Journal of King Saud University – Computer and Information Sciences (2016) **28**, 303–310

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

Topic-oriented community detection of rating-based social networks

Ali Reihanian [a,*], **Behrouz Minaei-Bidgoli** [b], **Hosein Alizadeh** [b]

[a] *Department of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran*
[b] *Department of Computer Engineering, Iran University of Science and Technology, Tehran, Iran*

Received 30 April 2015; revised 12 June 2015; accepted 3 July 2015
Available online 2 November 2015

KEYWORDS

Content analysis;
Topical community;
Community detection;
Modularity;
Purity

Abstract Nowadays, real world social networks contain a vast range of information including shared objects, comments, following information, etc. Finding meaningful communities in this kind of networks is an interesting research area and has attracted the attention of many researchers. The community structure of complex networks reveals both their organization and hidden relations among their constituents. Most of the researches in the field of community detection mainly focus on the topological structure of the network without performing any content analysis. In recent years, a number of researches have proposed approaches which consider both the contents that are interchanged in networks, and the topological structures of the networks in order to find more meaningful communities. In this research, the effect of topic analysis in finding more meaningful communities in social networking sites in which the users express their feelings toward different objects (like movies) by means of rating is demonstrated by performing extensive experiments.
© 2015 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

With the advance of information technology, online communications between people have increased significantly. This kind of communications have become more organized subsequent to the emergence of social networks. For example, folksonomies are social tagging sites which their users collaboratively express their feelings and sentiments toward a special resource like a movie or music by means of descriptive keywords (tags) (Chakraborty et al., 2012) or ratings. Finding meaningful communities in this kind of networks is an interesting research area and has attracted the attention of many researchers. The community structure of complex networks reveals both their organization and hidden relations among their constituents (Lancichinetti and Fortunato, 2012). A community (also sometimes referred to as a module or cluster (Leskovec et al., 2010)) is a dense sub network within a larger network, such as a close-knit group of friends in a social network or a group of interlinked web pages on the World Wide Web (Newman, 2011). As the people in the same community may usually have common hobbies and social functions, the identified communities can be used in

* Corresponding author.
E-mail address: areihanian@ustmb.ac.ir (A. Reihanian).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2015.07.001
1319-1578 © 2015 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

collaborative recommendation, information spreading, knowledge sharing and other applications that are beneficent to us (Zhao et al., 2012).

Most of the researches in the field of community detection mainly focus on the topological structure of a network. They just build a network of individuals without performing any content analysis. Most of these networks are built based on the number of communications between individuals. Actually, these researches just consider the graph structure of a network for finding communities and no content analysis has been used in the process of their proposed approaches.

Despite of the original definition of the networks, nowadays, real world networks contain a vast range of information including shared objects, comments, following information, etc. It is unreasonable for a community to be explained by a single entity because the community members are generally interacting with each other via a large number of distinguishable ways in various domains.

One of the possible solutions is to find topical clusters in which the nodes have the same topic of interest. Each topical cluster represents one of the topics of interest in the network. Then, a community detection algorithm can be applied to these topical clusters to find the ultimate communities (Zhao et al., 2012). In this way, we can analyze and estimate the effect of topic consideration in community detection.

In this paper, the effect of topic analysis in finding more meaningful communities in social networking sites in which the users express their feelings toward different objects (like movies) by means of rating, is demonstrated by performing extensive experiments. Therefore, the network is partitioned into different topical clusters in which the nodes have the same topic of interest. Then, a community detection algorithm is applied to the topical clusters in order to find more meaningful communities. This will lead us to communities in which the nodes are tightly connected and have the same topic of interest. This process is called topic-oriented community detection (Zhao et al., 2012). At last, the results of community detection with topic consideration are compared with the results of community detection without considering the topics of interest. Quantitative evaluations reveal that the results of community detection will be improved when the topic of interest in the network is considered.

The remainder of the paper is outlined as follows. Section 2 explains the motivation of our research. In Section 3, related works are reviewed. Section 4 explains the topic-oriented community detection. In order to evaluate the effect of topic consideration in identifying communities of rating-based social networks, extensive experiments are conducted on real-life data sets. The descriptions of these data sets, the experimental results and their analyses are given in Section 5. Finally, the conclusions are given in Section 6.

2. Motivation

In this section, the motivation of our research is explained with an example. Look at the example illustrated in Fig. 1. Fig. 1(a) is a network of 8 nodes and 11 edges. We call this network a basic network. Each node is an individual in the network, and each edge is the social relation of interactions or communications. The weight of each edge represents the number of communications between the related nodes. For example, if node i finishes five communications with node j, the assigned weight of their related edge will be 5. Consider that the topics of interest for each node are assigned to them manually. These topics represent the domain of interest for each individual in the network. In this specific network, each node can be interested in discussions related to religion, irreligion or both of them.

Fig. 1(b) shows the identified communities after applying a community detection algorithm on the basic network. In this situation, no content analysis has been performed. The members of each identified communities are connected, but as you can see, the community that is located at the top of the Fig. 1(b) incorporates different topics. Two members in this community are interested in religion while three members are interested in irreligion.

Zhao et al. (2012) extracted topical clusters from the basic network in order to detect communities which have a unique topic of interest and connected members. Each topical cluster contains the nodes of the basic network which have the same topic of interest. Fig. 1(c) shows the partition of the network that has two topical clusters. For example, in the topical cluster that is located at the bottom of the figure, all of the members are interested in the discussions related to religion. Then, a community detection algorithm will be applied to each topical cluster. Fig. 1(d) shows the identified communities in this situation. Each community has members who are connected to each other and have the same topic of interest. This is the condition we aim to analyze in the rating-based social networks.

3. Related works

Many researches have been done in the area of community detection. Most of these researches mainly focus on the topological structure or linkage patterns of networks. They merely consider the graph structure of a network for finding communities, while no content analysis is used in the process of their proposed approaches.

According to the community detection strategies which were employed in these researches, their proposed methods can be classified into optimization-based methods and heuristic methods. Some of the optimization-based methods focus on optimizing an objective function (Zhao et al., 2012). One of the most important works in the literature was a research done by Newman and Girvan, in which they introduced modularity as an objective function (Newman and Girvan, 2004). A large amount of works have been done to optimize modularity such as the methods which were developed by Arenas et al. (2007), Leicht and Newman (2008), Newman (2004). This function has been influential in the literature of community detection, and has gained success in many applications. Modularity is used to evaluate the quality of a particular division of a network into communities (Zhao et al., 2012). On the other hand, heuristic methods such as the GN algorithm (Girvan and Newman, 2002) and the CPM algorithm (Palla et al., 2005) design a graph clustering algorithm based on intuitive assumptions (Zhao et al., 2012).

Even though these researches have gained success in some applications, since they mainly focus on the topological structure of the networks, they ignored the contents interchanged between members. As a result, the relationships between the

Download English Version:

https://daneshyari.com/en/article/483982

Download Persian Version:

https://daneshyari.com/article/483982

Daneshyari.com