Physica A 394 (2014) 346–357

Contents lists available at ScienceDirect

Physica A

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

Weighting links based on edge centrality for community detection

Peng Gang Sun *

School of Computer Science and Technology, Xidian University, Xi'an 710071, China
Institute of Computational Bioinformatics, Xidian University, Xi'an 710071, China

HIGHLIGHTS

- Links and link weights are closely associated for the emergence of communities.
- Edge centralities can distinguish internal links of communities and external ones.
- Results on weighted networks with greater internal link weights outperform others.

ARTICLE INFO

Article history:
Received 10 March 2013
Received in revised form 2 August 2013
Available online 8 September 2013

Keywords:
Community
Edge centrality
Weighted function

ABSTRACT

Link weights have the equally important position as links in complex networks, and they are closely associated with each other for the emergence of communities. How to assign link weights to make a clear distinction between internal links of communities and external links connecting communities is of vital importance for community detection. Edge centralities provide a powerful approach for distinguishing internal links from external ones. Here, we first use edge centralities such as betweenness, information centrality and edge clustering coefficient to weight links of networks respectively to transform unweighted networks into weighted ones, and then a weighted function that both considers links and link weights is adopted on the weighted networks for community detection. We evaluate the performance of our approach on random networks as well as real-world networks. Better results are achieved on weighted networks with stronger weights of internal links of communities, and the results on unweighted networks outperform that of weighted networks with weaker weights of internal links of communities. The availability of our findings is also well-supported by the study of Granovetter that the weak links maintain the global integrity of the network while the strong links maintain the communities. Especially in the Karate club network, all the nodes are correctly classified when we weight links by edge betweenness. The results also give us a more comprehensive understanding on the correlation between links and link weights for community detection.

1. Introduction

One of the key properties for complex networks characterized by the presence of groups of nodes, called communities or modules with more internal links between nodes of the same group and comparatively fewer external links between nodes of different groups has been found in most complex systems many years ago [1–3]. The study of communities is of great

* Correspondence to: School of Computer Science and Technology, Xidian University, Xi'an 710071, China. Tel.: +86 29 88207344.
E-mail address: psun@mail.xidian.edu.cn.

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.08.048

importance, since they not only can reveal the organization characteristic of complex networks, but also uncover the hidden correlations among their components [4]. Therefore, community detection has attracted the attention of lots of scholars from various fields such as mathematics, physics, computer science, biological science and social sciences et al. [1–3] in recent years.

Plenty of techniques have been applied on community detection in the past several years [3–9]. An interesting approach is that Newman and Girvan [2,10,11] transform the problem of community detection into an optimization problem [2,10,11] by considering a measure, *modularity*, or Q as the objective function. However, the modularity optimization exposes to the resolution limits problem [12] in that the communities depend on the total size of links of the network, in the extreme case that cliques connected by a single link may not be correctly revealed [12–14]. Li and Zhang proposed [13] a quantitative function, *modularity density*, or the D-value to solve the resolution limit problem by maximizing this function, and their method got better results for community detection. However, the emergence of communities not only has close relationships with network links, but also link weights. Years ago, Granovetter [15,16] gave his findings that weak and strong links play different roles in real-world networks, that weak links maintain the global integrity of the network while strong links maintain the communities. For this reason, a weighted function, *modularity intensity*, that both considers links and link weights was proposed for community detection [17], and the results showed that maximizing this function obtained better performance in detecting communities when internal links of communities are stronger than external. Nevertheless, a new problem about how to assign link weights to make a clear distinction between internal links of communities and external ones arises subsequently, which is of vital importance for community detection. Edge centralities provide a powerful tool for distinguishing internal links of communities from external ones, and are successfully used by divisive algorithms, which remove internal links iteratively prior to external ones for splitting networks into groups [2,3,18,19]. Recently, many measures of edge centralities such as edge betweenness [18], information centrality [19] and edge clustering coefficient et al. [20] were proposed and achieved better performance for community detection.

Based on the discussion above, how to weight network links becomes a very important problem that will influence the results of community detection. In this paper, we use edge centralities such as betweenness, information centrality and an edge clustering coefficient to weight links of networks respectively to transform unweighted networks into weighted, and a weighted function, *modularity intensity* that both considers links and link weights is adopted on the weighted networks for community detection. We evaluate the performance of our approach on random networks as well as real-world networks.

The rest of the paper is organized as follows. In Section 2, we give an introduction to modularity intensity. In Section 3, we weight network links by edge centralities and give a comparison analysis among them. Section 4 presents the experimental results on both random networks and real-world networks. A conclusion is provided in Section 5.

2. Modularity intensity

Recently, a weighted function, *modularity intensity* was proposed [17] for evaluating the cohesiveness of a community. This measure is different from the modularity density in that it not only considers links between vertices, but also link weights. Given $G = (V, E)$ is a network, V and E are its sets of vertices and edges respectively. $G_1(V_1, E_1), G_2(V_2, E_2), \ldots, G_n(V_n, E_n)$ are n induced subnetworks for a partition of G, where V_i, E_i are sets of vertices and edges of G_i respectively, $i = 1, 2, \ldots, n$, $\forall i, j \in \{1, 2, \ldots, n\}$, $V_i = V_j$ or $V_i \cap V_j = \emptyset$. The modularity intensity of the partition of the network is defined as follows:

$$MI = \sum_{i=1}^{n} mi(G_i) = \sum_{i=1}^{n} \frac{\alpha \cdot \sum\limits_{\forall s \in V_i, \forall t \in V_i} A_{st} \cdot B_{st} - \beta \cdot \sum\limits_{\forall s \in V_i, \forall t \in \bar{V}_i} A_{st} \cdot B_{st}}{|V_i|} \tag{1}$$

where $mi(G_i)$ is the modularity intensity of G_i. $\bar{V}_i = V - V_i$. A is the adjacency matrix of G; B is the link weights matrix of G. α, β are the tuning parameters ($\alpha, \beta \in (0, 1], \alpha \geq \beta$). Here, we set $\alpha = \beta = 1$. Eq. (1) can be simplified as:

$$MI = \sum_{i=1}^{n} mi(G_i) = \sum_{i=1}^{n} \frac{\sum\limits_{\forall s \in V_i, \forall t \in V_i} A_{st} \cdot B_{st} - \sum\limits_{\forall s \in V_i, \forall t \in \bar{V}_i} A_{st} \cdot B_{st}}{|V_i|}. \tag{2}$$

For comparison, we rewrote the modularity density of the partition of the network by the following equation [13]:

$$D = \sum_{i=1}^{n} d(G_i) = \sum_{i=1}^{n} \frac{\sum\limits_{\forall s \in V_i, \forall t \in V_i} A_{st} - \sum\limits_{\forall s \in V_i, \forall t \in \bar{V}_i} A_{st}}{|V_i|} \tag{3}$$

where $d(G_i)$ is the modularity density of G_i.