Physica A 392 (2013) 6125–6134

Contents lists available at ScienceDirect

Physica A

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

Detecting overlapping communities by seed community in weighted complex networks

Junqiu Li, Xingyuan Wang *, Justine Eustace

Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China

HIGHLIGHTS

- Introduce seed communities and absorbing degree.
- Propose a new algorithm.
- Find overlapping vertices.
- Give excellent experimental results.

ARTICLE INFO

Article history:
Received 4 March 2013
Received in revised form 24 June 2013
Available online 7 August 2013

Keywords:
Seed community
Weighted networks
Overlapping community
Absorbing degree

ABSTRACT

Detection of community structures in the weighted complex networks is significant to understand the network structures and analysis of the network properties. We present a unique algorithm to detect overlapping communities in the weighted complex networks with considerable accuracy. For a given weighted network, all the seed communities are first extracted. Then to each seed community, more community members are absorbed using the absorbing degree function. In addition, our algorithm successfully finds common nodes between communities. The experiments using some real-world networks show that the performance of our algorithm is satisfactory.

1. Introduction

Systems that comprise many interacting parts with the ability of generating a new quality of macroscopic collectively can be characterized as complex systems [1]. Complex systems are commonly modeled as complex networks or graphs [2–6]. Here the entities of the system are represented by the vertices of the graph and the interactions between the vertices are represented by the links.

Recently, the characterization of community (or module) structures in complex networks has received considerable attention, and it is widely believed that large networks in the real world are composed of many communities, whose nodes often cluster into tightly knit groups with high density of within-group edges and low density of between-group edges [7–10]. The vertices can be into group or cliques according to their structural position in the networks. In some community, vertices sharing a large number of links with other vertices may have important role enforce stability within the community; vertices settled at the boundaries between communities play an important role of mediation and lead to the relationships and exchange between different communities [7].

Up to now, a large number of computer algorithms have been proposed to detect community structures in complex networks. Loosely we can classify community detection algorithms into greedy algorithm, spectral graph partitioning and clustering methods, for example, the Kernighan–Lin algorithm which divides the networks according to the optimization

* Corresponding author.
E-mail addresses: meiliqiutian@126.com (J. Li), wangxy@dlut.edu.cn (X. Wang), justineustace@yahoo.com (J. Eustace).

0378-4371/$ – see front matter
http://dx.doi.org/10.1016/j.physa.2013.07.066

of the number of within and between-community edges using a greedy algorithm [11] and a spectral bisection algorithm based on the eigenvectors of the Laplacian matrix of a graph [12]. Another category is Hierarchical clustering which detects a community based on the similarity or intensity between vertices, where it is divided into agglomerative method and divisive method by adding or removing edges [13].

In 2002, the most popular algorithm for community detection called GN algorithm was proposed by Girvan and Newman [8]. This algorithm recursively removes edges with the highest edge betweenness of shortest path between pairs of vertices that run along the edge, and then constructs a hierarchical tree. To estimate the accuracy of this method, they introduced a quantitative measure for the quality of network division, called modularity (represented by the Q function) [14]. In order to deal with large real-world networks, Newman proposed a fast algorithm based on the optimization of the modularity in Ref. [15] with the time complexity $O(n^2)$.

Various other related approaches have been proposed based on the analysis of successive neighborhoods, extracting the initial seed and many different ideas in Refs. [16–24]. The computational complexity of more accurate algorithms is often high. Based on this deduction, here we propose a fast and efficient algorithm for detecting overlapping community structures in the weighted networks. The key strategy of our algorithm is to mine all the seed communities. Using the absorbing degree function, our algorithm computes the absorbing degree between the neighboring vertices of the seed community and the seed community. Then according to the value of the absorbing degree, the initial seed communities constantly enlarge. The other advantage of our algorithm is that common vertices can be found accurately during the community structure detection.

We organized this paper as follows. In Section 2 we explain the crucial concepts of our new overlapping community detection algorithm. In Section 3 we describe how to extract the seed community and propose a new algorithm in detail. Experimental results of synthetic and some real-world networks are shown in Section 4. Finally, our conclusions appear in Section 5.

2. Basic concepts

2.1. Community structure and common vertex

Along with the physical significance of the network and statistical properties, the community structure is a common property of many complex networks [8]. The most distinguishing feature of complex networks is connectivity degree density. Generally community structures in real networks, the nodes within a group have higher edge density than the nodes among groups. Two communities may overlap when a node joins different communities. For example, in Fig. 1, there are four communities, denoted by circle, square, triangle, and octagon, respectively. Node 6 belongs to the circle community as well as the square community. It is a common phenomenon that a vertex should belong to several communities in a real-world complex network at the same time. We call these kinds of vertices and their corresponding communities as common vertices and overlapping community structures respectively. In unweighted networks, a common vertex has the same number of edges to different communities. For example Fig. 2 shows the sketch of a weighted network with such a community structure, where the dotted circles represent a vertex, and each edge characterizes the relations between a pair of vertices with edge weight.

2.2. Edge weight

A network can be modeled as $G(V, E)$, where V represents the set of n vertices or nodes, and E is a set of m links. In this work we shall use unweighted or weighted, but undirected complex networks. For given vertices $i \in V$ and $j \in V$, the edge weight E_{ij} is defined as

$$E_{ij} = \begin{cases} w_{ij}, & \text{if } i \text{ and } j \text{ are connected in weighted networks} \\ 1, & \text{if } i \text{ and } j \text{ are connected in unweighted networks} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

2.3. Vertex weight

The vertex weight V_i of vertex i in networks is defined as

$$V_{i(i\in V)} = \begin{cases} \sum_{j\in V} E_{ij}, & \text{if } j \text{ is the neighboring vertices of } i \text{ in weighted networks and unweighted networks} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

For example, in Fig. 2, the vertex weight of vertex 3 equals $0.2 + 0.5 + 0.6 + 0.7 = 2.0$, the vertex weight of vertex 6 is 1.4, etc. If the network is unweighted, the vertex weight of vertex 3 is 4. The vertex weight of a vertex in an unweighted network is equal to the total number of edges connected with this vertex, which is identical to the vertex degree.