Physica A 393 (2014) 560–570

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

The similarity of weights on edges and discovering of community structure

Yi Shen *

College of Information Science and Technology, Nanjing Agricultural University, Nanjing 210095, China

HIGHLIGHTS

- A weighted modularity based on the similarity of weights is proposed.
- Stable relationships between nodes are revealed by our method.
- Simulation results show the functions of the modularity.

ARTICLE INFO

Article history:
Received 1 August 2012
Received in revised form 27 May 2013
Available online 2 September 2013

Keywords:
Weighted networks
Similarity of weights
Community structure
Modularity function

ABSTRACT

In this paper, we propose a weighted modularity Q^W based on the similarity of weights on edges and a threshold coefficient ζ to evaluate the equivalence of edge weights. Simulations on benchmark networks and real networks show that optimization on the modularity enable us to obtain groups of nodes within which the edge weights are distributed uniformly but at random between them. The communities can reveal the uniform connections (stable relationships measured by the similarity of weights on edges) between nodes or some similarity between nodes' functions. Furthermore, with the dynamical moving of ζ, we observe that optimization on the Q^W allows for the discovering of a special hierarchical organization which reveals different levels of uniform connections between nodes in networks. The substructures revealed by the hierarchical organization enable us to obtain more information of networks, and give a potential way for partly remedying the intrinsic resolution problem of modularity.

1. Introduction

Complex network is an effective tool to study the real systems such as the social networks [1], communication networks [2], citation networks [3], etc. There is growing interest in the investigation of properties of such networks [4–9]. A common property of these networks is that they are structured in terms of modules or communities which are groups of nodes defined based on different concepts such as node similarity or comparison of initial graph with random graphs and so on [10–13]. Since the networks described at the level of communities are quite different from their properties at the level of the entire network, analyses that focus on communities is important for obtaining deep features of network topologies. Recently, many efforts [14–19] have been devoted to the detection of community structure in complex networks.

The community structure is corresponding to meaningful understandings of functions of groups [20] in networks. In unweighted networks, the community structure is groups of nodes in which edges are denser than the external between them [20]. The unweighted community structure reflects well on the close connections between functional units, and a large volume of detection algorithms have been proposed—the details can be seen in two recent comparisons, Refs. [21,22].

* Tel.: +86 02584395042.
E-mail addresses: shen_yi1979@126.com, shen_yi1979@njau.edu.cn.

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.08.063

Besides the topological issues that the interaction between two nodes is characterized by the existence of a link, it is important to realize that many complex networks are intrinsically weighted. For example, in social networks, the social ties between individuals may be strong or weak [23], and on the World Wide Web (WWW) the data traffic between websites may be heavy or light [24]. Moreover, in many cases the weights on links affect significantly the properties or functions of these networks, e.g., virus and disease spreading [25], synchronization dynamics of oscillators [26,27], and motif statistics [28,29]. It is clear that the edge weights, which indicate the strength of the interaction represented by a network, constitute an important variable for networks. Much useful information regarding the structure of networks is contained in the edge weights.

Generally, weights on edges taking greater values mean stronger connections for nodes pairs. According to it, Newman proposed the weighted community structure that groups nodes in which the edge weights are relatively larger than the external between them [11]. The weighted community structure proposed by Newman can be detected by optimizing the weighted modularity Q^w [11].

$$Q^w = \frac{1}{2T}\sum_{ij}\left[w_{ij} - \frac{T_iT_j}{2T}\right]\delta(c_i, c_j) \tag{1}$$

where w_{ij} represents the weight on the edge between nodes i and j, T_i is the strength of node $i : T_i = \sum_j w_{ij}$, $T = \sum_i T_i$ is the sum of all the edge weights in a network, δ is the Kronecker delta symbol that $\delta(c_i, c_j) = 1$ if $c_i = c_j$ and 0 otherwise, and c_i is the label of the community to which node i is assigned. Recently, several algorithms, such as the random walk-based method [30], the WGN algorithm [11], and the WEO algorithm [31], have been proposed for detecting the weighted community structure based on optimizing the modularity Q^w [11]. The random walk-based method [30] relies on a transition probability matrix that is calculated by evaluating the conductance between any two nodes. Consequently, problems with slowness and inaccuracy inevitably occur when the method is applied to dense weighted networks [30]. The WGN algorithm is generalized from its unweighted version, the GN algorithm [10], by calculating the weighted edge betweenness [11], and the WEO algorithm is generalized from its unweighted version, the EO algorithm, by replacing Q [20] by Q^w [11]. The results in Ref. [32] show that the WEO algorithm has the best performance among the three methods in terms of accuracy. It is worth pointing out that the distribution of weights on edges contains substantial potentially useful information. Therefore, the criteria for defining and evaluating community structure in weighted networks are not only limited to maximizing the edge weights within groups. For example, in social networks, the strength of social ties between individuals described by the weights on the edges between nodes are more evenly distributed within groups than those between groups, which indicates the existence of stable relationships between individuals or some similarity between their functions in groups.

In this paper, we propose a new community structure for weighted networks, i.e., groups of nodes within which the edge weights are distributed uniformly but between which they are distributed at random. The community structure can be detected by optimizing the weighted modularity Q^W, which is based on the similarity of weights, proposed by us. However, modularity has a serious resolution limit problem [33,34] that modularity optimization may fail to identify modules smaller than a certain size which depends on the number of links and on the degree of interconnectedness of the modules, even in cases where modules are unambiguously defined [33]. The performance is even worse for large networks [33] because larger networks usually have a wide distribution of modules size. Studies [33,34] show that simple optimization on modularity might miss meaningful substructures of a network, as confirmed in many real world examples [33]. In our proposed modularity Q^W, we define an extra threshold coefficient ζ for evaluating the equivalence of edge weights. With different value of ζ, we can obtain a special hierarchical organization that can describe different levels of the substructures of a network in the sense of uniform connections. The hierarchical organization cannot be achieved by optimization on the conventional weighted modularity Q^w, and it gives a potential way for partly remedying the intrinsic resolution problem of modularity.

2. The modularity function based on the similarity of weights

The detection of community structure for a network via modularity optimization depends greatly on the configuration model of the modularity [10–12]. Here, we consider an understandable case of a weighted network. If nodes i and j have more edges with the same weights, a weighted edge is more likely to be assigned between them. Otherwise, if they have no edge with the same weights, there should be no edge between them. Therefore, when we observe that there is an edge between nodes i and j in a real weighted network, it should be surprising if they have no edge with the same weights, and should make a bigger contribution to the modularity since modularity should be high for statistically surprising configurations. We propose a new configuration model according to this fact. We define w_{id} and w_{je} to be the weights on the dth edge of node i and the eth edge of node j, respectively, and we define $\Phi(a, b) = I_{(||a|-|b||\le\zeta)}$ as an event indicator function that returns $\Phi(a, b) = 1$ when $||a| - |b|| \le \zeta$ and 0 otherwise. ζ is a threshold coefficient for evaluating the equivalence of edge weights. We present the configuration model as follows:

$$P_{ij}^W = \frac{\sum_{ij}\sum_{d\in k_i}\sum_{e\in k_j}\Phi(w_{id}, w_{ie})}{2m}. \tag{2}$$