COMMUNICATIONS IN COMPUTATIONAL PHYSICS
Vol. **3**, No. 3, pp. 724-733

Commun. Comput. Phys.
March 2008

Spatial Correlation Function in Modular Networks

Ke Ping Li* and Zi You Gao

State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China.

Received 2 June 2007; Accepted (in revised version) 16 July 2007

Communicated by Dietrich Stauffer

Available online 27 November 2007

Abstract. Due to the complexity of the interactions among the nodes of the complex networks, the properties of the network modules, to a large extent, remain unknown or unexplored. In this paper, we introduce the spatial correlation function G_{rs} to describe the correlations among the modules of the weighted networks. In order to test the proposed method, we use our method to analyze and discuss the modular structures of the ER random networks, scale-free networks and the Chinese railway network. Rigorous analysis of the existing data shows that the spatial correlation function G_{rs} is suitable for describing the correlations among different network modules. Remarkably, we find that different networks display different correlations, especially, the correlation function G_{rs} with different networks meets different degree distribution, such as the linear and exponential distributions.

PACS: 89.75.Fb, 89.75.Da

Key words: Correlation function, weighted networks, modular structure.

1 Introduction

In many real world networks, it was found that there exist significant subnetworks (so-called network modules) in their structures, such as the metabolic networks [1], food webs [2], social networks [3] and Internet [4]. Many theoretical and experimental results indicate that the network modules perform specific tasks in the functional properties of such networks. A major current challenge is to understand their topological structures and the roles they playing in the networks. What are the mechanisms by which network modules emerge in the network? How to describe the interactions among the network modules?

*Corresponding author. *Email addresses:* `kpli44@188.com` (K. P. Li), `zygao@center.njtu.edu.cn` (Z. Y. Gao)

Recently, a number of methods have been proposed to understand the structure properties of complex networks, such as the statistical, mathematical and model-based analysis methods. Newman and Girvan proposed the clustering algorithm in [5, 6], where the quantitative definition of modularity was firstly introduced. In [7], the network is mapped into a spin system. Here, the problem of finding the modularity of a network is analogous to the standard statistical mechanism problem. Donetti and Muñoz proposed an algorithm that combines spectral methods with clustering techniques [8]. Ziv, Middendorf and Wiggins outlined an information theoretical algorithm by applying the information bottleneck on probability distributions [9]. In [10], a weighted network is proposed to model and simulate the Dutch railway network. For a review, see [11, 12].

Correlation is an important property of networks, which is of special interest [13–15]. For example, whether gregarious people are more likely to contact with gregarious people? or whether an old web site is more likely to connect to old ones? These questions often have fundamental importance in reality. The correlation function provides a tool to give a better insight into the problem mentioned above. In this paper, we introduce a new correlation function and use it to capture an explicit and obvious relationship between the network modules and that have highly connected nodes. The main aim is to get a better understanding of the network modules and their correlations. The paper is organized as follows: we introduce the proposed method in Section 2; the numerical and analytical results are presented in Section 3; finally, conclusions are presented.

2 The proposed method

The appearance of the network modules represents a broad range of natural phenomena. To describe these phenomena requires an understanding of the basic topological structures of such networks. These topological structures are based on the links of the nodes. A reasonable assumption is that the correlations among the network modules are describe by the links of the nodes on a coarse-grained level. In this paper, to assume that the modular structures of networks have been determined in advance, we describe the properties of the network modules and their correlations by introducing the spatial correlation function G_{rs}.

Our method is as follows. A weighted network which has N nodes is considered. Let w_{ij} be the weight of the edge that links the node i and the node j. We introduce the spatial correlation function $G_{rs}(i)$. Assume that the node i is within the module r and the node j is within the module s. Then the definition of $G_{rs}(i)$ is as follows:

$$G_{rs}(i) = \frac{1}{n_s} \sum_{j=1}^{N_s} d_i d_j w_{ij}, \tag{2.1}$$

where n_s represents the number of the links between the node i and the nodes within the module s, N_s represents the number of the neighbours of the node i and d_i is the degree of the node i. The degree d_i of the node i is defined as the number of edges linked to the node i. According to the definition of $G_{rs}(i)$, if the $G_{rs}(i)$-value is higher, the node i has more correlations with the nodes that are within the module s.

In complex networks, the variations of the existing weights induced by new links can be represented by the following rule [16],

$$w_{ij} \longrightarrow w_{ij} + \delta \frac{w_{ij}}{s_i}. \tag{2.2}$$

Here δ is the fraction of weight which is induced by the new link onto the others, and s_i is called the node strength, which is defined as [17,18]

$$s_i = \sum_{j=1}^{d_i} w_{ij}. \tag{2.3}$$

In the proposed method, the variations of the existing weights are caused by the values of weights. By assuming that these variations comply with the same rule described by Eq. (2.2), we obtain

$$G_{rs}(i) \longrightarrow G_{rs}(i) + \frac{1}{n_s} \sum_{j=1}^{N_s} d_i d_j \delta \frac{w_{ij}}{s_i}. \tag{2.4}$$

Here δ is the fraction of weights rearranging between the node i and its neighbors j, which is induced by the variations of weights with the vertex i. In addition, in the proposed method, the parameter δ can be chosen by experience, or be determined by numerical simulations. From Eq. (2.4), we can measure how the spatial correlation function $G_{rs}(i)$ varies with the time t.

3 Numerical simulations

In order to test the proposed method, as different examples, we analyze and discuss the modular structures of the ER random networks, scale-free networks and the Chinese railway network. In Subsection 3.1, we first construct the ER random networks and scale-free networks, with weights 1, and then let the weights vary with time on such networks. In Subsection 3.2, as a reality network, the Chinese railway network will be considered. Its structure is fixed in advance, and its weights are calculated by the Chinese railway time table.

3.1 ER random networks and scale-free networks

Both the numerical and analytical results indicated that random networks and scale-free networks have modularity [7]. In order to test the performance of the method proposed in this paper, we analyze and discuss the spatial correlation functions in the Erdös-Rényi (ER) random networks and scale-free networks. Using the proposed method, we need to detect the modular structure of weighted networks in advance. Here we firstly construct the ER random networks and scale-free networks, and then determine their modular structure using the method proposed by Newman and Girvan [5,6].

An ER random network having 40 nodes is considered. Each node is linked to other nodes with probability p. To each link, we assign a weight w_{ij}=1. Using the method proposed by Newman and Girvan, the considered network is divided into two modules. We select that module which has the largest number of nodes as our investigative object. By averaging the values of $G_{rr}(i)$ over the nodes with degree k in the module r, we obtain the degree distribution $P_r(k)$. Fig. 1 shows the correlation degree distribution $P_r(k)$ of the selected module. From Fig. 1, it can be seen that several nodes which have higher degrees have higher values of $P_r(k)$. The simulation result indicates that the module has a "hub-like" core. Several nodes within the "hub-like" core have more correlations with other nodes.

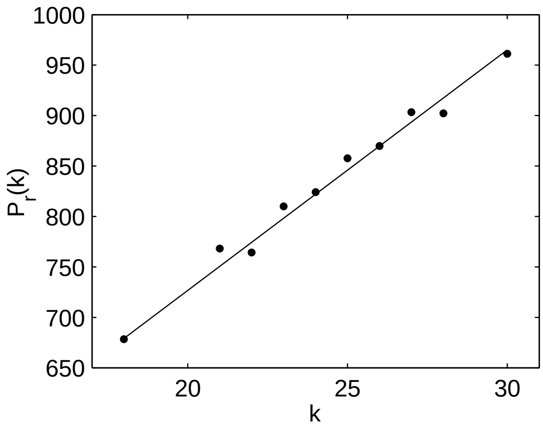

Figure 1: A plot of $P_r(k)$ vs k for p=0.4.

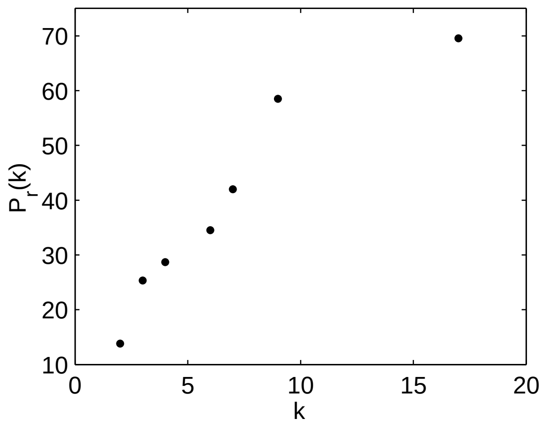

Figure 2: A plot of $P_r(k)$ vs k for m_0=4 and m=2.

As the second example, we consider a scale-free network which has 40 nodes. Using the similar steps mentioned above, we firstly construct a scale-free network. We choose m_0 nodes as an initial set of nodes. When a new node is added to the set, m edges are linked between the new node and the existing nodes. This procedure is iterated until the number of nodes reached a given number. To each link, we assign a weight w_{ij}=1. We select the module having the largest number of nodes as our investigative object. Fig. 2 shows how the correlation degree distribution $P_r(k)$ of the selected module varies with the degree k. From Fig. 2, it can be seen that there is a similar "hub-like" core observed in the ER random networks.

For each node i within the selected module r, we measure the correlation function $G_{rs}(i)$ $(r \neq s)$. The measured results are shown in Fig. 3. Fig. 3(a) is the result for the ER random network, and Fig. 3(b) is the result for scale-free network. Here the symbols "square" are the values of G_{rs} $(r \neq s)$, and the symbols "cross" are the values of G_{rr}. From Fig. 3(a), we observe that the data which have higher G_{rs} have higher G_{rr}. In Fig. 3(b), there is only a fraction of data whose G_{rr} is greater than zero. The nodes whose G_{rs} are

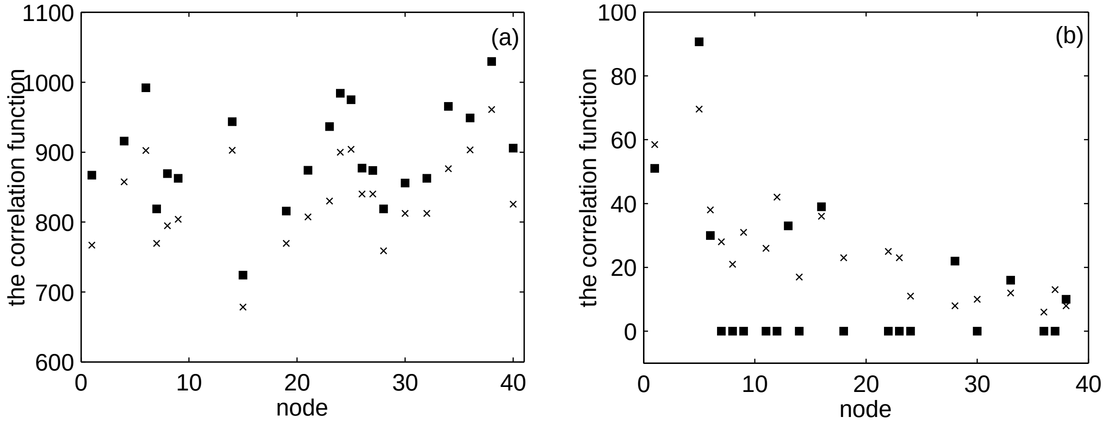

Figure 3: The correlation function $G_{rs}(i)$ $(r \neq s)$. (a) ER random network; (b) scale-free network. Here '■' and 'x' represent G_{rs} and G_{rr}, respectively.

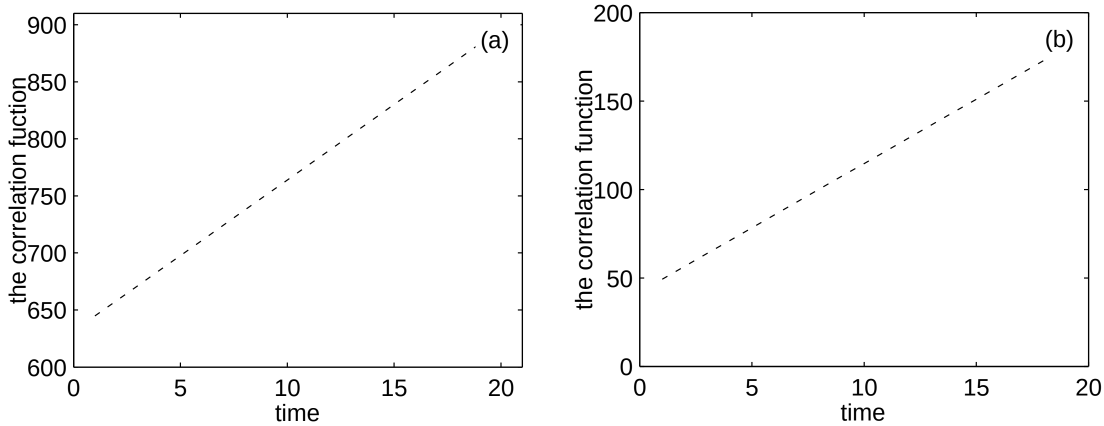

Figure 4: Time dependence of correlation function $\bar{G}_{rs}$ for δ=0.5. (a) ER random network; (b) scale-free network.

equal to zero are not directly correlated with the nodes which belong to other modules. The simulation results demonstrate that the ER random networks and the scale-free networks have different correlations. In the ER random networks, most of nodes participate in different modules at the same time. But only a fraction of nodes participate in different modules in the scale-free networks.

In order to investigate how the correlations among different modules vary with the time t, we further define the function $\bar{G}_{rs}$. This average value $\bar{G}_{rs}$ is calculated by averaging $G_{rs}(i)$ within the module r. In the beginning, each link is assigned an initial weight w_{ij}=1. As the time proceeds, the weights of all links are updated according to Eq. (2.2). Each updated step is regarded as an evolution time step. At each evolution time step, we record the averaging correlation function $\bar{G}_{rs}$. Then we obtain a time series of $\bar{G}_{rs}$. This time series describe how the correlation function varies with time. Fig. 4 shows the results. Here δ is set to be δ=0.5. Fig. 4(a) is the result for the ER random network, and Fig. 4(b) is the result for scale-free network. From Fig. 4, we can see that the correlation functions $\bar{G}_{rs}$ observed both in Figs. 4(a) and 4(b) vary with time linearly.

3.2 Chinese railway network

In the real-world networks, the railway network is one of the most important networks. In order to test our method proposed in this paper, we collect the data of northeast railway network on a coarse-grained level following the recent Chinese railway time table [19]. Fig. 5 shows the map of the northeast railway network, which contains a total of 69 stations. In Fig. 5, the nodes of the network represent the stations, the links among nodes represent the rail lines, and the weights represent the numbers of different passenger trains running on these rail lines. All nodes shown in Fig. 5 are labeled by a number, which only represents the order of nodes.

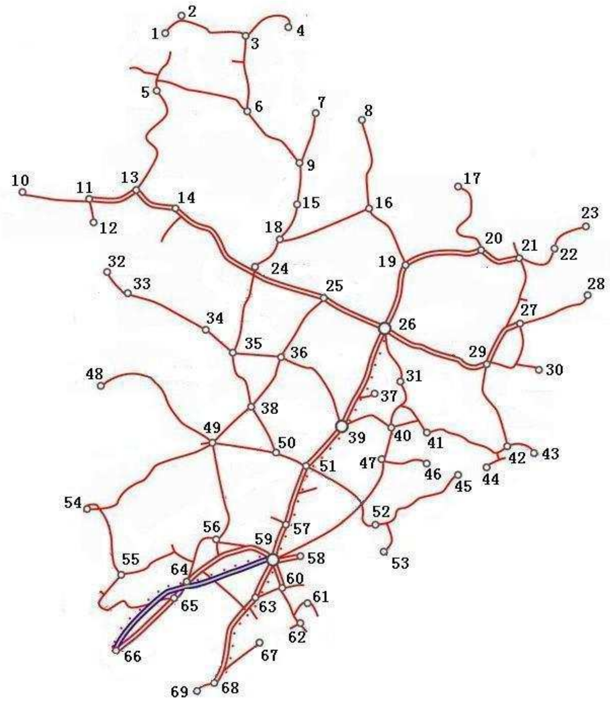

Figure 5: The northeast railway network in China.

First, we classify the nodes of the network into different modules according to the definition of the module. Here, the module is defined as that if some nodes belong to the same railway regional administration, they constitute a module. Fig. 6 plots the result. From Fig. 6, we can find that the nodes of the network can be divided into two modules, which are labeled by the number 1 and 2. Here, we use a dotted line to distinguish them. In modules 1 and 2, there are 31 nodes and 38 nodes, respectively.

Fig. 7 shows how $P_r(k)$ varies with the degree k. Here the horizontal axis represents the node degree in the module r and the vertical axis represents the values of $P_r(k)$. From Fig. 7, it can be seen that several nodes which have higher degrees have higher $P_r(k)$-values. The simulation results indicate that in each module of the network, several nodes are core nodes which have more correlations than other nodes. This result is similar to that observed in the ER random networks and scale-free networks. Further, we can use a mathematical function to fit the data shown in both Figs. 7(a) and 7(b), and find

$$P_r(k) \sim k^{\beta}, \quad \beta \approx 2.72.$$

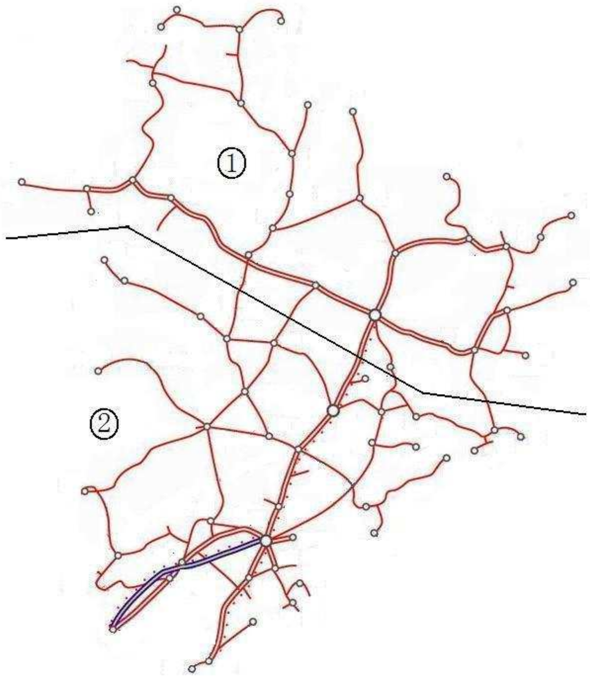

Figure 6: The division of the nodes of the network shown in Fig. 5.

Using the proposed method, we measure the distribution of the correlation function $G_{rs}(i)$. The measured results labeled by circles are shown in Fig. 8. In Fig. 8, there are only 5 nodes whose values of the correlation function are greater than zero. From Fig. 5, we see that these nodes are the boundary nodes between two modules. On the other hand, the nodes whose values of G_{rs} are equal to zero are not directly correlative with the nodes belonging to other modules. In general, if a node has a higher G_{rs}-value, the weights of the links among this node and its neighbor nodes have higher values. This result indicates that the correlation function G_{rs} represents the strength of the interaction between two different modules.

We choose one node which belongs to the module r, and calculate its G_{rs}. From the definition of the correlation function G_{rs}, the only variable in the Chinese railway network is the weight of links. Following the recent Chinese railway time table [18], every

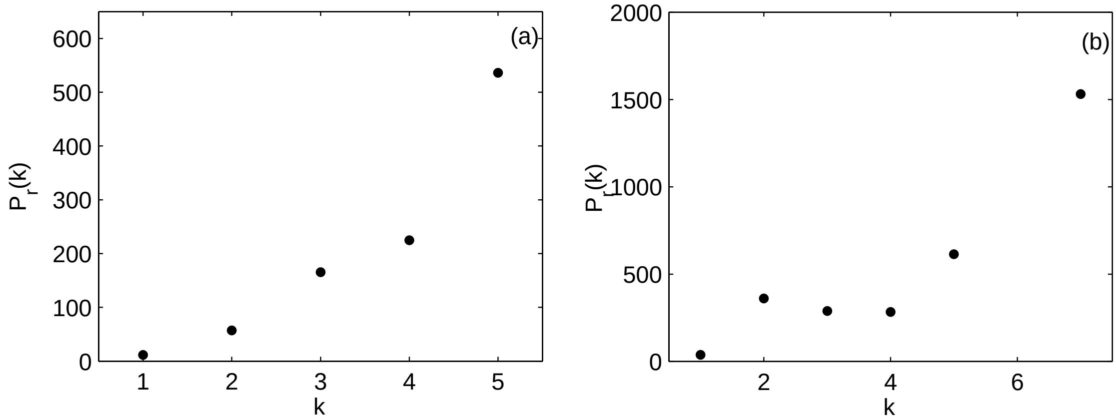

Figure 7: A plot of $P_r(k)$ vs k. (a) the 1st module; (b) the 2nd module.

one hour, we calculate the numbers of the different passenger trains running on the chosen rail lines. The calculated results are regarded as the weights of links. Then we obtain a time series of weights. Based on the time series of weights, we obtain the time series of the correlation function G_{rs}. Fig. 9 is the result for $r=1$ and $s=2$. From Fig. 9, we can see that sometimes the correlation function has higher values, and sometimes it has lower ones. In the former case, the number of the different passenger trains running on the rail line is higher. This means that there is a strong correlation between two different modules. Fig. 10 shows how the parameter δ used in Eq. (2.4) varies with the time t. To a large extent, δ represents the variations of the correlation function G_{rs}.

4 Conclusions

In this work, we outline a new method to describe the modules in weighted networks. The main improvement is that we introduce the spatial correlation function G_{rs}, and use it to describe the correlations among the modules of the weighted networks. The numerical simulations demonstrate that module has a "hub-like" core, which is formed by connecting high-degree nodes to each other. Moreover, different network displays different correlations. In the ER random networks, most of nodes participate in different modules at the same time. But only a fraction of nodes participate in different modules in scale-free networks.

The dimensionless parameter δ is an important factor. It governs the variation of the correlation function G_{rs}. When δ is a constant, as shown in Fig. 4, the correlation function G_{rs} varies with time linearly. When δ is similar to a random number, as shown in Fig. 9, the correlation function G_{rs} varies with time randomly. Although our study is focused on the ER random networks, the scale-free networks and the Chinese railway network, it is easy to extend our study to other networks, such as the Internet, social and ecological systems etc.

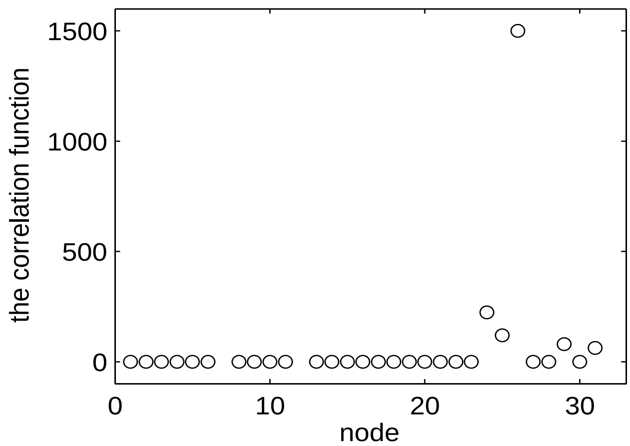

Figure 8: The distribution of the correlation function G_{rs} for $r=1$ and $s=2$.

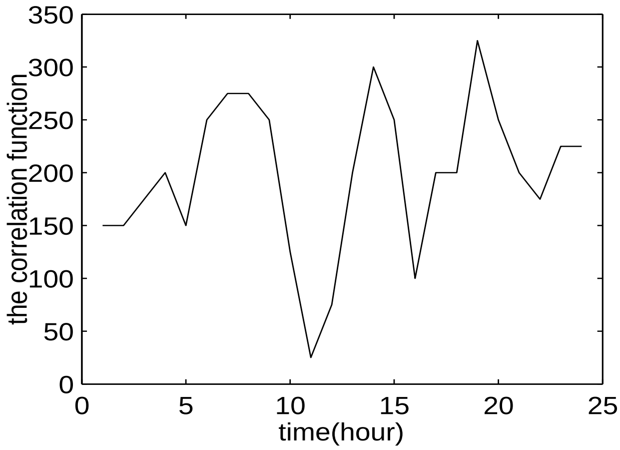

Figure 9: The correlation function G_{rs} as a function of the time t.

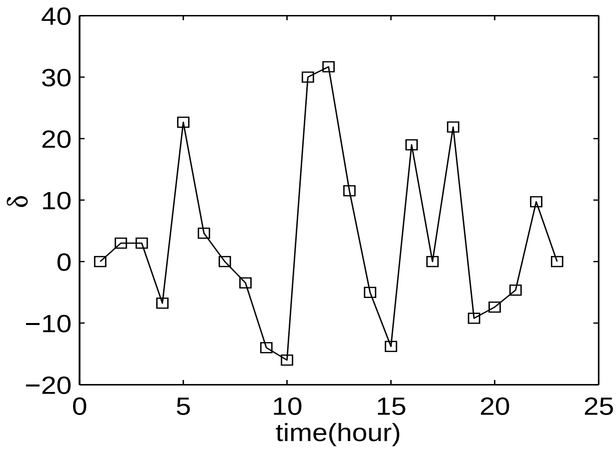

Figure 10: Parameter δ as a function of time t.

Acknowledgments

This project is supported by Changjiang Scholars and Innovative Research Team in University under Grant No. IRT0605, the National Natural Science Foundation of China under Grant Nos. 60634010 and 60776829, New Century Excellent Talents in University under Grant No. NCET-06-0074 and the Key Project of Chinese Ministry of Education under Grant No. 107007.

References

[1] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabási, Science 297 (2002) 1551.
[2] A. E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, W. W. Taylor, Nature 426 (2003) 282.
[3] M. Girvan, M. E. J. Newman, Publ. Natl. Acad. Sci. USA 99 (2002) 7821.
[4] K. A. Eriksen, I. Simonsen, S. Maslov, K. Sneppen, Phys. Rev. Lett. 90 (2003) 148701.
[5] M. E. J. Newman, M. Girvan, Phys. Rev. E 69 (2004) 026113.
[6] M. E. J. Newman, Phys. Rev. E 69 (2004) 066133.
[7] R. Guimerà, M. Sales-Pardo, L. A. N. Amaral, Phys. Rev. E 70 (2004) 025101(R).
[8] L. Donetti, M. A. Muñoz, J. Stat. Mech. Theor. Exp. (2004) P10012.
[9] E. Ziv, M. Middendorf, C. H. Wiggins, Phys. Rev. E 71 (2005) 046117.
[10] K. P. Li, Z. Y. Gao, Int. J. Mod. Phy. C 17 (2006) 1339.
[11] M. E. J. Newman, SIAM Rev. 45 (2003) 167.
[12] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Phys. Rep. 424 (2006) 175.
[13] J. Berg, M. Lässig, Phys. Rev. Lett. 89 (2002) 228701.
[14] K.-I. Goh, E. Oh, B. Kahng, D. Kim, Phys. Rev. E 67 (2003) 017101.
[15] M. Barthelemy, B. Gondran, E. Guichard, Phys. Rev. E 66 (2002) 056110.
[16] A. Barrat, M. Barthélemy, A. Vespignani, Phys. Rev. Lett. 92 (2004) 228701.
[17] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. USA 101 (2004) 3747.
[18] S. H. Yook, H. Jeong, A.-L. Barabási, Y. Tu, Phys. Rev. Lett. 86 (2001) 5835.
[19] http://qq.ip138.com/train/