Knowledge-Based Systems 26 (2012) 164–173

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Topic oriented community detection through social objects and link analysis in social networks

Zhongying Zhao [a,b,c,*], Shengzhong Feng [b], Qiang Wang [b], Joshua Zhexue Huang [b], Graham J. Williams [b], Jianping Fan [b]

[a] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
[b] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[c] Graduate School of Chinese Academy of Sciences, Beijing 100080, China

ARTICLE INFO

Article history:
Received 23 June 2010
Received in revised form 28 July 2011
Accepted 28 July 2011
Available online 4 August 2011

Keywords:
Social networks
Community detection
Link analysis
Social objects clustering

ABSTRACT

Community detection is an important issue in social network analysis. Most existing methods detect communities through analyzing the linkage of the network. The drawback is that each community identified by those methods can only reflect the strength of connections, but it cannot reflect the semantics such as the interesting topics shared by people. To address this problem, we propose a topic oriented community detection approach which combines both social objects clustering and link analysis. We first use a subspace clustering algorithm to group all the social objects into topics. Then we divide the members that are involved in those social objects into topical clusters, each corresponding to a distinct topic. In order to differentiate the strength of connections, we perform a link analysis on each topical cluster to detect the topical communities. Experiments on real data sets have shown that our approach was able to identify more meaningful communities. The quantitative evaluation indicated that our approach can achieve a better performance when the topics are at least as important as the links to the analysis.

1. Introduction

With social networks becoming popular (such as Flickr, YouTube, LiveJournal, Facebook, Digg, MySpace, DBLP collaboration network, etc.), analyzing such network data has become an increasingly important research issue. Community detection, as a major topic in social network analysis, has received a great deal of attention [1–3]. Discovering inherent community structures can help us understand the networks more deeply and reveal interesting properties shared by the members. People belonging to the same community are more likely to have common hobbies, social functions, occupations, interests on some topics, viewpoints etc. Therefore, the identified communities can be used in collaborative recommendation [4], information spreading [5], knowledge sharing [6], and other applications, which can benefit us greatly.

Existing studies on community detection mainly focus on link analysis or topological structure of the network [7–14]. Communities identified by those works often incorporate different topics since stronger connections represent the interactions that occur across several different topics, which confuses the meanings of the community. Look at the example illustrated in Fig. 1. Fig. 1(a) is a social network consisting 9 nodes and 11 edges. The nodes represent the members involved in the social activities and the edges represent the social relations of interactions or communications. The weight attached to each edge represents the strength of connections between the corresponding members. In addition, we assume the topics of each member have been extracted from social objects through clustering, and the topics are labeled at each node, such as football, music or both. Fig. 1(b) shows the result of discovered communities based on link analysis. We can see that members within a community are connected, but they have different topics of interest. In the left community of Fig. 1(b), there are 4 members interested in 'football' and other 3 members interested in 'music'. This shows that the results from link analysis have ambiguous meanings of communities.

Social objects like emails or blogs often imply the topics that are shared by people. This has motivated research on community discovery through analyzing the contents of the social objects [15–17]. Each community identified by this kind of method often has one common topic. However, such community often contains weakly connected people since it is common that some people are often not connected, especially in distributed environments, i.e., they do not know each other and never communicate. Fig. 1(c) shows the results from clustering of social objects on the network of Fig. 1(a). We can see that members within a community have a

* Corresponding author at: Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.
E-mail addresses: zy.zhao@siat.ac.cn, zy.zhao10@gmail.com (Z. Zhao), sz.feng@siat.ac.cn (S. Feng), qiang.wang1@siat.ac.cn (Q. Wang), zx.huang@siat.ac.cn (J.Z. Huang), Graham.Williams@togaware.com (G.J. Williams), jp.fan@siat.ac.cn (J. Fan).

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.07.017

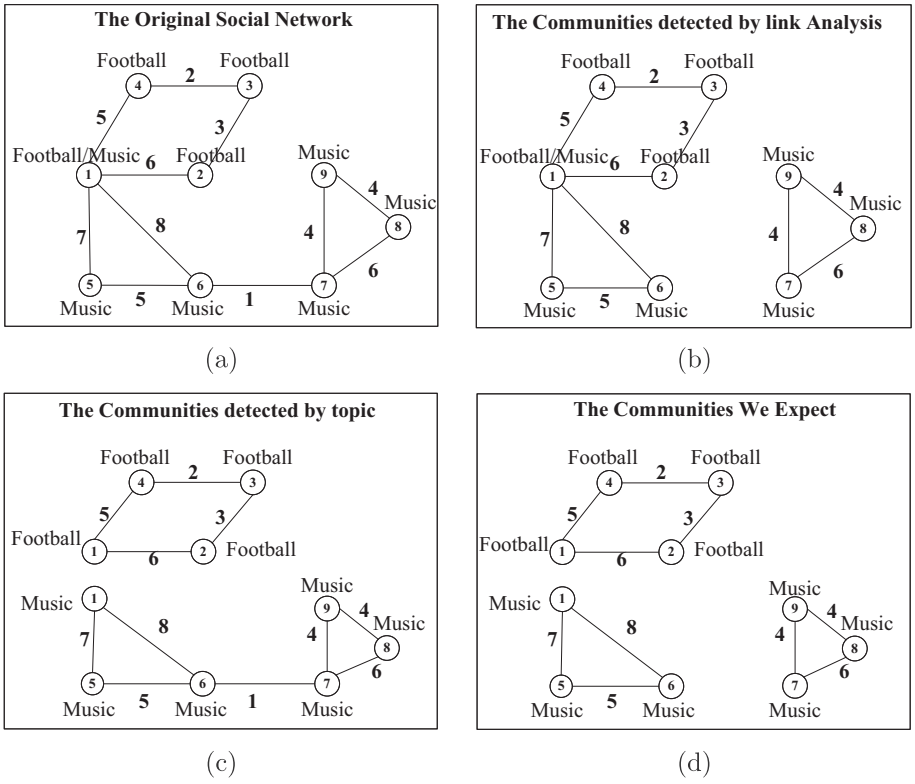

Fig. 1. An example to illustration the motivation of our work.

common topic, but they are not closely connected. In other words, the resulting communities formed on topics only can not reflect the strength of social relations.

To sum up, neither link analysis nor social objects clustering alone is sufficient in determining meaningful communities. Fig. 1(d) shows an ideal result. The members within one community are closely connected, meanwhile they have the same interested topics. This is the result we aim to achieve in this paper.

In this paper, we propose a topic oriented community detection approach which combines social objects clustering and link analysis. Firstly, all social objects are clustered into different topics. Then, the members involved in those social objects are divided into different topical clusters, each corresponding to a certain topic. Finally, the link-based community detection for each topical cluster is performed to differentiate the strength of connections between members.

Compared to the existing work, our approach can identify communities from the perspective of both topics and the link structures. From this result, we can easily find which people are attracted to which community, and by what topic. Such findings can be used to improve the performance of direct marketing and collaborative recommendations.

The rest of the paper is organized as follows. Section 2 reviews the related work. In Section 3, we present the topic oriented community detection approach based on social objects clustering and link analysis. In order to verify our approach, we conducted extensive experiments on real life data sets. The experimental design and results analysis are given in Section 4. Finally, we conclude the paper in Section 5.

2. Related work

A wide range of works have been done to discover communities in a network [7–14]. According to the community detection strategy, the previous works can be classified into optimization based methods and heuristic methods. The optimization based methods include spectral methods and local search based methods. Spectral methods aim to minimize the defined cut-function (e.g., [7]), while local search based methods aim to optimize an objective function such as the function of 'modularity' (e.g., [8,18–23]). The modularity function is used to evaluate the quality of a particular division of a network into communities [24,25]. The heuristic methods often design a graph clustering algorithm based on intuitive assumptions (e.g., MFC algorithm [9], HITS algorithm [10], GN algortihm [11], WH algorithm [12], CPM algorithm [13], FEC algorithm [14]).

Those works have gained success in some applications but they mainly focus on the topological structures or linkage patterns of networks, ignoring the interested topics shared by members. As a result, a community often contains members interested in different topics, which misleads or mixes the meanings of the community.

Another related work is topic modeling through analyzing the contents of social objects. There are several topic models, such as pLSI [26], LDA [27], AT [28], etc. Interactive applications or social network analysis motivate research on topic modeling or topic-based community detection. Zeng et al. [15] proposed a framework for analysis of user activity on an interactive website. User activity analysis tasks, such as user group discovery, can be performed in the framework. McCallum et al. [16] presented the Author-Recipient-Topic model to discover the discussion topics of social networks. Tian et al. [17] proposed OLAP-style aggregation strategies to partition the graph according to attribute similarity, so that nodes within one community share the same attribute values.

The above methods aim to group members interested in common topics into one community. However, they ignore the relationships between members. As shown in Fig. 1(c), the generated communities tend to have very low connectivity.

In fact, both the topological structure and social objects associated with members are important for community detection. The existing approaches mentioned above consider only one aspect but ignore the other. As a result, the identified communities either