Physica A 405 (2014) 226–234

Contents lists available at ScienceDirect

Physica A

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

Incorporating profile information in community detection for online social networks☆

W. Fan*, K.H. Yeung

Department of Electronic Engineering, City University of Hong Kong, Hong Kong

HIGHLIGHTS

- We construct additional networks with users' profile information.
- We propose methods to improve values of modularity.
- With our methods, partitions are more correlated with users' characteristic features.

ARTICLE INFO

Article history:
Received 26 March 2013
Received in revised form 24 February 2014
Available online 5 March 2014

Keywords:
Community structure
Online social networks
Profile information

ABSTRACT

Community structure is an important feature in the study of complex networks. It is because nodes of the same community may have similar properties. In this paper we extend two popular community detection methods to partition online social networks. In our extended methods, the profile information of users is used for partitioning. We apply the extended methods in several sample networks of Facebook. Compared with the original methods, the community structures we obtain have higher modularity. Our results indicate that users' profile information is consistent with the community structure of their friendship network to some extent. To the best of our knowledge, this paper is the first to discuss how profile information can be used to improve community detection in online social networks.

1. Introduction

Online social networking service is a popular web-based service which provides a platform for people to create profiles and share these profiles with other people. With this service, people can share their daily life with their friends on Internet. They can also take part in some activities together or join groups online. Some online social networks such as MySpace, Facebook, Google+, and Bebo are popular and have attracted millions of users [1]. Due to their popularity, researchers have studied the topology structure of online social networks, such as MySpace, orkut, and cyworld [2–4]. It is found that these networks are scale-free networks, which have been demonstrated to be vulnerable and the well-connected nodes are crucial in epidemic spreading [5,6].

For better understanding of online social networks, researchers have tried to find more structural features of these networks. Community structure is one of these features. In the past few years, the topic of community detection has attracted many researchers. They have proposed methods to detect communities in networks based on networks' connectivity structures. These detection methods have been applied to real-world networks, such as biological networks [7] and

☆ This work is supported by City University Strategic Research Grant No. 7002734.
* Corresponding author. Tel.: +852 27889869.
E-mail addresses: fanwei.fw@gmail.com (W. Fan), eeayeung@cityu.edu.hk (K.H. Yeung).

http://dx.doi.org/10.1016/j.physa.2014.02.075
0378-4371/

social networks [8]. Traud et al. detect communities in Facebook networks of university students [8]. They find that the online communities of students are related to their offline activities. For instance, some communities depend strongly on students' houses affiliation. In other words, the work concludes that dormitory locations of users have strong impact on the community structure of the users. This is actually the case in online social networks, as will be discussed later in this paper.

In online social networks, people often share a lot of profile information besides their friend lists. The information may have correlation with community structure of these networks. For example, if two users are friends and share common interests, it is quite possible that they are in the same community. So it is possible that the profile information can be utilized to improve the quality of community detection in friendship network. To verify this, we will propose community detection methods for online social network incorporating the profile information. We will take Facebook as an example. It will be shown that profile information of users is useful in community detection.

2. Background

2.1. Profiles of Facebook

In most online social networks, each user has a profile to record his/her basic information. For example, when people create accounts on Facebook, they are required to provide their names, email addresses, birthdays, and gender information. If they like, users may also provide additional information (e.g. their locations, colleges and so on) to help their friends find them easily. All these pieces of information are shown on their profiles. With the information in one's profile, we can obtain some features of this user. As members of the same community may share some common features, it is therefore possible that information on profiles is helpful for community detection.

2.2. Community structure

Communities are defined as special clusters of nodes. The concentration of edges within these clusters is high and that between clusters is relatively low [9]. It has been shown that community structures were found in real-world networks, such as Zachary's network [10], and the collaboration network between scientists as mentioned in Ref. [7]. Researchers have proposed lots of community detection methods. In this paper, our methods are based on Newman's modularity matrix method [11] and the Louvain method [12].

3. Methods on community detection

3.1. Dataset

We design a program to download the data we need automatically from online social networks. First, the friendship network is necessary. So we start from several source users, and get the list of their friends. Then the program will read the friend lists of these friends, and create a new list of users which will be read in the next round. This breadth-first method is like a "snowball". Second, since the information shown in users' profiles will be utilized to assist community detection in our methods, we will record the information of each user in a database. The database stores users' basic information, online activities they attend, number of photos uploaded, favorite pages and so on. "Like" is one of Facebook's most known features. If a user thinks the content on a page is interesting, he/she can click the button "like" on this page. And users can "like" the pages of their colleges, favorite movies, and so on. These pages are also called "favorite pages" in this paper.

In this paper, each user of Facebook is described as a "node" of a network, and the friendship between two users can be defined as an "edge". And the number of friends of a user i is its degree, labeled with k_i. These users and friendship construct a network with n nodes and m edges, which can be presented as an adjacency matrix $\boldsymbol{A}$. A_{ij} is the weight of the edge between nodes i and j. In the friendship network, $A_{ij} = 1$ if there is an edge connecting these two nodes, and $A_{ii} = 0$. So we have $m = \frac{1}{2}\sum_{ij} A_{ij}$ and $k_i = \sum_j A_{ij}$.

3.2. Newman's modularity matrix method

In the area of community detection, the function "modularity", defined by Newman and Girvan [13], is a popular quality function to determine the quality of a partition of a network. Here "partition" is the community assignment of nodes. In a partition, each node can belong to only one community. Modularity is defined by Q =(number of edges within communities) −(expected number of such edges):

$$Q = \frac{1}{2m}\sum_{ij}[A_{ij} - P_{ij}]\delta(c_i, c_j). \tag{1}$$

Here $P_{ij} = \frac{k_i k_j}{2m}$ is the expected number of edges between nodes i and j. c_i is the index of the community which node i is in, and $\delta(u, v) = 1$ if $u = v$ and $\delta(u, v) = 0$ otherwise. The value of modularity is between -1 and 1. And a larger value of modularity means a better performance of a partition.