Available online at www.sciencedirect.com
ScienceDirect
Physica A 378 (2007) 583–590
PHYSICA A
www.elsevier.com/locate/physa

The effect of weight on community structure of networks

Ying Fan[a], Menghui Li[a], Peng Zhang[a], Jinshan Wu[b,*], Zengru Di[a]

[a]*Department of Systems Science, School of Management, Beijing Normal University, Beijing 100875, PR China*
[b]*Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada V6T 1Z1*

Received 2 October 2006; received in revised form 21 November 2006
Available online 23 December 2006

Abstract

The effect of weight on community structures is investigated in this paper. We use weighted modularity Q^w to evaluate the partitions and weighted extremal optimization algorithm to detect communities. Starting from empirical and idealized weighted networks, the matching between weights and edges are disturbed. Then using similarity function S to measure the difference between community structures, it is found that the redistribution of weights does strongly affect the community structure especially in dense networks. This indicates that the community structure in networks is a suitable property to reflect the role of weight.

Keywords: Weighted network; Community structure; Role of link weight

1. Introduction

Link weights, as strength of the interaction represented by networks, are believed to be an important variable in networks. It gives more information about networks besides its topology properties dominated by links. Recently more and more study in complex networks focus on the weighted networks. The problems involve the definition of weight and other quantities which characterize the weighted networks [1–3], the empirical studies of its statistical properties [4–7], evolving models [8–13], and transportation or other dynamics on weighted networks [14–17].

However, how important is the weight, or what significant changes on network structures are induced when weight is changed? This question is related with the role of weight. It should be a fundamental question in the study of weighted networks. But it has not been investigated deeply in the previous studies.

The role of weight should be first investigated by analyzing the correlation between link weight and other properties. In this way, it attempt to answer the question that whether there is some internal mechanism strongly determining weights or not. For example, one may image that link betweenness affects link weight largely because the larger link betweenness implies that the link has more important role in communication on networks, so that the weight on the edge might be also larger. If this is true, the weight should be less

*Corresponding author.
E-mail address: jinshanw@phas.ubc.ca (J. Wu).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2006.12.021

important for the networks because the weight is dependent on links or the weight is determined by the network topology.

In some networks [18,19], the average link weight is proportional to the product of end-point degrees as $\langle w_{ij}\rangle \sim (k_i k_j)^{\theta}$. For instance, link weight depend on the end-point degrees in *E. coli* metabolic network, where the link weights represent the optimal metabolic fluxes [18], and in World Airport Network, where the link weights represent the number of available seats on any given connection for the year 2002 [19]. There are actual flux on these two networks, node degrees could affect link weights by the flux. Here we have investigated such correlations in some social networks. The results are different from metabolic network and World Airport Network. For example, the link weights and the vertex degrees of the scientist collaboration network lack correlations [19].

For the database of our scientific collaboration networks of Econophysicists (EP-SCN) [3], BNU-Email network [20], and Rhesus monkey societies [2,21], based on the standard method for linear correlation analysis in mathematical statistics, we get the correlation results for the link weights and link betweenness. The results are shown in Table 1. All the coefficients of correlation for links are less than 0.25. These negative results reveal that the weight is really an independent variable for social networks. By the way, we also get the correlation coefficients for the vertex weight (strength of vertex) and its degree. They are 0.79, 0.44 and 0.71 for above three networks. The results are rational because both strength and degree are quantities over vertex and related with the number of edges connected onto. In addition, we investigate the correlation between link weight w_{ij} and product of the degrees of the end-point nodes k_i and k_j. As shown in Fig. 1, we found no visible correlation between link weight w_{ij} and the end-point degree $k_i k_j$.

From the above negative conclusion on correlation analysis, we know that link weight is an independent variable at some level. This makes the work on the role of weight more attractive: since it is somehow independent, then how significant is it?

We are going to discuss its significance by considering the difference of network quantities when edge weights are disturbed. First, we can get rid of weights to get an binary corresponding network and compare it

Table 1
Correlation coefficients for weight and other quantities

		EP-SCN	BNU-Email	Monkey
Links	Weight-betweenness	0.0055	0.028	0.19
Vertices	Strength-degree	0.79	0.44	0.71

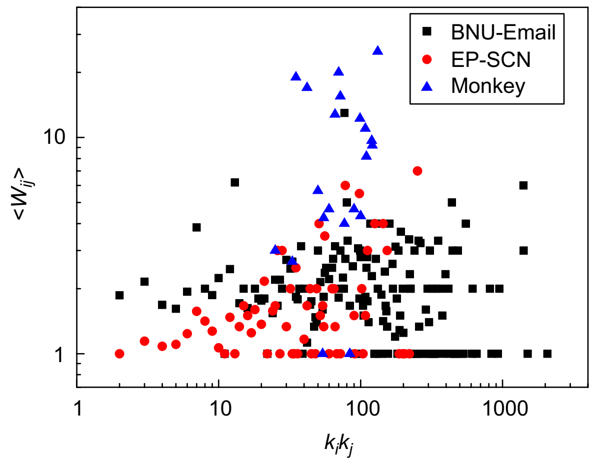

Fig. 1. Scatter plot of link weight and the product of end-point degree. There is no visible correlation and linear analysis between $\ln(w_{ij})$ and $\ln(k_i k_j)$ gives us correlation coefficient 0.0.

with the original weighted network. Besides comparing weighted network and its binary correspondence, we can also study the effects of weight by redistributing weights onto different edges to mix the matching between weights and edges [7]. But our previous investigation shows that this redistribution has little effect on the single vertex statistics, neither significant effect on distance [7]. This negative conclusion seems that weight does not have significant effect on network structures. However, we strongly suspect that it is only because we have not found the proper measurement to present its effect.

The effects of weight on the network structures can be investigated on two classes of properties: single vertex statistics and correlation statistics. The former includes vertex based properties such as degree, clustering coefficient, and the latter includes global properties such as distance, betweenness and especially community structure. An analogy between networks and condensed matter may give us some clues for insightful investigation. In condensed matter, sometimes an effective single electron picture is well enough for a large number of phenomena. An effective field is used to represent effect from all electrons and lattice ions in condensed matter. However, there are something beyond this single particle scheme so that it requires to consider the correlation between electrons. Single vertex statistics naturally belong to the former class. But community structure measures directly the correlation among vertices. Therefore, in this work, we will use the community structure, a statistics on correlation between vertices, as a measurement on the role of weight.

In binary networks, community structure is groups of network vertices where links within same group are much denser than links between different groups [22]. For weighted network, as we have mentioned above, link weight is somewhat an independent variable and should have some important effects on structure of networks. As for community structures, the definition of the community must integrate links with link weights. Newman has generalized the modularity Q to weighted modularity Q^w for evaluation the partitions of weighted networks [2]:

$$Q^w = \frac{1}{2T}\sum_{ij}\left[w_{ij} - \frac{T_i T_j}{2T}\right]\delta(c_i, c_j), \tag{1}$$

where w_{ij} represents the weight in edge between nodes i and j, T_i is the weight of node i: $T_i = \sum_j w_{ij}$, T is the sum of all link weight in networks, and c_i shows that vertex i belongs to community c_i. It takes both links and link weights into account. Usually, groups separated with the link weights should be different from the result based only on topological linkage. Given the same topological structure, different assignments of link weights may result in different community structures. Our basic question is how will the community structure change when the weights are disturbed.

There are several questions that should be answered before the realization of the above ideas. First, what are the networks for this investigation. Our previous analysis uses networks of Econophysicists as our typical networks [3]. Recently, we have got more datas on BNU-Email networks and Rhesus monkey societies [2,21]. Hopefully, dense weighted networks will give us more affirmative conclusions. Besides these real networks, we can also construct idealized *ad hoc* weighted networks for our investigation. Second, how to extract and to evaluate community structure from a given network. Currently, there are several typical algorithms in the literature: Hierarchical Clustering [23,24], betweenness based GN algorithm [25], Potts model based algorithm [26], Extremal Optimization (EO) algorithm [27], and so on. We have investigated the performance of several approaches, and found weighted extremal optimization (WEO) is a reliable algorithm [28]. So in this paper, we use WEO to detect the community structure. WEO is an extend case of EO algorithm [27]. The approach of WEO is directly related with the definition of weighted modularity Q^w. In WEO algorithm, the contribution of node i to the weighted modularity is defined as $\lambda_i^w = q_i^w/T_i = T_{r(i)}/T_i - a_{r(i)}^w$, where $T_{r(i)}$ is the summation of link weight that a node i belonging to a community r has with nodes in the same community, $a_{r(i)}^w$ is the fraction of summation of vertex weight of community r, and T_i is the vertex weight of node i. The process of detecting community structure by WEO algorithm is same to EO algorithm [27]. It performs well in weighted networks. Then the third, how to compare different community structures among the same set of vertices. We have proposed similarity function S in Ref. [20] to measure the difference between partitions. The function S integrates the information about the proportion of nodes co-appearance in pair groups of A, B and the total number of communities. Starting from two community structure $\{A_1, A_2, \ldots, A_K\}$ and $\{B_1, B_2, \ldots, B_M\}$ over the same set N, firstly, we need to identify the correspondence between As and Bs. This is to calculate the similarities of all pairs of A_i and B_j,

$$s_{ij} = \frac{|A_i \cap B_j|}{|A_i \cup B_j|}, \tag{2}$$

and then match the pairs with highest similarities. Then, for each pair of groups, after necessary re-ordering, the similarity of A_j and B_j is recorded as s_j. When the two partition has different number of communities, say, $k > M$, some different A_j will match to the same B_j. But this can be took care of automatically. Then the total similarity can be calculated as

$$S = \frac{\sum_{i=1}^{\max(K,M)} s_i}{\max(K,M)}. \tag{3}$$

Here, we use the similarity function S to quantify the difference of different partitions.

The paper is organized as follows. In Section 2, based on empirical data of several weighted networks, we compared community structures of weighted and corresponding binary, disturbed weighted networks. The results demonstrate that the weight has effects on communities, especially in dense networks. In Section 3, we did the same investigation but on idealized *ad hoc* weighted networks. The results give us systematic view about the effects of weight on community structures. Finally, we give some conclusion remarks.

2. Community structures in real and binary, inverse weighted networks

In this section, we focus on the effect of weight on community structure in real weighted networks by comparing the communities of empirical weighted networks, corresponding binary networks, and inverse weighted networks. Here inverse weighted network means the matching between edges and edge weights are inverted so that the edge with highest weight now get the lowest weight. Empirical networks include Econophysicists collaboration network [3], BNU-Email network and Rhesus monkey network [2,21]. For detecting and comparing community structure, we take the largest connected cluster of above networks. For the Econophysicists collaboration network, it includes 271 nodes and 371 edges. In order to distinguish the network with different proportion of possible links, we define the denseness of network as the ratio of existing links to all possible links among the nodes. The denseness for Econophysicists collaboration network is 0.01. The database for BNU-Email network includes the times of Emails between any two mailboxes (*@bnu.edu.cn) in a week. The network includes 740 nodes and 1400 links. We also use its largest cluster, which includes 620 nodes and 1117 links. The denseness of BNU-Email network is 0.006. The Rhesus monkey network include 16 nodes and 69 edges. Link weight w_{ij} is the total number of instances of grooming of each monkey by each other during the period of observation, $w_{ij} = w_{i\to j} + w_{j\to i}$, where $w_{i\to j}$ represents the number of i groomed j. It is a connected network with denseness equals 0.575. So it is a relatively dense network.

As mentioned in the Introduction, besides considering the binary and weighted networks, an important way to investigate the effects of weight is to study the impact of disturbing weight to the network properties. We have introduced the way to re-assign weights onto edges with certain probability p for weighted networks [7]. There we defined two special cases that $p = 1$ represents the original weighted network in a decreasing order of link weight,

$$W(p=1) = (w_{i_1j_1} = w^1 \geqslant w_{i_2j_2} = w^2 \geqslant \cdots \geqslant w_{(i_L)(j_L)} = w^L), \tag{4}$$

and $p = -1$ is defined as the inverse order,

$$W(p=-1) = (w_{i_1j_1} = w^L \leqslant \cdots \leqslant w_{(i_{L-1})(j_{L-1})} = w^2 \leqslant w_{(i_L)(j_L)} = w^1). \tag{5}$$

For this empirical networks, we compare only the community structures of original, binary and inverse weighted networks.

We apply WEO algorithm 20 times for each network. Then the community structure is shown by the corresponding co-appearance matrix. Element of this matrix Co_{ij} corresponds to the fraction of times that nodes i and j belong to the same group over all those 20 times of running. In our plots it is represented by the grey scale at the position (i,j). The label of vertices in the original one is also used in the plot of the other two: binary and inverse networks. For a given network, we can also find its final communities by the most probable partitions. Then the difference between any pair of partitions is given by the dissimilarity function S. In Table 2, we show the comparison of the communities which formed in original, binary and inverse weighted networks.

First, we show the comparison on the Monkey network (Fig. 2). Both Fig. 2 and similarity measure S shows a big difference between binary/inverse and original networks. Then we did the same investigation on network of Econophysicists (Fig. 3) and BNU-Emailbox (Fig. 4). Although there is not such big difference as in Monkey network, we still can see that in binary and inverse networks some nonzero co-appearances appear among the clearly divided vertices in the original network. From Figs. 2–4, we notice that some vertices are not in the same group as in the original networks could be in the same group in the binary and inverse networks. As shown in Table 2, we measured the similarity between community structure of original and binary/inverse networks. We can see that the community structures are quite different. This indicates that link weight is an important variable. Considering link weight may help us to comprehend and analyze the characteristics of systems better. It shows a even larger difference in Monkey network than the others. We believe the reason is because the Monkey network is much denser. Intuitively this may be understood in the following way. When the topological connection is dense it is hard to distinguish vertices therefore the

Table 2
Similarities between communities of original and binary, inverse networks

	EP-SCN	BNU-Email	Monkey
S (original-binary)	0.63	0.42	0.33
S (original-inverse)	0.54	0.31	0.25

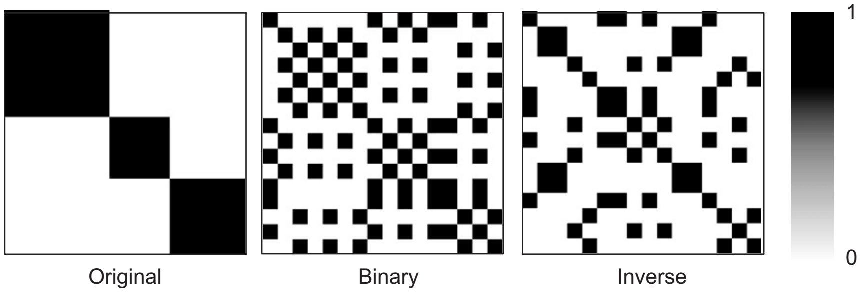

Fig. 2. The normalized co-appearance matrix for original and corresponding binary, inverse weighted Rhesus monkey networks. Community structure of those three networks looks quite different. The grey scale of the position (i, j) corresponds to the fraction of times that nodes i and j belong to the same partition.

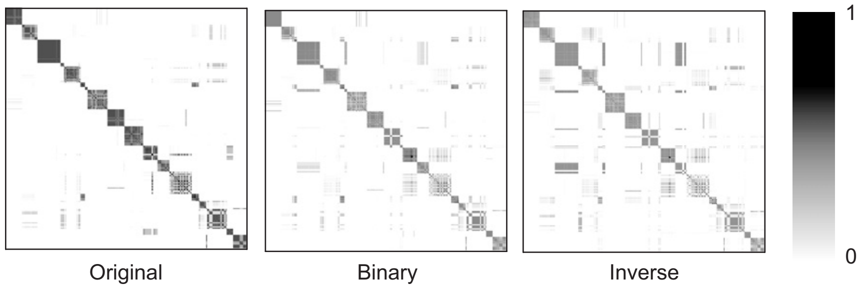

Fig. 3. The normalized co-appearance matrix for original, corresponding binary, and inverse weighted Econophysicists collaboration networks. Some nonzero co-appearances appear among the clearly divided vertices in the original network.

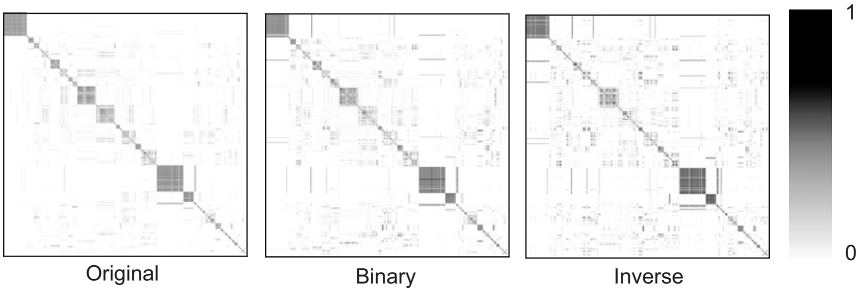

Fig. 4. The normalized co-appearance matrix for original and corresponding binary, and inverse weighted BNU-Email networks. Some nonzero co-appearances appear among the clearly divided vertices in the original network.

community structure is fuzzy. In that case, including information from link weight will help to distinguish the vertices so that we can get much clearer community structure.

3. Results on idealized weighted networks

In order to get more systematic conclusions about the effects of weight, we did the same analysis on more idealized *ad hoc* weighted networks. The idealized networks is firstly introduced by Newman [25] and used by many other authors [26,27,29]. Here we set each network with $n = 128$ vertices divided into four groups of 32 nodes. Vertices are assigned to groups and are randomly connected to vertices of the same group by an average of $\langle k_{intra}\rangle$ links and to vertices of different groups by an average of $\langle k_{inter}\rangle$ links with constrain $\langle k_{intra}\rangle + \langle k_{inter}\rangle = 16$. Generally $\langle k_{intra}\rangle > \langle k_{inter}\rangle$. While $\langle k_{intra}\rangle$ decreases, the difference between vertices inside and outside a group become smaller. Therefore the communities become more diffuse. For a given network topology, here we assign similarity weight to each link. The intragroup link weight is assigned as w_{intra}, while the intergroup link weight is assigned as w_{inter}. In practise, the relationship among the nodes in groups is usually much closer than the relationship between groups. So w_{intra} is normally bigger than w_{inter}, with a constrain

$$\langle w_{intra}\rangle + \langle w_{inter}\rangle = 2, \tag{6}$$

where $\langle w_{intra}\rangle$ ($\langle w_{inter}\rangle$) is the average of all intragroup (intergroup) link weights.

We use *ad hoc* networks with uniform distribution of link weights here. For a given network topology with certain $\langle k_{inter}\rangle$, weights are taken randomly from a 0.5 interval around $\langle w_{intra}\rangle$ and $\langle w_{inter}\rangle$, respectively, for intragroup connections and intergroup connections. That is $[\langle w_{intra}\rangle - 0.25, \langle w_{intra}\rangle + 0.25]$ and $[\langle w_{inter}\rangle - 0.25, \langle w_{inter}\rangle + 0.25]$, respectively. In the following simulations, we take $\langle w_{intra}\rangle = 1.6$ and so that $\langle w_{inter}\rangle = 0.4$.

Now we exam the effects of weight on community structures based on idealized weighted networks. Here, instead of presenting the co-appearance matrix, which is too large to fit in here, we only show value of S, the similarity between difference structures. First, we applied WEO method to the binary and weighted *ad hoc* networks. Then both community structures are compared with the presumed communities to calculate the similarity function S. We see that in both curves of Fig. 5, beyond certain value of $\langle k_{inter}\rangle$, S decreases sharply. Such value is about 6 and 10 for binary and weighted network, respectively. This means that though it is hard to tell the right community structure for binary networks in the regime, the community structure of weighted networks is still clear enough for the algorithm to discover it. That is the regime $\langle k_{inter}\rangle \in [6, 10]$. Intuitively, we can understand this in the following way. When we can only make use of the information about existence of link, not its weight, then when $\langle k_{inter}\rangle > 6$, it is almost indistinguishable between intra and inter group links. However, if further information such as edge weights can be taken into account, we will still be able to tell the difference between intra and inter links. This means that when the community structure based on links are fuzzy, the role of link weight is more obvious. Here it shows the effect of weight on community structure from a different but related point view. All the results in this section are the average of 20 network realizations and 10 runs each. Fig. 5 also tell us about the difference between binary and weighted network. Both of them

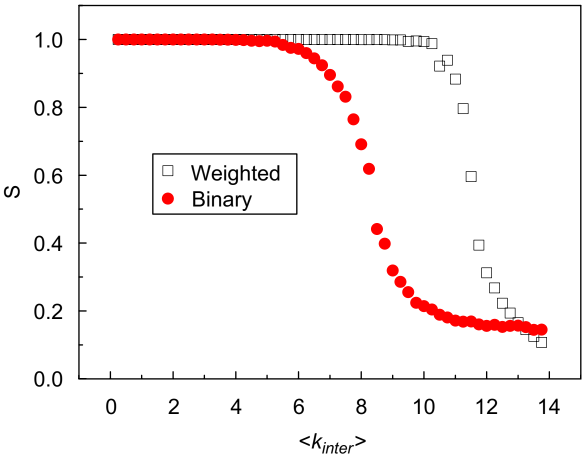

Fig. 5. The similarity between groups found by WEO algorithm and the presumed community structure. Community structure of the binary networks becomes indistinguishable when $\langle k_{inter} \rangle$ beyond about 6, while the one of weighted networks are still clear enough to detect by the algorithm for $\langle k_{inter} \rangle$ up to 10. This clearly shows the effect of weight on community structures. And this diagram also tell us about the difference between binary and weighted network. Both of them compared with the presumed structure. We see before $\langle k_{inter} \rangle = 6$ they both agree with the presumed structure, and after $\langle k_{inter} \rangle = 10$ they both do not agree with it. And in the middle, we see one agrees but the other not, so there it is a big difference between those two. Therefore, in Fig. 6 only difference between weighted and inverse networks are included.

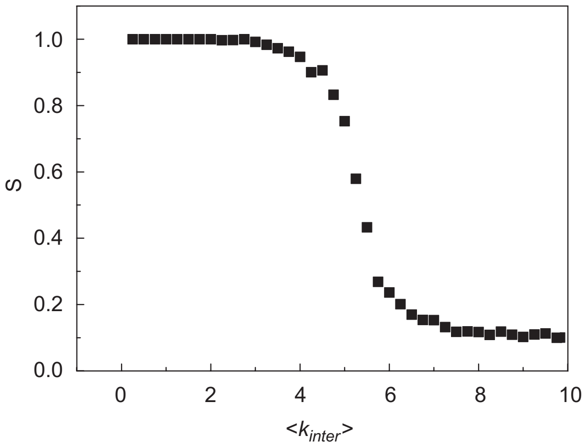

Fig. 6. The similarity between original weighted and the inverse networks. The communities are found by WEO algorithm. We only listed comparison in the regime WEO could detect the right structure of the original weighted networks. We can see the difference between community structure of weighted and inverse networks are very large when $4 < \langle k_{inter} \rangle < 10$, where weights becomes important for community structure.

compared with the presumed structure. We see before $\langle k_{inter} \rangle = 6$ they both agree with the presumed structure, and after $\langle k_{inter} \rangle = 10$ they both do not agree with it. And in the middle, we see one agrees but the other not, so there it is a big difference between those two.

In Fig. 6, we come back to our former way to present the difference among original and inverse networks, the similarity measure S. It is interesting to notice that S is around 1 when $\langle k_{inter} \rangle$ is small although it decreases gradually up to 0 with the increasing of $\langle k_{inter} \rangle$. This reveals that both link and link weight are two factors that determine the structure of networks. In other words, link weight have important function for some networks having fuzzy community structures. Especially in dense networks such as Rhesus monkey network, link weight is crucial to the network structures. When the network is sparse topological linkage is likely dominant while for dense network link weight is more crucial.

4. Conclusion

It is well believed that weight plays important role in networks. However, it has not been shown that a slight change in weight will significantly change the structure of network. Furthermore, one even does not know how to measure such difference. In this work, the influence of the weight on community structures is used to investigate the role of weight. We discussed both empirical networks including Rhesus monkey, Econophysicist collaboration and BNU-Email network, and idealized *ad hoc* weighted networks. We compare the original networks with its binary correspondence and the inverse weighted networks. It is found that weight do have big influence on communities structure, especially on dense networks. Effects of weight on network quantities other than community structure can also be discussed. Or one may discuss its effect by disturbing the weight distribution and comparing behaviours of physical models.

Acknowledgements

The authors want to thank Dr. Newman for his cooperation data. This work is partially supported by 985 Project and NSFC under the Grant Nos. 70431002, 70371072 and 70471080.

References

[1] M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Physica A 346 (2005) 34–43.
[2] M.E.J. Newman, Phys. Rev. E 70 (2004) 056131.
[3] Y. Fan, M. Li, J. Chen, L. Gao, Z. Di, J. Wu, Int. J. Mod. Phys. B 18 (17–19) (2004) 2505–2511.
[4] E. Almaas, P.L. Krapivsky, S. Redner, Phys. Rev. E 71 (2005) 036124.
[5] G. Bianconi, arXiv:cond-mat/0412399.
[6] W. Jezewski, Physica A 354 (2005) 672–680.
[7] M. Li, Y. Fan, J. Chen, L. Gao, Z. Di, J. Wu, Physica A 350 (2005) 643–656.
[8] S.H. Yook, H. Jeong, A. Barabasi, Y. Tu, Phys. Rev. Lett. 86 (2001) 5835–5838.
[9] A. Barrat, M. Barthelemy, A. Vespignani, Phys. Rev. Lett. 92 (2004) 227801.
[10] T. Antal, P.L. Krapivsky, Phys. Rev. E 71 (2005) 026103.
[11] A. Barrat, M. Barthelemy, A. Vespignani, Phys. Rev. E 70 (2004) 066149.
[12] S.N. Dorogovtsev, J.F.F. Mendes, arXiv:cond-mat/0408343.
[13] W.-X. Wang, B.-H. Wang, B. Hu, G. Yan, Q. Ou, Phys. Rev. Lett. 94 (2005) 188702.
[14] G. Bagler, arXiv:cond-mat/0409773.
[15] K.I. Goh, B. Kahng, D. Kim, Phys. Rev. E 72 (2005) 017103.
[16] K.I. Goh, J.D. Noh, B. Kahng, D. Kim, Phys. Rev. E 72 (2005) 017102.
[17] G. Yan, T. Zhou, J. Wang, Z. Fu, B. Wang, Chin. Phys. Lett. 22 (2005) 501.
[18] E. Almaas, B. Kovács, T. Vicsek, Z.N. Oltvai, A.-L. Barabási, Nature 427 (2004) 839.
[19] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. USA 101 (2004) 3747.
[20] P. Zhang, M. Li, J. Wu, Z. Di, Y. Fan, Physica A 367 (2006) 577–585.
[21] D.S. Sade, Folia Primatologica 18 (1972) 196–223.
[22] M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. USA 99 (2002) 7821–7826.
[23] M.E.J. Newman, Eur. Phys. J. B 38 (2004) 321–330.
[24] J. Scott, Social Network Analysis: A Handbook, 2nd ed., Sage, London, 2000.
[25] M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. USA 99 (2002) 7821–7826.
[26] J. Reichardt, S. Bornholdt, Phys. Rev. Lett. 93 (2004) 218701.
[27] J. Duch, A. Arenas, Phys. Rev. E 72 (2005) 027104.
[28] Y. Fan, M. Li, P. Zhang, J. Wu, Z. Di, Physica A (2007), in press, doi:10.1016/j.physa.2006.11.036.
[29] L. Danon, A. Díaz-Guilera, J. Duch, A. Arenas, J. Stat. Mech. (2005) P09008.