PHYSA: 17428 Model 3Gsc pp. 1–18 (col. fig: NIL)

ARTICLE IN PRESS

Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

Detecting community structure in complex networks using an interaction optimization process

Paul Kim, Sangwook Kim *

School of Computer Science and Engineering, Kyungpook National University, Republic of Korea

HIGHLIGHTS

- Improving the technique for detecting community structures is important for understanding and controlling complex networks.
- Most community detection methods have a high computational complexity and are sensitive to network forms and types.
- We propose an algorithm that uses an interaction optimization process to detect community structures in complex networks.
- We find that the structure quality and coverage resulting from our algorithm surpass the results of other methods.

ARTICLE INFO

Article history:
Received 4 May 2016
Received in revised form 2 August 2016
Available online xxxx

Keywords:
Community detection
Interaction-based community
Interaction optimization process
Network structure analysis
Complex networks

ABSTRACT

Most complex networks contain community structures. Detecting these community structures is important for understanding and controlling the networks. Most community detection methods use network topology and edge density to identify optimal communities; however, these methods have a high computational complexity and are sensitive to network forms and types. To address these problems, in this paper, we propose an algorithm that uses an interaction optimization process to detect community structures in complex networks. This algorithm efficiently searches the candidates of optimal communities by optimizing the interactions of the members within each community based on the concept of greedy optimization. During this process, each candidate is evaluated using an interaction-based community model. This model quickly and accurately measures the difference between the quantity and quality of intra- and inter-community interactions. We test our algorithm on several benchmark networks with known community structures that include diverse communities detected by other methods. Additionally, after applying our algorithm to several real-world complex networks, we compare our algorithm with other methods. We find that the structure quality and coverage results achieved by our algorithm surpass those of the other methods.

1. Introduction

Complex networks that describe real-world systems and concepts are graphs with non-trivial topological features compared with regular or random graphs. These networks are naturally generated from big data using graph theory. By analyzing both the big data and the networks, we can develop new functions, services, and platforms. The analysis results are applicable to many applications. In particular, when we know the structure of complex networks in real time, making

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

E-mail address: kimsw@knu.ac.kr (S. Kim).

http://dx.doi.org/10.1016/j.physa.2016.08.012
0378-4371/

Please cite this article in press as: P. Kim, S. Kim, Detecting community structure in complex networks using an interaction optimization process, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.08.012

accurate and speedy analyses for understanding the networks is possible. Generally, these networks can be decomposed into communities or groups. The communities are typically subgraphs; the density of edges within the community is greater than the density of edges between communities [1,2]. Such communities often exist in social networks, user-product networks, biological networks, and infrastructure networks. In such networks, the community structures can be exploited for a diverse range of applications such as information sharing and diffusion, recommendation, classification, and so forth [3].

Many methods and algorithms that can detect communities in complex networks are already available. However, because real-world complex networks are large-scale graphs composed of millions of nodes and tens of millions of edges, the computational complexity of the available algorithms for detecting communities has become an important factor. Most methods fail to find community structures in real time because they have a non-linear computational complexity. Furthermore, the results of these methods are strongly influenced by the edge density because they identify communities based on the network topology without taking interaction structures into account. 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 [4].

To analyze such networks in real time, an ideal community detection algorithm should be both faster and more accurate. Moreover, the algorithm should be able to analyze the structures of communities using weighted networks as well as directed networks. Therefore, we propose an algorithm to detect community structures using an interaction optimization process. To achieve this, we first define a model of an interaction-based community to measure the quality of the community structures in a network. Then, we propose an interaction optimization process based on a greedy algorithm that efficiently finds the optimal structure of the interaction-based communities among candidate structures.

Various techniques including modularity optimization, detection of dense subgraphs, and statistical inference, among many others, have been used to detect community structures in complex networks. The Girvan–Newman (GN) algorithm [5] is a well-known method for detecting non-overlapping communities. The GN algorithm, which is based on divisive hierarchical clustering, probes the community structure by moving high levels of space between edges. The optimal communities identified by this method consist of non-overlapping community structures selected by means of GN modularity measurements [6,7]. To reduce the computational complexity in large-scale networks, Newman's greedy optimization (NGO) [8] based on modularity maximization, the Clauset–Newman–Moore (CNM) algorithm [9], and the Louvain method [10] have been proposed to decrease the time required to measure GN modularity. In particular, because Louvain is a multilevel aggregation method for optimizing modularity, it has been shown to outperform other non-overlapping community detection methods in terms of computation time. In large-scale networks, however, the Louvain method requires more time in two processes: reconfiguring multilevel graphs and tracking node memberships. Complex network cluster detection (Conclude) [11] is a clustering method that couples the accuracy of global approaches with the scalability of local methods. Conclude generates random, non-backtracking walks of finite length to compute the importance of each edge in keeping the network connected. It is computationally efficient because its cost is nearly linear with respect to the number of edges in the network.

However, an important property of real-world communities is that a member can simultaneously belong to several groups [12]. The cluster-overlap Newman–Girvan algorithm (CONGA), which has been used to extend the GN algorithm, is based on divisive hierarchical clustering. It detects communities only in undirected and unweighted networks [13]. This method searches candidates of overlapping community structures; each community structure is evaluated using Nicosia Q3 modularity [14], which is based on GN modularity. CPM combines one or more communities that share $k-1$ nodes after identifying the maximal k-cliques in the network. In addition, the community overlap propagation algorithm (COPRA) [15] and the speaker–listener label propagation algorithm (SLPA) [16] based on the label propagation algorithm (LPA) [17] have been studied for finding overlapping communities. These methods iteratively propagate the unique label of each node to the node's neighbors without calculating community quality. Infomap [18,19] can detect community structures in both weighted and directed networks using an information-theoretic approach. The method decomposes a network into modules by optimally compressing a description of information flows on the network. Both SLPA and Infomap detect non-overlapping and overlapping communities according to specific parameters.

Overlapping community detection methods can discover node duplications between communities, but because they are highly dependent on network topology and node centrality, they do not guarantee the quality of the overlapping community structure. Finally, the methods are slower than non-overlapping community detection methods because the operations used to manage overlapping nodes have a high computational complexity. In principle, methods based on node iteration are faster than methods based on edge iteration, even when the computational complexity of the methods based on edge iteration is lower, because the number of edges in most complex networks is far larger than the number of nodes. In addition, because the quality of communities detected by conventional methods is determined by community topology, the quality of each community is different. To increase the quality of each community, community topology and interaction structure should be considered at the same time.

To find an optimal community structure, most methods use two processes: searching community structure candidates and calculating the quality of each candidate. Accordingly, to reduce the time required to detect an optimal community structure, these searching and calculating operations should have a linear computational complexity. Moreover, the target of these operations should be nodes rather than edges. To achieve this, we define a model of interaction-based community to efficiently measure the quality of each community structure candidate from its nodes. Using this model, we propose an