Physica A 389 (2010) 4177–4187

Contents lists available at ScienceDirect

Physica A

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

Detecting overlapping communities of weighted networks via a local algorithm

Duanbing Chen[a], Mingsheng Shang[a], Zehua Lv[b,*], Yan Fu[a]

[a] Web Sciences Center, School of Computer Science, University of Electronic Science and Technology of China, Chengdu 611731, PR China
[b] School of Computer Science, Huazhong University of Science and Technology, Wuhan 430074, PR China

ARTICLE INFO

Article history:
Received 15 January 2010
Received in revised form 20 May 2010
Available online 4 June 2010

Keywords:
Weighted networks
Overlapping community
Local algorithm
Node strength

ABSTRACT

Identification of communities is significant in understanding the structures and functions of networks. Since some nodes naturally belong to several communities, the study of overlapping communities has attracted increasing attention recently, and many algorithms have been designed to detect overlapping communities. In this paper, an overlapping communities detecting algorithm is proposed whose main strategies are finding an initial partial community from a node with maximal node strength and adding tight nodes to expand the partial community. Seven real-world complex networks and one synthetic network are used to evaluate the algorithm. Experimental results demonstrate that the algorithm proposed is efficient for detecting overlapping communities in weighted networks.

1. Introduction

Many complex systems in nature and society can be described as graphs or networks [1–3]. A community detecting problem is to detect community structures, which may overlap each other, in a weighted or unweighted complex network with n nodes and m edges. The edge weight can be set to 1 if the network is unweighted.

In recent years, the detection and analysis of community structures in complex networks has attracted a great deal of attention in many applications [4]. Community structure, whose nodes often cluster into tightly knit groups with high density of within-group edges and low density of between-group edges, has been paid much attention because of its significance in analyzing complex networks.

So far, many algorithms have been proposed to detect communities. Two classical algorithms are the spectral bisection algorithm on the basis of eigenvectors of the Laplacian matrix of a graph [5] and the Kernighan–Lin algorithm which improves on an initial division of network by optimization of the number of within- and between-community edges using a greedy strategy [6]. In recent years, various community detecting algorithms on the basis of modularity [7] have been presented. In 2004, Newman proposed a fast algorithm to detect community structures [8] based on modularity. Excellent results are obtained, especially for sparse networks, and it is typically thousands of times faster than the algorithm in [7]. Ruan and Zhang [9] proposed an efficient heuristic algorithm *QCUT*, which combines spectral graph partitioning and local searching to optimize the modularity Q. They applied *QCUT* to study a protein–protein interaction network and reveal some interesting biological results. Duch and Arenas [10] presented a method to find communities in complex networks based on

* Corresponding author. Tel.: +86 153 2726 0979.
E-mail address: hustcdb@yahoo.com.cn (Z. Lv).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2010.05.046

extremal optimization of modularity. Experimental results show that it is feasible to be used for the accurate identification of communities in large complex networks. Wang et al. [11] proposed a very fast algorithm for detecting the community. The algorithm used local information and local modularity to analyze community structures in complex networks. It is based on a table that describes a network and a virtual cache similar to the cache in a computer structure. Other community structure detecting algorithms based on different strategies also have been presented. For example, Clauset et al. [12] presented a hierarchical agglomeration algorithm to detect the community in very large networks. Its running time on a network with n nodes and m edges is $O(md \log n)$, where d is the depth of the dendrogram describing the community structure. Clauset [13] proposed a local community detection algorithm based on local modularity defined by the author. The algorithm adds a node into partial community C and updates the neighbors of C at each step. In 2006, Newman proposed a community structure detecting algorithm using eigenvectors of matrices [14]. Based on the previous benefit function known as modularity, the author presented another benefit function–modularity matrix which plays a role in community detection similar to the Laplacian matrix in graph partitioning calculations. Chen et al. [15] presented a fast and efficient algorithm by adding a node into a partial community recursively until obtaining a local optimal community. An even faster and more accurate algorithm based on subgraph similarity was proposed by Xiang et al. [16].

All the above-mentioned algorithms have a strong assumption that each node belongs to one and only one community. However, a node may belong to several communities in real-world networks, and furthermore, some networks are weighted. Because of this, researchers pay much attention to studying weighted networks and detecting overlapping communities. Barrat et al. [17] studied two real-world weighted networks, i.e., a scientific collaboration network and a world-wide air-transportation network. They defined appropriate metrics combining weighted and topological observables that enable one to characterize the complex statistical properties and heterogeneity of the actual strength of the edges and vertices. The weights characterizing the various connections exhibit complex statistical features with highly varying distributions and power-law behavior. Their study showed that the analysis of the weighted quantities and the study of the correlations between weights and topology provide a complementary perspective on the structural organization of the network that might be undetected by quantities based only on topological information. Ou et al. [18] proposed a model for resource-allocation dynamics to investigate the dynamic behavior of resource/traffic flow on weighted scale-free networks. They found that the dynamical system will evolve into a kinetic equilibrium state, where the strength, defined by the amount of resource or traffic load, is correlated with the degree in a power-law form with tunable exponent. In recent years, overlapping community structures have been widely studied [19–27]. Baumes et al. [19] proposed two efficient local algorithms, *RaRe* and *IS*, to detect the overlapping community structures in networks, and both synthetic and real-world networks were used to test the performance of algorithms. Lancichinetti et al. [20] presented a local algorithm based on local optimization of a fitness function to find overlapping communities which are revealed by peaks in the fitness histogram. In [21], the community structure is detected by k-clique percolation and overlaps between communities are guaranteed by the fact that one node can participate in more than one clique. Evans and Lambiotte [22] used a partition of the links of a network to uncover its overlapping community structure in the complex network. Shen et al. [23] proposed a new measure named Q_c, which is based on a maximal clique view of the original network, to quantify the overlapping community structure and presented an identifying method to detect the overlapping community by finding an optimal cover, i.e., the one with the maximal Q_c. Zhang et al. [24] designed a novel algorithm to identify overlapping communities in complex networks by first mapping the network nodes into Euclidean space and then applying fuzzy c-means clustering. Chen et al. [25] presented another algorithm to detect the overlapping community structure in the network by expanding a partial community which is started from a special single node. Gregory [26] proposed a two-phase method of detecting overlapping communities. In the first phase, a network is transformed to a new one by splitting nodes, using the idea of split betweenness; in the second phase, the transformed network is processed by a disjoint community detection algorithm. Airoldi et al. [27] presented a new overlapping community detecting method, i.e., a Bayesian model, to investigate the protein–protein interactions, including their typical interaction patterns, and the degree of membership of objects to groups. Very recently, Shang et al. [28] proposed an efficient overlapping community detection algorithm based on merging the community cores and expanding the community.

Inspired by the above approaches, in this paper, a new local algorithm based on node strength is proposed to detect the overlapping community structures. The main strategies are to find an initial community from a node with maximal node strength and to expand the partial community from the initial one by adding nodes that are tight with the community. Some real-world and synthetic networks are used to evaluate the presented algorithm and acceptable results are obtained.

2. Fundamental concepts

2.1. Community

A community consists of nodes and edges between these nodes, where nodes often cluster into tightly knit groups with high density of within-group edges and low density of between-group edges. It is noted that two communities may overlap each other since a node can join different communities. For example, as shown in Fig. 1, there are three communities in this network, denoted by circle, square and triangle, respectively. Node 6 is a common node since it should belong to the circle community as well as the square community.