Empirical analysis of dependence between stations in Chinese railway network

Yong-Li Wang [a], Tao Zhou [b,c], Jian-Jun Shi [a], Jian Wang [a], Da-Ren He [a,*]

[a] *College of Physics Science and Technology, Yangzhou University, Yangzhou, 225002, PR China*

[b] *Department of Modern Physics and Nonlinear Science Center, University of Science and Technology of China, Hefei 230026, PR China*

[c] *Department of Physics, University of Fribourg, Chemin du Muse 3, CH-1700 Fribourg, Switzerland*

A railway transportation system can be represented by a bipartite network consisting of trains and stations, where a train is connected to all stations where it stops. In this paper, motivated by the resource-allocation process taking place on networks, we design a method to project a Chinese train-station bipartite network into a weighted station network. A new metric is proposed to quantify the dependence between pairs of stations, which is shown to follow a shifted power-law distribution. In addition, we compare the resource-allocation method and the well-known multiple-edge method, and the results indicate that our proposed method is more reasonable.

1. Introduction

After the pioneering discoveries of small-world phenomenon [1] and scale-free property [2], study of complex networks quickly becomes a joint focus of many branches of science [3–5]. Therein, transportation networks, such as the airport networks [6–8], urban public transport networks [9–11] and railway networks [12,13], have attracted increasing attention. Among them, the railway networks, having an enormous impact on the national economy and being closely related to the mobility of millions of citizens, are frequently considered in the priority projects of the governments.

Railway network (RN) is a bipartite network (see Refs. [14–19] for some typical bipartite networks), which contains two sets of nodes, stations and trains, where a train is connected to all stations where it stops. To show directly the relations between stations, we usually map this bipartite network onto a one-mode network containing stations only, where two stations are connected if they connect to at least one common train. This projected network is called the station network (SN) (see Fig. 1 for an illustration). Note that the unweighted projection loses much information. For example, a pair of stations sharing many common trains is of no difference from the case of sharing only one train. An alternative is to project a bipartite network onto a multiple graph [20], in which the number of edges connecting two stations is defined as the number of trains shared by them (see Fig. 1(c) for an illustration). It is also a straightforward way to generate a weighted station network, where the weight is directly defined as the number of multiple edges.

* Corresponding address: College of Physics Science and Technology, Yangzhou University, Si-Wang-Ting Road 180, Yangzhou, 225002, PR China. Tel.: +86 514 87931059; fax: +86 514 87975467.

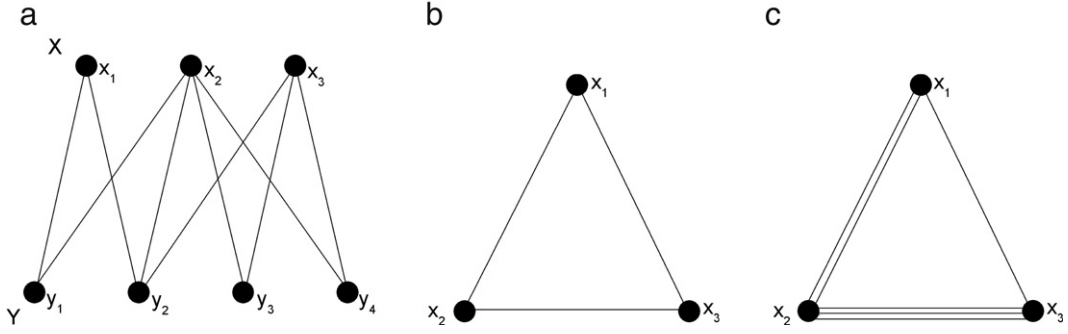*E-mail address:* darendo10@yahoo.com.cn (D.-R. He).

Fig. 1. (a) A typical railway network consists of 3 stations (X nodes) and 4 trains (Y nodes). (b) A one-mode projection onto stations, which is also a network in Space P. (c) Taking into account multiple edges.

The basic function of RN is to transport resources (e.g., passengers and goods). The dependence between two stations is not symmetrical, and usually a small station is highly dependent on a central station as a large fraction of resource it received or sent out is related to this central station. The weighted network obtained by the multiple-edge method cannot reflect the asymmetrically functional relations. Recently, Zhou et al. suggested a new weighting method based on a resource-allocation process on bipartite networks [21,22]. In this paper, we propose a modified resource-allocation method, and investigate the resulting weighted SN, with particular emphasis on the statistics of dependences between station pairs. It is worth mentioning that some scientists have suggested the use of *Space P* and *Space L* to characterize the bus transportation networks [13,23,24]. Using the language of the railway system, our projected station network is in Space P, while in Space L, two stations are connected if they are consecutive stops for at least one train. Therefore, all results in this paper are also relevant to other transportation networks in Space P.

The rest of this paper is organized as follows: In Section 2, we briefly introduce the original resource-allocation method, and make a certain modification to better describe SN. In Section 3, we investigate the dependence between stations based on the resource-allocation method. In Section 4 we define dependence by using multiple-edge method, and in Section 5 we compare the multiple-edge method with the resource-allocation method. Summary and discussion are finally presented in Section 6.

2. Resource-allocation method

A bipartite network, consisting of two node sets as $X = \{x_1, x_2, x_3 \ldots, x_m\}$ and $Y = \{y_1, y_2, y_3 \ldots, y_n\}$, can be described by an adjacent matrix A, where $a_{ij} = 1$ if x_i and y_j are connected, otherwise, $a_{ij} = 0$. Then we define x_i's degree in bipartite network as:

$$h_i = \sum_{j=1}^{n} a_{ij}, \tag{1}$$

and y_j's degree as:

$$t_j = \sum_{i=1}^{m} a_{ij}. \tag{2}$$

Without loss of generality, we consider the projection onto X nodes. The resource-allocation process consists of two steps [21]: we first equally allocate the resource of every X node to its Y neighbors, and then equally allocate the received resource of every Y node back to its neighbors in X. Merging these two steps, each X node actually gets some resource from other X nodes. Denote the initial and final resources located in X nodes by two vectors, $\vec{X}$ and $\vec{X}'$, respectively. Then the resource-allocation process can be written as $\vec{X} = W\vec{X}'$, where

$$w_{ij} = \frac{1}{h_j} \sum_{l=1}^{n} \frac{a_{il}a_{jl}}{t_l}, \quad i, j = 1, 2, \ldots, m. \tag{3}$$

The entry w_{ij} represents the ratio of resource j allocates to i to the total amount of j's resource, which can be considered as i's importance from the viewpoint of j. Accordingly, we define w_{ij} as the strength that j depends on i. Generally speaking, $w_{ij} \neq w_{ji}$ since $h_j \neq h_i$. Fig. 2 illustrates the resource-allocation method onto X nodes for Fig. 1(a), resulting in:

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} 5/12 & 5/24 & 1/9 \\ 5/12 & 11/24 & 4/9 \\ 1/6 & 8/24 & 4/9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{4}$$
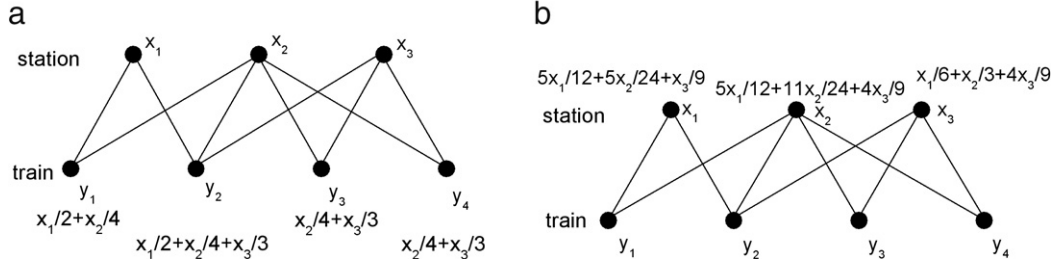
Fig. 2. Illustration of the resource-allocation process of Fig. 1(a). (a) The resource flows from stations to trains; (b) The resource returns to stations.
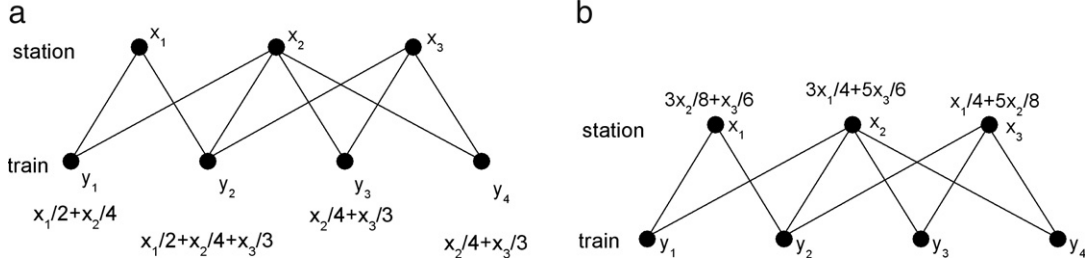

Fig. 3. Illustration of the modified resource-allocation process of Fig. 1(a). (a) The resource flows from stations to trains; (b) The resource returns to stations.

In the resource-allocation method, the nonzero diagonal element w_{ii} represents the ratio of resource i allocates to itself. This is reasonable for some networks. For example, in co-authorship networks [25–29], w_{ii} relates to the ratio of i's contribution to all his/her publications. However, in SN, one train departing from station i will not stop in i, namely i is not allowed to transport resources back to it. Therefore, we modify Eq. (3) to

$$w_{ij} = \frac{1}{h_j} \sum_{l=1}^{n} \frac{a_{il}a_{jl}}{t_l - 1} \quad (i \neq j), \quad \text{and} \quad w_{ij} = 0 (i = j). \tag{5}$$

The modified process is illustrated in Fig. 3, resulting in:

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} 0 & 3/8 & 1/6 \\ 3/4 & 0 & 5/6 \\ 1/4 & 5/8 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{6}$$

3. Statistics of dependence

The Chinese railway system consists of 2502 trains and 2978 stations (http://www.chinamor.cn.net). By using the resource-allocation method, we get a weighted matrix $W = \{w_{ij}\}_{2978 \times 2978}$, where the element w_{ij} represents how much station j depends on station i. The larger the w_{ij}, the higher fraction of j's resource will be transported to i. We calculate dependences of 80530 connected station pairs in SN, and the result shows 87% are asymmetrical. The maximal dependence is *Dali* (a rather small city near to the provincial capital *Kunming*, famous in tour) on *Kunming*, accordingly to the numerical result, 85% of *Dali*'s resource is transported to *Kunming*. Actually, *Dali* has 10 trains, 8 of which only stop in *Kunming*, and the other two stop at five stations including *Kunming*.

We first introduce the so-called *shifted power-law* (SPL) function [30], which can be expressed as:

$$P(k) \propto (k + \alpha)^{-\gamma}, \tag{7}$$

where k is the variable, α and γ are constants. This function follows a straight line with a slope $-\gamma$ on the $\ln P(k) - \ln(k + \alpha)$ plane. For $\alpha \ll k$, Eq. (7) can be approximated as

$$P(k) \propto k^{-\gamma}, \tag{8}$$

indicating a power law, while if $\alpha \gg k$, it approximately tends to an exponential distribution, as

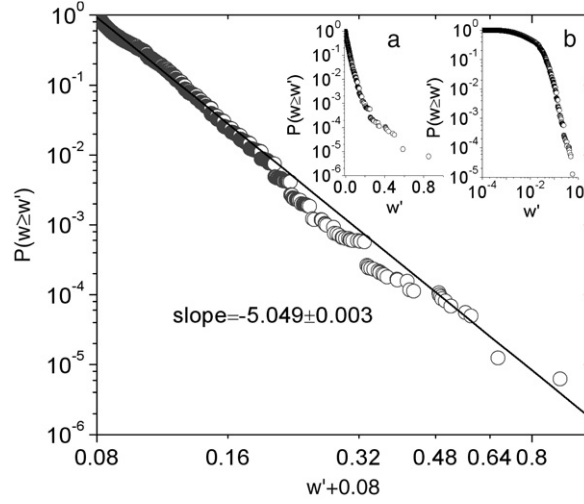$$P(k) \propto \exp(-k\gamma). \tag{9}$$

3

Fig. 4. Accumulative distribution of dependences between stations, which obeys a shifted power law. The insets show such a distribution in (a) log–linear and (b) log–log plots.

When the parameter α changes from 0 to ∞, the distribution varies from a power-law to an exponential distribution. As shown in Fig. 4, the accumulative distribution of dependences can be well fitted by an SPL function, as:

$$P(w \geq w') \propto (w' + \alpha)^{-\gamma}, \tag{10}$$

where $\alpha \approx 0.08$ and $\gamma \approx 5.05$. Note that, as shown in the insets of Fig. 4, the distribution cannot be well described by either exponential or power-law distribution.

Two stations, i and j, usually have different functional importance to each other, namely $w_{ij} \neq w_{ji}$. To measure the strength of asymmetry on an edge, we define the difference of edge as

$$d_{ij} = \left| w_{ij} - w_{ji} \right|. \tag{11}$$

Clearly, the larger the d_{ij}, the higher the edge's asymmetry. The most asymmetrical edge, corresponding to the highest difference, is the one between *Dali* and *Kunming*. The dependence of *Dali* on *Kunming* (i.e., how much *Dali* depends on *Kunming*) is 0.85, of *Kunming* on *Dali* is 0.13, so the difference is 0.72. As mentioned above, *Dali* is very dependent on *Kunming*, while *Kunming* is the capital of *Yunnan* province and does not depend much on *Dali*. Fig. 5 shows the accumulative distribution of edge differences, which can also be well fitted by an SPL function:

$$P(d \geq d') \propto (d' + \alpha)^{-\gamma}, \tag{12}$$

where $\alpha \approx 0.06$ and $\gamma \approx 4.49$. This result indicates that the SN is a very asymmetrical network from a functional viewpoint.

4. Multiple-edge method

Note that, the definition of dependence by resource-allocation method is normalized, that is, $\sum_i w_{ij} = 1$. This normalization essentially corresponds to the conservation law. Analogously, we can define a normalized dependence based on the multiple-edge method. Firstly, we calculate the number of multiple edges between i and j, as:

$$q_{ij} = \sum_{l=1}^{n} a_{il} a_{jl}, \tag{13}$$

which can be considered as the edge weight. Accordingly, the strength of a node j reads

$$s_j = \sum_{i \neq j} q_{ij}. \tag{14}$$

In the railway system, q_{ij} is the number of different trains connecting i and j. If each station equally distributes its resource, the ratio of resource j transports to i to the total amount of j's resource is:

$$e_{ij} = \frac{q_{ij}}{s_j} = \frac{q_{ij}}{\sum\limits_{i \neq j} q_{ij}} \quad (i \neq j), \quad \text{and} \quad e_{ij} = 0 \quad (i = j). \tag{15}$$

We define e_{ij} as the strength that station j depends on station i. Generally speaking, $e_{ij} \neq e_{ji}$.
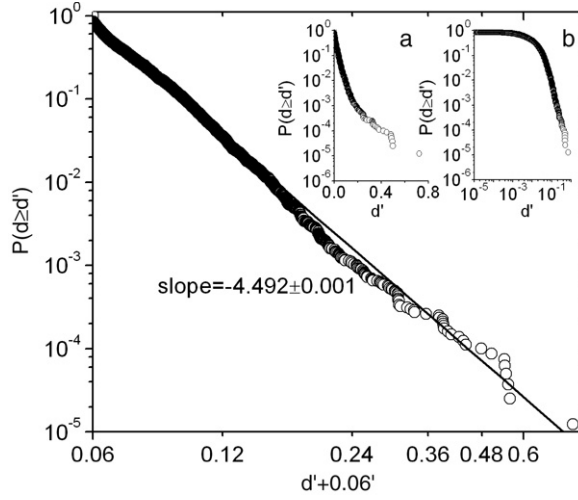
4

Fig. 5. Accumulative distribution of edge differences, which obeys a shifted power law. The insets show such a distribution in (a) log-linear and (b) log–log plots.

5. Comparison between resource-allocation and multiple-edge methods

Substituting Eqs. (13) and (2) into Eq. (14), one has

$$s_j = \sum_{i \neq j} q_{ij} = \sum_{l=1}^{n} a_{jl}(t_l - 1), \tag{16}$$

and thus

$$e_{ij} = \frac{\sum_{l=1}^{n} a_{il}a_{jl}}{\sum_{l=1}^{n} a_{jl}(t_l - 1)} \quad (i \neq j). \tag{17}$$

If t_l is a constant t, then Eq. (17) will become

$$e_{ij} = \frac{\sum_{l=1}^{n} a_{il}a_{jl}}{(t-1) \times \sum_{l=1}^{n} a_{jl}} = \frac{\sum_{l=1}^{n} a_{il}a_{jl}}{h_j \times (t-1)} \quad (i \neq j). \tag{18}$$

At the same time, Eq. (5) can be rewritten as

$$w_{ij} = \frac{\sum_{l=1}^{n} a_{il}a_{jl}}{h_j \times (t-1)} \quad (i \neq j). \tag{19}$$

Clearly, under the condition that every train stops in the same number of stations, the two methods – resource-allocation and multiple-edge – give exactly the same result. Consider two trains, denoted by T_1 and T_2, stopping in N_1 and N_2 stations, respectively. Based on the multiple-edge method, before normalization, T_1 and T_2 respectively contribute to $N_1 (N_1 - 1)$ and $N_2 (N_2 - 1)$ edges (here we consider the directions of edges). Therefore, on a statistical level, we may approximate that the weight contribution of a train is proportional to the square number of its stopping stations. In contrast, when adopting the resource-allocation method, the total weights contributed by T_1 is:

$$W(T_1) = \sum_j \frac{a_{j1}}{h_j}, \tag{20}$$

due to the fact that each station j, contributes $1/h_j$ to the train weight. Neglecting the degree–degree correlation in the bipartite train-station network, the total weight contribution of T_1 is about $N_1/\langle h \rangle$, approximately proportional to N_1, where $\langle h \rangle$ denotes the average degree of stations in the bipartite train-station networks. In a word, the multiple-edge method assigns higher weights to the trains containing many stations, compared with the resource-allocation method. However, at

Table 1
The h, $\langle t \rangle$ and $h(\langle t \rangle - 1)$ of *Beijing West Station*, *Baoji Station*, *Xinyang Station* and *Wuxi Station*.

	Beijing West	Baoji	Xinyang	Wuxi
h	241	100	99	163
$\langle t \rangle$	7.42	18.95	18.43	14.01
$h(\langle t \rangle - 1)$	1547.22	1795.00	1725.57	2120.63

least in China, a train stopping in many stations is not guaranteed to be important. Instead, it is usually a very slow train that stops in many small stations between the start and the destination.

To compare the resource-allocation method and the multiple-edge method, we concentrate on the node pairs assigned inconsistent dependences by these two methods. Accordingly, a node pair, i and j, is involved only if $w_{ij} > w_{ji}$ ($w_{ij} < w_{ji}$) while $e_{ij} < e_{ji}$ ($e_{ij} > e_{ji}$). In the present network, about 5% of connected station pairs are assigned inconsistent dependences, of which a detailed investigation may reveal the essential difference between the resource-allocation method and the multiple-edge method. We focus on one of the largest stations in China, namely the *Beijing West Station*. There are three inconsistent pairs adjacent to the *Beijing West Station*, where the three involved stations are *Baoji Station*, *Xinyang Station* and *Wuxi Station*. *Baoji*, *Xinyang* and *Wuxi* are all middle-size cities in China, and of course, they depend much on *Beijing West Station*. The resource-allocation method shows that these three smaller-size stations are much more dependent on *Beijing West Station* than the inverse, while the multiple-edge method gives the opposite result, which is in conflict with the common sense. We therefore think the resource-allocation method is more reasonable than the multiple-edge method, at least for the specific Chinese railway system. The reason is exactly what we mentioned in the last paragraph, that is, the multiple-edge method assigns unfairly high weights to the trains stopping in many stations.

We can present a further explanation for this phenomenon in a more microscopic viewpoint. On average, we can approximately rewrite Eq. (18) into

$$e_{ij} \simeq \frac{\sum\limits_{l=1}^{n} a_{il} a_{jl}}{h_j \times (\langle t^j \rangle - 1)},$$ (21)

where $\langle t^j \rangle$ represents the averaged number of stopping stations over all the trains adjacent to the station j. Then we have

$$\frac{w_{ij}}{w_{ji}} = \frac{h_i}{h_j}$$ (22)

and

$$\frac{e_{ij}}{e_{ji}} \simeq \frac{h_i}{h_j} \times \frac{\langle t^i \rangle - 1}{\langle t^j \rangle - 1}.$$ (23)

If $h_i > h_j$ while $h_i(\langle t^i \rangle - 1) < h_j(\langle t^j \rangle - 1)$, we shall see $w_{ij} > w_{ji}$, but $e_{ij} < e_{ji}$. The data of h, $\langle t \rangle$ and $h(\langle t \rangle - 1)$ of the stations are listed in Table 1, which clearly highlights the essential difference of those two methods and again indicates that the multiple-edge method prefers the trains stopping in many stations.

6. Conclusion and discussion

In this paper, based on a modified resource-allocation method, we investigated the statistics of functional dependencies between station pairs. Empirical results showed that 87% of connected station pairs are asymmetrical. We further employed a measure named edge difference to quantify the strength of asymmetry, and found that both the accumulative distributions of dependences and edge differences obey the shifted power law. Furthermore, by adding a simple normalization operation, we extended the well-known multiple-edge weighting method to an asymmetrical version. Although mostly the resource-allocation and multiple-edge methods are qualitatively in accordance with each other, the microscopic investigation on the inconsistent results indicates that the resource-allocation method is more reasonable. We deem that the multiple-edge method fails for giving unfairly high weights to the trains stopping in many stations.

Why do we distribute resources equally? Actually, to investigate the dependence between stations, the perfect conditions are: (1) Get all information about topology of the bipartite railway network. (2) Know an idiographic amount of every station's initial resource. (3) Know the idiographic ratio of resource one station transports to every passing train and one train transports to every station at which it stops. If so, we can accurately calculate the idiographic amount of resource station j transport to i. However, we often only know the topology of the railway network, and cannot get the information mentioned in (2) and (3). To overcome the difficulty in (2), we regard the initial resource of every station as a unit, and pay more attention to the ratio of resource station j transport to i to the total amount of station j's resource. Then the dependence is only a ratio. It is very crucial but hard to overcome the difficulty in (3), and one straightforward way, as in the current paper, is to assume the stations equally allocate resources to their passing trains, and trains equally allocate resources to all stations it stops.

It is worth mentioning that our studies are in Space P, thus are also relevant for bus transport networks. In addition, we can investigate the functional relationship between nodes in many other networks by the resource-allocation method, such as the Hollywood movie actor network [1,2,30,31], the traditional Chinese herb prescription formulation network [30], and so on. These investigations will help us understand in depth the functional relations between nodes in many real networks. Note that, in these real networks, the resource of one node may not be passengers or goods; it can be understood as importance, intensity, wealth, etc. For example, in the recommender systems [21,32,33], the resource diffused between users and objects can be considered as a kind of recommending capacity.

Acknowledgment

References

[1] D.J Watts, S.H. Strogatz, Nature 393 (1998) 440.
[2] A.L. Barabási, R. Albert, Science 286 (1999) 509.
[3] R. Albert, A.L. Barabási, Rev. Modern Phys. 74 (2002) 47.
[4] S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. 51 (2002) 1079.
[5] M.E.J. Newman, SIAM Rev. 45 (2003) 167.
[6] L.A.N. Amaral, A. Scala, M. Barthelémy, H.E. Stanley, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 11149.
[7] W. Li, X. Cai, Phys. Rev. E 69 (2004) 046106.
[8] A. Barrat, M. Barthelémy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 3747.
[9] V. Latora, M. Marchiori, Phys. Rev. Lett. 87 (2001) 198701.
[10] V. Kalapala, V. Sanwalani, A. Clauset, C. Moore, Phys. Rev. E 73 (2006) 026130.
[11] Y.Z. Chen, N. Li, D.R. He, Physica A 376 (2007) 747.
[12] P. Sen, S. Dasgupta, A. Chatterjee, P.A. Sreeram, G. Mukherjee, S.S. Manna, Phys. Rev. E 67 (2003) 036106.
[13] M. Kurant, P. Thiran, Phys. Rev. Lett. 96 (2006) 138701.
[14] P. Holme, F. Liljeros, C.R. Edling, B.J. Kim, Phys. Rev. E 68 (2003) 056107.
[15] J. Ohkubo, K. Tanaka, T. Horiguchi, Phys. Rev. E 72 (2005) 036120.
[16] R. Lambiotte, M. Ausloos, Phys. Rev. E 72 (2005) 066107.
[17] P.G. Lind, M.C. Gonzalez, H.J. Herrmann, Phys. Rev. E 72 (2005) 056127.
[18] E. Estrada, J.A. Rodrıguez-Velazquez, Phys. Rev. E 72 (2005) 046105.
[19] J.J. Ramasco, S.N. Dorogovtsev, R. Pastor-Satorras, Phys. Rev. E 70 (2004) 036106.
[20] M.E.J. Newman, Phys. Rev. E 70 (2004) 056131.
[21] T. Zhou, J. Ren, M. Medo, Y.C. Zhang, Phys. Rev. E 76 (2007) 046115.
[22] Q. Qu, Y.D. Jin, T. Zhou, B.H. Wang, B.Q. Yin, Phys. Rev. E 75 (2007) 021102.
[23] M. Marchiori, V. Latora, Physica A 285 (2000) 539.
[24] M. Kurant, P. Thiran, Phys. Rev. E 74 (2006) 036114.
[25] M. Li, J. Wu, D. Wang, T. Zhou, Z. Di, Y. Fan, Physica A 375 (2007) 355.
[26] M.E.J. Newman, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 404.
[27] M.E.J. Newman, Phys. Rev. E 64 (2001) 016131.
[28] M.E.J. Newman, Phys. Rev. E 64 (2001) 016132.
[29] M.E.J. Newman, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 5200.
[30] H. Chang, B.B. Su, Y.P. Zhou, D.R. He, Physica A 383 (2007) 687.
[31] P.P. Zhang, K. Chen, Y. He, T. Zhou, B.B. Su, Y. Jin, H. Chang, Y.P. Zhou, L.C. Sun, B.H. Wang, D.R. He, Physica A 360 (2006) 599.
[32] Y.C. Zhang, M. Medo, J. Ren, T. Zhou, T. Li, F. Yang, Europhys. Lett. 80 (2007) 68003.
[33] T. Zhou, L.L. Jiang, R.Q. Su, Y.C. Zhang, Europhys. Lett. 81 (2008) 58004.