Physica A 417 (2015) 46–56

Contents lists available at ScienceDirect

Physica A

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

Detecting overlapping and hierarchical communities in complex network using interaction-based edge clustering

Paul Kim, Sangwook Kim *

School of Computer Science and Engineering, Kyungpook National University, 401 IT-4, 80 Daehakro, Bukgu, 702-701 Daegu, Republic of Korea

HIGHLIGHTS

- Most community detection methods use network topology and edge density.
- These methods decompose nodes connected by high weights into different communities, even when they intuitively belong to a single community.
- We propose a method of detecting overlapping and hierarchical communities in complex networks using interaction-based edge clustering.
- We find that the community quality and the overlap quality for our method surpass the results of the other methods.

ARTICLE INFO

Article history:
Received 6 January 2014
Received in revised form 10 September 2014
Available online 20 September 2014

Keywords:
Community detection
Edge clustering
Overlapping communities
Hierarchical structure
Complex network

ABSTRACT

Most community detection methods use network topology and edge density to identify optimal communities. However, in these methods, several objects that are connected by high weights may be decomposed into different communities, even when they intuitively belong to a single community. In this case, it is more effective to classify the objects into the same community because they perform important roles in controlling and understanding the network. To achieve this goal, in this paper, we propose a method of detecting optimal community structures in a complex network using interaction-based edge clustering. Our approach is to consider network topology as well as interaction density when identifying overlapping and hierarchical communities. Additionally, we measure the differences between the quantity and quality of intra- and inter-community interactions to evaluate the quality of the community structure. We test our method on several benchmark networks with known community structures. Additionally, after applying our method to several real-world complex networks, we evaluate our method through comparison with other methods. We find that the community quality and the overlap quality for our method surpass the results of the other methods.

1. Introduction

Networks that describe complex systems or concepts can be decomposed into communities or groups. Communities are usually subgraphs: the density of edges within the community is greater than the density of edges between communities [1]. The detection of community structures can be easy to understand, and a network can often be analyzed efficiently by dividing it into several groups [2]. Such communities often exist in social networks, biological networks and infrastructure

* Corresponding author. Tel.: +82 53 940 8881.
E-mail address: kimsw@knu.ac.kr (S. Kim).

http://dx.doi.org/10.1016/j.physa.2014.09.035
0378-4371/

networks. There are many methods available to detect communities in complex networks. However, because most methods and algorithms identify communities based on the network topology, the community structure is strongly influenced by the edge density. Specifically, the identification of real communities with different structures and edge densities is difficult. Additionally, although most complex networks are directed and weighted graphs, because each member interacts with other members, most conventional methods do not consider the edge direction and weight simultaneously [3]. Therefore, we propose a community detection method based on the interactions of the members. This method evaluates and maximizes the difference between the quantity and quality of intra- and inter-community interactions to identify optimal communities. We refer to the communities identified by our method as interaction communities (IC).

Interaction communities are a specific type of community structure that maximizes the internal interaction within the community and minimizes the external interaction between communities. This concept differs in several ways from the conventional community definition. The candidate community structures identified using our method are different from the structures identified using modularity or edge density. When the weights of the edges in a network differ, the difference between interaction communities and the ideal community structure is greater. The primary reason for this difference is that we use the interaction density instead of the network topology to determine the community structure. This means that our quality function for evaluating the community structure in the clustering process returns a high value if the weights of the intra-community edges are maximized. However, when the weights of all edges in the network are the same, our method identifies a community structure that is consistent with a structure based on the edge density, such as modularity. Additionally, because we determine the clustering order in terms of edge directions and weights, a structure based on interaction communities is fundamentally different from the results of other clustering methods.

Various techniques, including hierarchical clustering, modularity optimization, detection of dense subgraphs, and statistical inference, among many others, have been used to detect community structures. The Girvan–Newman Algorithm (GN) is a well-known method [4]. The GN algorithm, which is based on divisive hierarchical clustering, probes the community structure by removing high levels of space between edges. The optimal communities identified by this method consist of hierarchical structures selected by means of modularity measurements [5,6]. This method detects non-overlapping communities.

However, an important property of communities in the real world is that a node can belong to several communities [7]. Methods for detecting overlapping communities have been the subject of extensive study. These methods attempt to allow a node to be shared among several groups. The cluster-overlap Newman–Girvan algorithm (CONGA), which has been used to extend the GN algorithm, is a divisive hierarchical algorithm that clusters undirected and unweighted networks [8,9]. In this method, an overlapping node is divided into several nodes, and the overlapping communities are evaluated using Nicosia's modularity [10]. The clique percolation method (CPM) is a popular method of detecting overlapping communities, but it has a non-hierarchical structure [11,12]. CPM combines two communities that share $k - 1$ nodes after identifying the maximal k-cliques in the network. Agglomerative hierarchical clustering based on maximal cliques (EAGLE) can identify both overlapping communities and hierarchical structures [13]. In this algorithm, overlapping communities are evaluated using a quality function that extends Girvan–Newman modularity. A link-community detection method (LC) based on agglomerative hierarchical clustering has been proposed. This method identifies overlapping communities and hierarchical structures by grouping two links that share one node [14].

Edges in complex networks can have directions and weights because the members interact directly with a measurable frequency and duration. Additionally, collaboration or communication events between the same members in a social network can be repeated, and a higher frequency of collaboration or communication usually indicates a closer relationship [12]. In this context, CPMd (CPM with directions) and CPMw (CPM with weights) can account for the directions and weights of edges, respectively, in the detection of overlapping communities [15,16]. CPMdw (CPM with directions and weights) mixes CPMd and CPMw and can detect communities in a weighted and directed network. Likewise, LC can be extended to LCd (LC with directions), LCw (LC with weights) and LCdw (LC with directions and weights). These methods can detect only dense communities while accounting for edge directions and weights.

To find communities in real-world complex networks, it is important to consider several approaches to community detection. These approaches include the detection of overlapping communities and hierarchical structure as well as consideration for directed and weighted edges. However, most methods for community detection do not use these approaches simultaneously. Moreover, most methods decompose nodes connected by high weights into different communities, even when they intuitively belong to a single community. The primary reason for this behavior is that these methods treat the edges connecting nodes as inter-community edges if they are bridges. In this case, it is more effective to classify the objects into the same community because they perform important roles in controlling and understanding the network. Therefore, in this paper, we consider network topology as well as interaction density in determining edge weights for the identification of overlapping and hierarchical communities. To achieve this goal, we propose a method of detecting optimal community structures in a complex network using interaction-based edge clustering. This method is based on single-linkage hierarchical clustering and searches for overlapping community structures in a weighted and directed network. Fig. 1 illustrates the differences between our method and previously proposed methods.

2. Methods

Our objective is to identify optimal communities while minimizing the influence of edge density and maximizing the quantity and quality of internal interactions. Consequently, we propose a method consisting of two processes. First, we