NNTDM 17 (2011), 3, 10-12

A characterization of modularity in graphs

Yilun Shang

Institute for Cyber Security, University of Texas at San Antonio
San Antonio, Texas 78249, USA
e-mail: $shylmath@hotmail.com$

Abstract: We present alternative expressions for modularity in graphs. Modularity is used as a measure to characterize the community of networks, which is one of the most important features in real-world networks, especially social networks.
Keywords: Modularity, Community structure.
AMS Classification: 05C50

Graph communities, also called modules or groups, are groups of vertices which probably share common properties and/or play similar roles within the graph. Such a structure reveals important features of the topology and the evolution of processes on the graph. Social networks are paradigmatic examples of graphs with communities [3]. A metric known as modularity, proposed by Newman [7], has led to numerous researches in community structure detection and analysis. Great efforts have been made to characterize the modularity metric and investigate the partition resolution in networks, see e.g. [1, 2, 4, 5, 6]. We refer the reader to [3] for a recent excellent survey on modularity and its various applications.

In this brief paper, we present a new expression for modularity to get further insight on community structure in networks. In what follows, we only consider simple and undirected graphs.

Let $G = (V, E)$ be a graph with vertex set V and edge set $E \subseteq V \times V$. Let $|V| = n$ be the number of vertices, and $|E| = m$ be the number of edges. Let $A = (a_{ij})_{n \times n}$ be the adjacency matrix of G, where $a_{ij} = a_{ji} = 1$ if $(i, j) \in E$, and $a_{ij} = a_{ji} = 0$, otherwise. The degree of vertex $i \in V$ is denoted by d_i. The modularity [7] is a measure of the quality of a particular division of the network, and is defined as

$$Q = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(a_{ij} - \frac{d_i d_j}{2m} \right) 1_{[i \text{ and } j \text{ belong to the same group}]}, \tag{1}$$

which is proportional to the number of edges falling within groups minus the expected number in a network with edges placed at random. In fact, if edges are placed at random, the expected number of edges between a pair of vertices i and j is $d_i d_j / 2m$.

Suppose that there are g groups (or communities) $G_1, \cdots, G_g$ in the network G. Thus, we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} 1_{[i \text{ and } j \text{ belong to the same group}]} = 2 \sum_{k=1}^{g} m_k, \tag{2}$$

where m_k is the number of edges in group G_k. Let m_{int} be the number of edges that are cut by partitioning the network into g groups, i.e., the number of inter-group edges. It is clear that

$$m = \sum_{k=1}^{g} m_k + m_{int}. \tag{3}$$

Likewise, we observe that

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} d_i d_j 1_{[i \text{ and } j \text{ belong to the same group}]} &= \sum_{k=1}^{g}\Big(\sum_{i\in G_k} d_i\Big)\Big(\sum_{j\in G_k} d_j\Big) \\ &= \sum_{k=1}^{g} d_{G_k}^2, \end{aligned}\tag{4}$$

where $d_{G_k} = \sum_{i\in G_k} d_i$ is the sum of degrees of vertices in group G_k. By using the Binet-Cauchy identity [8], we may rewrite the righthand side of (4) as

$$\begin{aligned}\sum_{k=1}^{g} d_{G_k}^2 &= \frac{1}{g}\Big(\sum_{k=1}^{g} d_{G_k}^{1-\alpha}\Big)\Big(\sum_{k=1}^{g} d_{G_k}^{1+\alpha}\Big) \\ &+ \frac{1}{g}\sum_{1\le i<j\le n}(d_{G_i}^{1-\alpha} - d_{G_j}^{1-\alpha})(d_{G_i}^{1+\alpha} - d_{G_j}^{1+\alpha}),\end{aligned}\tag{5}$$

for any $\alpha > 0$.

Combining (5) with (1)-(4), we obtain

$$\begin{aligned}Q &= 1 - \frac{m_{int}}{m} - \frac{1}{g(2m)^2}\Big[\Big(\sum_{k=1}^{g} d_{G_k}^{1-\alpha}\Big)\Big(\sum_{k=1}^{g} d_{G_k}^{1+\alpha}\Big) \\ &+ \sum_{1\le i<j\le n}(d_{G_i}^{1-\alpha} - d_{G_j}^{1-\alpha})(d_{G_i}^{1+\alpha} - d_{G_j}^{1+\alpha})\Big].\end{aligned}\tag{6}$$

In particular, if we take $\alpha = 1$ and exploit the fact that $\sum_{k=1}^{g} d_{G_k} = 2m$, we may recover the expression given in [5]

$$Q = 1 - \frac{m_{int}}{m} - \frac{1}{g}\Big[1 + \sum_{1\le i<j\le n}\Big(\frac{d_{G_i} - d_{G_j}}{2m}\Big)^2\Big]. \tag{7}$$

References

[1] Arenas, A, A. Fernández, S. Gómez. Analysis of the structure of complex networks at different resolution levels. *New J. Phys.*, 10 (2008) 053039

[2] Danon, L., A. Díaz-Guilera, J. Duch, A. Arenas. Comparing community structure identification. *J. Stat. Mech.: Theory Exp.*, 2005 P09008

[3] Fortunato, S. Community detection in graphs. *Phys. Rep.*, 486 (2010) 75–174

[4] Kumpula, J. M., J. Saramäki, K. Kaski, J. Kertész. Limited resolution in complex network community detection with Potts model approach. *Eur. Phys. J. B*, 56 (2007) 41–45

[5] Mieghem, P. V., X. Ge, P. Schumm, S. Trajanovski, H. Wang. Spectral graph analysis of modularity and assortativity. *Phys. Rev. E*, 82 (2010) 056113

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

[7] Newman, M. E. J., M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69 (2004) 026113

[8] Weisstein, E. W. *CRC Concise Encyclopedia of Mathematics*. CRC Press, 2003