A Novel Algorithm for Hierarchical Community Structure Detection in Complex Networks

Chuan Shi, Jian Zhang, Liangliang Shi, Yanan Cai, and Bin Wu

Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia
Beijing University of Posts and Telecommunications, Beijing, China, 100876
{shichuan,strubby,stona126,caiyanan,wubin}@bupt.edu.cn

Abstract. Networks have in recent years emerged as an invaluable tool for describing and quantifying complex systems in many branches of science. Recent studies suggest that network often exhibit hierarchical organization, where vertices divide into groups that further subdivided into groups of groups, and so forth over multiple scales. In this paper, we introduce a novel algorithm that searches for the hierarchical structure. The method iteratively combines the similar communities with the elaborate design of community similarity and combination threshold. The experiments on artificial and real networks show that the method is able to obtain reasonable hierarchical structure solutions.

Keywords: hierarchical structure, community detection, similarity matrix, possibility matrix.

1 Introduction

Community detection in complex networks has attracted a lot of attentions in recent years. The communities are groups of nodes that are densely interconnected but only sparely connected with the rest of the network [1] [2]. Since a decisive advance was made by Newman and Girvan [3] in 2002, the community detection algorithms are experiencing a surge. The contemporary community detection algorithms can be roughly classified into two categories: optimization based methods and heuristic methods. These algorithms can effectively reveal the community structures; however, they cannot solve the networks with hierarchical structure.

Traditionally, hierarchical structure is represented by a tree or dendrogram in which closely related pairs of vertices have lowest common ancestors that are lower in the tree than those of more distantly related pairs. This structure can be modeled mathematically using a probabilistic approach. Although these methods could create hierarchical structure, these hierarchical structures cannot be efficiently identified.

More recently, several schemes have been designed to identify hierarchical organization of complex systems. Marta Sales-Pardon et al. [4] proposed a universal method that comprises two major steps: (1) estimating the "proximity" in the hierarchy between all pairs of nodes; (2) uncovering the overall hierarchical organization

L. Cao, J. Zhong, and Y. Feng (Eds.): ADMA 2010, Part I, LNCS 6440, pp. 557–564, 2010.

of nodes with an unsupervised algorithm. Ravasz et al. [5] studied the hierarchical structure of metabolic networks, but in their analysis the authors put emphasis on detecting "global signatures" of hierarchical network architecture. More direct methods to investigate the hierarchical organization of the nodes in a network have also been recently proposed [6-8]. Although useful in some contexts, these methods do not clearly identify hierarchical levels and thus fail to identify the different levels in the hierarchy as well as the number of modules and their composition at each level.

In this paper, we propose a simple but effective algorithm to detect hierarchical structure, which is called Iteration Hierarchical Community Detection algorithm (IHCD). IHCD first considers each node as a community, and then repeatedly combines "similar" communities into larger one. With the elaborated design of the community similarity and the combination condition, IHCD can automatically reveal the valuable hierarchical structure. The experiments give excellent results.

2 Preliminary

A network *N* can be modeled as a graph $G=(V,E)$ where *V* is a set of objects, called nodes or vertices, and *E* is a set of links, called edges, that connect two elements of *V*. Network also can be equivalently represented as a connectivity matrix *A*. In the case of unweighted and undirected networks, the adjacent matrix can be used to illustrate the network.

$$A[i,j]=\begin{cases}0 & V(i,j)\in G\\ 1 & V(i,j)\notin G\end{cases}. \tag{1}$$

The set of node i's adjacency nodes can be noted as N_i, and the number of elements in N_i is $|N_i|$. A definite community partition is to find a partition of nodes, i.e., $P=\{\{C_1,C_2,\cdots\}\,|\,C_i\subset V, C_i\bigcap C_j=\varnothing\}$. The number of elements (i.e., vertices) in C_i is $|C_i|$. An ideal community (also call cluster or module) in a network is a group of vertices (i.e. a sub-graph) having a high density of edges within them, and a lower density of edges between groups. A hierarchical structure in a network means that there are several partitions $P_1,P_2,\cdots$. For any two partitions $P_i=\{C_{i1},C_{i2},\cdots\}$, $P_j=\{C_{j1},C_{j2},\cdots\}(i<j)$ if any one community in P_j is constituted by several communities in P_i (i.e., $C_{jk}=C_{ik1}\cup C_{ik2}\cup\cdots$), we can say that there are hierarchical structure in the graph, and $P_1<P_2<\cdots$. That is, the resulting network will have a larger density of connections between nodes grouped in the same sub-module at the first level, a smaller density of connections between groups of nodes grouped in the same module at the second level, and an even. Thus, the network has by construction a hierarchical organization.