Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 18 (2008) 1043–1047
Progress in Natural Science
www.elsevier.com/locate/pnsc

Short communication

A measure of centrality based on modularity matrix

Gaoxia Wang, Yi Shen *, Enjie Luan

The Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Received 4 March 2008; received in revised form 27 March 2008; accepted 28 March 2008

Abstract

In this paper, a kind of measure of structural centrality for networks, called modularity centrality, is introduced. This centrality index is based on the eigenvector belonging to the largest magnitude eigenvalue of modularity matrix. The measure is illustrated and compared with the standard centrality measures using a classic dataset. The statistical distribution of modularity centrality is investigated by considering large computer generated graphs and two networks from the real world.

Keywords: Centrality; Modularity centrality; Modularity matrix

1. Introduction

Centrality measures serve to quantify that in a network some nodes are more important (central) than others [1]. The idea of centrality was firstly introduced in the context of social systems, where a relation was assumed between the location of an individual in the network and its influence and/or power in group processes [2,3]. Various centrality measures have been proposed over the years to quantify the importance of an individual in a social network [3]. And the issue of structural centrality has attracted the attention of physicists [4,5], who have extended its applications to the realm of biological [6], technological [7] and geographical networks [1,8].

The standard centrality measures can be divided into two classes: those based on the idea that the centrality of a node in a network is related to its distance to the other nodes, and those based on the idea that central nodes stand between others [3]. Degree [9] and closeness [10] are examples of measures of this first kind, while shortest-path [11] or flow [12] betweenness is the measure of the second kind [13].

Recently, Latora and Marchiori proposed a new class of centrality measures, the so-called delta centralities [13]. This class of centralities measure the contribution of a node to a network cohesiveness property, from the observed variation in such property when the node is deleted. The information centrality [13], based on the concept of efficient propagation of information over the network [14,15], is the special case of delta centralities.

In Ref. [16], Newman studied the problem of detecting the community structure of complex networks. He introduced the modularity matrix of networks and proposed a spectra method that used the eigenspectrum of matrix to find the community structure. The modularity matrix contains the structural information of networks [17], so we can use the values of eigenvector adhered to nodes to measure the importance of nodes in the whole network. In this paper, based on the first leading eigenvectors of modularity matrix, we propose a new centrality measure, the so-called modularity centrality.

2. Classic measures of centrality

The following is a list of the classic measures in social networks, a newly proposed information centrality by Latora is listed as well. The definitions are given in terms

* Corresponding author. Tel.: +86 27 87543630; fax: +86 27 87543130.
E-mail address: yishen64@163.com (Y. Shen).

1002-0071/$ - see front matter
doi:10.1016/j.pnsc.2008.03.015

of an undirected, unweighted graph G, of N nodes and K edges. The graph is described by an $N \times N$ adjacency matrix $\mathbf{A}$, whose entry A_{ij} is equal to 1 when there is an edge between i and j and 0 otherwise.

The degree centrality ($C^{\mathbf{D}}$) is based on the idea that important nodes are those with the largest number of ties to other nodes in the graph. The degree centrality of a node i is defined as [2]:

$$C_i^{\mathrm{D}} = \frac{k_i}{N-1} = \frac{\sum_{j \in G} A_{ij}}{N-1} \tag{1}$$

where k_i is the degree of node i.

The closeness centrality ($C^{\mathbf{C}}$) measures to an extent a node i is near to all the other nodes along the shortest paths and is defined as [3]:

$$C_i^{\mathrm{C}} = (L_i)^{-1} = \frac{N-1}{\sum_{j \in G, j \neq i} d_{ij}} \tag{2}$$

where d_{ij} is the shortest path length between i and j, and L_i is the average distance from i to all the other nodes.

The betweenness centrality ($C^{\mathbf{B}}$) is based on the idea that node is central if it lies between many other nodes, in the sense that it is traversed by many of the shortest paths connecting couples of nodes. The betweenness centrality of node i is [11]:

$$C_i^{\mathrm{B}} = \frac{1}{(N-1)(N-2)} \sum_{j \in G, j \neq i} \sum_{k \neq i, k \neq j} \frac{n_{jk}(i)}{n_{jk}} \tag{3}$$

where n_{jk} is the number of the shortest path between j and k, and $n_{jk}(i)$ is the number of the shortest path between j and k that contains node i.

The information centrality ($C^{\mathbf{I}}$) relates the node centrality to the ability of the network to respond to the deactivation of the node. The information centrality of node i is defined as the relative drop in the network efficiency $E(G)$ caused by the removal from G of the edges incident in i [13]:

$$C_i^{\mathrm{I}} = \frac{\Delta E}{E} = \frac{E(G) - E(G')}{E(G)} \tag{4}$$

where the efficiency of a graph G is defined as:

$$E(G) = \frac{1}{N(N-1)} \sum_{i,j \in G, i \neq j} \frac{1}{d_{ij}} \tag{5}$$

in which G' is the graph with N nodes and $K - k_i$ edges obtained by removing the k_i edges linked with node i from the original graph G. An advantage of using the efficiency to measure the performance of a graph is that $E(G)$ is finite even for disconnected graphs.

3. Modularity centrality

The various kinds of matrices of networks, such as adjacency matrix, Laplacian matrix, normalized matrix, often contain some useful information about the structure of the network. For finding the efficient method of detecting the community structure, Newman introduces the modularity function and modularity matrix to avoid the influences of random factors so as to obtain the better divisions of the community structure.

The modularity function Q is the number of edges falling within communities minus the expected number in an equivalent network with edges placed at random. This quantity is high for good community divisions and low for poor ones [16]. Consider the partition of a network G into p non-overlapping communities, Q can be expressed as

$$Q = \frac{1}{2m} \mathbf{Trace}\left(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{M} \boldsymbol{X}\right) \tag{6}$$

where the assignment matrix $\boldsymbol{X} = (x_{ih}), x_{ih} = 1$ if vertex i belongs to community h and $x_{ih} = 0$ otherwise. $\boldsymbol{M}$ is the so-called modularity matrix and in vector notation it can be written as $\boldsymbol{M} = \boldsymbol{A} - \frac{\boldsymbol{k}\boldsymbol{k}^{\mathrm{T}}}{2m}$, where $\boldsymbol{k}$ is the n-element vector whose elements are degrees k_i of the nodes. We can see that $\boldsymbol{M}$ is the real symmetric matrix with elements $M_{ij} = A_{ij} - \frac{k_i k_j}{2m}$. For any network, due to

$$\sum_j M_{ij} = \sum_j A_{ij} - k_i \sum_j \frac{k_j}{2m} = k_i - k_i = 0 \tag{7}$$

holds, it implies that the vector $(1, 1, \ldots, 1)$ is an eigenvector of the modularity matrix with eigenvalue zero. The eigenvalues of the modularity matrix are not all of one sign and the matrix usually has both positive and negative eigenvalues.

Writing $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{T}}$, where $\boldsymbol{U} = (u_1|u_2|\ldots)$ is the matrix of eigenvectors of $\boldsymbol{M}$ and $\boldsymbol{D}$ is the diagonal matrix of eigenvalues $D_{ii} = \beta_i$, Eq. (6) can be transformed to

$$Q = \sum_{j=1}^{n} \sum_{k=1}^{p} \beta_j \left(u_j^{\mathrm{T}} x_k\right) \tag{8}$$

From (8) we can see that the eigenvalues and eigenvectors of the modularity matrix are closely tied to the value of Q, thus closely tied to the community structure of the network. In fact, every eigenvalue and the corresponding eigenvector, either positive or negative, give their contributions to the value of Q. In particular, the largest positive eigenvalue with its eigenvector has the most positive contribution to the modularity, and the most negative eigenvalue with its eigenvector make substantial negative contribution to the modularity. In the meantime, those vertices, as a consequence of their situation within the network, have the power to make substantial positive or negative contributions to the overall modularity of the network. This contribution is embodied by the magnitudes of elements corresponding to the vertices. In Ref. [16], Newman pointed out that vertices with the greatest capacity for making positive contributions to the modularity have the greatest capacity to make negative contributions. So the magnitudes of the elements of the eigenvector corresponding to the largest magnitude eigenvalue of the modularity matrix give a measure of the "strength" with which vertices belong to their assigned communities. Thus, these magnitudes define a kind of centrality index that quantifies how central vertices are in communities. We call this centrality measure modularity centrality and use the norm C^{M} to represent it. We define it to be equal to the magnitudes of the elements of the eigenvector belonging to the largest magnitude eigenvalue. That is, let $\beta = max(|\beta_i|), i = 1, 2, \ldots, n$, the vector $\boldsymbol{u}$ is the eigenvector belonging to β, and the modularity centrality score of node i, is the magnitude of the ith elements of vector $\boldsymbol{u}$, i.e., $|u_i|$.

3.1. An example

The modularity centrality agrees with the standard measures on assignment of extremes. For instance, it gives the maximum importance to the central node of a star and equal importance to the nodes of a complete graph. However, the agreement breaks down between these extremes, such as the bridge node connecting the two main parts of a network. We show here a simple example to illustrate this.

Consider two graphs G_1 and G_2 with no cycles (Fig. 1). The six centrality scores for G_1 are shown in Table 1, where nodes are ordered in descending order of C^{M}. Although all the six measures attribute the highest centrality to node 2, there are some differences worth mentioning.

We can see that, C^{M} has the same results as C^{I} for both ranks and resolution. As better shown in Table 2, C^{M}, as well as C^{I}, assigns the top score in G_1 to node 2, the second score to nodes 1, 3, the third score to nodes 7, 12, and is also able to disentangle nodes 9, 10, 11 (fourth score) from the remaining ones. The only other measure that operates such a distinction is C^{C} which, on the other hand, assigns the second score to nodes 7, 12 and the third score to nodes 1, 3 inverting to the result of C^{M} and C^{I}. Neither C^{D} nor C^{B} has the resolution of C^{M}, C^{I} and C^{C}. In fact, C^{D} assigns the top score to the three nodes 1, 2, 3, all having five neighbors, and the second score to nodes 7, 12, both with two neighbors; while C^{B} assigns the top score to node 2 and the second score to nodes 1, 3, 7, 12. Both C^{D} and C^{B} do not distinguish nodes 9, 10, 11 from the remaining ones.

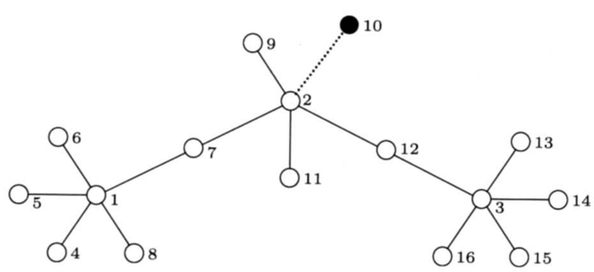

Fig. 1. The graph G_1 with $N = 16$ nodes, and the graph G_2, obtained from G_1 by disconnecting node 10 from the rest of the graph.

Table 1
Centrality values for the nodes of G_1

Node	C^{M}	C^{I}	C^{D}	C^{C}	C^{B}
2	0.5493	0.5904	0.3333	0.4545	0.7143
3	0.4970	0.4439	0.3333	0.3488	0.4762
1	0.4970	0.4439	0.3333	0.3488	0.4762
12	0.2113	0.3893	0.1333	0.4054	0.4762
7	0.2113	0.3893	0.1333	0.4054	0.4762
9	0.1140	0.1160	0.0667	0.3191	0.0000
10	0.1140	0.1160	0.0667	0.3191	0.0000
11	0.1140	0.1160	0.0667	0.3191	0.0000
13	0.0973	0.1064	0.0667	0.2631	0.0000
14	0.0973	0.1064	0.0667	0.2631	0.0000
15	0.0973	0.1064	0.0667	0.2631	0.0000
16	0.0973	0.1064	0.0667	0.2631	0.0000
4	0.0973	0.1064	0.0667	0.2631	0.0000
5	0.0973	0.1064	0.0667	0.2631	0.0000
6	0.0973	0.1064	0.0667	0.2631	0.0000
8	0.0973	0.1064	0.0667	0.2631	0.0000

Table 2
Centrality rankings in G_1

Rank	C^{M}	C^{I}	C^{D}	C^{C}	C^{B}
1	2	2	1, 2, 3	2	2
2	1, 3	1, 3	7, 12	7, 12	1, 3, 7, 12
3	7, 12	7, 12	Others	1, 3	Others
4	9, 10, 11	9, 10, 11		9, 10, 11	9, 10, 11
5	Others	Others		Others	Others

Table 3
Centrality rankings in G_2

Rank	C^{M}	C^{I}	C^{D}	C^{C}	C^{B}
1	1, 3	2	1, 3	2	2
2	2	1, 3	2	7, 12	1, 3
3	7, 12	7, 12	7, 12	1, 3	7, 12
4	4–8, 13–16	9, 11	Others	9, 11	Others
5	Others	Others		Others	

The node ranking obtained in G_2 is reported in Table 3. In graph G_1, nodes 2, 1 and 3 have the same number of neighbors; in G_2, node 2 has less neighbors than nodes 1 and 3. This affects the node ranking based on C^{D} and C^{B}, while it does not change the rankings based on C^{I} and C^{C}.

We can show that C^{M} has the same resolution as the C^{I}, C^{C}. In addition, C^{M} assigns the most important nodes to node 1 and node 3 rather than node 2 which is assigned by C^{I}. In the meantime, the 4th rank and 5th rank exchange the order for C^{M} and C^{I}.

3.2. Large networks

For large networks, the analysis emphasis of the centrality is now shifted from the role of central nodes and their identification to the distribution of centrality values through all nodes [8].

So in this section we investigate how the modularity centrality is statistically distributed among the nodes of large networks. In order to reduce the statistical fluctuations, we have computed the cumulative distribution $P(C^{M})$ defined in terms of the (differential) distribution $p(C^{M})$ as:

$$P(C^{M}) = \int_{C^{M}}^{+\infty} p(X)\,\mathrm{d}X = \int_{C^{M}}^{+\infty} \frac{N(X)}{N}\,\mathrm{d}X \tag{9}$$

where $p(X)\mathrm{d}X$ is the probability to find a node with a modularity centrality ranging in the interval $(X, X + \mathrm{d}X)$, while $N(X)\mathrm{d}X$ is the number of nodes with a modularity centrality ranging in the interval $(X, X + \mathrm{d}X)$, and N is the total number of nodes in the network.

First, we consider two kinds of artificial generated graphs, namely, Erdős–Rényi (ER) random graphs and the generalized random graphs with a given degree distribution. We have generated ER random graphs, which has $N = 1000$ nodes and the average degree $\langle k \rangle = 9.884$, and the generalized random graphs with $N = 1000$ nodes and a power-law degree distribution $p_k \sim k^{-\gamma}$ with exponent $\gamma = 3$. Fig. 2 shows the cumulative distributions of modularity centrality scores obtained in the two cases (circles). (a) is the semilog plot figure, the solid line is the exponential fit to the points, while the straight line in the log–log plot of (b) is a power law fit, $P(C^{\mathrm{M}}) \sim (C^{\mathrm{M}})^{-\mu}$, with an exponent $\mu = 1.42$. The latter result indicates that in a random graph with a power-law degree distribution, the modularity centrality is also distributed as a power law. In the case considered we have found that $p(C^{\mathrm{M}}) \sim (C^{\mathrm{M}})^{-2.42}$.

We have also considered two networks from the real world. The first network we considered is the electrical power grid of the western United States [19]. For this network, nodes represent generators, transformers and substations, and edges represent high-voltage transmission lines between them. The degree distribution in this network is consistent with an exponential and is thus relatively homogeneous. The distribution of betweenness centrality is more skewed than that displayed by semirandom networks with the same distribution of links, indicating that the power grid has structures that are not captured by these models [20]. This is demonstrated by the distribution of modularity centrality (semi-log plot in Fig. 3(a)). The fit line is a piecewise linear function, which means that the distribution function is a piecewise exponential function.

In Fig. 3(b), we report the cumulative distribution of modularity centrality in the network of US airports in 1997 (The data are available at http://vlado.fmf.uni-lj.si/pub/networks/default.htm). The network has $N = 332$ nodes representing the airports and $K = 2126$ links representing the flights. This network exhibits a broad scale modularity distribution as the artificial network considered in Fig. 2(b). The solid line in the log-log plot of Fig. 3(b) is a power law fit to the empirical distribution, $P(C^M) \sim C^{-\mu}$, with $\mu = 1.4986$.

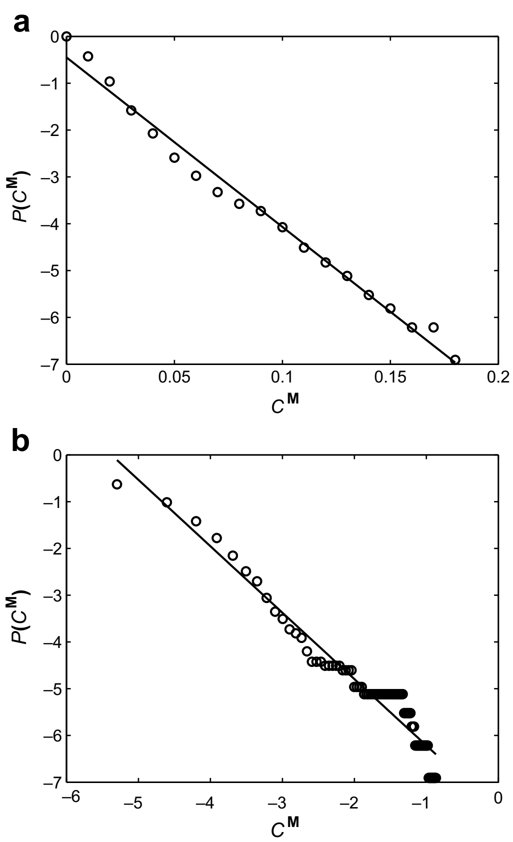

Fig. 2. Cumulative distribution of modularity centrality in two computer generated networks. (a) ER random graph with $N = 1000$ nodes and the average degree $\langle k \rangle = 9.884$; (b) generalized random graph with $N = 1000$ nodes and a scale-free degree distribution $p_k \sim k^{-3}$. The results are averages over an ensemble of 30 graphs. The solid line in (a) is the exponential fit to the distribution, while in (b) is the power fit.

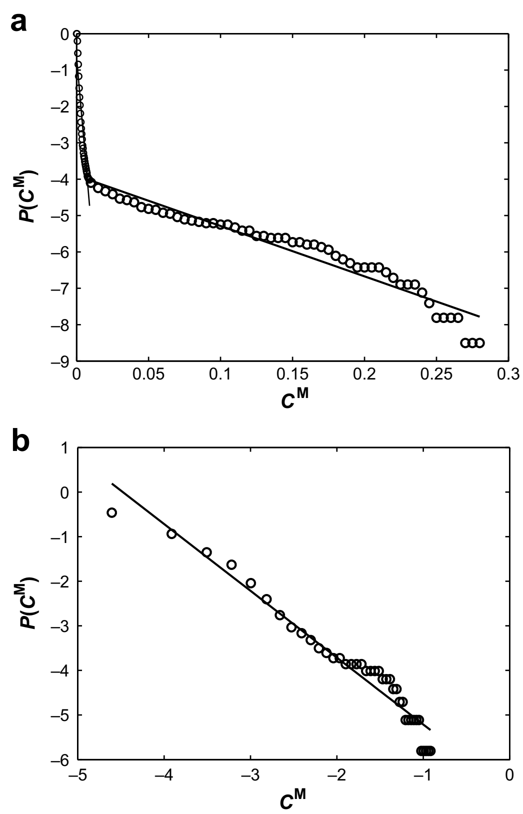

Fig. 3. Cumulative distribution of modularity centrality in real world networks. (a) The western US power transmission grid, which has $N = 4941$ and $\langle k \rangle = 2.67$ [18]; (b) the US airport networks in 1997, which has $N = 332$ and $K = 2126$ lines. The solid line in (a) is a piecewise exponential function that fits to the distribution, while in (b) is a power fit.

4. Conclusions

In this paper, we have introduced a new kind of centrality measures, the so-called modularity centrality. We have illustrated similarities and dissimilarities with respect to the standard measures adopted in sociometry and information centrality presented in technological networks by considering some small networks. From the standard example, we can find that modularity centrality has better resolution to the key nodes. We have also investigated how the modularity centrality is statistically distributed among the nodes of large graphs, by considering artificial generated graphs and networks from the real world.

Now, centrality is a fundamental concept in the network analysis, and the new centrality measures have been proposed for various uses. The modularity centrality is inspired by the study of the detection of community structure of complex networks, and the modularity centrality would be a good measure of the actual strength of the corresponding community of nodes. However, it remains to be seen, in the light of further empirical work, if and in which cases the new measure is more appropriate than the others.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant Nos. 60574025 and 60740430664) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20070487052). The authors thank Dr. Adilson E. Motter for providing the power-grid data.

References

[1] Crucitti P, Latora V, Porta S. Centrality measures in spatial networks of urban streets. Phys Rev E 2006;73(3):036125.

[2] Wasserman S, Faust K. Social networks analysis: methods and applications. Cambridge: Cambridge University Press; 1994.

[3] Scott J. Social network analysis: a handbook. 2nd ed. London: Sage Publications; 2000.

[4] Newman MEJ. The structure and function of complex networks. SIAM Rev 2003;45:167–256.

[5] Boccaletti S, Latora V, Moreno Y, et al. Complex networks: structure and dynamics. Phys Rep 2006;424:175–308.

[6] Jeong H, Mason SP, Barabasi AL, et al. Lethality and centrality in protein networks. Nature 2001;411(6833):41–2.

[7] Pastor-Satorras R, Vazquez A, Vespignani A. Dynamical and correlation properties of the internet. Phys Rev Lett 2001;87(25):258701.

[8] Crucitti P, Latora V, Porta S. Centrality in networks of urban streets. Chaos: an interdisciplinary journal of nonlinear science 2006;16(1):015113.

[9] Nieminen J. On the centrality in a graph. Scand J Psychol 1974;15:322–36.

[10] Sabidussi G. The centrality index of a graph. Psychometrika 1966;31:581–603.

[11] Freeman LC. A set of measures of centrality based upon betweenness. Sociometry 1977;40:35–41.

[12] Freeman LC, Borgatti SP, White DR. Centrality in valued graphs: a measure of betweenness based on network flow. Social Networks 1991;13(2):141–54.

[13] Latora V, Marchiori M. A measure of centrality based on network efficiency. New J Phys 2007;9(6):188.

[14] Latora V, Marchiori M. Efficient behavior of small-world networks. Phys Rev Lett 2001;87(19):198701.

[15] Latora V, Marchiori M. Economic small-world behavior in weighted networks. Eur Phys J B 2003;32(2):249–63.

[16] Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. Phys Rev E 2006;74(3):036104.

[17] Wang GX, Shen Y, Ouyang M. A vector partitioning approach to detecting community structure in complex networks. Comput Math Appl 2008;55(12):2746–52.

[18] Motter AE, Lai YC. Cascade-based attacks on complex networks. Phys Rev E 2002;66(6):065102.

[19] Amaral LAN, Scala A, Barthelemy M, et al. Classes of small-world networks. Proc Natl Acad Sci USA 2000;97(21):11149–52.

[20] Motter AE, Nishikawa T, Lai YC. Range-based attack on links in scale-free networks: are long-range links responsible for the small-world phenomenon? Phys Rev E 2002;66(6):065103.