Physica A 415 (2014) 398–406

Contents lists available at ScienceDirect

Physica A

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

Uncovering the overlapping community structure of complex networks by maximal cliques

Junqiu Li, Xingyuan Wang *, Yaozu Cui

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

HIGHLIGHTS

- Introduce the maximal cliques, overlapping vertex, bridge vertex and isolated vertex.
- Propose a new algorithm to find all the maximal cliques.
- Expand the maximal cliques by some rules given in this paper.
- Give satisfactory experimental results.

ARTICLE INFO

Article history:
Received 5 March 2014
Received in revised form 9 July 2014
Available online 20 August 2014

Keywords:
Complex networks
Overlapping communities
Maximal cliques

ABSTRACT

In this paper, a unique algorithm is proposed to detect overlapping communities in the un-weighted and weighted networks with considerable accuracy. The maximal cliques, overlapping vertex, bridge vertex and isolated vertex are introduced. First, all the maximal cliques are extracted by the algorithm based on the deep and bread searching. Then two maximal cliques can be merged into a larger sub-graph by some given rules. In addition, the proposed algorithm successfully finds overlapping vertices and bridge vertices between communities. Experimental results using some real-world networks data show that the performance of the proposed algorithm is satisfactory.

1. Introduction

In nature and society, many real networks can be described as complex networks [1–3], e.g. relationship in social system, metabolism in biological systems, protein interaction network, literature reference network, and scientific research cooperation network, etc. Complex networks are a collection of vertices and edges, the vertices of the graph represent the entities of the system and the links represent the interactions between the entities. As the further study on the characters and the physical meanings of the complex networks in the microscopic and mesoscopic level, researchers found that a common feature which is called community structure exists in many real networks. Communities are groups or clusters of vertices within which the network connections are densely connected to each other, but between which they are sparsely connected with the rest of the network [4–7].

In recent years, many algorithms for detecting communities have been proposed. Two classical algorithms are the Kernighan–Lin algorithm that uses a greedy algorithm and classifies the networks by optimizing the number of within and between-community edges [8] and the spectral bisection algorithm based on the eigenvectors of the Laplacian matrix of graph [9]. In Ref. [10], Newman and Girvan proposed a divisive algorithm called the GN algorithm that recursively removes edges with the highest betweenness of the edges, constructed a hierarchical tree, and then we can obtain various divided

* Corresponding author.
E-mail addresses: meiliqiutian@126.com (J. Li), wangxy@dlut.edu.cn (X. Wang), cyz3471@sina.com (Y. Cui).

http://dx.doi.org/10.1016/j.physa.2014.08.025

communities by cutting the hierarchical tree at different points. This method makes up for the deficiency of the traditional algorithms, but also has its own defects. This algorithm has high time complexity and has not the defining quantity of community structure in the network. To measure the quality of network partitioning, the modularity [10] represented by the Q function was introduced. Recently, many algorithms are proposed based on modularity. In 2004, Newman proposed a fast greedy algorithm [11] to maximize the modularity, starting with the initial network of n independent communities, with each community having one vertex. Clauset and Newman described a new algorithm for inferring community structure from networks by greedily optimizing the modularity [12], which runs in time $O(md \log n)$ for a network with n vertices and m edges where d is the depth of the dendrogram. This algorithm is considerably faster than most previous traditional algorithms. The same algorithm was presented by Duch and Arenas [13] based on an extremal optimization of the value of modularity. Newman proposed an algorithm using the eigenvectors of matrices [14]. Wang et al. [15] proposed a very fast algorithm only based on local information to detect community. Chen et al. [16] presented a fast and efficient algorithm by mining a node with the closest relations with the community recursively until obtaining a local optimal community.

All the above algorithms mentioned are unable to find overlapping communities, and assume that each vertex belongs to only one community. However, there are some vertices belonging to more than one community in some real networks. So there are a lot of algorithms to explore overlapping community structures in the complex networks recently [17–20]. In addition there is another special type of complex network, which is made up of a lot of maximal cliques. A maximal clique is a complete sub-graph which is not a subset of any other complete sub-graph in a complex network. The number of vertex in a maximal clique is more than two and any two vertices have a link connecting to each other. All the maximal cliques are connected by some overlapping vertices or bridge vertices (in Sections 2.1 and 2.2). This special type of network can be seen in real life. For instance, in scientific collaboration network [21], some authors belong to different research groups characterized by different research directions. Such inter-community authors are often of the most interest. In Chinese medical prescription network (see Fig. 1), a number of Chinese medicinal materials constitute some Chinese medical prescriptions. Fig. 1 is a small medical prescription network, which is an un-weighted network with 20 Chinese medicines as vertices and 6 maximal cliques representing the number of the Chinese medical prescriptions. According to Chinese medical prescriptions, these 6 maximal cliques cannot be merged into a larger sub-graph, so 6 communities and 6 overlapping vertices are obviously detected from Fig. 1.

In this paper, a new algorithm based on the deep and bread searching for extracting all the maximal cliques (complete sub-graphs) is proposed firstly. According to the characteristics of real networks and some given rules, some maximal cliques connected by some overlapping vertices can be merged into a larger sub-graph. All maximal cliques are the community structure in some real networks which are completely composed of the maximal cliques connected by some overlapping vertices. Last, some real world networks are used to test the performance of the presented algorithm.

2. Preliminary concepts

2.1. Overlapping vertex

In the past few years, there are many algorithms to detect non-overlapping communities in complex network. But many real networks have overlapping communities and some vertices belong to more than one community. So this vertex is called the overlapping vertex, and this network having the overlapping vertices is called overlapping neighboring network in this paper. In Fig. 1, this is overlapping neighboring network with six overlapping vertices and six maximal cliques.

2.2. Bridge vertex

Recent studies revealed that some real networks cannot be able to be divided into two communities without allowing bridge vertices connecting multiple groups or communities. The concept of "bridge" is defined as the vertices that cross structural holes between discrete groups of people [22].

To emphasize the importance of bridge vertices in community detection and to illustrate the concept, we take a simple graph shown in Fig. 2 as an example. A visual inspection of this graph most likely suggests three maximal cliques {1, 2, 3}, {4, 5, 6, 7, 8}, {14, 15, 16, 17}. Two maximal cliques are connected by vertex 11, which does not belong to any maximal clique. In this paper, similar vertex 11 is called bridge vertex.

2.3. Isolated vertex

In addition to extracting maximal cliques, overlapping vertex, and bridge vertices, some vertices with low vertex degree do not belong to any of maximal cliques. These vertices are called isolated vertex in this paper. In different real networks, isolated vertices can be isolated from the entire network during the early stage of extracting maximal cliques. At the last stage, isolated vertices can be divided into some communities by some given rules. In Fig. 2, vertex degree of vertices 9, 10, 12 and 13 is 1 and these vertices do not belong to any maximal cliques. So these vertices are called isolated vertices.