From the influence of paths connected to node  $x$ , the effectiveness of common neighbors between endpoints  $x$  and  $y$  can be defined as:

$$\text{IF}(y|x) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \left( \frac{c_{xz}}{k_z} \right)^\delta \quad (10)$$

where  $c_{xz}$  denotes the number of different paths between node  $x$  and  $z$  with length no more than 2, and  $\delta$  is the adjust parameter.

**Definition 2** On an unweighted undirected network  $G(V, E)$ . The effective common neighbor index of endpoints  $x$  and  $y$ , is defined as:

$$s_{xy}^{\text{ECN}} = \text{IF}(y|x) + \text{IF}(x|y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \left( \frac{c_{xz}^\delta + c_{yz}^\delta}{k_z^\delta} \right) \quad (11)$$

Here, the  $\delta (\delta \geq 0)$  aims to adjust the strength of effectiveness. When  $\delta=0$ , the ECN index is the same as the CN index 12. Take Fig.1 for example, the similarity of two endpoints is  $s_{xy}^{\text{ECN}} = (3^\delta + 2^\delta) / 9^\delta$ .

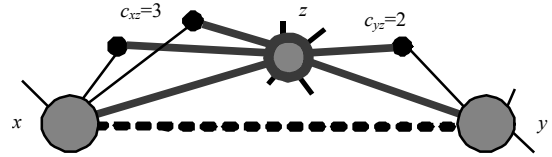


Fig. 1 Effectiveness of common neighbors between two endpoints.

## 3 Metric and Datasets

To quantify the prediction accuracy, the standard metric called precision<sup>[27]</sup> is introduced. Precision is defined as the ratio of relevant links to the top  $L$  predicted links (here we set  $L=100$ ). If there are  $m$  relevant links appeared in the probe set, the precision value is:

$$\text{Precision} = \frac{m}{L} \quad (12)$$

Clearly, the higher precision value means higher prediction accuracy.

Our experiments are performed on fifteen different real networks: 1) Jazz<sup>[28]</sup>: the network of Jazz musicians. 2) Caenorhabditis elegans (CE)<sup>[29]</sup>: the neural network of the nematode worm. 3) Kohonen<sup>[30]</sup>: the citing network of articles with topic “self-organizing maps” or references to “Kohonen T”.

4) USAir<sup>[31]</sup>: the USA airline network. 5) Hamster<sup>[32]</sup>: a friendship network of users on the website hamsterster.com. 6) Metabolic<sup>[33]</sup>: the metabolic network of *Caenorhabditis elegans*. 7) Email network (Email)<sup>[34]</sup>: the internal email communication network between employees of a mid-sized manufacturing company. 8) Yeast PPI (Yeast)<sup>[35]</sup>: a protein-protein interaction network of yeast. 9) Political blogs (PB)<sup>[36]</sup>: a network of US political blogs. 10) Openflights (Flight)<sup>[37]</sup>: the network of flights between air-ports of the world. 11) Scientometrics (SciMet)<sup>[30]</sup>: the citing network of the articles from scientometrics. 12) Food Web of Everglades ecosystem (FWEW)<sup>[38]</sup>: the network of carbon exchanges occurring during the wet season in the Everglades Graminoid Marshes of South Florida. 13) Route topology of internet (Router)<sup>[39]</sup>: the topology of internet in router-level. 14) Power Grid (Power)<sup>[29]</sup>: an electrical power grid network of western US; 15) Food Web of Florida Bay ecosystem (FWFB)<sup>[38]</sup>: the network of carbon exchanges occurring during the wet season in Florida Bay.

**Table 1 Basic topological features of real networks**

Data	$ V $	$ E $	$\langle k \rangle$	$\langle d \rangle$	$r$	$C$	$H$
Jazz	198	2 742	27.7	2.24	0.02	0.633	1.4
CE	297	2 148	14.46	2.46	-0.163	0.308	1.8
Kohonen	3 704	12 683	3.4	5.64	-0.121	0.252	11.2
USAir	332	2 128	12.81	2.74	-0.208	0.749	3.36
Hamster	1 858	12 534	13.49	3.39	-0.085	0.090	3.36
Metabolic	453	2 025	8.94	2.67	-0.226	0.647	4.49
Email	167	5 784	69.26	1.87	-0.295	0.541	1.66
Yeast	2 375	11 693	9.85	5.10	0.469	0.378	3.48
PB	1 222	16 717	27.36	2.74	-0.221	0.361	2.97
Flight	2 939	30 501	20.75	4.18	0.051	0.255	5.22
SciMet	3 084	10 399	6.74	4.18	-0.033	0.151	2.78
FWEW	69	880	25.51	1.64	-0.298	0.552	1.27
Router	5 022	6 258	2.49	5.99	-0.138	0.033	5.5
Power	4 941	6 594	2.67	18.9	0.003	0.107	1.45
FWFB	128	2 075	32.42	1.78	-0.112	0.335	1.24

Table 1 shows the basic topological features of these networks.  $|V|$  is the number of nodes,  $|E|$  denotes the number of links,  $\langle k \rangle$  indicates the average degree.  $\langle d \rangle$  denotes the average distance.  $r$  is the assortativity coefficient<sup>[40]</sup>.  $C$  represents the clustering coefficient<sup>[29]</sup>.  $H$  is the degree heterogeneity.

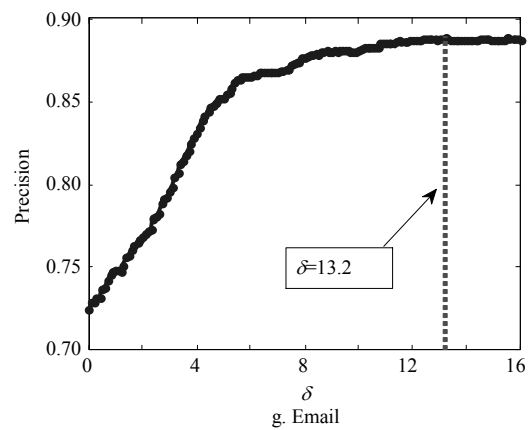
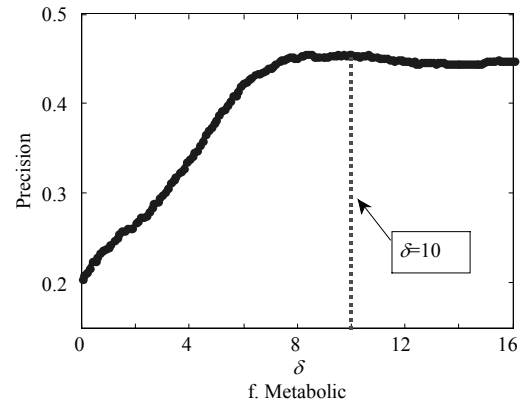
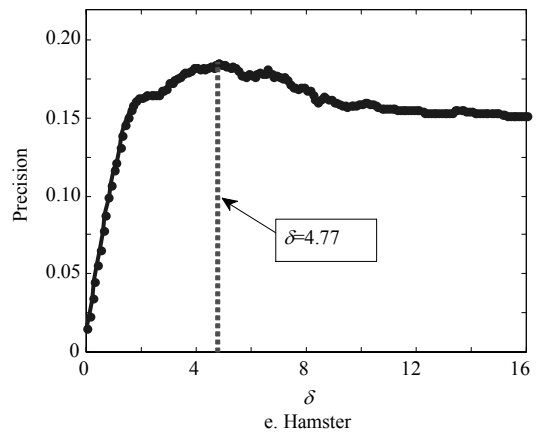
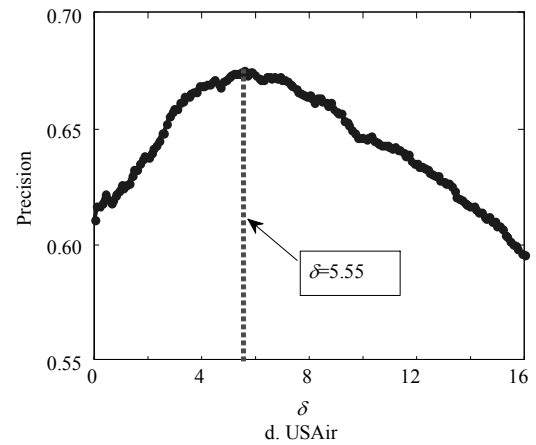
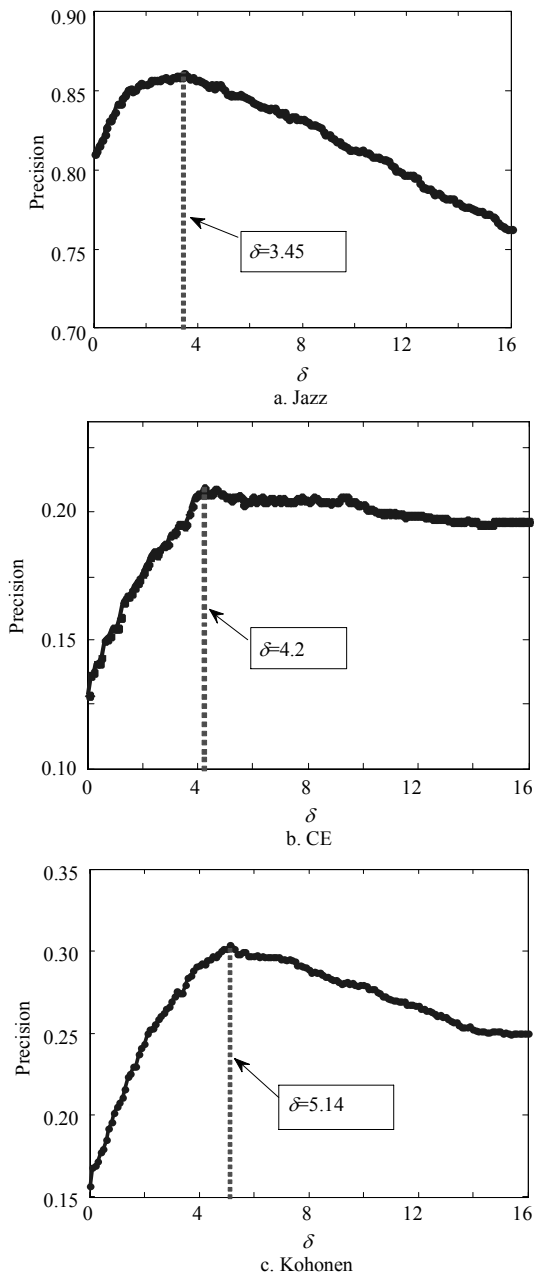
Each original data is randomly divided into training set containing 90% of links, and the probe set containing the remaining 10%.

## 4 Result and Discussion

Firstly, we report the impact of adjust parameter  $\delta$  on prediction accuracy in different networks, and each precision value is the average of 20 realizations. As shown in Fig. 2, all the precision curves can achieve the optimal values when  $\delta > 1$ , and each of them exceeds the precision value at  $\delta = 1$  without enhancement. It demonstrates that higher effective neighbors have a much greater influence on the similarity between endpoints than expected and the contribution of different effective common neighbors for similarity is not linear (for different networks, its power exponent is the optimal  $\delta$ ). The precision values show a rising trend with an increase of  $\delta$  before reaching the highest point, which indicates that the effective common neighbors can improve the prediction accuracy compared with CN (when  $\delta = 0$ ). The measurement of effective common neighbors and their enhancement are both beneficial to the description of similarity between two endpoints. For some of datasets, the precision curves decline gradually after peak point. However, the precision values are stable around the maximum value in some networks such as metabolic, Email, Yeast, FWEW, Router and FWFB. In these networks, the role that the lower effective common neighbors play is very limited in link prediction of missing links.

Table 2 shows the comparison between ECN index (optimal values) and 9 similarity indices in 15 networks, and each precision data is the average of 20 realizations. In 10 out of 15 datasets, ECN index obtains the best performance, but worse than the SPM index in Jazz, Hamster, PB, FWEW and FWFB. For some lower clustering network such as Hamster and Router, the improvement of precision is significant for ECN index, which indicates that common neighbors with different topological structures perform different roles in calculating the similarity between endpoints in these lower clustering networks. With the consideration of longer path information, the precision values of path-based indices (LP and Katz) are higher than neighbor-based indices (CN, RA and AA) in most of dataset. However, in some high clustering network

such as Jazz, USAir, Metabolic and Email, the neighbor-weighted indices (RA and AA) perform better than most of the global indices. Also, the local-community-based index CAR can perform better than some global indices in Jazz, Kohonen, PB, Flight and SciMet. Having considered the effectiveness of common neighbors, ECN index can get a good performance by fusing some topology information used in neighbor-weighted indices and path-based indices. With the strength parameter, ECN index enhance the performance of effectiveness by adjusting the influence of different effective common neighbors



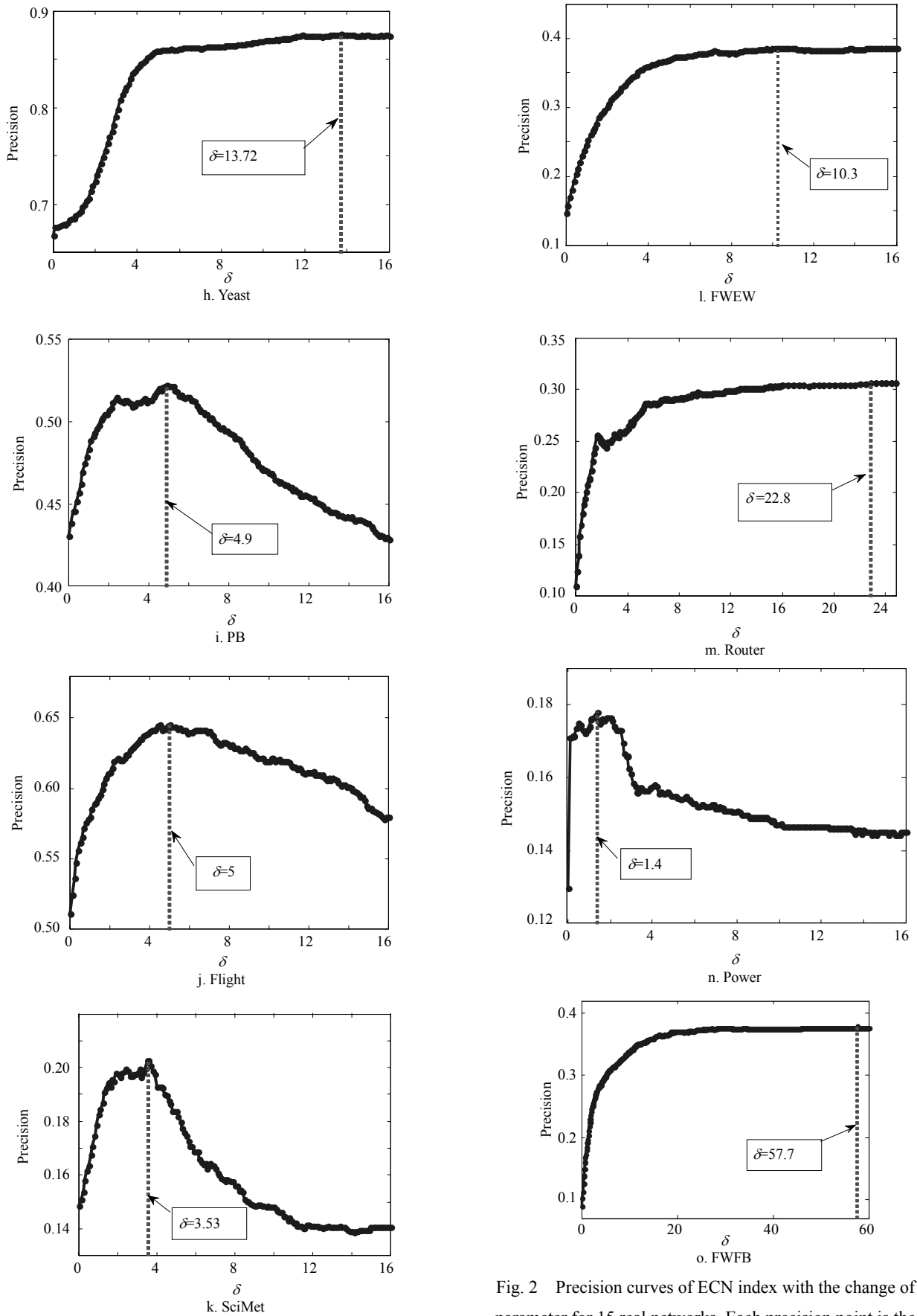


Fig. 2 Precision curves of ECN index with the change of parameter for 15 real networks. Each precision point is the average of 20 realizations.



**Table 2 Comparison of the precision value between ECN (optimal value) and some similarity indices. Each value is the average of 20 realizations, each of which corresponds to an independent division of  $E^T$  and  $E^P$**

Data	CN	RA	AA	CAR	LP <sup>a</sup>	LP <sup>b</sup>	Katz <sup>a</sup>	Katz <sup>b</sup>	ACT	Cos+	SPM	ECN
Jazz	0.810	0.809	0.830	0.851	0.798	0.775	0.798	0.740	0.258	0.356	0.895	<b>0.861</b>
CE	0.129	0.131	0.137	0.129	0.138	0.138	0.138	0.138	0.064	0.069	0.207	<b>0.210</b>
Kohonen	0.157	0.145	0.154	0.211	0.159	0.159	0.159	0.155	0.010	0.020	0.089	<b>0.304</b>
USAir	0.611	0.636	0.632	0.612	0.615	0.612	0.615	0.604	0.492	0.104	0.589	<b>0.676</b>
Hamster	0.015	0.004	0.011	0.037	0.017	0.054	0.017	0.075	0.089	0.020	0.898	<b>0.186</b>
Metabolic	0.206	0.305	0.253	0.195	0.201	0.200	0.201	0.200	0.120	0.110	0.445	<b>0.456</b>
Email	0.724	0.733	0.727	0.691	0.726	0.722	0.726	0.714	0.633	0.637	0.809	<b>0.889</b>
Yeast	0.668	0.495	0.693	0.670	0.671	0.733	0.670	0.721	0.565	0.239	0.699	<b>0.876</b>
PB	0.431	0.253	0.389	0.472	0.436	0.465	0.436	0.467	0.127	0.342	0.669	<b>0.522</b>
Flight	0.511	0.354	0.443	0.631	0.521	0.559	0.522	0.554	0.344	0.030	0.127	<b>0.646</b>
SciMet	0.149	0.130	0.147	0.165	0.149	0.150	0.149	0.147	0.000	0.000	0.079	<b>0.203</b>
FEW	0.147	0.163	0.154	0.142	0.158	0.189	0.158	0.197	0.271	0.000	0.543	<b>0.387</b>
Router	0.110	0.092	0.113	0.109	0.132	0.132	0.131	0.130	0.000	0.019	0.015	<b>0.306</b>
Power	0.130	0.071	0.097	0.131	0.148	0.148	0.146	0.146	0.000	0.086	0.017	<b>0.178</b>
FWFB	0.090	0.087	0.086	0.089	0.096	0.133	0.096	0.146	0.178	0.030	0.718	<b>0.378</b>

<sup>a</sup>Parameter  $\alpha=0.001$ . <sup>b</sup>Parameter  $\alpha=0.01$ .

on similarity in different networks. Having considered the structural consistency of a network, SPM performs very well for all the datasets. Surprisingly, some global indices (ACT and Cos+) perform worse than expected, and their precision values are close to zero in SciMet, Router, and Power networks. To summarize, the common neighbors between endpoints seem to be more effective for similarity calculating under the standard metric precision, and the consideration of local information around common neighbors can significantly improve the precision value.

## 5 Conclusions

There are a lot of similarity indices considering local information between endpoints such as common neighbors and shorter paths. However, most of them ignore the effectiveness of common neighbors. Based on local information, an ECN index is proposed, which weights the common neighbors by the local topology information around them. Empirical results demonstrate that ECN index can significantly improve the prediction accuracy in 15 real networks, compared with 7 similarity indices. The measurement of effective common neighbors and their enhancement are both beneficial to the improvement of prediction accuracy or description of similarity between two endpoints. The

ECN index is more suitable for the link prediction of networks concerning more about the prediction accuracy of the top  $L$  predicted links (such as protein-protein interaction networks). In addition, the complexity of ECN is between  $O(N\langle k \rangle^2)$  (CN) and  $2O(N\langle k \rangle^2)$ . Therefore, it can also be applied to predict missing links of large-scale networks.

## References

- [1] BARABÁSI A L. Scale-free networks: A decade and beyond[J]. *Science*, 2009, 325(5939): 412-413.
- [2] LI K, LI H J, WANG H. Situation analysis of relationship in social networks based on link entropy[J]. *Modern Physics Letters B*, 2015, 29(13): 1550061.
- [3] BULLMORE E, SPORNS O. Complex brain networks: Graph theoretical analysis of structural and functional systems[J]. *Nature Reviews Neuroscience*, 2009, 10(3): 186-198.
- [4] SCHWEITZER F, FAGIOLO G, SORNETTE D, et al. Economic networks: The new challenges[J]. *Science*, 2009, 325(5939): 422-425.
- [5] LIN J, BAN Y. Complex network topology of transportation systems[J]. *Transport Reviews*, 2013, 33(6): 658-685.
- [6] WANG P, XU B W, WU Y R, et al. Link prediction in social networks: The state-of-the-art[J]. *Science China Information Sciences*, 2015, 58(1): 1-38.
- [7] VON M C, JENSEN L J, SNEL B, et al. STRING: Known and predicted protein-protein associations, integrated and transferred across organisms[J]. *Nucleic Acids Research*, 2005, 33(suppl.1): 433-437.



- [8] SCELLATO S, NOULAS A, MASCOLO C. Exploiting place features in link prediction on location-based social networks[C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, CA, USA: ACM, 2011: 1046-1054.
- [9] HOLLAND P W, LASKEY K B, LEINHARDT S. Stochastic blockmodels: First steps[J]. *Social Networks*, 1983, 5(2): 109-137.
- [10] LIU S, XU X F, XIE R G, et al. Anomalous heat conduction and anomalous diffusion in low dimensional nanoscale systems[J]. *The European Physical Journal B*, 2012, 85(10): 337.
- [11] LÜ L, ZHOU T. Link prediction in complex networks: A survey[J]. *Physica A: Statistical Mechanics and its Applications*, 2011, 390(6): 1150-1170.
- [12] LORRAIN F, WHITE H C. Structural equivalence of individuals in social networks[J]. *Social Networks*. 1977(1): 67-98.
- [13] ZHOU T, LÜ L, ZHANG Y C. Predicting missing links via local information[J]. *The European Physical Journal B*, 2009, 71(4): 623-630.
- [14] ADAMIC L A, ADAR E. Friends and neighbors on the web[J]. *Social Networks*, 2003, 25(3): 211-230.
- [15] LÜ L, JIN C H, ZHOU T. Similarity index based on local paths for link prediction of complex networks[J]. *Physical Review E*, 2009, 80(4): 046122.
- [16] LIU S, JI X, LIU C, et al. Extended resource allocation index for link prediction of complex network[J]. *Physica A: Statistical Mechanics and its Applications*, 2017, 479: 174-183.
- [17] LIU S, JI X, LIU C, et al. Similarity indices based on link weight assignment for link prediction of unweighted complex networks[J]. *International Journal of Modern Physics B*, 2017, 31(2): 1650254.
- [18] YAO Y, ZHANG R, YANG F, et al. Link prediction based on local weighted paths for complex networks[J]. *International Journal of Modern Physics C*, 2017, 28(4): 1750053.
- [19] MA Y, CHENG G, LIU Z, et al. Clustering-based link prediction in scientific coauthorship networks[J]. *International Journal of Modern Physics C*, 2017, 28(06): 1750082.
- [20] FENG X, ZHAO J C, XU K. Link prediction in complex networks: A clustering perspective[J]. *The European Physical Journal B*, 2012, 85(1): 1-3.
- [21] LÜ L, PAN L, ZHOU T, et al. Toward link predictability of complex networks[J]. *Proceedings of the National Academy of Sciences*, 2015, 112(8): 2325-2330.
- [22] KATZ L. A new status index derived from sociometric analysis[J]. *Psychometrika*, 1953, 18(1): 39-43.
- [23] LEICHT E A, HOLME P, NEWMAN M E J. Vertex similarity in networks[J]. *Physical Review E*, 2006, 73(2): 026120.
- [24] KLEIN D J, RANDIĆ M. Resistance distance[J]. *Journal of Mathematical Chemistry*, 1993, 12(1): 81-95.
- [25] FOUSS F, PIROTTE A, RENDERS J M, et al. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2007, 19(3): 355-369.
- [26] CANNISTRACI C V, ALANIS-LOBATO G, RAVASI T. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks[J]. *Scientific Reports*, 2013, 3: 1613-1618.
- [27] HERLOCKER J L, KONSTAN J A, TERVEEN L G, et al. Evaluating collaborative filtering recommender systems[J]. *ACM Transactions on Information Systems (TOIS)*, 2004, 22(1): 5-53.
- [28] GLEISER P M, DANON L. Community structure in jazz[J]. *Advances in Complex Systems*, 2003, 6(4): 565-573.
- [29] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393(6684): 440-445.
- [30] ZHANG Q M, LÜ L, WANG W Q, et al. Potential theory for directed networks[J]. *PloS One*, 2013, 8(2): e55437.
- [31] BATAGELJ V, MRVAR A. Pajek-program for large network analysis[J]. *Connections*, 1998, 21(2): 47-57.
- [32] LÜ L, PAN L, ZHOU T, et al. Toward link predictability of complex networks[J]. *Proceedings of the National Academy of Sciences*, 2015, 112(8): 2325-2330.
- [33] DUCH J, ARENAS A. Community detection in complex networks using extremal optimization[J]. *Physical Review E*, 2005, 72(2): 027104.
- [34] MICHALSKI R, PALUS S, KAZIENKO P. Matching organizational structure and social network extracted from email communication[C]//International Conference on Business Information Systems. Berlin, Heidelberg: Springer, 2011: 197-206.
- [35] BU D, ZHAO Y, CAI L, et al. Topological structure analysis of the protein-protein interaction network in budding yeast[J]. *Nucleic Acids Research*, 2003, 31(9): 2443-2450.
- [36] ADAMIC L A, GLANCE N. The political blogosphere and the 2004 US election: Divided they blog[C]//Proceedings of the 3rd International Workshop on Link Discovery. Chicago, USA: ACM, 2005: 36-43.
- [37] OPSAHL T, AGNEESSENS F, SKVORETZ J. Node centrality in weighted networks: Generalizing degree and shortest paths[J]. *Social Networks*, 2010, 32(3): 245-251.
- [38] ULANOWICZ R E, DEANGELIS D L. Network analysis of trophic dynamics in south florida ecosystems[J]. *US Geological Survey Program on the South Florida Ecosystem*, 2005, 114: 45-48.
- [39] SPRING N, MAHAJAN R, WETHERALL D. Measuring ISP topologies with Rocketfuel[J]. *ACM SIGCOMM Computer Communication Review*, 2002, 32(4): 133-145.
- [40] NEWMAN M E J. Assortative mixing in networks[J]. *Physical Review Letters*, 2002, 89(20): 208701.