Link prediction based on a semi-local similarity index*

Bai Meng(白 萌), Hu Ke(胡 柯), and Tang Yi(唐 翌)†

Department of Physics, Xiangtan University, Xiangtan 411105, *China*

(Received 21 April 2011; revised manuscript received 20 July 2011)

Missing link prediction provides significant instruction for both analysis of network structure and mining of unknown links in incomplete networks. Recently, many algorithms have been proposed based on various node-similarity measures. Among these measures, the common neighbour index, the resource allocation index, and the local path index, stemming from different source, have been proved to have relatively high accuracy and low computational effort. In this paper, we propose a similarity index by combining the resource allocation index and the local path index. Simulation results on six unweighted networks show that the accuracy of the proposed index is higher than that of the local path one. Based on the same idea of the present index, we develop its corresponding weighted version and test it on several weighted networks. It is found that, except for the USAir network, the weighted variant also performs better than both the weighted resource allocation index and the weighted local path index. Due to the improved accuracy and the still low computational complexity, the indices may be useful for link prediction.

Keywords: link prediction, resource allocation, local path

PACS: 89.75.Fb, 89.75.Hc, 89.20.Ff **DOI:** 10.1088/1674-1056/20/12/128902

1. Introduction

Predicting missing links involves estimating the likelihood of existing yet unknown connections between agents. This problem has attracted increasing attention due to the theoretical significance and potential applications in mining data about network structures.[1−3] For example, in the protein–protein interaction network, there are a certain number of interactions not tested, that may be very important. In order to find the missing links, especially when the network is huge and sparse, to check every potential interaction blindly in the laboratory is time-consuming as the number of all possible links increases squarely with that of nodes. And large numbers of samples often mean high cost which is usually not possible. If we can detect the underlying links which are most likely to exist by adopting some link prediction methods, the experiments will be well targeted to identify the missing links.

Besides helping to find missing links, the link prediction methods are also beneficial for the understanding of the evolution of real networks.[4] For instance, there are many models to explain the mechanisms of the evolution of complex networks, but it is difficult to judge which one is best. If we treat each model as one link prediction algorithm, the matrices of link prediction accuracy can be used to evaluate the evolving mechanism for given networks.

In addition, the link prediction algorithms can be extended to broader applications. For instance, some prediction algorithms are used to filter out information in recommendation systems.[5−8] Recently, Zhang *et al.*[9] have applied similarity-based link prediction algorithms to find hidden links to solve the classification problem in partially labelled networks.

To date, many link prediction algorithms have been proposed, and most of them are designed based on node similarity solely associated with the network structure.[4] One important reason is that topology can reflect useful information of structure similarity or, in other words, the links in a network indicate certain similarity between the nodes they connect.[10,11] According to the domain of required topological information, the links can be grouped into local, semi-local and global ones.[4] For instance, the Katz index,[12] hitting time,[13] commute time,[14] rooted page rank,[15] SimRank,[16] blondel index,[17] average commute time index,[13] and random walk with restart index,[15] which require full knowledge of

*Project supported by the National Natural Science Foundation of China (Grant No. 30570432), the Young Research Foundation of Education Department of Hunan Province of China (Grant No. 11B128), and partly by the Doctor Startup Project of Xiangtan University (Grant No. 10QDZ20).

†Corresponding author. E-mail: yitang16@yahoo.com

© 2011 Chinese Physical Society and IOP Publishing Ltd

the network topology, are global. The common neighbour index,[18] Salton index,[19] Jaccard index,[20] Sϕrensen index,[21] Hub promoted index,[22] Hub depressed index,[23] preferential attachment index,[24] Adamic–Adar index,[25] and the recently proposed Resource Allocation index[5,23,26,27] are local, which consider only the immediate neighbourhood information; in other words, all these methods depend on the triangle closing model. The local path index[23,28] and local random walk index[29] use additional information by taking account of nodes a few more steps away from the latent similarity ones, and are therefore referred to as semi-local indices. Among those algorithms mentioned above, the global ones have higher accuracy, while the local ones have computing superiority. For large networks, global search means considerable CPU time and large memory space. This is less effective and sometimes intolerable. To maintain relatively high accuracy and simultaneously to take less computing effort, the trade-off choice is the semi-local index.

Along with a complex topological structure, plenty of real networks display a large heterogeneity in the capacity and intensity of the connections which can be described by corresponding link weights in the networks. Due to the topology of networks not being independent of their weights,[30] it is reasonable to develop algorithms based on link weights to predict link existence. As a pioneer work, Murata and Moriyasu[31] proposed a weighted version of several local indices, and obtained good prediction performance in weighted networks. Lü and Zhou[32] further found that the weak ties may play a more important role than strong ties in the prediction.

In this paper, we propose a semi-local index in unweighted networks by introducing the resource allocation process into the Local Path index, and develop its corresponding weighted version in weighted networks. The two indices consider the difference of nodes with different degrees along 3-order paths, this consideration may be a helpful exploration for link prediction.

2. Link prediction in unweighted networks

2.1. Data sets

Six real networks are drawn from different fields, which include a wide scope of the network's classes. For weighted and directed networks, we neglect the link weights and link direction, and consider only the existence of links. The data sets are simply described as follows:

(i) US air network (USAir): the weighted network of the US air transportation system.[33]

(ii) Email network—Enron email network, in which there are two reversed links between each node pair. It indicates that email is exchanged and the network is undirected in fact.[34]

(iii) Political blogs (Pblogs): a directed network of hyperlinks between weblogs on US politics, recorded in 2005 by Adamic and Glance.[35]

(iv) Internet: a symmetrized snapshot of the structure of the Internet at the level autonomous systems, reconstructed from BGP tables posted by the University of Oregon Route Views Project.[36]

(v) Neural network (Neural): a directed and weighted neural network of C. Elegans is compiled by D. Watts and S. Strogatz.[37]

(vi) Power grid network (Grid): an undirected, unweighted network representing the topology of the Western States Power Grid of the United States. The data were also compiled by D. Watts and S. Strogatz. [37] The data of Grid as well as Pblog, Internet, and Neural can be downloaded from http://www-personal.umich.edu/~ mejn/netdata/.

Table 1. Basic topological features of six real networks. N denotes the number of nodes, M is the total undirected links, $N_{\rm C}$ represents the size of giant component and the number in parentheses is the number of components, $\langle k\rangle$ is the average degree, e and r denote the network efficiency[38] and assortative coefficient,[39] respectively, H is the degree heterogeneity, defined as $\langle k^2\rangle/\langle k\rangle^2$, C represents the clustering coefficient.[37]

Networks	N	M	$N_{\rm C}$	$\langle k\rangle$	e	r	H	C
USAir	332	2126	332(1)	12.807	0.409	–0.208	3.464	0.749
Email	36692	183831	33696(1065)	10.020	0.221	–0.111	13.980	0.716
Pblog	1224	16718	1222(2)	27.317	0.397	–0.221	2.975	0.362
Internet	22963	48436	22963(1)	4.219	0.276	–0.199	61.978	0.350
Neural	297	2148	297(1)	14.465	0.445	–0.163	1.801	0.308
Grid	4941	6594	4941(1)	2.669	0.056	0.003	1.450	0.107

2.2. Similarity indices

Our similarity index is designed according to the resource allocation index and the local path index, while the resource allocation index is based on the common neighbour one. Thus, we first introduce these three indices, and then give the definition of our index.

(I) Common neighbour (CN) index: If let $\Gamma(x)$ and $\Gamma(y)$ denote the sets of nearest neighbours of nodes x and y respectively, the CN may be written as

$$S_{xy} = |\Gamma(x) \cap \Gamma(y)|. \tag{1}$$

This measure rests on the fact that many networks display high clustering or transitivity. Therefore, two nodes sharing common neighbours are connected with high probability.

(II) Resource allocation (RA) index: Motivated by the resource allocation process taking place on networks, Zhou *et al.* designed the resource allocation index to refine the simple count of common neighbours, which is defined as

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

where z is the common neighbour between nodes x and y, $k(z)$ is the degree of node z. In fact, the RA index is not only closely related to the resource allocation process, but also can successfully suppress the contribution of common neighbours with high degree by assigning them less weight. This refinement for the CN index means that the RA index has particularly good performance on networks with large clustering coefficient, high degree of heterogeneity and absence of a strongly assortative linking pattern.[23]

(III) Local path (LP) index: Although the neighbourhood-based measures CN index and RA index have very low computational complexity, they are over-localized and do not allow nodes to be similar without sharing neighbours, which leads to a very large fraction of node pairs being assigned zero similarity score (see Table 2). Of course, one can consider some global measures, such as the Katz index, to relax the limitation of the over-localization. However, these global measures have relatively high computational complexity, resulting in limitation of algorithmic implementation for huge networks. The LP index can provide a good tradeoff by using a little more information involving the next-nearest neighbours, which is defined as

$$S_{xy} = (A^2)_{xy} + \varepsilon (A^3)_{xy}, \tag{3}$$

where $(A^2)_{xy}$ denotes the number of different paths of length 2 connecting nodes x and y. Its value is the number of common neighbours between the two nodes. Similarly, $(A^3)_{xy}$ is the path counts of length 3, ε is a tunable parameter, controlling the contribution of paths of length 3 to the similarity.

Table 2. The ratio of number of node pairs with different similarity scores in six real networks. The S_0 denotes the ratio of number of zero-score node pairs to that of all node pairs. $S_1/S_{>0}$ denotes the ratio of number of one-score node pairs to that of non-zero-score node pairs.

Score	USAir	Email	Pblog	Internet	Neural	Grid
S_0	0.597	0.977	0.604	0.958	0.454	0.999
$S_1/S_{>0}$	0.465	0.787	0.402	0.913	0.520	0.929

(IV) Resource allocation along local path (RALP) index: By introducing the resource allocation process into the LP index, one can obtain the RALP index, which is written as

$$S_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z)} + \varepsilon \sum_{(i,j) \in l_{x \to y}} \frac{1}{k(i)k(j)}\ , \tag{4}$$

where $l_{x \to y}$ is the path of length three from node x to node y, i and j are intermediate nodes along the path $l_{x \to y}$. The RALP index not only retains the advantage of LP and RA, but also has low computational complexity. It is expected to perform better.

2.3. Method

Let $G(V, E)$ denote a simple undirected network, which is described by the sets of nodes V and links E. Multiple links and self-connections are excluded from E. Every index referred to in this paper assigns a similarity matrix S whose real element S_{xy} represents how similar two nodes x and y are. We say that S_{xy} is their similarity 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 an algorithm's accuracy, a fraction of the observed links E (in this paper, always 90 % of the whole) is randomly singled out as the training set E_T, the remaining links (10%) are used as the probe set E_P to be predicted and no information in this set is allowed to be used for prediction. Clearly, $E = E_T \cup E_P$ and $E_T \cap E_P = \phi$. The prediction quality was evaluated by a standard matrix, the area under the receiver operating characteristic (ROC) curve.[40] In the present case, this matrix 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 $U\backslash E$, where U denotes the universal set). In the implementation, among n independent comparisons, if there are n' occurrences of the missing link having a higher score and n'' occurrences of the missing link and nonexistent link having the same score, the accuracy is defined as:

$$\mathrm{AUC} = \frac{n' + 0.5n''}{n} \ . \tag{5}$$

If all the scores are generated from an 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.

2.4. Performance evaluation

According to the definition of these indices mentioned above, the RA index is designed by introducing the resource allocation into the CN index; the LP and RALP indices may be viewed as the extensions of the CN and RA indices, respectively. Thus, they all have a foundation on the CN index, which means the clustering coefficient of a network plays a significant role on link prediction by these indices. Usually, the prediction has high accuracy when the clustering coefficient is high. As is easily found from Table 3, these indices have a good performance in networks with high clustering coefficient such as USAir, Email and Pblogs. However, the predictive accuracy has no simple relation to the clustering coefficient, and multiple ingredients can jointly affect the results. For example, although the clustering coefficient of Internet is higher than that of Neural, the predictive accuracies of these indices in Internet are lower than those in Neural. This abnormality has an important implication to the influence of another noticeable property: the network efficiency e. Generally, the lower the network efficiency, the less predictive accuracy of these neighbourhood-based measures. Thus, if a network possesses both high clustering coefficient and high network efficiency, these indices have good performance generally. The other factors, such as the assortative coefficient r and the heterogeneity H, may have indirect and weak influence on the link prediction, but we have not found a definite correlation of them to the prediction accuracy.

Comparing the CN index and RA index, we find that the latter performs better, especially for networks (i.e., Plogs and USAir) with large clustering coefficient, high degree of heterogeneity and absence of a strongly assortative linking pattern. It is observed that the RA index exhibits particularly good performance on USAir, even better than the LP and RALP indices. This may be derived from the specific structure of USAir(the hierarchical organization consisting of hubs, local centers and small local airports).[23]

Table 3. The predictive accuracies based on the four indices on six real networks. Every accuracy is obtained by averaging 10 implementations with independently random partitions of testing set and probe set. In LP and RALP, ε is set as 0.001.

Measures	USAir	Email	Pblog	Internet	Neural	Grid
CN	0.937	0.954	0.919	0.710	0.846	0.591
RA	**0.956**	0.955	0.922	0.711	0.868	0.591
LP	0.928	0.957	0.929	0.800	0.857	0.639
RALP	0.951	**0.959**	**0.932**	**0.810**	**0.883**	**0.639**

Secondly, as seen in Table 3, the LP index performs better than the CN index in most networks. When taking into account the contribution of longer paths to similarity score, some nodes assigned the same zero score by the CN index are now distinguishable under the similarity definition of the LP index, and hence the prediction accuracy is increased. In particular, for a network with low efficiency, such as the Power grid, the improvement is remarkable. In fact, the LP index can relax the limitation of over-localization of the CN index for networks with low efficiency. However, although the LP index considers the contribution of both path counts and path length, it neglects the difference of nodes along the paths of the same length. That is to say, the LP index simply considers that every path of the same length has an equivalent contribution to the similarity score. In fact, this simplification have a certain negative effect on link

prediction, and can be readily revised. For example, the CN index neglects the difference of nodes along paths of length 2. In contrast, the RA index counts it by penalizing the contribution of common-neighbour nodes with high degree, and the results show that the penalty of the RA index is significant for link prediction. However, the RA index does not overcome the disadvantage of over-localization of the CN index. Here we design a new index, the RALP index, by introducing the resource allocation process into the LP index (see Table 3); this combination improves the prediction performance.

However, the LP and RALP indices only consider paths up to 3-order; one may think that it would be better to extend them a little further to higher-order paths. This does not seem necessary when the network efficiency is not very low. The reason is twofold: i) the computational complexity grows rapidly with the increase of path order; ii) the LP index aims at both relaxing the limitation of over-localization of the CN index and breaking the "degeneracy of the states". For a network with high efficiency, the higher-order paths will have little effect on further relaxing the limitation of the CN's over-localization or breaking the "degeneracy of the states". Indeed, it is sufficient for the LP index to consider paths no more than 3-order, so it seems enough for the RALP index to adopt an analogous consideration. To illustrate this, we study the 4-order extension of the RALP index, and the similarity index takes the form

$$S_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{k(z)} + \varepsilon \sum_{(i,j)\in l_{x\to y}} \frac{1}{k(i)k(j)} + \varepsilon^2 \sum_{(i,j,h)\in l_{x\to y}} \frac{1}{k(i)k(j)k(h)}, \quad (6)$$

where i, j, and h are the sequential intermediate nodes along the path, and ε is set as 0.001. The results in Table 4 show that the accuracies of most networks are not enhanced, and some even go down. The sole exception is power grid network. Its low clustering coefficient and efficiency improve the performance when extending RALP to longer paths.

Table 4. Comparison of the accuracies of the RALP index (path length $L = 3$) and the 4-order extension ($L = 4$) of the RALP index measured by AUC. Each number is obtained by averaging over 10 different realizations.

Measure	USAir	Email	Pblog	Internet	Neural	Grid
$L = 3$	0.951	0.959	0.932	0.810	0.883	0.639
$L = 4$	0.949	0.948	0.932	0.785	0.883	**0.677**

3. Link prediction in weighted networks

3.1. Weighted similarity indices

Based on the weighted CN index proposed by Murata *et al.*,[31] we first develop the weighted LP index. And then according to the weighted RA index suggested by Lü and Zhou,[32] the weighted RALP index can be easily obtained. The four weighted similarity indices are summarized as below.

I) Weighted common neighbour (WCN) index

$$S_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} (w_{xz} + w_{zy}) , \quad (7)$$

where $w_{xz} = w_{zx}$ denotes the link weight between nodes x and z.

II) Weighted resource allocation (WRA) index

$$S_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{w_{xz} + w_{zy}}{s(z)} , \quad (8)$$

where $s(z) = \sum_{x\in\Gamma(z)} w_{zx}$ is the strength of node z.

III) Weighted local path (WLP) index

$$S_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} (w_{xz} + w_{zy}) + \varepsilon \sum_{(i,j)\in l_{x\to y}} (w_{xi} + w_{ij})(w_{ij} + w_{jy}). \quad (9)$$

IV) Weighted RALP (WRALP) index

$$S_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{(w_{xz} + w_{zy})}{s(z)} + \varepsilon \sum_{(i,j)\in l_{x\to y}} \frac{(w_{xi} + w_{ij})(w_{ij} + w_{jy})}{s(i)s(j)}. \quad (10)$$

3.2. Data sets and results

The weighted networks are: 1) USAir: the US air transportation network, in which the link weight denotes the distance between two airports;[33] 2) Neural: the neural network of C. Elegans, where the weights indicate the interaction strength;[37] 3) CG-science: the co-authorship network in computational geometry till February 2002. It has 6158 nodes and 11898 edges, and its clustering coefficient C and efficiency e are 0.728 and 0.072, respectively.[33] Each link weight equals the cooperation times of two authors. The cooperation may be a co-authored paper or book.

As can be seen in Table 5, for all three networks, the WRA index shows great improvement by penalizing the contribution of common-neighbour nodes with

high strength; in particular, for USAir, WRA performs best. It may be that resource allocation process was originally proposed to explain the nonlinear correlation between transportation capacity and connectivity of each airport.[23] Similarly as on unweighted networks, the WRALP index, considering the difference of nodes along the paths, performs better than the WLP index. As to the effect of weight distribution on the efficiency of weighted networks,[41] and thereby on the accuracy of link prediction, more investigation is required. For Neural and the CGscience, the WRALP index breaks the "degeneracy of the states", which obtains better performance than the WRA index. However, why the above weighted algorithms perform worse on USAir and Neural networks than their corresponding unweighted cases needs further exploration.

Table 5. The accuracy of three weighted networks measured by AUC. Each number is obtained by averaging over 10 implementations with independently random divisions of testing set and probe set. In WLP index and WRALP index, ε is set as 0.001, too.

USAir	0.911	**0.929**	0.907	0.924
Neural	0.790	0.847	0.768	**0.863**
CGscience	0.892	0.893	0.902	**0.903**

4. Conclusion and discussion

We have proposed a semi-local similarity index and its corresponding weighted variant for predicting missing links in both unweighted and weighted networks. This index in unweighted networks integrates the LP index and the RA one, and has good performance in most sample networks. The RALP index, beyond increasing the prediction accuracy, can also enlarge the application scope of the LP index.

However, introducing the resource allocation process into the LP index can be viewed as only a possible, but not an optimal, approach to improve it. Although it has been proved that introducing the resource allocation process into paths of length 2 (i.e., the RA index) is of considerable significance,[18] further extending it to paths of length 3 may be moderate. Perhaps RA's penalty for high-degree nodes along paths of length 3 is excessive. Proper formulation of penalty for high-degree nodes along paths of length 3 or longer requires taking many more detailed factors into account.

Acknowledgement

We thank Dr. L. Lü for help with the data.

References

[1] Krebs V E 2001 *Connections* **24** 43
[2] Kautz H, Selman B and Shah M 1997 *Commun. ACM* **40** 63
[3] Chakrabarti S, Dom B E, Gibson D, Kleinerg J, Kumar R, Raghavan P, Rajagopalan S and Tomkins A 1999 *IEEE Computer* **32** 60
[4] Lü L and Zhou T 2011 *Physica* A **390** 1150
[5] Zhou T, Ren J, Medo M and Zhang Y C 2007 *Phys. Rev.* E **76** 046115
[6] Zhou T, Jiang L L, Su R Q and Zhang Y C 2008 *Eur. Phys. J.* B **81** 58004
[7] Zhang Y C, Medo M, Ren J, Zhou T, Li T and Yang F 2007 *Eur. Phys. J.* B **80** 68003
[8] Zhang Y C, Blattner M and Yu Y K 2007 *Phys. Rev. Lett.* **99** 154301
[9] Zhang Q M, Shang M S and Lü L 2010 *Int. J. Mod. Phys.* C **21** 813
[10] Leicht E A, Holme P and Newman M E J 2006 *Phys. Rev.* E **73** 026120
[11] Liben-Nowell D and Kleinberg J 2007 *J. Am. Soc. Inf. Sci. Tec.* **58** 1019
[12] Katz L 1953 *Psychometrika* **18** 39
[13] Göbel F and Jagers A 1974 *Stoch. Processes Appl.* **2** 311
[14] Fouss F, Pirotte A, Renders J M and Saerens M 2007 *IEEE Trans. Knowl. Data Eng.* **19** 355
[15] Brin S and Page L 1998 *Comput. Netw. ISDN Syst.* **30** 107
[16] Jeh G and Widom J 2002 *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA p. 538
[17] Blondel V D, Gajardo A, Heymans M, Senellart P and Dooren P V 2004 *SIAM Rev.* **46** 647
[18] Lorrain F and White H C 1971 *J. Math. Soc.* **1** 49
[19] Salton G and McGill M J 1983 *Introduction to Modern Information Retrieval* (Auckland: MuGraw-Hill) pp. 201–215
[20] Jaccard P 1901 *Bulletin de la Societe Vaudoise des Sciences Naturelle* **37** 547
[21] Sϕrensen T 1948 *Biol. Skr.* **5** 1
[22] Ravasz E, Somera A L, Mongru D A, Oltvai Z N and Barabási 2002 *Science* **297** 1553
[23] Zhou T, Lü L and Zhang Y C 2009 *Eur. Phys. J.* B **71** 623
[24] Barabási A L and Albert R 1999 *Science* **286** 509
[25] Adamic L A and Adar E 2003 *Social Networks* **25** 211
[26] Ou Q, Jin Y D, Zhou T, Wang B H and Yin B Q 2007 *Phys. Rev.* E **75** 021102
[27] Sun D, Zhou T, Liu J G, Liu R R, Jia C X and Wang B H 2009 *Phys. Rev.* E **80** 017101
[28] Lü L, Jin C H and Zhou T 2009 *Phys. Rev.* E **80** 046122
[29] Liu W P and Lü L 2010 *Europhys. Lett.* **89** 58007
[30] Barrat A, Barthélemy, Pastor-Satorras R and Vespignani A 2004 *Proc. Natl. Acad. Sci.* **101** 3747
[31] Murata T and Moriyasu S 2007 *ACM International Conference on Web Intelligence*, Fremont, November 2–5, 2007 p. 85

[32] Lü L and Zhou T 2010 *Europhys. Lett.* **89** 18001
[33] The data was released by Batagelj V and Mrvar A in 2006, available at http://vlado.fmf.uni-lj.si/pub/networks/data
[34] Leskovec J, Kleinberg J and Faloutsos C 2005 *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Chicago, August 21–24, 2005 p. 177
[35] Adamic L A and Glance N 2005 *Proceedings of the 3rd International Workshop on Link Discovery*, Chicago, August 21–24, 2005 p. 36
[36] This snapshot was created by Newman M E J from data for July 22, 2006 and is not previously published
[37] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[38] Latora V and Marchiori M 2001 *Phys. Rev. Lett.* **87** 198701
[39] Newman M E J 2002 *Phys. Rev. Lett.* **89** 208701
[40] Hanely J A and Mcneil B J 1982 *Radiology* **143** 29
[41] Tian L, Di Z and Yao H 2011 *Acta Phys. Sin.* **60** 028901 (in Chinese)