Information Sciences 299 (2015) 296–311

Contents lists available at ScienceDirect

Information Sciences

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

An integer programming approach and visual analysis for detecting hierarchical community structures in social networks ☆

Chun-Cheng Lin *, Jia-Rong Kang, Jyun-Yu Chen

Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 300, Taiwan

ARTICLE INFO

Article history:
Received 14 May 2014
Received in revised form 6 November 2014
Accepted 3 December 2014
Available online 16 December 2014

Keywords:
Social network
Community detection
Hierarchical community structure
Integer programming
Visual analysis

ABSTRACT

Detecting community structures in social networks is a very important task in social network analysis as these community structures explain relationships among individuals and can be used to predict social behavior. The relationship among subcommunities in each community can further be identified as hierarchical community structures, in which each super node at each hierarchical level represents a nested structure of communities or nodes. Most previous studies attempting to detect hierarchical community structures focused on new metaheuristics that are computationally efficient but do not guarantee the optimal community partition. As a result, this work applies a novel integer programming (IP) approach to detect hierarchical community structures in social networks. This approach has flexible community capacity limits, does not limit the community numbers at different levels, and maximizes a quality measure for hierarchical community partition. The proposed IP approach can use existing software solvers to detect hierarchical community structures without implementing an algorithm. Visual analysis of experimental results shows that the proposed model with different settings for level numbers can analyze reasonable and sophisticated hierarchical community structures, such that the relationships between communities at different levels can be elucidated clearly.

1. Introduction

A social network can be regarded as a graph in which each individual is represented as a node, and the social relationship between two individuals is represented as a link between the two corresponding nodes [48]. Fig. 1(a) shows an example of nodes and their connections. Recent social network analysis has become increasingly popular, as it helps elucidate the social behavior and backgrounds of individuals [2,41,46]. Some previous studies applied the cluster analysis to social networks, in which the nodes within each cluster are strongly connected with each other (*i.e.*, interactions within the same cluster are strong), and the links between two different clusters are connected weakly [10,20,47]. Extended from cluster analysis, the individuals with similar backgrounds or interests interact frequently and generally gather together to form one or several communities. A structure with multiple communities for the social network is called a *community structure* [6,7,27], *e.g.*,

☆ A preliminary work was presented at the 2013 IEEE Conference on Industrial Engineering and Engineering Management (IEEM 2013), December 10–13, 2013, Bankok, Thailand.

* Corresponding author. Tel.: +886 3 5731758; fax: +886 3 5729101.
E-mail address: cclin321@nctu.edu.tw (C.-C. Lin).

http://dx.doi.org/10.1016/j.ins.2014.12.009
0020-0255/

nodes with the same color that belong to the same community, which is encircled[1] by blue dots (Fig. 1(a)). Hence, differing from previous studies using cluster analysis, the degree of similarity is regarded as the evaluation criterion for detecting/partitioning community structures in social or complex networks [28].

Recent works further investigated the hierarchical relationships in community structures. Each community may have subcommunities based on the strong similarities among individuals, such that the hierarchical relationship forms a nested structure called a *hierarchical community structure* [5,11,35] (Fig. 1(a)), and the hierarchical relationship is represented by a tree-like dendrogram (Fig. 1(b)), in which each community is represented by a node and the hierarchical relationship is represented as a tree. The hierarchical community structure provides information about a community partition and the hierarchical level division for large-scale complex social networks, such that sophisticated social behaviors and interactive relationships among each node can be realized. Hence, this hierarchical community structure has garnered considerable attention and has been applied in both science and engineering.

Most previous works were lacking in three important ways. First, to the best of our knowledge, no mathematical programming methods have been developed to detect hierarchical community structures. Most previous works designed heuristics or metaheuristics to generate approximate solutions for hierarchical community detection in social or complex networks [11,32,33,36,37,45]. Although those approaches are computationally efficient, they cannot guarantee that exact optimal solutions will be obtained. Notably, Li [21] and Xu et al. [43] developed the mathematical programming methods for detecting community structures without any hierarchy.

Second, most previous works (*e.g.*, [32,36,37]) did not apply flexible community capacity limits for different hierarchical levels. In a hierarchical community structure, the community capacity limit markedly affects the quality of detecting the community structure at each hierarchical level. For larger networks, when the community capacity limit is too small, a hierarchical community structure with too many similar communities at the same hierarchical level is typically detected; for small networks, when the community capacity limit is too large, a community structure with different community sizes will be detected. Although some studies (*e.g.*, [43]) used the community capacity limit to avoid excessively large differences between communities, the community capacity limit was fixed for all hierarchical levels.

Third, many works applied an upper bound for the number of communities at each hierarchical level. Generally, the number of communities at different hierarchical levels differs, and, hence, most works (*e.g.*, [11,33,45]) set an upper bound for the number of communities at each hierarchical level according to experience. However, an upper bound that is too large or too small tends to generate unreasonable community structures, adversely affecting the structure's quality.

This work applies a novel integer programming (IP) approach to detect hierarchical community structures in social networks. This approach has flexible community capacity limits for different hierarchical levels and the number of communities at each hierarchical level is not restricted, such that reasonable hierarchical community structures are detected efficiently and effectively according to a predetermined number of hierarchical levels. The hierarchical community structure facilitates observation of the information for the community partition and the hierarchical level divisions. Additionally, the proposed approach overcomes the three main limitations. Finally, visual analysis of detected hierarchical community structures for three real social network instances helps elucidate detection results. Comparison with approaches in previous studies shows that the proposed approach performs better.

The remainder of this paper is organized as follows. Section 2 reviews related studies, including those focused on detecting community structures and hierarchical community structures. Section 3 describes the proposed IP approach. Section 4 analyzes experimental results for two benchmark network instances as well as a real social network instance collected from the Facebook website. Section 5 gives conclusion and directions for future work.

2. Related works

This section discusses the detection of community structures and hierarchical community structures in social networks or complex networks. In addition, the similarity or quality measures that assess community structures are also examined.

2.1. Detection of community structures

Various methods for detecting community structures based on conventional graph clustering methods have been proposed, including the hierarchical clustering algorithm [12,8,25], graph partitioning [1], k-means clustering [23], particle swarm optimization [3], and bee colony optimization [18]. Subsequently, novel methods for detecting community structures were proposed. For example, Xu et al. [43] proposed an IP method that applied a community capacity limit and minimal difference for numbers of nodes in different communities. In [13], a weight was assigned to each node in a network, and then the conventional k-means clustering method was applied to classify nodes into k communities. A nonlinear programming model was established in [21], which also adopted the Lagrangian method to reduce time complexity of computing the model. In [40], a novel algorithm was proposed to detect dynamic communities in social network, such that the effect of noisy data was eliminated, and the real community structure and abnormal events were discovered. In [46], a method with supervised learning mechanism was proposed to incorporate prior information into community structure detection.

[1] For interpretation of color in 'Figs. 1, 4, 6, and 7', the reader is referred to the web version of this article.