OFFPRINT

Link prediction in weighted networks: The role of weak ties

Linyuan Lü and Tao Zhou

Please visit the new website
`www.epljournal.org`

TARGET YOUR RESEARCH WITH EPL

Sign up to receive the free EPL table of contents alert.

www.epljournal.org/alerts

Link prediction in weighted networks: The role of weak ties

LINYUAN LÜ[1] and TAO ZHOU[1,2(a)]

[1] *Department of Physics, University of Fribourg - Chemin du Musée 3, Fribourg 1700, Switzerland*
[2] *Department of Modern Physics, University of Science and Technology of China - Hefei 230026, PRC*

Abstract – Plenty of algorithms for link prediction have been proposed and were applied to various real networks. Among these algorithms, the weights of links are rarely taken into account. In this letter, we use local similarity indices to estimate the likelihood of the existence of links in weighted networks, including *Common Neighbor*, *Adamic-Adar Index*, *Resource Allocation Index*, and their weighted versions. We have tested the prediction accuracy on real social, technological and biological networks. Overall speaking, the resource allocation index performs best. To our surprise, sometimes the weighted indices perform even worse than the unweighted indices, which reminds us of the well-known *Weak-Ties Theory*. Further experimental study shows that the weak ties play a significant role in the link prediction, and to emphasize the contributions of weak ties can remarkably enhance the prediction accuracy for some networks. We give a semi-quantitative explanation based on the motif analysis. This letter provides a start point for the possible weak-ties theory in information retrieval.

Introduction. – Many complex systems can be well described by networks with nodes representing individuals or agents and links denoting the relations or interactions between nodes [1–5]. Recently, the link prediction in complex networks has attracted increasing attention from computer scientists [6,7] and physicists [8,9]. Link prediction aims at estimating the likelihood of the existence of a link between two nodes, based on the observed links and the attributes of the nodes. For example, classical information retrieval can be viewed as predicting missing links between words and documents [10], and the process of recommending items to users can be considered as a link prediction problem in the user-item bipartite networks [11,12]. The problem of link prediction can be categorized into two classes: one is the prediction of existing yet unknown links, such as food webs, protein-protein interaction networks and metabolic networks; the other is the prediction of links that may appear in the future of evolving networks, like on-line social networks. For the former task, since the discovery of links is costly in the laboratory or the field, to make predictions based on the links already known and to focus on those links which are most likely to exist, instead of blindly checking all possible links, may reduce the experimental costs. For the latter

task, very likely but not yet existent links can be recommended as promising friendships, which can help users in finding new friends and thus enhance their loyalties to the web sites.

Some algorithms based on Markov chains [13] and machine learning [14] have been proposed recently, and another group of algorithms are based on the measurements of node similarity. In this letter, we concentrate on the latter. Node similarity can be defined by using the essential attributes of nodes, namely two nodes are considered to be more similar if they have many common features. However, the essential features of nodes are usually not available, and thus the mainstream of similarity-based link prediction algorithms consider only the observed network structure. Liben-Nowell and Kleinberg [15] systematically compared some structure-based node similarity indices for link prediction problem in co-authorship networks, and Zhou *et al.* [16,17] studied nine well-known local similarity indices on six real networks extracted from disparate fields, as well as proposed two new local indices.

Up to now, most studies of link prediction do not take weights of links into consideration. Murata and Moriyasu [18] proposed three weighted similarity indices, as variants of the Common Neighbors, Adamic-Adar and Preferential Attachment indices, respectively. They

(a)E-mail: `zhutou@ustc.edu`

applied these indices to the networks of *Question-Answer Bulletin Boards System*, and the results show that with the consideration of weights the prediction accuracy can be enhanced. To our surprise, when we apply the weighted indices to the co-authorship network and the US air transportation network, we find that the weighted indices perform even worse than the unweighted ones. Actually, Liben-Nowell and Kleinberg [15] reported the similar observation for weighted Katz index. These unexpected results remind us of the well-known *Weak Ties Theory* [19,20]. Further experimental study shows that in some networks the weak links play a significant role in the link prediction, and to emphasize the contributions of weak links can remarkably enhance the prediction accuracy.

Data and method. – Considering an undirected simple network $G(V, E)$, where V is the set of nodes and E is the set of links. The multiple links and self-connections are not allowed. For each pair of nodes, $x, y \in V$, we assign a score, s_{xy}, according to a given similarity measure. Higher score means higher similarity between x and y, and vice versa. Since G is undirected, the score is supposed to be symmetry, say $s_{xy} = s_{yx}$. All the nonexistent links are sorted in a descending order according to their scores, and the links at the top are most likely to exist. To test the algorithm's accuracy, the observed links, E, is randomly divided into two parts: the training set, E^T, is treated as known information, while the probe set, E^P, is used for testing and no information therein is allowed to be used for prediction. Clearly, $E = E^T \cup E^P$ and $E^T \cap E^P = \varnothing$. In this paper, the training set always contains 90% of links, and the remaining 10% of links constitute the probe set[1]. To quantify the prediction accuracy, we use a standard metric called *precision*, which is defined as the ratio of relevant items selected to the number of items selected. We focus on the top L predicted links[2], if there are L_r relevant links (*i.e.*, the links in the probe set), the precision equals L_r/L. Clearly, higher precision means higher prediction accuracy.

The empirical data used in this paper include i) USAir —the US air transportation network, which contains 332 airports and 2126 airlines (see *Pajak Datasets*). The weight of a link is the frequency of flights between two airports. ii) *C. elegans* —the neural network of the nematode worm *C. elegans*, in which an edge joins two neurons if they are connected by either a synapse or a gap junction [22]. This network contains 297 neurons and 2148 links. iii) CGScience —the co-authorship network

---

[1]We have checked that the choice of such a proportion will not qualitatively affect the phenomena reported in this letter unless the fraction of data in the probe set is too large (then the known information is too sparse to give reasonable predictions) or too small (then the fluctuation is too large and the statistical regularities are hardly to be observed).

[2]In the literature, L usually ranges from 10 to 100, and the precision tends to decrease with the increasing of L [21]. We have checked that within such a range, the varying of L has little effects on the results reported in this letter, but for small L the fluctuation of precision is very large. We therefore set $L = 100$ in this letter.

in computational geometry till February 2002 (see *Pajek Datasets*). This network contains 7343 authors and 11898 links. Two authors are linked if they co-authorized at least one paper/book, and the weight of a link is the number of co-authorized papers/books.

Unweighted similarity indices based on local information. – Among many similarity indices, Liben-Nowell and Kleinberg [15] showed that the *Common Neighbors* (CN) and *Adamic-Adar* (AA) index [23] perform best, which has been further demonstrated by systematically comparing CN, AA index with seven other well-known local similarity indices [16]. In addition, Zhou *et al.* [16] proposed a new index named *Resource Allocation* (RA) index, which can beat both CN and AA index. Therefore, in this paper, we concentrate on CN, AA index and RA index, whose definitions are as following.

i) CN. In common sense, two nodes, x and y, are more likely to form a link if they have many common neighbors. Let $\Gamma(x)$ denote the set of neighbors of node x. The simplest measure of the neighborhood overlap is the directed count:

$$s_{xy} = |\Gamma(x) \cap \Gamma(y)|, \tag{1}$$

where $|Q|$ is the cardinality of the set Q.

ii) AA index. It refines the simple counting of common neighbors by giving the lower-connected neighbors more weights, as

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log k(z)}, \tag{2}$$

where $k(z)$ is the degree of node z, namely $k(z) = |\Gamma(z)|$.

iii) RA index. Considering a pair of nodes, x and y, which are not directly connected. The node x can send some resource to y, with their common neighbors playing the role of transmitters. In the simplest case, we assume that each transmitter has a unit of resource, and will averagely distribute to all its neighbors. As a results the amount of resource y received is defined as the similarity between x and y, which is

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z)}. \tag{3}$$

A common drawback of all the common-neighbor–based indices (including CN, AA, RA, etc.) is that their effectiveness depends on the clustering of networks. If the target network has a low clustering coefficient, for most node pairs their similarities are zero. As reported by Zhou *et al.* [16], those common-neighbor–based indices can give relatively accurate predictions only if the target network is of a high clustering coefficient. Clearly, to increase the links in the probe set will reduce the clustering coefficient for the training network (we have checked it for the three real networks used in this letter), and thus depress the precision. That is to say, if the known information is little

Table 1: Algorithm's accuracy, measured by precision. Each number is obtained by averaging over 100 implementations with independently random divisions of the training set and probe set. The numbers inside the brackets denote the standard derivations. For example, 0.592(48) means the precision is 0.592, and the standard derivation is 0.048. The abbreviations, WCN*, WAA* and WRA*, represent the highest precisions obtained by eqs. (7)–(9), respectively. The corresponding optimal values of α are shown in table 2. US, CE and CS stand for USAir, *C. elegans* and CGScience, respectively.

	CN	WCN	WCN*	AA	WAA	WAA*	RA	WRA	WRA*
US	0.592(48)	0.443(48)	0.617(45)	0.606(49)	0.517(50)	0.639(48)	0.626(39)	0.558(48)	0.633(41)
CE	0.132(26)	0.162(38)	0.182(40)	0.136(31)	0.170(35)	0.188(39)	0.128(32)	0.155(31)	0.164(36)
CS	0.625(59)	0.299(45)	0.782(57)	0.780(49)	0.292(51)	0.917(37)	0.963(15)	0.938(17)	0.969(16)

(corresponding to a small training set and a low clustering coefficient for the training network), the prediction is hardly to be accurate, which is an explicit instance of the so-called *sparsity problem* [24] in information retrieval.

Empirical analysis shows that [16] comparing with CN and AA, RA can enhance the prediction accuracy measured by the area under a receiver operating characteristic curve (AUC) [25], especially for the networks with large average degrees (in such cases, the difference between RA and AA is big). AUC takes into account the whole ranking, while precision only concentrates on the top L predicted links. As shown in table 1, subject to precision, RA still performs overall better than CN and AA. Here comes a simple but significant result, the RA index outperforms the CN and AA indices, and thus can find its applications in better characterizing the proximity of nodes in networks.

Weighted similarity indices. – The above-mentioned similarity indices only consider the binary relations among nodes, however, in the real world, links are naturally weighted, which may represent the transportation load between two airports in an airline network or the number of co-authorized papers in a co-authorship network. We expect the similarity indices taking into account link weights can give better predictions. Murata and Moriyasu [18] proposed a simple way to extend a similarity index for binary networks to a weighted index. Following this method, the weighted CN, weighted AA index and weighted RA index (denoted by WCN, WAA and WRA, respectively) are

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} w(x,z) + w(z,y), \qquad (4)$$

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x,z) + w(z,y)}{\log(1 + s(z))}, \qquad (5)$$

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x,z) + w(z,y)}{s(z)}. \qquad (6)$$

Here, $w(x,y) = w(y,x)$ denotes the weight of the link between nodes x and y, and $s(x) = \sum_{z \in \Gamma(x)} w(x,z)$ is the strength of node x. Note that, since $s(z)$ may be smaller

than 1 we use $\log(1 + s(z))$ in eq. (5) to avoid a negative score.

To our surprise, when we apply the weighted indices to the three real networks, as shown in table 1, we find that except the *C. elegans*, the weighted indices perform even worse than the corresponding unweighted ones. Especially for CN in USAir and CGScience, with consideration of the weights the precisions are sharply decreased. These unexpected results remind us of the well-known *Weak-Ties Theory* [19,20], which states that people usually obtain useful information or opportunities through the acquaintances but not their close friends, namely the weak ties in their friendship networks play a significant role. Recently, Onnela *et al.* [26] demonstrated that the weak ties mainly maintain the connectivity in mobile communication networks, and Csermely [27] found that the weak ties could maintain the stability of biological systems. In contrast, the role of weak ties in link prediction problem has not been investigated yet.

Role of weak ties. – In this section, we provide a starting point to investigate the role of weak ties in link prediction by introducing a free parameter, α, to control the relative contributions of weak ties to the similarity measure. The parameter-dependent indices for WCN, WAA and WRA are

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} w(x,z)^\alpha + w(z,y)^\alpha, \qquad (7)$$

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x,z)^\alpha + w(z,y)^\alpha}{\log(1 + s(z))}, \qquad (8)$$

$$s_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x,z)^\alpha + w(z,y)^\alpha}{s(z)}, \qquad (9)$$

where $s(x) = \sum_{z \in \Gamma(x)} w(x,z)^\alpha$. When $\alpha = 0$, $s(x)$ is the degree of node x, and the indices degenerate to the unweighted cases. When $\alpha = 1$, the indices are equivalent to the simply weighted indices, as shown in eqs. (4)–(6). The numerical results are given in fig. 1, table 1 and table 2. Except *C. elegans*, the optimal values of α are all smaller than 1. That is to say, for some weighted networks, the weak links play a more important role in the link prediction than indicated by their weights. A big surprise is that sometimes the optimal values of α are
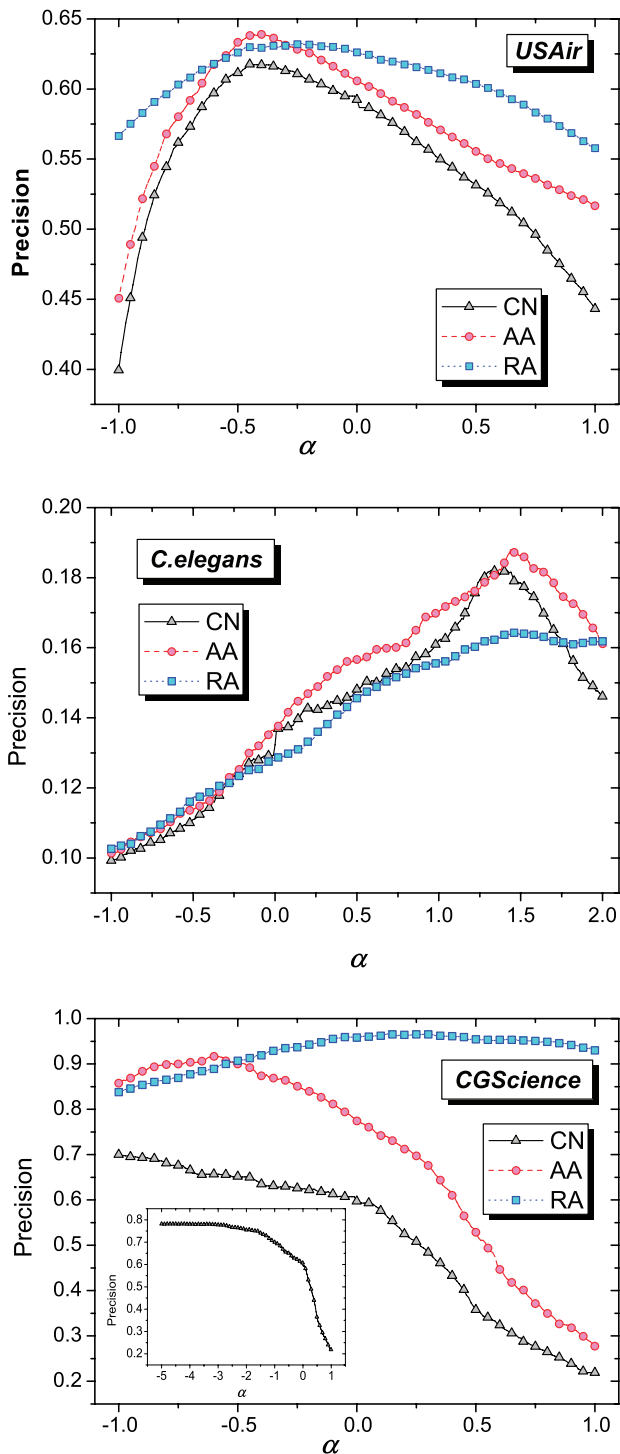
Fig. 1: (Color online) Precision as a function of α for USAir, *C. elegans* and CGScience. The inset in the plot for CGScience shows the precision of CN for $\alpha \in [-5, 1]$. Each data point is obtained by averaging over 100 realizations, each of which corresponds to an independent division of the training set and probe set.

negative. In these cases the weak links actually play a more important role than the strong links. Although it is well-known that the weak ties mainly maintain the network connectivity [26], this result is still striking for us.

Table 2: Optimal values of the parameter α subject to the highest precisions. For CGScience, with the decreasing of α the precision increases monotonously and eventually reaches a stable value, 0.782, at the point $\alpha = -4.15$.

	WCN*	WAA*	WRA*
USAir	-0.41	-0.40	-0.24
C. elegans	1.41	1.44	1.56
CGScience	-4.15	-0.60	0.13

Motif analysis. – It is observed that the weak ties play different roles in link prediction for different networks. For the scientific collaboration network, CGScience, and the transportation network, USAir, the weak ties play a more significant role than the strong ties, while the situation is opposite for the neural network of *C. elegans*. To get an in-depth understanding of the role of weak ties, we apply a motif analysis inspired by Milo *et al.* [28,29]. However, different from the method proposed by Milo *et al.* [28,29], we consider not only the connecting pattern, but also the weights of links. We concentrate on the motifs consisted of three connected nodes. Each link is classified into two categories: weak links and strong links. As shown in fig. 2, there are in total seven different motifs under our consideration.

The basic idea in the neighborhood-based similarity indices (*e.g.*, CN, AA and RA) for link prediction is to estimate the likelihood between two nodes based on the common neighbors connecting them. Here we go one step further to see how the common neighbors connect to the two nodes: through strong links or weak links? In table 3 we report the number of motifs, where all the links are ranked in a descending order of weights with the first 50% are considered to be the strong links and the other 50% the weak links[3]. To see clearly, as shown in fig. 2, each number is normalized by the number of motifs-1 (*i.e.*, triangles consisted of three strong links) in the corresponding network. We here do not use the Z-score based on null models since we are not interested in the comparison with random ensembles, but the relative abundance of different motifs in real networks.

As indicated by fig. 2, the reasons why weak links play a more significant role in USAir and CGScience are different. For USAir, it is because the very small number of motifs-6 compared with motifs-4 and motifs-5. That is to say, if two nodes are connected by a path consisted of

---

[3]In general, all the links can be classified into three categories: a fraction p for weak links, a fraction q for strong links and $1 - p - q$ others. However, this method is complicated and the information of a fraction $1 - p - q$ of links is waste. Or we can set a fraction p of links to be weak links while the others are strong links. We here set $p = 0.5$ because it is a natural and fair choice, otherwise some bias may occur. For example, if we choose $p = 0.9$ (*i.e.*, 90% of links are weak links), for all the three networks, $p_s > p_w$ (see later analysis), which is not meaningful or distinguishable. It is because the number of the sixth subgraph, N_6, becomes too large and leads to a relatively small p_w.

Table 3: Number of motifs in USAir, CGScience and *C. elegans*. The IDs of motifs (1–7, from left to right) are in accordance with the order shown in fig. 2 (*e.g.*, the motif consisted of three strong links is of ID 1). N_i ($i = 1, 2, \ldots, 7$) denotes the number of the i-th motifs (*i.e.*, motifs-i).

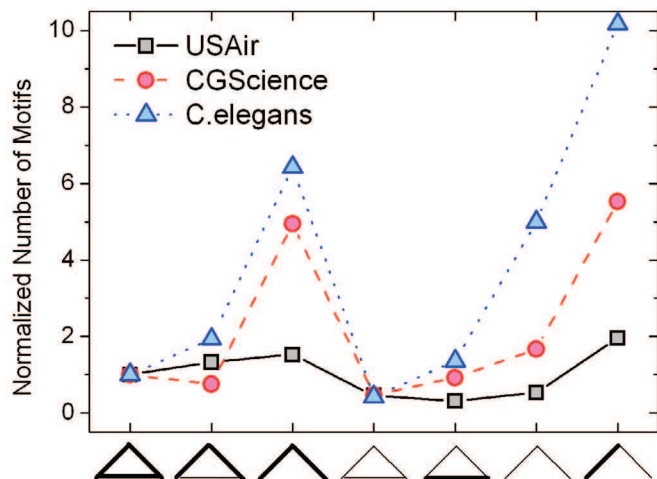| Network/Number of motifs	N_1	N_2	N_3	N_4	N_5	N_6	N_7
USAir	3943	5246	6020	1784	1208	2104	7731
CGScience	4327	3259	21397	2023	3981	7220	23914
C. elegans	687	1329	4412	291	934	3433	6991

Fig. 2: (Color online) Normalized number of motifs for USAir (square), CGScience (circle) and *C. elegans* (triangle). There are in total seven motifs under consideration (below the X-axis), with thick lines representing strong links and thin lines representing weak links.

two weak links, the probability that they are not directly connected (expressed by motif-6) is low. For CGScience, a path consisted of two strong links is not a good indicator because the motifs-3 are relatively abundant. Let us consider a simple question: given that (x, z) and (y, z) are two strong links, what is the probability that x and y are neighboring? Obviously, this probability, denoted by p_s, is

$$p_s = \frac{3N_1 + N_2}{3N_1 + N_2 + N_3}. \tag{10}$$

Analogously, the probability that x and y are neighboring provided that (x, z) and (y, z) are two weak links is

$$p_w = \frac{3N_4 + N_5}{3N_4 + N_5 + N_6}. \tag{11}$$

Table 4 reports the values of p_s and p_w for the three real networks. For USAir and CGScience, in which the weak ties are more significant in link prediction, one can see that $p_w > p_s$. In contrast, $p_s > p_w$ for *C. elegans* where the strong ties are more significant.

Conclusions and discussion. – In this letter, we applied three local similarity indices, Common Neighbor, Adamic-Adar index and Resource Allocation index, to the link prediction problem in three real networks, USAir, *C. elegans* and CGScience. We found that our previously

Table 4: p_s and p_w for the three real networks.

	USAir	CGScience	*C. elegans*
p_s	0.7393	0.4315	0.4345
p_w	0.7572	0.5819	0.3442

proposed index, RA [16], performs overall best. Furthermore, with the consideration of weights, we tested three weighted variants of CN, AA and RA, denoted by WCN, WAA and WRA. To our surprise, except *C. elegans*, the weighted indices perform even worse than their corresponding unweighted versions. These unexpected results remind us the weak-ties theory [19,20] which claims that the links with small weights yet play a more important role in social networks. Our experimental study shows that in some networks the weak ties play a more significant role in link prediction, and to emphasize the contributions of weak ties can remarkably enhance the prediction accuracy. In another word, the weak links in such network are not as weak as their weights suggested.

Although the prediction accuracies of both the unweighted indices (eqs. (1)–(3)) and the simply weighted indices (eqs. (4)–(6)) can be further improved by introducing the parameter α (eqs. (7)–(9)), this letter does not aim at highlighting these parameter-dependent indices. Instead, we attempted to uncover the role of weak ties in the link prediction problem. We gave a semi-quantitative explanation for the observed experimental results based on the motif analysis. Despite of the lack of a full understanding of the role of weak ties, this letter provides some insights about the link prediction of weighted networks. In addition, the motif analysis method for weighted networks may be further extended to detect the correlations between link weights and local topologies, as a complementary method to some other approaches [30–32].

In the early stage of the studies of weak-ties theory in social science, whether a tie is weak or strong is mostly determined qualitatively [33]. For example, Lin *et al.* [34] identified acquaintances or friends of friends as weak whereas friends, relatives and neighbors were considered to be strong, and Friedkin [35] treated asymmetrical contact as a weak tie and reciprocal contact as a strong tie. Recently, the availability of automatically saved data makes it possible to quantitatively analyse the role of tie strength in huge-size social networks. For example, the strength of a tie can be measured by the cumulative time spent in telephone communication in a time period [26,36]. And the concept of weak ties has been applied in many

branches of science beyond the social-network analysis, such as the biological science [27] and the information science [37]. Although the majority of previous works focused on the role of weak ties on dynamics, such as the information flow [38], rumor spreading [36] and knowledge transfer [39], some recent works suggested that the weak ties also play a key topological role, like to maintain the connectivity of a network [26] and to affect the emergence of community structure [40]. In contrast to the above-mentioned works, the definition of the tie strength in this letter is not limited to social science or to a specific method. Indeed, the strength of a tie can be obtained by a standard method that maps a multi-edge network to a weighted network (for *C. elegans* and CGScience, see the method reported in [41]) or by a functionally relevant measurement (the real traffic between two airports for USAir). The concept of weak ties in this letter is less explicit but may provide a wider horizon than the traditional studies on weak ties. Especially, this concept is new in the information filtering problem and thus we do not want to make it specific, explicit but narrow. In a word, we hope this work can provide a start point for the possible *weak-ties theory in information retrieval*.

$$* * *$$

REFERENCES

[1] ALBERT R. and BARABÁSI A.-L., *Rev. Mod. Phys.*, **74** (2002) 47.

[2] DOROGOVTSEV S. N. and MENDES J. F. F., *Adv. Phys.*, **51** (2002) 1079.

[3] NEWMAN M. E. J., *SIAM Rev.*, **45** (2003) 167.

[4] BOCCALETTI S., LATORA V., MORENO Y., CHAVEZ M. and HUANG D.-U., *Phys. Rep.*, **424** (2006) 175.

[5] DA F. COSTA L., RODRIGUES F. A., TRAVIESO G. and BOAS P. R. U., *Adv. Phys.*, **56** (2007) 167.

[6] GETOOR L. and DIEHL C. P., *SIGKDD Explor. Newslett.*, **7**, issue No. 2 (2005) 3.

[7] HUANG Z., LI X. and CHEN H., *Link prediction approach to collaborative filtering*, in *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries* (ACM Press, New York) 2005.

[8] CLAUSET A., MOORE C. and NEWMAN M. E. J., *Nature*, **453** (2008) 98.

[9] REDNER S., *Nature*, **453** (2008) 47.

[10] SALTON G. and MCGILL M. J., *Introduction to Model Information Retrieval* (McGraw-Hill, Auckland) 1983.

[11] ZHOU T., REN J., MEDO M. and ZHANG Y.-C., *Phys. Rev. E*, **76** (2007) 046115.

[12] ZHOU T., JIANG L.-L., SU R.-Q. and ZHANG Y.-C., *EPL*, **81** (2008) 58004.

[13] BILGIC M., NAMATA G. M. and GETOOR L., *Combining collective classification and link prediction*, in *Proceedings of the 2007 seventh IEEE International Conference on Data Mining* (IEEE Press, Washington DC) 2007, pp. 381–386.

[14] WANG C., SATULURI V. and PARTHASARATHY S., *Local Probabilistic Models for Link Prediction*, in *Proceedings of the 2007 seventh IEEE International Conference on Data Mining* (IEEE Press, Washington DC) 2007, pp. 322–331.

[15] LIBEN-NOWELL D. and KLEINBERG J., *J. Am. Soc. Inf. Sci. & Technol.*, **58** (2007) 1019.

[16] ZHOU T., LÜ L. and ZHANG Y.-C., *Eur. Phys. J. B*, **71** (2009) 623.

[17] LÜ L., JIN C.-H. and ZHOU T., *Phys. Rev. E*, **80** (2009) 046122.

[18] MURATA T. and MORIYASU S., *Link prediction of social networks based on weighted proximity measures*, in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence* (ACM Press, New York) 2007.

[19] GRANOVETTER M. S., *Am. J. Sociol.*, **78** (1973) 1360.

[20] CSERMELY P., *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks* (Springer-Verlag, Berlin/Heidelberg) 2006.

[21] HERLOCKER J. L., KONSTAN J. A., TERVEEN K. and RIEDL J. T., *ACM Trans. Inf. Syst.*, **22** (2004) 5.

[22] WATTS D. J. and STROGATZ S. H., *Nature*, **393** (1998) 440.

[23] ADAMIC L. A. and ADAR E., *Soc. Netw.*, **25** (2003) 211.

[24] HUANG Z., CHEN H. and ZENG D., *ACM Trans. Inf. Syst.*, **22** (2004) 116.

[25] HANLEY J. A. and MCNEIL B. J., *Radiology*, **143** (1982) 29.

[26] ONNELA J.-P., SARAMÄKI J., HYVÖNEN J., SZABÓ G., LAZER D., KASKI K., KERTÉSZ J. and BARABÁSI A.-L., *Proc. Natl. Acad. Sci. U.S.A.*, **104** (2007) 7332.

[27] CSERMELY P., *Trends Biochem. Sci.*, **29** (2004) 331.

[28] MILO R., SHEN-ORR S., ITZKOVITZ S., KASHTAN N., CHKLOVSKII D. and ALON U., *Science*, **298** (2002) 824.

[29] MILO R., ITZKOVITZ S., KASHTAN N., LEVITT R., SHEN-ORR S., AYZENSHTAT I., SHEFFER M. and ALON U., *Science*, **303** (2004) 1538.

[30] BARRAT A., BARTHÉLEMY M., PASTOR-SATORRAS R. and VESPIGNANI A., *Proc. Natl. Acad. Sci. U.S.A.*, **101** (2004) 3747.

[31] LI M.-H., FAN Y., WANG D.-H., LI D.-Q., WU J.-S. and DI Z.-R., *Phys. Lett. A*, **364** (2007) 488.

[32] SERRANO M. Á., BOGUÑÁ M. and VESPIGNANI A., *Proc. Natl. Acad. Sci. U.S.A.*, **106** (2009) 6483.

[33] GRANOVETTER M., *Sociol. Theory*, **1** (1983) 201.

[34] LIN N., ENSEL W. and VAUGHN J., *Am. Sociol. Rev.*, **46** (1981) 393.

[35] FRIEDKIN N. E., *Soc. Netw.*, **2** (1980) 411.

[36] LAI G. and WONG O., *Soc. Netw.*, **24** (2002) 49.

[37] STEFFES E. M. and BURGEE L. E., *Internet Res.*, **19** (2009) 42.

[38] FRIEDKIN N. E., *Soc. Netw.*, **3** (1982) 273.

[39] LEVIN D. Z. and CROSS R., *Manag. Sci.*, **50** (2004) 1477.

[40] KUMPULA J. M., ONNELA J. P., SARAMAKI J., KASKI K. and KERTESZ J., *Phys. Rev. Lett.*, **99** (2007) 228701.

[41] NEWMAN M. E. J., *Phys. Rev. E*, **70** (2004) 056131.