Knowledge-Based Systems 73 (2015) 227–235

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Enhanced link clustering with observations on ground truth to discover social circles

Yanmei Hu, Bo Yang *

School of Computer Science and Engineering, University of Electronic Science and Technology of China, 611731 Chengdu, China

A R T I C L E I N F O

Article history:
Received 17 January 2014
Received in revised form 18 April 2014
Accepted 5 October 2014
Available online 12 October 2014

Keywords:
Online social network
Social circles identifying
Edge clustering
Ego network
Node profile

A B S T R A C T

Along with the development of internet, our personal online social networks become bigger and more jumbled than before, and it is necessary to provide a good way to organize them. Social networking sites allow users to manually categorize their friends into social circles (e.g., 'list' on Facebook and Twitter), but it is laborious. The problem of *social circles identifying* is thus posed on a user's ego network, while there are currently few efficient as well as effective methods to identify user's social circles. In this paper, we propose a new method, named enhanced link clustering, for social circles identifying on ego networks. The proposed method integrates node profile and network structure by constructing an edge profile for each edge. Utilization of both node profile and network structure information makes the proposed method more effective. Taking edge similarity instead of node similarity to discriminate nodes into different circles allows us to detect overlapping circles. Moreover, we observe that nodes in one circle appear transitive similarity and some nodes are only densely connected, or share common properties. These observations make the process of edge clustering efficient. Experiments on several real datasets demonstrate that our method is much faster, and also more effective compared with maximization likelihood-like method, which has been proved to dominate most methods.

1. Introduction

With the fast development of internet, online social networks play a more and more important role in our daily life. We make friends, interact with world, and even study on online social networks. Huge amount of information is generated on online social networks in each day. If all kinds of the information are flowing to us, we would be lost, just like a little shrimp in sea. Fortunately, online social networks allow us to construct our own social network to filter information. Even so, there are overwhelming volumes of content generated by friends. In order to deal with this, our personal social networks are necessary to be organized. One of the practical mechanisms to organize social network and information generated in social network is to categorize friends into social circles [19]. According to social circles, we can set up distinct activities, for instance, information filtering (e.g. just focusing status updates posted by classmates), privacy protection (e.g. hiding some personal information from unfamiliar friends), etc. Almost all online social networks provide such function as social circles, e.g., 'friend lists' on Facebook, Friends, Renren, and Twitter, 'circles' on Google+.

Circles are usually user-specific since each user organizes his/her own ego network. A user's ego network is constituted by his/her friends and the friendships among these friends, and so is independent of all other users who are not connected to him/her. Therefore, the problem of circles identifying is formulated as node clustering problem on ego network [19]. For visualization, let us see an example. Fig. 1 shows an ego network, central user u is named as ego while other users v_i, friends of user u, are named as *alters*. Then, the task is to identify circles to which each *alter* v_i belongs, such as those labeled circles in Fig. 1.

There are generally two ways to identify circles for users in major online social networks. One is to let users to indicate circles manually; the other is to automatically provide circles but in a naïve way of identifying friend-circle relationships by sharing certain attribute. Neither approach is particularly satisfactory: the former is time-consuming and does not update automatically when new friends add in, while the latter fails to capture individual aspects of users' communities, and may lead to poor results when user profile information is missing or unobservable. It is thus very necessary to find a proper method to solve the problem of circle identifying, i.e., finding a method to help users identify satisfying circles.

* Corresponding author. Tel.: +86 28 61831002.
E-mail addresses: huym260@126.com (Y. Hu), yangbo@uestc.edu.cn (B. Yang).

http://dx.doi.org/10.1016/j.knosys.2014.10.006
0950-7051/

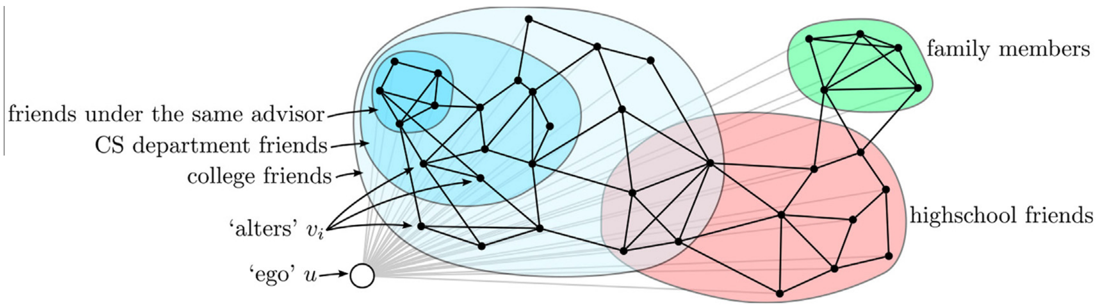

Fig. 1. An ego network with labeled circles coming from [19].

The task of social circles identifying in ego network can fall into the field of cluster/community[1] detection. Similar to cluster, there are generally two types of information which take effect in the identifying process, i.e., node profile and ego network structure. Nodes in one circle are expected to share common properties [23], so it is needed to explicitly model the similarity of node profile in different circles. Meanwhile, circles should be formed by nodes which are densely connected. On the other hand, as observed in [1], circles are overlapped and some circles are hierarchically nested in larger ones, such as shown in Fig. 1. This means that a node may have multiple memberships to circles. Thus it is also necessary to extract information from ego network structure to model this.

In this paper, we propose an enhanced link clustering method, which integrates two types of information: node profile (i.e., the properties of nodes) and network structure, to cope with the problem of automatically identifying users' social circles. Our method is based on link/edge[2] similarity and the knowledge extracted from observations on ego networks. In particular, we establish a model by taking link similarity instead of node similarity and discriminate nodes into different circles by clustering similar links, since link clustering can naturally captures the significant fact that nodes have multiple memberships to circles. In addition, we also transfer node profile to link profile to make links also carry profile information. Moreover, based on observations from ground-truth data, we find that nodes in one circle represent transitionally similarity, but not necessary to be very similar to each other or densely connected. In view of transitional similarity, we extend link clustering algorithm by integrating node profile and network structure to identify circles. Our method has three merits: First, because we treat ego network from the perspective of edges rather than nodes, our method can naturally assign a node to multiply circles. Second, by shifting node profile to edge profile, it is also convenient to integrate node profile and network structure for link clustering. Due to consideration of both types of information, our method can generate better circles aligning to ground-truth. Third, our method is multi-granularity. We allow identifying circles at different scale by establishing different similarity thresholds. It means that our method is hierarchical and can generate a large range of meaningful circles. Finally, we experiment the proposed method on several real datasets provided by Stanford University [19]. For each ego network in these datasets, ground-truth circles are hand-labeled or collected. This provides us a standard to evaluate given models. Experiments show that the proposed method is not only effective, but also more efficient in comparison with state-of-the-art method.

The remained of this paper is organized as following. Section 2 surveys related work. Section 3 presents the established model for social circles. Section 4 illustrates the similarity definition between edges, followed by presentation of the enhanced link clustering algorithm in Section 5. In Section 6, we experiment the proposed enhanced link clustering method. Finally, conclusions and future work are given out in Section 7.

[1] In this paper, we use cluster and community interchangeably.

[2] We use link and edge interchangeably.

2. Related work

In recent years, online social networks have taken a big change to us. In academic realm, not only are the existing research problems rechallenged, e.g., information diffusion [3,14], information maps [34], community detection and evolution [37,38,39,15,20,21,5], recommendation [22,11]), but also new research problems are raised. Friends organization is one of those new research problems. Friends origination is user-specific and information generated even only by friends is rich and diverse, which cause the task of friends organization nontrivial. In most online social networks, user's friends are categorized manually or naïvely by a common attribute, there is no automatic and effective method. In 2012, McAuley and Leskovec firstly studied this problem and formulated it as social circles identifying problem [19]. It can be noted that social circles identifying on ego networks can fall into domain of community detection, although a circle is not exactly the same as a community [20,21]. There are a lot of studies on community detection [16,33,18,29,24] While earlier algorithms are designed under the supposition of disjoint communities [2,9,30,25], many researchers have observed that in real-world networks communities may overlap [28,27,17,10], or have hierarchical structure [31,32,4]. This means that nodes may have multi-memberships to communities. Leskovec et al. further obtained observation on a wide number of large networks that overlaps among communities are more densely connected [37,39], and also proposed a Community-Affiliation Graph Model (AGM) to explicitly model nodes' dense connections in overlap area among communities.

Most methods in the literature attempts to identify communities based on network structure [26,35,8,30,7], or node profile information [12], while rarely use both in a sound way. In order to find communities in social networks that are not completely observable, Yoshida [40] proposes an embedding method by embedding user profile into the well-known modularity function [26]. Similarly, Dang and Viennet [6] combine node properties into modularity function to obtain a composite modularity, and for reducing time complexity, a K-NN graph is constructed from the obtained composite modularity, on which the final communities are detected. Xia and Bu [36] construct a semantic network by extracting semantic information from comment content in Tianya.com. In order to reduce computational complexity, they only focus on "giant component" of the semantic network, and then traditional community detection algorithms are used. McAuley and Leskovec [19] propose a maximization likelihood-like method based on both of node profile and network structure. They treat circle memberships as latent variables and model profile similarity parameters to encode how each circle emerges. Circle memberships and profile similarity parameters are jointly optimized by