



Detecting community structure via the maximal sub-graphs and belonging degrees in complex networks



Yaozu Cui, Xingyuan Wang*, Justine Eustace

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

HIGHLIGHTS

- Introduce the maximal sub-graphs and the belonging degrees.
- Propose BASH (Based on mAximal Sub-graphs) algorithm.
- Find overlapping vertices.
- Give excellent experimental results.

ARTICLE INFO

Article history:

Received 23 April 2014

Received in revised form 28 July 2014

Available online 30 August 2014

Keywords:

Complex networks

Community structure

Maximal sub-graph

Belonging degree

ABSTRACT

Community structure is a common phenomenon in complex networks, and it has been shown that some communities in complex networks often overlap each other. So in this paper we propose a new algorithm to detect overlapping community structure in complex networks. To identify the overlapping community structure, our algorithm firstly extracts fully connected sub-graphs which are maximal sub-graphs from original networks. Then two maximal sub-graphs having the key pair-vertices can be merged into a new larger sub-graph using some belonging degree functions. Furthermore we extend the modularity function to evaluate the proposed algorithm. In addition, overlapping nodes between communities are founded successfully. Finally we report the comparison between the modularity and the computational complexity of the proposed algorithm with some other existing algorithms. The experimental results show that the proposed algorithm gives satisfactory results.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Many types of systems in nature and society can be represented as complex networks or graphs [1,2], where nodes (vertices) represent entities in real systems and edges represent the interactions between the vertices of the graphs. Examples include social relationships [3], spreading of viruses and diseases [4], the Internet and the WWW (World Wide Web) [5,6], and many others. The interaction of entities and topological properties of the networks actually determine the complex systems' dynamics. In the past decade, one property that attracted much attention is community structure, whose nodes often cluster into tightly knit groups with high density of within-group edges and low density of between-group edges [7–9]. Therefore, an efficient approach is need that can detect community structure of complex networks. Some vertices usually share common properties or play similar roles in the complex networks [7]. So detecting overlapping community structures is important to understand the structures and functional properties of the networks. It mainly

* Corresponding author.

E-mail addresses: cyz3471@sina.com (Y. Cui), wangxy@dlut.edu.cn (X. Wang), justineustace@yahoo.com (J. Eustace).

studies relative qualities on vertex degree, clustering coefficient of each vertex, betweenness and modularity [8] in complex networks. Recently a large number of algorithms have been proposed to detect the community structures in complex networks, and they can be roughly grouped into two categories.

The first category includes heuristic graph partitioning algorithms. This category of algorithms partitions the networks into some groups while minimizing the number of edges between the groups. And each vertex belongs to only one community. For example, Kernighan–Lin algorithm is a famous heuristic algorithm, which classifies the networks according to optimization of the number of within and between-community edges using a greedy algorithm [10]. This algorithm detects the community structures after we know the number of community. The other well-known algorithm is the spectral bisection algorithm based on the eigenvectors of Laplacian matrix of graph [11,12], which can divide the network into two communities with the same size. If the network would be divided into more than two communities, this algorithm must repeatedly execute in one sub-community.

Algorithms in the second category adopt hierarchical clustering which devote to discover the connected similarity or intensity between vertices, where they can be divided into agglomerative method and divisive method by adding or removing edges [13]. In this kind of category, overlapping vertices will exist, that is to say, some vertices may belong to more than one community. Girvan and Newman [7] proposed a most popular divisive algorithm in 2002, named GN algorithm. According to the highest edge betweenness of shortest path between pairs of vertices that run along the edge, this algorithm recursively removes edges to get the communities in a network, and then constructs hierarchical tree. If we do not know the number of communities in a real network, GN algorithm also will not know which step to terminate this divisive. In order to solve this problem, Newman et al. [14] introduced the modularity (represented by a Q function) to measure the quality of network division. And a fast algorithm based on greedy algorithm in Ref. [15] was proposed by Newman in 2004. In addition, recently there are some algorithms [16–23] to find overlapping community structures in networks.

BASH (Based on mAXimal Sub-graphs) algorithm is proposed in this paper to detect community structures in the complex networks. The key strategy of this algorithm is to mine all the maximal sub-graphs and the key pair-vertices. Then two maximal sub-graphs having the key pair-vertices are merged into a new larger sub-graph using some belonging degree functions. The other advantage of this approach is that the overlapping vertices can be found accurately during the community structures detection. The rest of this paper is organized as follows. In Section 2 we explain the maximal sub-graphs and the key pair-vertices and propose BASH algorithm for extracting the communities in detail. We give a number of applications of BASH algorithm to synthetic and some real-world networks in Section 3. In Section 4, we give our conclusions.

2. Description of the algorithm

2.1. Main idea

For an un-weighted network, the edge weight is 1 when there is an edge between two vertices. So the vertex weight $[V_i]$ is the total weight of the edges that are connected with it in an un-weighted or weighted network. The vertex weight is the same as the vertex degree which is the total number of edges that is connected with it in an un-weighted network. For example, the vertex weight and the vertex degree of vertex 1 equals 16 in Fig. 1.

Through the analysis of some real networks, we find that some vertices (nodes) constitute a fully connected sub-graph. This indicates that this sub-graph usually has the highest link-density and all vertices are connected to each other in a fully connected sub-graph. Based on this observation, we can call such fully connected sub-graph as the maximal sub-graph $[G_s]$. And a maximal sub-graph does not belong to other maximal sub-graphs. For example, in Fig. 2, there are seven maximal sub-graphs $\{v_1, v_3, v_4\}$, $\{v_2, v_5, v_6\}$, $\{v_3, v_4, v_5, v_6\}$, $\{v_4, v_6, v_9\}$, $\{v_7, v_9, v_{10}\}$, $\{v_7, v_8, v_{10}, v_{11}\}$, $\{v_{11}, v_{12}\}$ from the original network. In these maximal sub-graphs, any one does not contain the others. But there are two vertices and the vertex degree of vertex 12 is 1 in maximal sub-graph $\{v_{11}, v_{12}\}$. This maximal sub-graph cannot affect the division of community structure. So in this paper, the number of vertex in a maximal sub-graph is more than two vertices. And the maximal sub-graphs can be merged into a new larger sub-graph using some belonging degree functions to detect the community structure.

2.2. The algorithm

Different from traditional agglomerative algorithms [15], BASH algorithm proposed is a hierarchical clustering algorithm to detect the community structure, which deals with the set of maximal sub-graphs rather than the set of vertices. In the BASH algorithm, we firstly need to extract all the maximal sub-graphs from the original networks. This can be done by many algorithms. Here we choose the well-known Bron–Kerbosch algorithm [24] to extract all the maximal sub-graphs.

Through analyzing all the maximal sub-graphs extracted from a given network, we find that two or more than two vertices connected may exist in some maximal sub-graphs at the same time. These vertices have an important role in detecting community structure. With this assumption, these vertices are called the key pair-vertices. The important feature of the key pair-vertices is that the number of these vertices is two or more than two, these vertices are connected to each other, and the key pair-vertices may exist in many sub-graphs. The number of the key pair-vertices differs in different real networks. The main stage after extracting all the maximal sub-graphs is to find the key pair-vertices.

Download English Version:

<https://daneshyari.com/en/article/7380219>

Download Persian Version:

<https://daneshyari.com/article/7380219>

[Daneshyari.com](https://daneshyari.com)