

Research Article

Link Prediction in Complex Network via Penalizing Noncontribution Relations of Endpoints

Xuzhen Zhu,¹ Yang Tian,² and Hui Tian¹

¹ State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

² School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, China

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

Received 21 July 2014; Revised 10 September 2014; Accepted 24 September 2014; Published 16 November 2014

Academic Editor: Bo Shen

Copyright © 2014 Xuzhen Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Similarity based link prediction algorithms become the focus in complex network research. Although endpoint degree as source of influence diffusion plays an important role in link prediction, some noncontribution links, also called noncontribution relations, involved in the endpoint degree serve nothing to the similarity between the two nonadjacent endpoints. In this paper, we propose a novel link prediction algorithm to penalize those endpoints' degrees including many null links in influence diffusion, namely, noncontribution relations penalization algorithm, briefly called NRP. Seven mainstream baselines are introduced for comparison on nine benchmark datasets, and numerical analysis shows great improvement of accuracy performance, measured by the Area Under roc Curve (AUC). At last, we simply discuss the complexity of our algorithm.

1. Introduction

Many systems can be properly described as complex networks with nodes representing individuals or organizations and links mimicking the relations among them [1–3]. Link prediction as a critical problem of forecasting potential relations between nonadjacent nodes meaningfully provides strong ability to explore the evolution of network and unveils the development mechanism of network, attracting so many researchers' interests [4]. In real-world applications, link prediction gives help to significantly discover potential network links [5], bring in many striking applications of recommending friends in online social networks [6], explore protein-to-protein interactions [7], reconstruct airline networks [8], and boost e-commerce scales [9, 10].

Most conventional methods model the task of link prediction in the form of estimating the probability that two nonadjacent nodes would be linked, which is believed to be positively correlated with the similarity between them [11]. Mainstream methods take into account topological similarity based on network structures and can be classified into three major classes [11]. The first class calculates topological

similarity with global structural information, such as the Katz Index that counts all paths connecting two nodes with shorter paths preferred [12]. Such global indices show fair performance in prediction but suffer from high computational complexity [11]. The second class, defined on local structures, typically includes *Common Neighbors* (CN) Index [13] counting the number of common nodes, *Preferential Attachment* (PA) Index [14] preferring the links between endpoints of high degrees, and the methods penalizing high-degree common neighbors, such as *Adamic-Adar* (AA) Index [15], *Resource Allocation* (RA) Index [16], *Optimized AA* (OAA) Index, *Optimized RA* (ORA) Index [17], and *Centrality based Index* (DC-CN, BD-CN, and CC-CN) [18]. In spite of successfully reducing the computational expense, local indices suffer from relatively poor prediction performance [11]. In order to find a nice tradeoff between performance and complexity, the third class of similarity indices is proposed on quasilocal structures. The *Local Path* (LP) Index ignores long-path terms in Katz Index [19], and its bounded version (BLP) relates local paths in an elaborate way [20]. The *Significant Path* (SP) Index comprehensively considers paths and intermediate node degree [21]. The *Local Random Walk* (LRW)

Index limits a random walk within a local range [22], while the *Superposed Random Walk* (SRW) Index continuously releases a random walker at the starting node to emphasize the nodes near the target node [22].

Influence of a node plays a very important role in link prediction, so that a node with powerful influence can attract and impact on other nodes in a network, such as a very important person (VIP) in a social network, possessing strong power to transmit his/her reputation to attract others to follow. In a network, a node with high degree is believed to have more powerful influence and more likely to arise links in the future. However, when carefully examining the node degree influence in link prediction, we find that not all the influence is significant to the counterpart endpoint. In other words, there are noncontribution relations in node degree influence. For example, in Figure 1, we define the influence degree as the number of endpoint relations, valid influence degree as the number of endpoint relations which can eventually reach the counterpart endpoint, and connectivity chance as the proportion the valid influence occupies in the total amount of an endpoint, featuring the total valid amount of influence emanating from the endpoint. Despite an apparently high influence degree of 3 in subgraph (b), v_4 and v_6 in fact possess valid influence degree only of 1 and connectivity chance only accounting for 1/3 because of two noncontribution relations, and, in contrast, v_1 and v_3 although just with lower influence degree of 1 cast all influence degree to each other and have greater connectivity chance of 1. Furthermore, in subgraph (c), v_7 and v_{11} obtain the highest valid influence degree of 3 and also greater connectivity chance of 1 without any useless relations as v_1 and v_3 in subgraph (a). Therefore, we reasonably deduce the link likelihood order (c) > (a) > (b), according to the valid influence degree and connectivity chance. From the theoretical analysis and real illustration, we extract a finding that obviously the high chance of connectivity can enhance link likelihood and on the contrary the noncontribution relations would harm it.

In this paper, we propose a novel similarity index considering penalization on noncontribution relations with a tunable penalization factor β and random walk based valid influence diffusion, namely, noncontribution relations penalization (NRP) index. In contrast with recently published SP Index [21] and also a link prediction algorithm, there are obvious differences: NRP focuses on the penalization on noncontribution relations in endpoints and meanwhile adds bidirectional valid influence diffusions together to form the model, but SP just considers single direction paths and focuses on the heterogeneity of different paths concerning the intermediate nodes without considering the endpoints. Besides novelty, compared with seven mainstream baselines via experiments on nine benchmark datasets, numerical analysis shows a great improvement on NRP.

The rest of the paper is organized as follows: in Section 2, the new model based on penalization on noncontribution relations and significant connectivity is introduced; in Sections 3 and 4, the experimental materials of nine benchmark datasets and methods including metrics and seven mainstream baselines are described, respectively; we present

the results and discussions in Section 5 and finally make a conclusion.

2. Method

An undirected network $G(V, E)$ is considered, where V and E stand for the sets of nodes and links, respectively. Multiple links and self-connections are not allowed. For each pair of nodes, $x, y \in V$, every algorithm referred to in this paper assigns a score s_{xy} . This score can be viewed as a measure of similarity between nodes x and y , and hereinafter we do not distinguish similarity and score. All the nonexistent links are sorted in decreasing order according to their scores, and the links at the top are most likely to exist. To test the algorithm's accuracy, the set of links E is randomly divided into two parts: the training set E^T is treated as known information, while the testing set E^P is used for testing and no information in this set is allowed to be used for training. Clearly, $E^T \cup E^P = E$ and $E^T \cap E^P = \emptyset$.

As described in the introduction and Figure 1, the probability of link between two nonadjacent endpoints is based on the similarity established via considering valid influence diffusion which can be regarded as a resource diffusion process [22]. Here, for the sake of emphasizing the common transferring capability in bidirectional resource diffusion, we first extract the intermediate connectivity C ignoring the transfer probability values $1/k_x$ and $1/k_y$ emanating from the endpoints in Definition 1 as follows.

Definition 1. On an undirected unweighted network $G(V, E)$, one defines the intermediate connectivity C of a l -step path $q = \{v_0 = x, v_1, \dots, v_{l-1}, v_l = y\}$; connecting x and y equals the product of transfer probability from v_1 to v_l or from v_{l-1} to v_0 ignoring the transfer probability from the endpoint, as

$$C(x, y)|_l^j = \prod_{i=1}^{l-1} P(v_{i+1} | v_i) = \prod_{i=1}^{l-1} P(v_{l-1-i} | v_{l-i}), \quad (1)$$

where $P(v_{i+1} | v_i) = 1/k(v_i)$ is the transfer probability from v_i to v_{i+1} with $k(v_i)$ denoting the degree of node v_i .

Although the numbers of links are different in networks, the total resources in different networks are the same, simply set as one, necessarily normalizing resource of endpoint as $k_x/2|E|$ with k_x denoting the influence degree of endpoint x and $|E|$ denoting the number of links [22]. And, importantly, the more the quantity of the valid resource from one endpoint can be delivered to the counterpart, the more similar the two endpoints are. Furthermore, we add the bidirectional resource diffusion quantities on paths with length from 2 to t together and simplify the ultimate formalism in Definition 2 as our similarity index—noncontribution relations penalization (NRP) index.

Definition 2. On an undirected unweighted network $G(V, E)$, with the initial resource usually assigned according to the importance of nodes, here, one simply sets the initial resource of node x proportional to its normalized degree $k_x/2|E|$ [22]. The link prediction similarity index $\text{sim}(x, y)|_l$ on

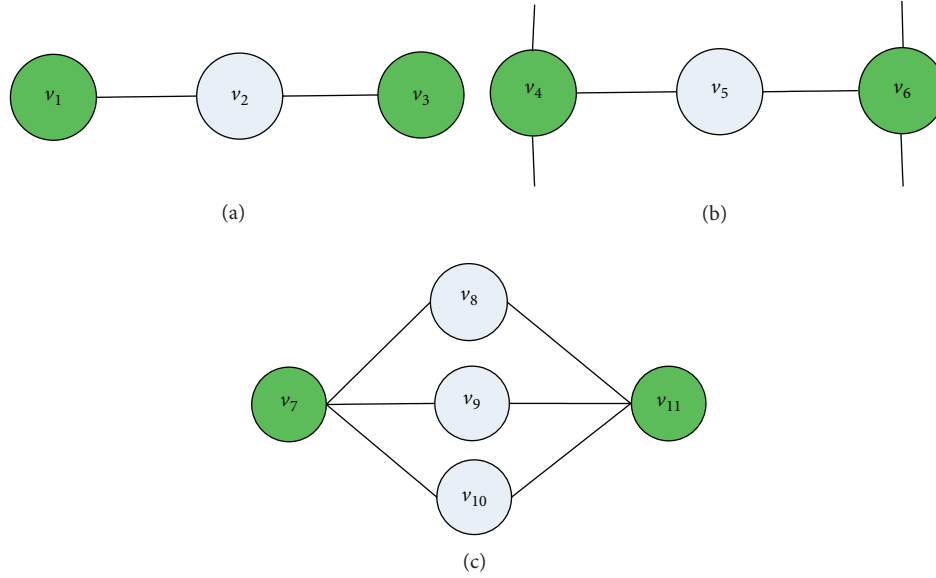


FIGURE 1: Illustration of noncontribution relations and valid influence diffusion. With the same valid influence degree of 1 and smaller connectivity chance of $1/3$ than 1 in subgraph (a), link likelihood in (b) is less than (a) because of two noncontribution relations attached on v_4 and v_6 ; meanwhile, subgraph (c) with the highest valid influence degree of 3 and the same connectivity chance in (a) obtains the greatest link likelihood. Totally, we rank (c) > (a) > (b) according to reasonable link likelihood.

a path $q = \{v_0 = x, v_1, \dots, v_l = y\}$ equals the sum of product of normalized and penalized node degrees and transfer probability between two ends in both directions from x to y and from y to x , as

$$\begin{aligned} \text{sim}(x, y)|_l &= \sum_{j=1}^{N(l)} \left[\left(\frac{k_x}{2|E|} \right)^\beta \frac{1}{k_x} C(x, y)|_l^j \right. \\ &\quad \left. + \left(\frac{k_y}{2|E|} \right)^\beta \frac{1}{k_y} C(x, y)|_l^j \right] \\ &= \sum_{j=1}^{N(l)} \left(\frac{1}{2|E|} \right)^\beta C(x, y)|_l^j \left[(k_x)^{\beta-1} + (k_y)^{\beta-1} \right]. \end{aligned} \quad (2)$$

After merging, we obtain the processed equation, where β is penalty factor in $[0, +\infty)$ and $N(l)$ is the number of paths of length l , $C(x, y)$ represents the intermediate connectivity deleting the impact from the endpoint, and $[(k_x)^{\beta-1} + (k_y)^{\beta-1}]$ emphasizes the depressed influence of endpoints. At last, we obtain the NRP index of summation on path lengths from 2 to t , saying that

$$s_{xy}^{\text{NRP}}(t) = \sum_{l=2}^t \text{sim}(x, y)|_l. \quad (3)$$

Since paths longer than three cost expensive computations contribute but little for predicting links, we just consider the paths with $l = 2$ and 3 in practice and later experiments [19, 22].

3. Experiments Data

Experiments are performed on nine real-world networks (datasets are freely downloaded from the following academic websites: <http://vlado.fmf.uni-lj.si/pub/networks/data>, <http://wiki.gephi.org/index.php?title=Datasets>, <http://lovro.lpt.fri.uni-lj.si/support.jsp>, <http://konect.uni-koblenz.de/networks/>, and <http://www.linkprediction.org/index.php/link/resource/data>). We converted arcs into undirected links and removed loops and multilinks to make them simple networks. (i) Network US Air97 (USAir) [23] is the network of the US air transportation system. (ii) Network Yeast PPI (Yeast) [24] is the protein-protein interaction network of yeast. (iii) Network NetScience (NS) [25] is the network of coauthorships between scientists publishing on the topic of networks. (iv) Network Jazz [26] is the network of Jazz musicians. (v) Network *C.elegans* (CE) [27] is the neural network of the nematode worm *C. elegans*. (vi) Network Slovka [28] is the Facebook friendship network of Slavko Žitnik. (vii) Network Email (Email) [29] is the email communication network of University of Rovira i Virgili (URV) in Tarragona, Spain. (viii) Network Infectious (Infec) [30] is the face-to-face contact network of people during the exhibition “Infectious: Stay Away” in 2009 at the Science Gallery in Dublin. (ix) Network EuroSiS (ES) [31] is the mapping network between Science in Society actors on the Web of 12 European countries. Table 1 reports the basic topological features of these networks.

Each dataset is randomly divided into a training set E^T containing 90% links and a testing set E^P containing the remaining 10% links.

TABLE 1: The basic topological features of the nine benchmark networks. $|V|$ denotes the number of nodes, $|E|$ is the number of links, $\langle k \rangle$ represents the average degree, $\langle d \rangle$ denotes the average distance, C indicates the clustering coefficient [27], r indicates the assortativity coefficient [33], and H is the degree heterogeneity, 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
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
CE	453	2025	8.94	2.66	0.655	-0.225	4.49
Slavko	334	2218	13.28	3.05	0.488	0.247	1.62
Email	1133	5451	9.62	3.61	0.254	0.078	1.94
Infec	410	2765	13.49	3.63	0.467	0.226	1.39
ES	1272	6454	10.15	3.86	0.382	-0.012	2.46

4. Experiments Metrics

Area under curve (AUC) [32], an accurate metric, can be interpreted as the probability that a randomly chosen missing link (a link in E^P) is given a higher score than a randomly chosen nonexistent link (a link in $U \setminus E$, where E denotes the universal link set). In the implementation, among n independent comparisons, if there are n' times the missing link has a higher score and n'' times they are the same score, AUC can be calculated as follows:

$$\text{AUC} = \frac{n' + 0.5n''}{n}. \quad (4)$$

AUC estimates the accuracy of the index globally, with the significance that if all the scores generated from independent and identical distribution, the accuracy should be about 0.5. Therefore, the degree to which the accuracy exceeds 0.5 indicates how much better the algorithm performs than pure chance.

4.1. *Baselines.* For comparison, we introduce six classical algorithms as follows.

- (1) Common Neighbors (CN) Index [13]: considering if two endpoints are similar, they may have many common neighbors with definition by calculating the number of common neighbors:

$$s_{xy}^{\text{CN}} = |\Gamma(x) \cap \Gamma(y)|, \quad (5)$$

where $\Gamma(x)$ is the set of neighbor nodes belonging to endpoint x and $\Gamma(x) \cap \Gamma(y)$ denotes the set of common neighbors of endpoints x and y .

- (2) Preferential Attachment (PA) Index [14]: believing the probability that the new link will connect x and y is proportional to $k_x \times k_y$:

$$s_{xy}^{\text{PA}} = k_x \times k_y, \quad (6)$$

where k_x and k_y denote the degree of node x and y , respectively.

- (3) Adamic-Adar (AA) Index [15]: based on CN, punishing the common neighbors with high degree by considering the logarithm of reciprocal of common neighbors' degrees:

$$s_{xy}^{\text{AA}} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(k_z)}. \quad (7)$$

- (4) Resource Allocation (RA) Index [16]: similar to AA, punishing the common neighbors with high degree just by considering the reciprocal of common neighbors' degrees:

$$s_{xy}^{\text{RA}} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z}. \quad (8)$$

- (5) Local Path (LP) Index [16, 19]: additionally counting the contribution of local paths with length 3 as follows:

$$S^{\text{LP}} = A^2 + \varepsilon A^3, \quad (9)$$

where A is the adjacency matrix and ε is a free parameter.

- (6) Bounded Local Path (BLP) Index [20]: bounding local paths with structural coefficients according to path lengths:

$$s_{xy}^{\text{BLP}} = \sum_{i=2}^{l_{\max}} \frac{1}{i-1} \cdot \frac{|P_i(v_x, v_y)|}{\prod_{j=2}^i (N-j)}, \quad (10)$$

where l_{\max} indicates the maximum length under consideration and $|P_i(v_x, v_y)|$ is the number of all paths connecting v_x and v_y with length i .

- (7) Superposed Random Walk (SRW) Index [22]: considering superposition of local paths less than t steps and the degrees of two endpoints as follows:

$$s_{xy}^{\text{SRW}}(t) = \sum_{\tau=1}^t [q_x \pi_{xy}(\tau) + q_y \pi_{yx}(\tau)], \quad (11)$$

where $q_x = k_x/2|E|$ and $\pi_{xy}(\tau)$ denote normalized initial resource and transfer probability from node x to y , respectively.

TABLE 2: Prediction accuracy measured by AUC values on the nine benchmark networks. Each data point is an average over 10 independent runs, each of which corresponds to a random 90%–10% division of training set and testing set. All the present results are optimal values. Numbers in brackets stand for the standard deviations.

AUC	CN	PA	AA	RA	LP	BLP	SRW	NRP
USAir	0.937 (0.008)	0.885 (0.017)	0.948 (0.007)	0.954 (0.007)	0.938 (0.007)	0.931 (0.010)	0.953 (0.011)	0.957 (0.011)
Yeast	0.723 (0.005)	0.494 (0.008)	0.724 (0.005)	0.723 (0.005)	0.732 (0.005)	0.733 (0.004)	0.735 (0.005)	0.736 (0.005)
NS	0.940 (0.011)	0.679 (0.012)	0.940 (0.011)	0.940 (0.011)	0.940 (0.011)	0.943 (0.009)	0.943 (0.009)	0.944 (0.009)
Jazz	0.953 (0.005)	0.762 (0.013)	0.961 (0.004)	0.970 (0.004)	0.954 (0.005)	0.951 (0.005)	0.960 (0.004)	0.970 (0.004)
CE	0.914 (0.011)	0.808 (0.021)	0.948 (0.010)	0.954 (0.010)	0.914 (0.011)	0.911 (0.007)	0.953 (0.009)	0.960 (0.009)
Slavko	0.941 (0.009)	0.775 (0.013)	0.945 (0.009)	0.946 (0.010)	0.944 (0.010)	0.943 (0.010)	0.949 (0.009)	0.952 (0.009)
Email	0.844 (0.007)	0.782 (0.007)	0.846 (0.007)	0.846 (0.007)	0.893 (0.006)	0.902 (0.005)	0.903 (0.007)	0.908 (0.006)
Infec	0.939 (0.009)	0.703 (0.013)	0.942 (0.009)	0.943 (0.009)	0.954 (0.012)	0.958 (0.006)	0.964 (0.006)	0.966 (0.006)
ES	0.910 (0.005)	0.820 (0.007)	0.912 (0.006)	0.912 (0.006)	0.936 (0.007)	0.938 (0.005)	0.945 (0.005)	0.946 (0.005)

5. Results and Discussions

Because of small contribution and expensive cost of computation on long paths, experiments on NRP just considering $t = 3$ in nine networks are implemented (referenced in [19, 22]). After obtaining AUC values experiments, in Section 5.1, we demonstrate important meaning of penalizing noncontribution relations in NRP in Figure 2 of nine subgraphs corresponding to nine different datasets. Furthermore, performance explanations about the nine subgraphs will be offered with different values of penalization factor β when searching for optimal values in $[0, 5]$ which is wide enough to find the optimal values. Then, we compare our algorithms with CN, PA, AA, RA, LP, BLP, and SRW via Table 2. At last, the computational complexity is discussed in Section 5.3.

5.1. Performances of Penalization on Noncontribution Relations. There are nine subgraphs in Figure 2 about AUC of NRP showing functions of penalization on noncontribution relations of endpoint degrees with penalization factor β varying in the range $[0, 5]$, measured by average AUC under 10 independent runs obeying random divisions of training sets and testing sets. The β -axis represents the penalization degree on endpoint degrees. According to (2) and (3), $\beta < 1$, that is, $\beta - 1 < 0$, suggests endpoints with noncontribution relations will be penalized and the smaller the β is, the more severely the noncontribution relations are suppressed, and vice versa. In Figure 2, the optimal β 's of all nine datasets locate between 0 and 1 when AUC curves reach the peaks, say, USAir at 0.38, Yeast at 0.77, NS at 0.86, CE at 0.11, Jazz at 0.36, Infec at 0.5, Slavko at 0.63, Email at 0.84, and ES at 0.63, which means CE suffers the most penalization and NS suffers the least. And when $\beta > 1$, the AUC performances decrease very severely compared with the optimal AUCs. The $\beta < 1$ consistently happening in nine various kinds of datasets means indeed that the noncontribution relations of endpoint degrees are penalized, and the performance will get worse if they are promoted when $\beta > 1$.

5.2. Comparison on AUC. To demonstrate the prediction ability, we report the performances of NRP index with

the optimal β values on nine datasets, respectively. Table 2 reports the average AUC values of NRP and baselines. The results in the table are all optimal and the best values are emphasized in bold font. NRP achieves the best performances in all nine datasets (see boldface in Table 2). Notice that those datasets represent different kinds of networks with heterogeneous topological features (see Table 1) and disparate organization principles; the comparison highlights that NRP works well consistently on various situations.

Analyzing the difference in performance between NRP and baselines, we realize that it is the penalization on large-degree nodes with many noncontribution relations and emphasis on connectivity based on random walk that explain the difference. PA simply assumes that only if the higher the degrees the two endpoints have, the more similar the two endpoints are, ignoring the existence of noncontribution relations in degrees of endpoints and the significant connectivity between endpoints, leading to the poorest accuracy performance, especially in Yeast. Instead, CN considers the connectivity via counting the number of common neighbors on 2-hop paths, but ignoring influence of endpoint degrees and connectivity of long paths, resulting in an improved but still worse performance. Further, AA and RA extend CN by similarly penalizing intermediate large-degree nodes and not surprisingly obtain better performance than CN. However, as the same as CN, ignorance of influence of endpoint degrees and long paths cause AA and RA to still work dissatisfiedly such as in Yeast. In contrast, LP and BLP take long paths into account and thus outperform on many networks such as in ES. However, lacking consideration of influence of endpoint degree, LP and BLP meet difficulty in accurately predicting missing links in contrast to NRP, such as in USAir. An exception for BLP is its worse performance especially than CN in USAir, Jazz, and CE because of inelegant formalism of coefficient which cannot hit an optimal value to help it overwhelmingly defeat CN. SRW, with consideration of influence of high endpoint degree and of long paths based on random walk, further improves the performance in contrast to the former algorithms, but neglect of penalization on endpoint degree with noncontribution relations preventing it from better prediction performance. Above all, NRP, outstandingly leveraging

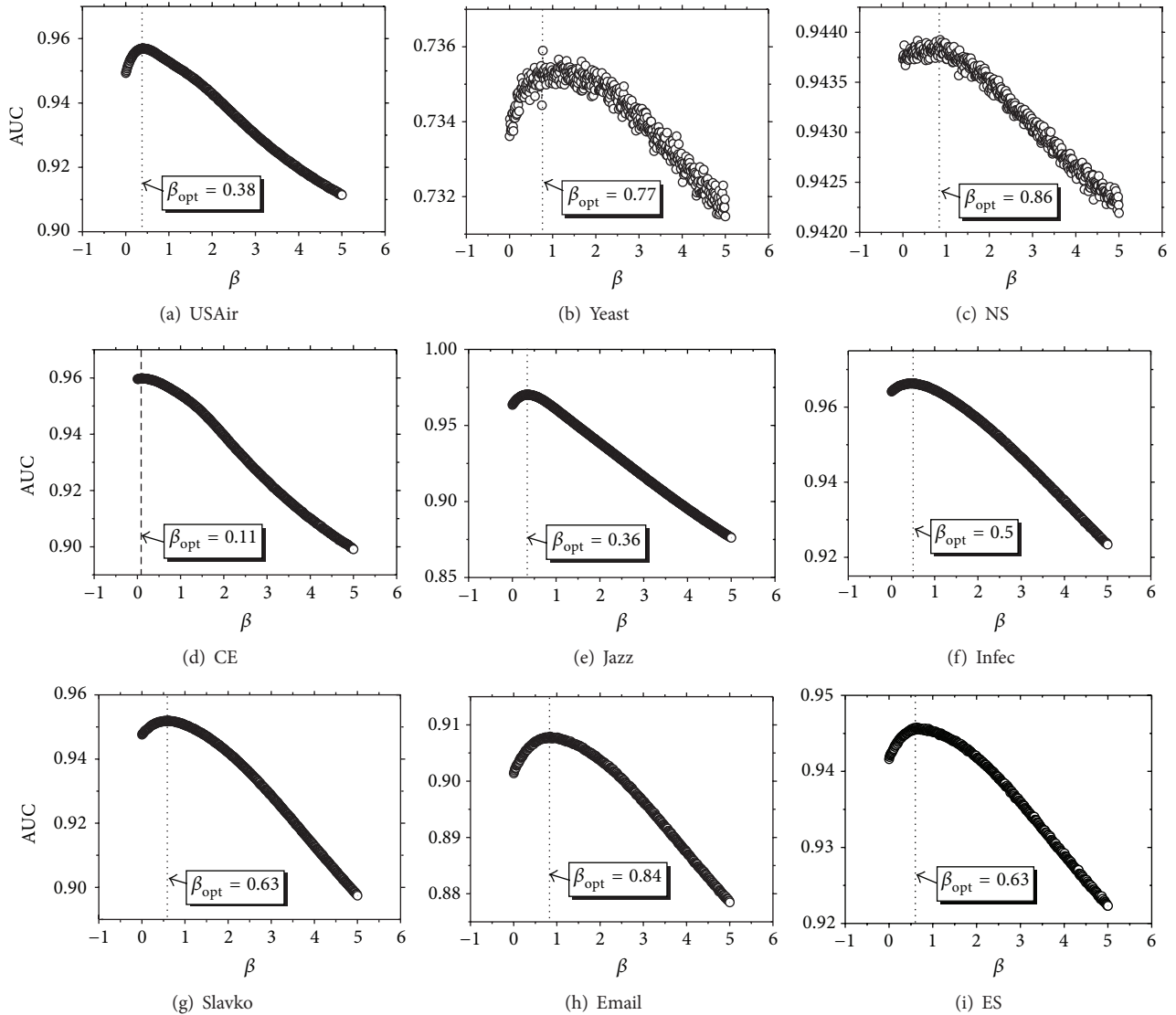


FIGURE 2: AUC of NPR, respectively, on nine different datasets with β in $[0, 5]$. The optimal AUC is obtained at β_{opt} . All AUC values are obtained by averaging over ten independently random divisions.

the valid influence of endpoint degree with penalization on noncontribution relations and connectivity based on random walk, outperforms all other mainstream algorithms on nine datasets.

5.3. Complexity Discussion. In application of algorithm, the low computation complexity is another very important concern in the design of prediction algorithm. As we know, the time complexity of product of two $N \times N$ matrices is $O(N^3)$. From the definitions of CN, PA, AA, RA, LP, BLP, and SRW, PA has the time complexity of $O(N^2)$, CN, PA, AA, and RA have the time complexity of $O(N^3)$, and LP, BLP, and SRW are of the complexity $M \times O(N^3)$ with coefficient $M \ll N^3$. In contrast, although with the same time complexity of $M \times O(N^3)$ of LP, BLP, and SRW, more than CN, PA, AA, and RA, our index shows stronger performance than them all.

Above all, our index, NPR, achieves the best performance with little increase in complexity.

6. Conclusions

In research on topological similarity based link prediction, we find that not all relations involved in the degree of endpoint make contribution to valid influence diffusion due to the existence of noncontribution relations. Accordingly, we expect, in case of penalizing the noncontribution relations in endpoints degree with a penalization factor β in the valid influence diffusion process, the link prediction performance could be improved. Therefore, in this paper a model named NPR based on noncontribution relations penalization of endpoint degree and random walk based valid influence diffusion is proposed. To detect our model, many experiments are implemented on

nine benchmark networks and the results compared with the CN, PA, AA, RA, LP, BLP, and SRW confirm our expectation that penalization on noncontribution relations of endpoint degree could greatly improve the accuracy performance in link prediction.

And due to the various structures and properties of experimental datasets, it is obvious that our model can be used in many applications of link prediction, such as transportation planning, biological reactions, friend recommendation in social network, and disease prevention.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 61471060), the National High Technology Research and Development Program of China (863 Program) (no. 2014AA01A706), the Funds for Creative Research Groups of China (no. 61121001) and BUPT Excellent Ph.D. Students Foundation (no. CX201433).

References

- [1] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, article 47, 2002.
- [2] M. E. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [4] L. Getoor and C. P. Diehl, "Link mining: a survey," *ACM SIGKDD Explorations Newsletter*, vol. 7, no. 2, p. 3, 2005.
- [5] W.-Q. Wang, Q.-M. Zhang, and T. Zhou, "Evaluating network models: a likelihood analysis," *EPL*, vol. 98, no. 2, Article ID 28004, 2012.
- [6] S. Scellato, A. Noulas, and C. Mascolo, "Exploiting place features in link prediction on location-based social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*, pp. 1046–1054, ACM, August 2011.
- [7] 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, article 1613, 2013.
- [8] R. Guimerà and M. Sales-Pardo, "Missing and spurious interactions and the reconstruction of complex networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 52, pp. 22073–22078, 2009.
- [9] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Physics Reports*, vol. 519, no. 1, pp. 1–49, 2012.
- [10] X. Z. Zhu, H. Tian, and S. M. Cai, "Personalized recommendation with corrected similarity," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2014, Article ID P07004, 2014.
- [11] L. L. Linyuan and T. Zhou, "Link prediction in complex networks: a survey," *Physica A*, vol. 390, no. 6, pp. 1150–1170, 2011.
- [12] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.
- [13] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Physical Review E*, vol. 64, no. 2, Article ID 025102, 2001.
- [14] Y.-B. Xie, T. Zhou, and B.-H. Wang, "Scale-free networks without growth," *Physica A: Statistical Mechanics and Its Applications*, vol. 387, no. 7, pp. 1683–1688, 2008.
- [15] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," *Social Networks*, vol. 25, no. 3, pp. 211–230, 2003.
- [16] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information," *European Physical Journal B*, vol. 71, no. 4, pp. 623–630, 2009.
- [17] X. Zhu, H. Tian, Z. Hu, and H. Liu, "Self-adaptive optimized link prediction based on weak ties theory in unweighted network," in *Proceedings of the International Conference on Social Computing (SocialCom '13)*, pp. 896–900, September 2013.
- [18] H. Liu, Z. Hu, H. Haddadi, and H. Tian, "Hidden link prediction based on node centrality and weak ties," *Europhysics Letters*, vol. 101, no. 1, Article ID 18004, 2013.
- [19] L. Lü, C.-H. Jin, and T. Zhou, "Similarity index based on local paths for link prediction of complex networks," *Physical Review E*, vol. 80, no. 4, Article ID 046122, 2009.
- [20] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, "Fast and accurate link prediction in social networking systems," *Journal of Systems and Software*, vol. 85, no. 9, pp. 2119–2132, 2012.
- [21] X. Zhu, H. Tian, S. Cai, J. Huang, and T. Zhou, "Predicting missing links via significant paths," *EPL (Europhysics Letters)*, vol. 106, no. 1, Article ID 18008, 2014.
- [22] W. Liu and L. Lu, "Link prediction based on local random walk," *EPL*, vol. 89, no. 5, Article ID 58007, 2010.
- [23] V. Batagelj and A. Mrvar, "Pajek-program for large network analysis," *Connections*, vol. 21, no. 2, p. 47, 1998.
- [24] D. Bu, Y. Zhao, L. Cai et al., "Topological structure analysis of the protein-protein interaction network in budding yeast," *Nucleic Acids Research*, vol. 31, no. 9, pp. 2443–2450, 2003.
- [25] M. E. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 3, Article ID 036104, 2006.
- [26] P. M. Gleiser and L. Danon, "Community structure in jazz," *Advances in Complex Systems*, vol. 6, no. 04, p. 565, 2003.
- [27] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [28] N. Blagus, L. Šubelj, and M. Bajec, "Self-similar scaling of density in complex real-world networks," *Physica A*, vol. 391, no. 8, pp. 2794–2802, 2012.
- [29] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, "Self-similar community structure in a network of human interactions," *Physical Review E*, vol. 68, no. 6, Article ID 065103, 2003.
- [30] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, "What's in a crowd? Analysis of face-to-face behavioral networks," *Journal of Theoretical Biology*, vol. 271, no. 1, pp. 166–180, 2011.
- [31] D. van Welden, "Mapping system theory problems to the field of knowledge discovery in databases," in *Proceedings of the 1st Future Business Technology Conference (Fubutec'04)*, pp. 55–59, EUROSIS, 2004.

- [32] J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology*, vol. 148, no. 3, pp. 839–843, 1983.
- [33] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, Article ID 208701, 2002.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

