Jrl Syst Sci & Complexity (2008) 21: 637–644

REMARKS ON NETWORK COMMUNITY PROPERTIES*

Jiguang WANG · Yuqing QIU · Ruisheng WANG · Xiangsun ZHANG

Received: Received: 25 July 2008

Abstract This paper discusses a popular community definition in complex network research in terms of the conditions under which a community is minimal, that is, the community cannot be split into several smaller communities or split and reorganized with other network elements into new communities. The result provides a base on which further optimization computation of the quantitative measure for community identification can be realized.

Key words Community, complex network, modularity, modularity density.

1 Introduction

Community detection problem is derived from the research of complex networks such as Word Wide Web, social relationship networks, protein-protein interaction networks, biological metabolic networks, etc. Many complex systems, such as social, ecological, biological systems, contain thousands of independent subunits which interact with some others to achieve specific functions. An appropriate tool to represent the relationship of these complex systems is the network research, in which nodes represent subunits and edges represent relationships between them. Studies on these networks reveal some topological properties, such as small-world, scale-free[1]. Besides these statistic properties, many networks have specific substructure called community or modules[2] which is believed to be the elementary substructure of complex networks to achieve fundamental functions. For example, in biological metabolic network, highly connected topological modules are closely overlapped with known metabolic functions[3]. Identification of communities using the topological information is important for research of the complex systems.

From a topological view of networks, a community is described as a group of nodes connecting densely inside and sparsely to the outside. The suggested concept has been playing very

Jiguang WANG · Yuqing QIU
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, *China*; *Graduate University of the Chinese Academy of Sciences, Beijing* 100049, *China.*
Ruisheng WANG
School of Information, Renmin University of China, Beijing 100872, *China.*
Xiangsun ZHANG
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, *China.*
Email: zxs@amt.ac.cn.
*The research is supported by the Ministry of Science and Technology of China under Grant No. 2006CB503905. Some authors are also supported by the National Natural Science Foundation of China under Grant Nos. 10631070 and 10701080, and the JSPS(Japan Society for the Promotion of Science)-NSFC(National Natural Science Foundation of China) collaboration project under Grant No. 10711140116.

important role in network research. Based on this concept, the quantitative measures modularity $Q^{[2]}$ and modularity density $D^{[4]}$ were adopted as the network community identification criteria. To strictly describe the concept of community, two kinds of communities were defined in [5]: in a strong sense, each node has more edges within the community than with the rest of the graph, while in a weak sense the sum of all degrees within the subnetwork is larger than the sum of all degrees toward the rest of the network. Here we focus on the stability (or divisibility) of communities in terms of the weak definition, and try to give a guideline for distinguishing indivisible communities. A community is called indivisible or minimal if it cannot be split into subnetworks which in turn satisfy the weak definition as sub-communities.

This paper is organized as follows. In Section 2, we discuss the indivisibility of a single community. A zero-one integer programming model is formulated to judge whether a graph is divisible. In Section 3 we study the community structure stability problem in two popular exemplary networks. In this case, the main concern is that under what condition the communities will not be split and recombined into a new community structure.

2 Indivisible Community

2.1 The Definition of Indivisible Communities

Generally, a network can be expressed as $N = (V, E)$ where V represents the nodes and E the edges. In this paper, we use the adjacency matrix A ($a_{ij} = 1$ if nodes i and j are connected and $a_{ij} = 0$ otherwise) to represent a network. Since a community is a subnetwork, it can also be expressed as a matrix form.

Let $N = (V, E)$ denote a community with $|V| = m$. For any $0 < k < m$, node index sets $I = \{i_1, i_2, \cdots, i_k\}$ and $\overline{I} = \{i_{k+1}, i_{k+2}, \cdots, i_m\}$ represent a partition of community N with k and $m - k$ nodes, respectively. We define the concept of indivisible communities as follows.

Definition 1 N is an indivisible community if and only if there is no partition I and $\overline{I}$, such that

$$\begin{cases} 2|E_I| > |E_{I,\overline{I}}| \\ 2|E_{\overline{I}}| > |E_{I,\overline{I}}|, \end{cases} \tag{1}$$

where $N_I = (V_I, E_I), N_{\overline{I}} = (V_{\overline{I}}, E_{\overline{I}})$, $E_{I,\overline{I}}$ is the set of edges between N_I and $N_{\overline{I}}$.

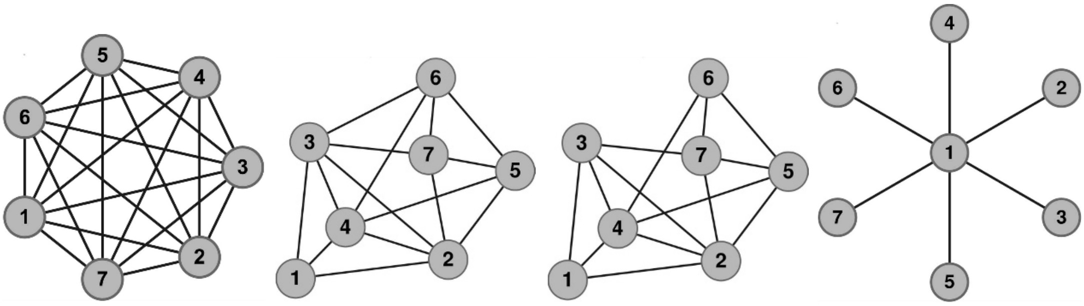
In other words, N can not be split into two subnetworks such that inside edges are more than half of outside ones(the definition of community in a weak sense in [5]). One thing should be emphasized is that it is enough to consider only the two-subnetwork partition when assessing the divisibility of graphs. In fact, it is easy to prove under this definition that if a graph can be partitioned into more than two communities, then it also can be partitioned into two communities.

A special example of indivisible communities is a clique (Figure 1(a)). To see this, we assume that N is a clique with m nodes. For some partition, $0 < k < m$, suppose that $k(k-1) > k(m-k)$ (and $(m-k)(m-k-1) > k(m-k)$), i.e., $k - 1 > (m - k)$, then

$$(m-k)(m-k-1) < (k-1)(m-k-1) < k(m-k) \tag{2}$$

which implies that inequations (1) cannot be satisfied.

Not all indivisible communities are cliques. One example is the so-called pseudo-clique (Figure 1(b) and (c)), i.e., delete some edges from a clique. We consider that a clique is cut into two modules N_1 with a nodes and N_2 with b nodes, then the number of edges in N_1, N_2, and between them are $a(a-1)/2$, $b(b-1)/2$, and ab, respectively. We suppose x edges are deleted from the clique. Then the divisible community at least satisfies: $ab - x < a(a - 1)$ and



(a) Clique (b) Pseudo-clique 1 (c) Pseudo-clique 2 (d) Star-shape

Figure 1 Diagrams of four kinds of communities

$ab - x < b(b-1)$. We assume that $a \geq b$, then we get $x > ab - b(b-1)$. So we confirm that the pseudo-clique with x edges deleted is indivisible when $x \leq ab - b(b-1)$. Suppose there are total n nodes,

$$x \leq -2b^2 + (n+1)b. \tag{3}$$

So we can give a sufficient condition for a network being indivisible:

Proposition 1 *Given a connected network $N = (V, E)$ with $n (n \geq 4)$ nodes, if $|E| \geq F(n)$, it is indivisible, where $F(n)$ is defined as*

$$F(n) = \begin{cases} C_n^2 - \frac{n}{2} & n \geq 4\ ,\ n\ \in \{\text{even}\}, \\ C_n^2 - n + 1 & n \geq 4\ ,\ n\ \in \{\text{odd}\}. \end{cases}$$

Proof Consider N as a pseudo-clique that has been deleted $C_n^2 - |E|$ edges from a clique. From the above discussion, we can determine $F(n)$ according to minimizing $f(b) = -2b^2 + (n+1)b$ (refer to condition (3)). In fact, $f(b)$ takes its the minimum at the boundary of integer b, $2 \leq b \leq \lfloor \frac{n}{2} \rfloor$, where the lower boundary is taken as 2 because b can not be one for the connectivity of the network. When $n \geq 4$, we take $b = \lfloor \frac{n}{2} \rfloor$, then we will get the $F(n)$ immediately. ∎

We take $n = 7$ as an example (Figure 1(b) and (c)), pseudo-clique 1 with six edges deleted is indivisible, while pseudo-clique 2 with seven edges deleted is divisible.

Corollary 1 *Given a network $N = (V, E)$ with n nodes, if $|E| \geq \frac{n(n-2)}{2}$, then N is indivisible.*

Not all indivisible communities are dense like a clique. One example is the star-shape graph (Figure 1(d)). It is easy to prove there is no partition satisfying conditions (1).

2.2 Distinguish Indivisible Communities

To judge a network N being indivisible or not is not an easy work, but we can transform the problem as following.

For a network $N = (V, E)$, given a partition represented as I and $\overline{I}$, let

$$x_i = \begin{cases} 1, & \text{if } i \in I, \\ 0, & \text{if } i \in \overline{I}, \end{cases}$$

where I is the part with less edges than $\overline{I}$ and if the given graph is divisible, we have

$$\sum_{(i,j)\in E}(x_i+x_j)(2-x_i-x_j) < \sum_{(i,j)\in E}(x_i+x_j)(x_i+x_j-1), \tag{4}$$

where the left part of the formula is the number of outside edges, while the right is the number of edges in I. Let $y_{ij}=x_i+x_j$, then we get

$$\sum_{(i,j)\in E} 2y_{ij}^2-3y_{ij}>0. \tag{5}$$

Then it can be formulated as the following zero-one integer programming:

$$\begin{aligned}
\max\ & \sum_{(i,j)\in E} 2y_{ij}^2-3y_{ij}\\
\text{s.t.}\ & \sum_{(i,j)\in E}(y_{ij}-1)\le 0,\\
& y_{ij}=x_i+x_j,\\
& x_i\in\{0,1\},
\end{aligned} \tag{6}$$

where the first constraint is added to ensure I has less edges than $\widetilde{I}$.

If the optimal value of the programming is larger than zero, the graph is divisible, otherwise it is indivisible. This problem can not be solved easily, but we can get some insights in looking for optimal partition.

Note that

$$y_{ij}=\begin{cases}2, & \text{if } \ i\in I, j\in I;\\ 1, & \text{if } \ i\in I,\ j\in\overline{I}, \quad \text{or } \ i\in\overline{I},\ j\in I;\\ 0, & \text{if } \ i\in\overline{I},\ j\in\overline{I}.\end{cases}$$

Let

$$z_{ij}=2y_{ij}^2-3y_{ij}=\begin{cases}2, & \text{if } \ i\in I,\ j\in I;\\ -1, & \text{if } \ i\in I,\ j\in\overline{I}, \quad \text{or } \ i\in\overline{I},\ j\in I;\\ 0, & \text{if } \ i\in\overline{I},\ j\in\overline{I}.\end{cases}$$

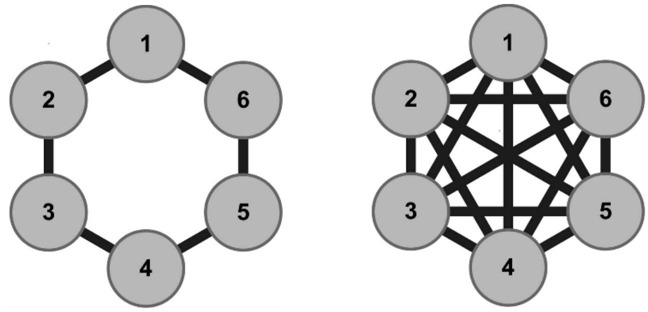
Thus, to maximize $\sum\limits_{(i,j)\in E} z_{ij}$ equals to maximize $|E_I|$, and minimize $|E_{I,\overline{I}}|$, where I is the part with less edges.

3 Community Stability in Special Environment

Two exemplary networks with known community structure are often used in network research. One is a ring of dense lumps (Figure 2(a)) whose adjacency matrix is defined by

$$A^L=\begin{pmatrix}
A_1 & M_l & 0 & \cdots & 0 & 0 & M_l\\
M_l & A_2 & M_l & \cdots & 0 & 0 & 0\\
0 & M_l & A_3 & \cdots & 0 & 0 & 0\\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\
0 & 0 & 0 & \cdots & A_{L-2} & M_l & 0\\
0 & 0 & 0 & \cdots & M_l & A_{L-1} & M_l\\
M_l & 0 & 0 & \cdots & 0 & M_l & A_L
\end{pmatrix}, \tag{7}$$

where $L\ge 4$, A_i is an $m\times m$ adjacency matrix to represent a connected subnetwork $N_i=(V_i,E_i)$ ($|E_1|=|E_2|=\cdots=|E_L|=|E|$), then A^L is an $Lm\times Lm$ matrix. M_l is a matrix with l elements



(a) The ring of N (b) The ad hoc network

Figure 2 Diagrams of exemplary networks

non-zero. When $|E_i| = \frac{1}{2}m(m-1),\ m > 3$ and $l = 1$, the network is a ring of cliques discussed in [4] and [6]. For convenience, we note this kind of networks as Ring of N.

The second exemplary network is a special version of the ad hoc network (a computer-generated network, see Figure 2(a)) originally appeared in [7] and discussed in [8]. Its adjacency matrix takes the form:

$$A^L = \begin{pmatrix} A_1 & M_l & M_l & \cdots & M_l & M_l & M_l \\ M_l & A_2 & M_l & \cdots & M_l & M_l & M_l \\ M_l & M_l & A_3 & \cdots & M_l & M_l & M_l \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ M_l & M_l & M_l & \cdots & A_{L-2} & M_l & M_l \\ M_l & M_l & M_l & \cdots & M_l & A_{L-1} & M_l \\ M_l & M_l & M_l & \cdots & M_l & M_l & A_L \end{pmatrix}. \tag{8}$$

Obviously, each N_i should be considered as a community. Here we discuss the condition on N_i and l under which the N_is will not split and no re-combinations of N_is will happen.

It is noted that the necessary condition for N_i being a community in the Ring of N and the ad hoc network are

$$l < |E|, \quad l < \frac{2}{(L-1)}|E| \tag{9}$$

respectively. The above condition can be understood as the necessary condition such that each subnetwork A is not re-combinable. To estimate the sufficient condition, some special structures are adopted. So we can only get some reference values for the sufficient condition.

Before this discussion, we firstly introduce modularity Q[2] and modularity density D[4]. For a given network $N = (V, E)$, and its partition $\{N_s = (V_s, E_s),\ s = 1, 2, \cdots, k\}$,

$$Q = \sum_{s=1}^{k} \left[\frac{|E_s|}{|E|} - \left(\frac{d_s}{2|E|} \right)^2 \right],$$

$$D = \sum_{s=1}^{k} \left(\frac{2|E_s|}{|V_s|} - \frac{|\overline{E_s}|}{|V_s|} \right),$$

where d_s represents total degrees of all nodes in N_s, $\overline{E_s}$ is all edges linking V_s and $V \setminus V_s$.

3.1 Ring of N

Consider Ns in the Ring of N (shortly denoted as R). Suppose that N has its elements distributed uniformly, and we can split N into two $\widetilde{N}$ between which there are $\widetilde{l}$ links:

$$|E| \ = \ 2|\widetilde{E}| \ + \ \widetilde{l}$$

and one $\widetilde{N}$ connects the left neighboring N in l links and the other connects the right neighboring N in l links (see Figure 3(a)). The re-combined "Ring of N", shortly denoted as R', is: two neighboring $\widetilde{N}$ which belong to two Ns become a new community. We will discuss the conditions with which the re-combination will not happen for the criteria Modularity Q and Modularity density D, respectively.

The "density" for the re-combined structure is

$$D_{R'} \ = \ L\Big(\frac{4|\widetilde{E}|+2l}{m} \ - \ \frac{2\widetilde{l}}{m}\Big). \tag{10}$$

From

$$\begin{aligned} D_R - D_{R'} &= \frac{L}{m}(2|E|-2l) - \frac{L}{m}(4|\widetilde{E}|+2l - \ 2\widetilde{l}) \\ &= \frac{L}{m}(4\widetilde{l}-4l) > 0, \end{aligned}$$

we have $l < \widetilde{l}$. To tighten the estimation up to the real sufficient condition, we take $\widetilde{l} =$ minicut$\{\widetilde{N},\widetilde{N}\}$, then

$$l < \text{minicut}\{\widetilde{N},\widetilde{N}\}. \tag{11}$$

If N is an m-clique, minicut$\{\widetilde{N},\widetilde{N}\} = \frac{m^2}{4}$. For the modularity Q,

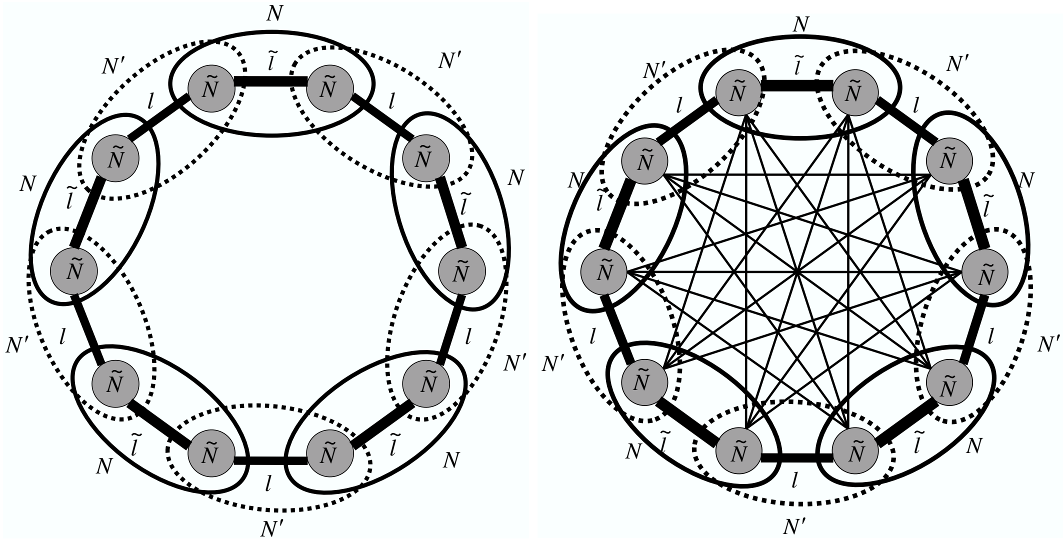
$$\begin{aligned} Q_R - Q_{R'} &= L\Big\{\frac{|E|}{L|E|+Ll} - \Big(\frac{|E|+l}{L|E|+Ll}\Big)^2\Big\} - L\Big\{\frac{2|\widetilde{E}|+l}{L|E|+Ll} - \Big(\frac{2|\widetilde{E}|+l+\widetilde{l}}{L|E|+Ll}\Big)^2\Big\} \\ &= L\Big\{\frac{|E|}{L|E|+Ll} - \frac{1}{L^2}\Big\} - L\Big\{\frac{2|\widetilde{E}|+l}{L|E|+Ll} - \frac{1}{L^2}\Big\} \\ &= L\Big\{\frac{|E|}{L|E|+Ll} - \frac{|E|+l-\widetilde{l}}{L|E|+Ll}\Big\} > 0, \end{aligned}$$

which implies $l < \widetilde{l} = \text{minicut}\{\widetilde{N},\widetilde{N}\}$ as in (11).

3.2 Ad Hoc Network

We now consider the Ad Hoc network (shortly denoted by H). Suppose that the same re-combination structure in (3.1) happens in the ad hoc network, that is, the re-combination is along a ring of Ns which include all Ns on the network. The distribution of links is illustrated in the Figure 3(b). The necessary condition for N being a community is

$$l \ < \ \frac{2|E|}{(L-1)}. \tag{12}$$



(a) The ring of N (b) The ad hoc network

Figure 3 Illustration for calculating Q and D in the two exemplary networks: dash line N' represents the possible re-combined communities. In (b), all edges without marking a weight are weighted by $\frac{l}{4}$

First we examine the quantitative measure D. Using D's definition:

$$D_H = L\left\{\frac{2|E|}{m} - \frac{(L-1)l}{m}\right\},$$

$$D_{H'} = L\left\{\frac{4|\widetilde{E}|+2l}{m} - \frac{2\widetilde{l}+(L-3)l}{m}\right\} = \frac{L}{m}(2|E|-(L-5)l-4\widetilde{l}),$$

then,

$$D_H - D_{H'} = \frac{L}{m}(4\widetilde{l}-4l) > 0$$

requires $l < \widetilde{l}$. When the modularity Q is used, then

$$\begin{aligned} Q_H - Q_{H'} &= L\left\{\frac{|E|}{L|E|+\frac{L(L-1)l}{2}} - \left(\frac{2|E|+(L-1)l}{2(L|E|+\frac{L(L-1)l}{2})}\right)^2\right\} \\ &\quad -L\left\{\frac{2|\widetilde{E}|+l}{L|E|+\frac{L(L-1)l}{2}} - \left(\frac{4|\widetilde{E}|+2l+2\widetilde{l}+(L-3)l}{2(L|E|+\frac{L(L-1)l}{2})}\right)^2\right\} \\ &= L\left\{\frac{|E|}{L|E|+\frac{L(L-1)l}{2}} - \left(\frac{2|E|+(L-1)l}{2(L|E|+\frac{L(L-1)l}{2})}\right)^2\right\} \\ &\quad -L\left\{\frac{|E|+l-\widetilde{l}}{L|E|+\frac{L(L-1)l}{2}} - \left(\frac{2|E|+(L-1)l}{2(L|E|+\frac{L(L-1)l}{2})}\right)^2\right\} \\ &= L\frac{\widetilde{l}-l}{L|E|+\frac{L(L-1)l}{2}} > 0 \end{aligned}$$

asks the same condition $l < \widetilde{l}$.

According to the above discussion, it is plausible there exists a bound $\bar{l}$, when $l < \bar{l}$ the As in both the Ring of N and the ad hoc network cannot be split and reorganized into a new community structure with lower D or Q value. Here, we choose a special link structure and conclude that when $l < \widetilde{l} = \text{minicut}\{\widetilde{N}, \widetilde{N}\}$, the community structure is not re-combinable.

4 Conclusion

We define a new concept "indivisible community" in this short paper, and give several examples to explain it. A sufficient condition and a zero-one integer programming model are formulated to judge whether a graph is divisible. Besides, a more difficult problem, community stability in networks is considered for the two popular exemplary networks. Although our discussion is based on strong assumptions, it provides insight for our understanding of the concept of communities. Further work is required to analyze the conditions in more general situations.

References

[1] S. H. Strogatz, Exploring complex networks, *Nature*, 2001, **410**(6825): 268–276.

[2] M. E. J. Newman, Modularity and community structure in networks, in *Proc. Natl. Acad. Sci. USA*, 2006, **103**(23): 8577–8582.

[3] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi, Hierarchical organization of modularity in metabolic Networks, *Science*, 2002, **297**(5586): 1551–1555.

[4] Z. Li, S. Zhang, R. S. Wang, X. S. Zhang, and L. Chen, Quantitative function for community detection, *Physical Review E*, 2008, **77**(3): 036109.

[5] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Defining and identifying communities in networks, in *Proc. Natl. Acad. Sci. USA*, 2003, **101**(9): 2658–2663.

[6] S. Fortunato and M. Barthelemy, Resolution limit in community detection, in *Proc. Natl. Acad. Sci. USA*, 2007, **104**(1): 36–41.

[7] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, *Physical Review E*, 2004, **69**(2): 026113.

[8] W. E, T. Li and E. Vanden-Eijnden, Optimal partition and effective dynamics of complex networks, in *Proc. Natl. Acad. Sci. USA*, 2008, **105**(23): 7907–7912.