Vol 16 No 8, August 2007
1009-1963/2007/16(08)/2304-06

Chinese Physics

© 2007 Chin. Phys. Soc.
and IOP Publishing Ltd

Detecting and describing the modular structures of weighted networks*

Li Ke-Ping(李克平) and Gao Zi-You(高自友)

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

(Received 14 November 2006; revised manuscript received 28 December 2006)

In the functional properties of complex networks, modules play a central role. In this paper, we propose a new method to detect and describe the modular structures of weighted networks. In order to test the proposed method, as an example, we use our method to analyse the structural properties of the Chinese railway network. Here, the stations are regarded as the nodes and the track sections are regarded as the links. Rigorous analysis of the existing data shows that using the proposed algorithm, the nodes of network can be classified naturally. Moreover, there are several core nodes in each module. Remarkably, we introduce the correlation function G_{rs}, and use it to distinguish the different modules in weighted networks.

Keywords: weighted networks, modular structure, railway network
PACC: 0590, 0250

1. Introduction

Recently, the study of complex networks has received renewed attention. A number of methods have been proposed to understand the structural properties of complex networks, such as the statistical, mathematical and model-based analysis methods.[1−4] However, many of the properties of networks remain unknown. One of them is the property of modular structure. Modular structure has been observed in many real world networks, including metabolic networks,[5] food webs[6] and social networks.[7] What are the mechanisms by which modularity emerges in complex networks? How to identify the modules of a given network? As a result of these questions, the problem of the detection and description of the modular structure in complex networks becomes the focus of many recent studies.

Modular structure usually represents functional units, such as the social groups and biological modules, etc. Conceptually, there are two types of the definition of a module: i.e., the self-referring and comparative definitions.[8−10] According to the self-referring definition, a module is defined as a subgroup of a graph containing more than two nodes where all the nodes are linked to each other. The comparative definition means that comparing the number of internal links with the number of external links, a module will be denser than its surroundings. In order to detect the modules in complex networks, many algorithms have been proposed, such as the clustering algorithms[8,11] and information theoretical algorithm,[12] etc. Despite the large amount of study in this area, a consensus on what is the definition of a module has not been reached. As written in Ref.[13], 'a generally accepted definition of a module does not exist and different authors use the concept in quite different ways'.

In complex networks, understanding the modular structure is potentially very useful. For example, in the web networks, modular analysis has uncovered thematic clusters.[14,15] In real complex networks, an important problem is that we often cannot determine how many modules we wish to discover. In this paper, a new methodology is proposed, which enables us to find the modules in the weighted networks. Using the proposed method, we do not consider the number of the modules in advance, and the final modular structure can be regarded as a natural division. 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 method of the detection of modules has been the subject of discussion in various disciplines, includ-

*Project supported by the National Basic Research Program of China (Grant No 2006CB705500), the National Natural Science Foundation of China (Grant No 60634010), New Century Excellent Talents in University (Grant No NCET-06-0074) and the Key Project of Chinese Ministry of Education (Grant No 107007).

ing applied mathematics and interdisciplinary physics, etc. However, most of them produce a division of the modules, regardless of whether the network has the natural division. In this paper, we design a new algorithm to detect the modules for the weighted networks, in which each link has a weight associated with it. In our algorithm, we do not consider how many modules we wish to discover. The division of the modules is completely based on the weights of links.

Our method is as follows. A network containing N nodes is considered. Let w_{ij} be the weight of the edge that links node i with node j. We introduce two parameters, Q and P. Q is defined as $Q_i = \sum_{j=1}^{d_i} w_{ij}$, where d_i is the degree of node i. The parameter P is calculated by $P_{ij} = w_{ij} / \sum_{k=1}^{d_j} w_{jk}$. At the first step, each node of the network is regarded as a module. This means that each node is the sole member of each of the modules. Starting with such a state, we repeatedly join the modules together in pairs. At a different step, the division of the network has smaller numbers of modules. Finally, we obtain a division from which we can investigate the modular properties of the network. Our algorithm consists of the following steps.

(1) According to the values of Q, all nodes of the network are arranged in a queue. This queue is ordered from high to low value.

(2) Beginning from the first node in the queue, we consider node i and the module that contains it, and determine which modules including the neighbours of node i will be joined.

(3) For node i, if P_{ij} has the highest value among $\{P_{jk}\}, k = 1, 2, \ldots, d_j$, then the module including node j will be joined into the module which includes node i.

(4) After node j is joined, the link from node i to node j will not be considered again.

Figure 1 shows an example. In Fig.1, there are 7 nodes, which are considered separately as a module. First, we arrange these nodes into a queue according to their values of Q. The result queue is 5, 7, 4, 6, 3, 2 and 1. According to the order of the nodes in this queue, we consider each node's neighbours, and determine which module will be joined. For example, we can assume that node 5 has been included by the module r. In Fig.1, there are 5 neighbours around node 5. For the neighbour 3 of node 5, since $P_{35} = 0.33$ has the highest value among P_{31}, P_{32}, P_{34} and P_{35}, the module including node 3 will be joined into the module r. These procedures can be applied to all nodes in the network.

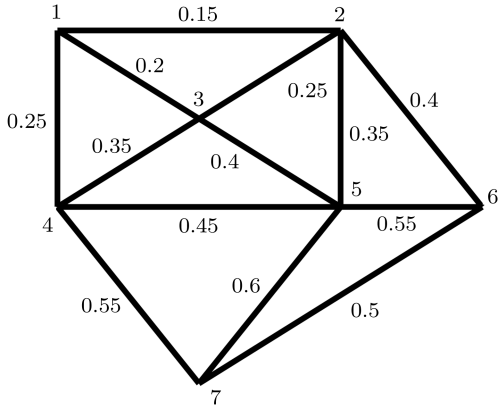


Fig.1. An example of the weighted networks.

When there are no edges between a pair of modules, the joining of them could not result in a change in the parameter P. So we need only to consider those pairs between which there are edges. The advantage of the proposed algorithm is that it is especially simple to find the modular structure. It should be pointed out that from the number of the links of network, some nodes cannot constitute a module, but from the distribution of the weights, they could constitute a module. This case is similar to the identification of traffic jams in networks.

3. Numerical simulations

Railway network is one of the most important networks in the real-world networks. It has very complex topological structure, and has attracted the attention of researchers in many different contexts.[16−18] The Chinese railway network is a dense network with hundred of thousands of trains plying in this network per day. In order to test our method proposed in this work, we collect the data of the Northeast railway network on a coarse-grained level following the recent Chinese railway timetable,[19] and use the proposed method to detect and analyse the modular structure of the railway network.

The Northeast railway network contains a total of 69 stations. Figure 2 shows the map of the Northeast railway network. In Fig.2, the nodes of the network represent the stations, the links among nodes represent the section tracks, and the weights are the numbers of the different passenger trains running on these section tracks. Based on the data of the Northeast railway network, we investigate the modular properties of the Northeast railway network. First, we detect

the modular structure, and classify the nodes of the network into different modules. Figure 3 plots the result. From Fig.3, we can find that the nodes of the network can be divided into three parts, which are labelled by 1, 2 and 3. Here, we use dotted lines to distinguish them.

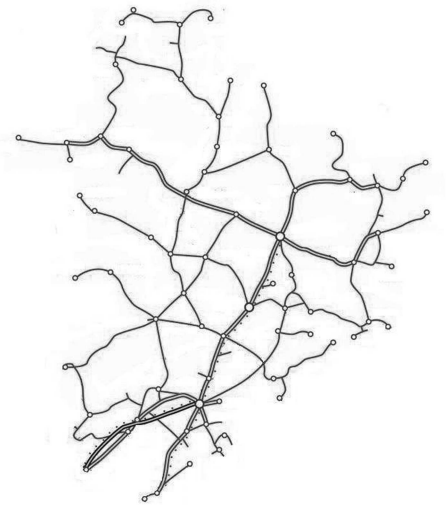


Fig.2. The Northeast railway network in China.

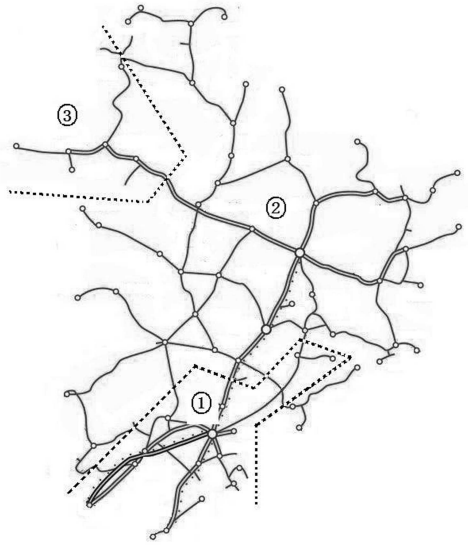


Fig.3. The division of the nodes of the network shown in Fig.2.

Finding and describing the modules of complex networks is an extremely important task. In our method, we give a new way to detect the modules of weighted networks. As mentioned in Section 2, each node in the weighted network can be characterized by the parameters Q and P. Here, we further define a parameter $C_s(k)$, which is the average of Q over the nodes with degree k in the module s. Figure 4 shows how $C_s(k)$ varies with the degree k. Here the horizontal axis is the node degree in the module s and the vertical axis is the value of $C_s(k)$.

In Fig.4(a), the distribution of $C_s(k)$ for the 1st module is shown ($s = 1$), the distribution of $C_s(k)$ for the 2nd module is shown in Fig.4(b) ($s = 2$), and in Fig.4(c) the distribution of $C_s(k)$ for the 3rd module is shown ($s = 3$). From Fig.4, it is obvious that several nodes which have higher degrees have higher values of $C_s(k)$. Basically, it can be inferred that the higher the degree, the higher the value of $C_s(k)$. The simulation results indicate that in each module of the network, several nodes are core nodes, which have closer relationships with other nodes. In practical problem, it is possible to control the evolution of the networks by controlling the states of these core nodes.

Another important parameter introduced in this work is J_s: it is defined as $J_s(i) = \sum_{j=1}^{d_{si}} w_{ij} / \sum_{j=1}^{d_i} w_{ij}$, where d_{si} is the number of links of node i in the module s. $J_s(i)$ measures how the links of node i are between the module itself and other modules. In general, the value of $J_s(i)$ is $0 \leq J_s(i) \leq 1$. If the links of node i are all within its own module s, the value of $J_s(i)$ is 1. Figure 5 plots the distributions of the parameter J_s for different modules. Here the horizontal axis represents the nodes in the module s, and the vertical axis represents the value of J_s. In Figs.5(a), 5(b) and 5(c), we respectively show the distributions of J_s for the 1st, 2nd and 3rd modules ($s = 1$, 2 and 3). From Fig.5, we can find that there are a number of nodes whose values of J_s are equal to 1. This result means that at the same time, only a few nodes participate in other modules. On the other hand, we can use the parameter J_s to analyse how the nodes participate in the different modules.

In order to investigate the relationships among different modules, we introduce the correlation function G_{rs}, which measures the correlation that occurs between the modules r and s. To assume that node i is within the module r, then the definition of $G_{rs}(i)$ is given as follows:

$$G_{rs}(i) = \frac{1}{n_s} \sum_{j=1}^{N_s} Q_i Q_j \theta_{ij},$$

where n_s represents the number of links between node i and the nodes within the module s. N_s represents the number of nodes in module s. In the above equation, θ_{ij} is defined as follows: if there is no link between nodes i and j, then $\theta_{ij} = 0$, otherwise $\theta_{ij} = 1$.

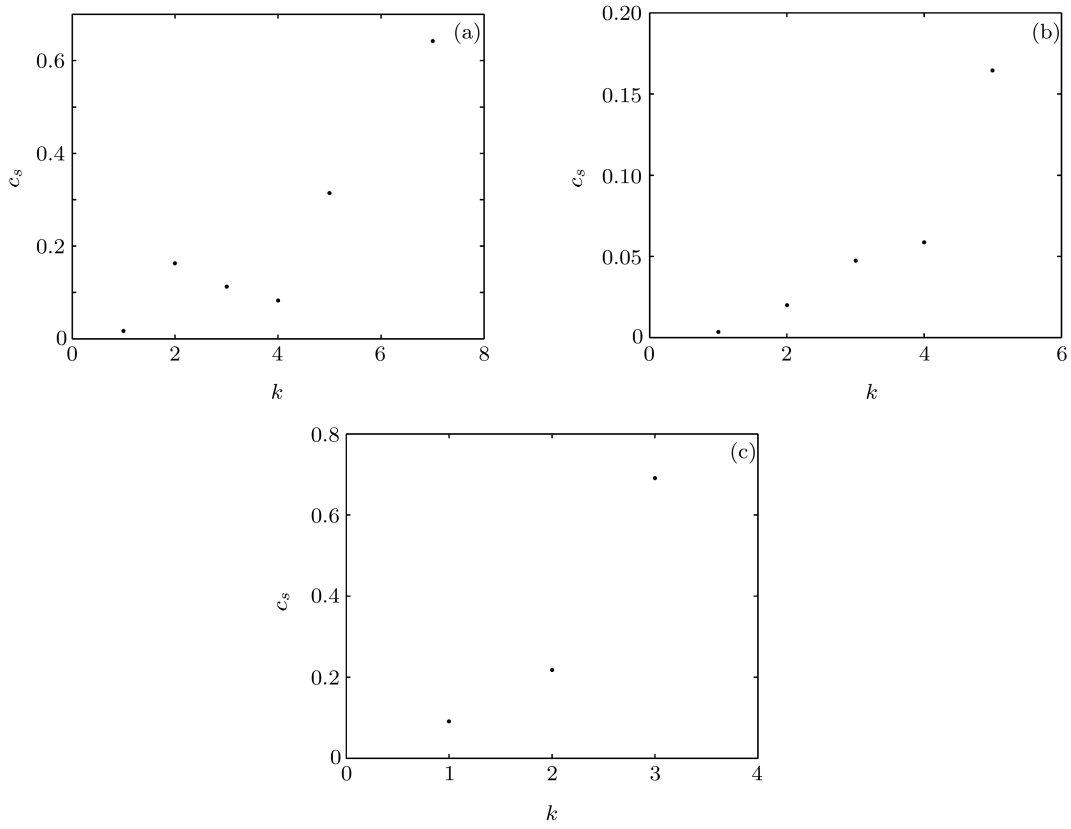


Fig.4. A plot of $C_s(k)$ vs k. (a), the 1st module; (b), the 2nd module; (c), the 3rd module.

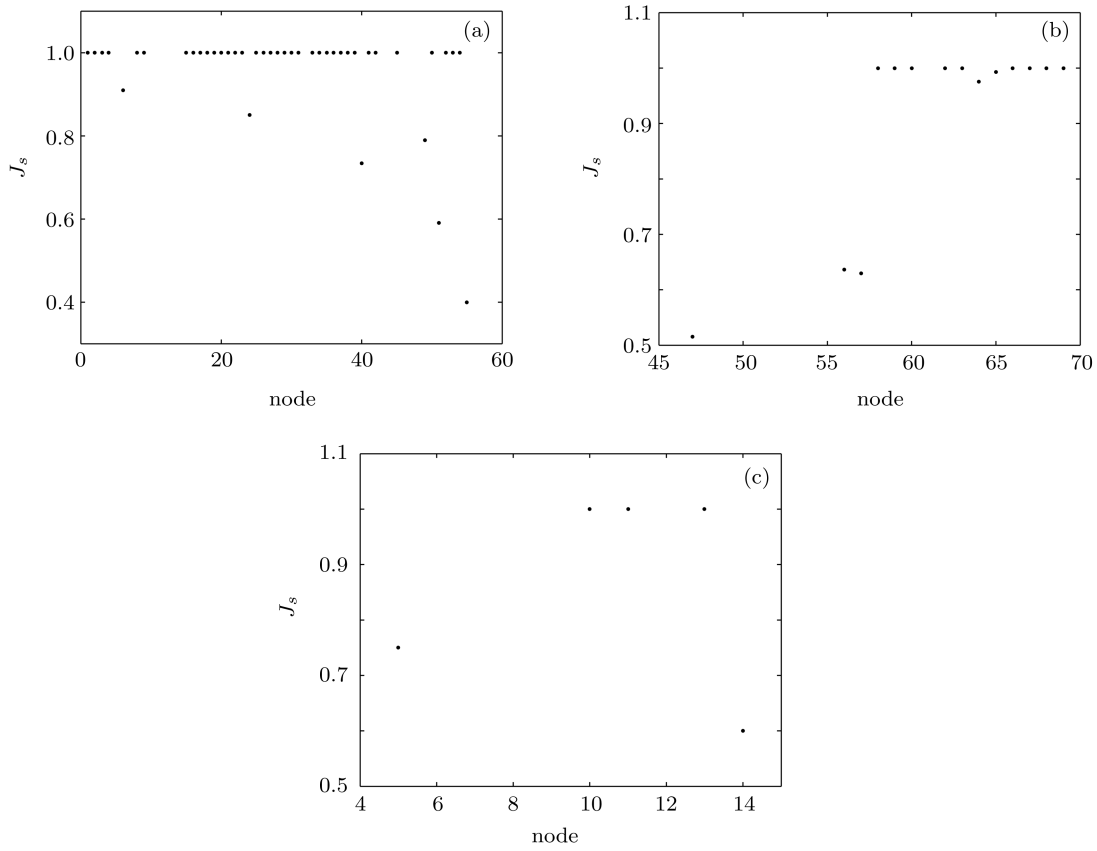


Fig.5. The distribution of J_s. (a) the 1st module; (b) the 2nd module; (c) the 3rd module.

According to the definition of G_{rs}, as $s = r$, $G_{rr}(i)$ measures the correlation that occurs between node i and the module r itself. Figure 6 displays the distributions of the correlation function G_{rs}. Here the horizontal axis represents the node, and the vertical axis represents the correlation function G_{rs}. In Fig.6, the points labelled by '$\times$' denote the correlations that occur between the nodes and the modules themselves, and the points labelled by '$\bigcirc$' denote the correlations that occur between the nodes and other modules. From Fig.6, it is obvious that the absolute majority of the correlations between the nodes and the modules themselves are larger than the correlations between these nodes and other modules. In Fig.6(b), we can find that the values of G_{rs} between the 1st module and the 3rd module are zero. This means that the nodes within the 1st module do not participate in the 3rd modules. It is postulated that we can use the correlation function G_{rs} to distinguish to which module node i belongs.

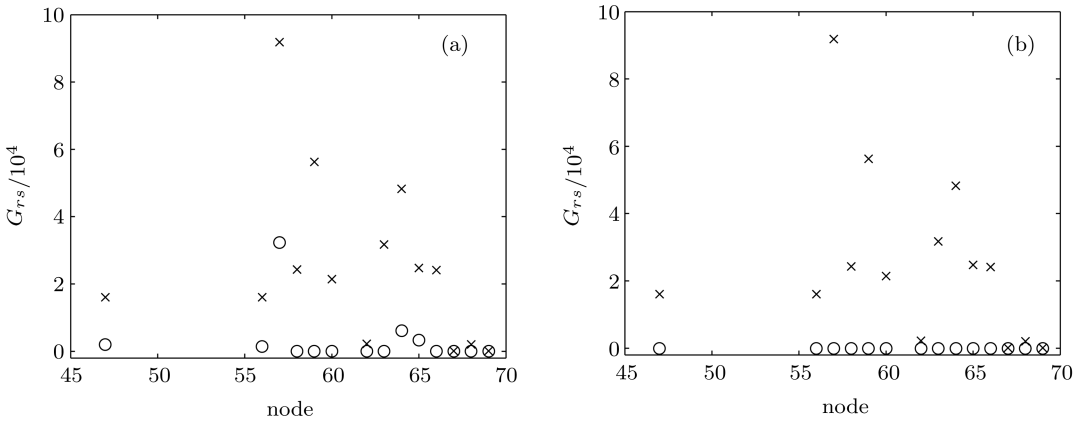


Fig.6. The distributions of the correlation function G_{rs}. (a), $r = 1$, $s = 2$; (b), $r = 1$, $s = 3$.

4. Conclusions

In summary, we present a new method to find the modules in weighted networks. Using the proposed method, a natural division of the modules can be obtained. Two parameters are introduced to analyse and describe the structure properties of the modules, i.e. $C_s(k)$ and J_s. In addition, we introduce the correlation function G_{rs}, which is used to measure the relationships among the modules of the weighted networks. A number of simulation results indicate that in the modules of the weighted networks, several nodes play the core role, which have far more relations with other nodes. Moreover, there are different correlations among the modules of the networks. The correlation function G_{rs} can be used to identify and describe the modules of the weighted networks. In this work, our example is focused on the Chinese railway network, but it is easy to extend our study to other networks, such as social and ecological systems, etc.

References

[1] Li Y, Liu Y, Shan X M, Ren Y, Jiao J and Qiu B 2005 *Chin. Phys.* **14** 2153

[2] Hu K and Tang Y 2006 *Chin. Phys.* **15** 2782

[3] Zhao W, He H S, Lin Z C and Yang K Q 2006 *Acta Phys. Sin.* **55** 3906 (in Chinese)

[4] Li J, Wang B H, Jiang P Q, Zhou T and Wang W X 2006 *Acta Phys. Sin.***55** 4051 (in Chinese)

[5] Ravasz E, Somera A L, Mongru D A, Oltvai Z N and Barabsi A L 2002 *Science* **297** 1551

[6] Krause A E, Frank K A, Mason D M, Ulanowicz R E and Taylor W W 2003 *Nature* **426** 282

[7] Girvan M and Newman M E J 2002 *Publ. Natl. Acad. Sci. USA* **99** 7821

[8] Newman M E J and Girvan M 2004 *Phys. Rev.* E **69** 026113

[9] Danon L, Duch J, Arenas A and Diaz-Guilera A, arXiv:cond-mat/0505245

[10] Radicchi F, Castellano C, Cecconi F, Loreto V and Parisi D 2004 *Publ. Natl. Acad. Sci. USA* **101** 2658

[11] Newman M E J 2004 *Phys. Rev.* E **69** 066133

[12] Ziv E, Middendorf M and Wiggins C H 2005 *Phys. Rev.* E **71** 046117

[13] Schlosser G and Wagner G P 2004 *Modularity in Development and Evolution* (Chicago: University of Chicago Press)

[14] Flake G W, Lawrence S, Giles C L and Coetzee F M 2002 *IEEE Computer* **35** 66

[15] Eckmann J P and Moses E 2002 *Publ. Natl. Acad. Sci. USA* **99** 5825

[16] Benguigui L and Daoud M 1991 *Geographical Analysis* **23** 362
Benguigui L 1995 *Environment and Planning* A **27** 1147

[17] Latora V and Marchiori M 2002 *Physica* A **314** 109

[18] Sen P, Dasgupta S, Chatterjee A, Sreeram P A, Mukherjee G and Manna S S 2003 *Phys. Rev.* E **67** 036106

[19] http://qq.ip138.com/train/