Neurocomputing 174 (2016) 605–616

Contents lists available at ScienceDirect

Neurocomputing

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

Mining community and inferring friendship in mobile social networks

Ke Xu [a,b], Keju Zou [c], Yan Huang [a], Xiaoyang Yu [a], Xinfang Zhang [a,*]

[a] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China
[b] College of Computer Science, South-Central University for Nationalities, Wuhan, China
[c] School of Engineering, Sun Yat-Sen University, Guangzhou, China

ARTICLE INFO

Article history:
Received 15 June 2015
Received in revised form
16 September 2015
Accepted 22 September 2015
Available online 1 October 2015

Keywords:
Mobile social network
Location-based sevice
Mining community
Inferring friendship

ABSTRACT

Along with the rapidly growth of mobile terminals and wireless technologies, mobile social networking services are very popular with peoples. Recently many mobile social platforms based on location-based service are developed to allow users to share their check-ins and events with friends. Check-ins data in location-based mobile social networks as well as call detail records (CDR) in mobile communication network may provide insight into community structure, relationships and members in the network. In this paper, we study the problem of community detection and friendship prediction in mobile social networks. We have presented a method to find community structure built on combination entropy, and evaluate modularity of a virtual campus mobile network (V-Net). The outcomes demonstrate that the proposed algorithm mine meaningful communities according to users' registration. We investigate the potential friendship among users by taking into account both users' links with friends and their check-ins at various positions in Gowalla. This work describes the probability distributions of friendships per number of friends, number of check-ins and number of visited places. The findings confirm that our approaches achieve well performance with aggregated features of user similarity and place entropy than other methods. Moreover, members reveal different social properties in the two networks, in the V-Net influence users tend to hold community together, while in Gowalla community members are likely to visit the common positions.

1. Introduction

The growing number of rich sensors and communication enabled devices brings opportunities for perceiving the environment in innovative ways. Nowadays smartphone not only serves as the key communication and computing mobile terminate of choice, but it also equips with an abundant set of embedded sensors including a camera, microphone, accelerometer, gyroscope, digital compass, and GPS. Generally, these sensors promote new applications across a wide variety of fields, such as social networking services [1,2], transportation [3,4], user mobility [5,6], touring routes [7,8], and business sites selection [9,10]. The most popular of those applications is location-based services platform in the mobile social networks.

Location-based social networks (LBSNs) become more popular and excite great interest among people, who are addressing daily opinions on each others web spaces, having a chat, running a blog, writing comments, posting and noticing pictures and videos, preparing activities and last but not least, updating their current status. With ubiquitous mobile devices including smartphones and GPS devices, social networking platforms built on location-based services, such as Facebook Places, Foursquare, Gowalla, and Brightkite, become more and more popular among teenagers. For an instance, as of January, 2014, Foursquare claims to have over 45 million people worldwide and over 5 billions of check-ins with millions more every day. Users of a location-based social networking platform can pick their friends list and obtain listed as friends to other users as well as traditional social network [11–13]. They can check in and share their positions and events to their friends via the social network. Furthermore, users may alert friends to their check-ins when visiting a venue (e.g., museum, hostel, and shopping mall) via their mobile phones [14–16]. Users are encouraged to check in at position to earn badges, venue mayorships and receive special offers.

Community detection and friendship prediction is a key topic in social networking field that provides special approach whereby the study can benefit scientists as well as businesses in a variety of domains. The community structure is an important property which can disclose several unknown characteristics for a given network. Members belong to the same group or community are possible to have common interests or similar properties. Social networks can represent a variety of relationships including

* Corresponding author. Tel.: +86 2787559507.
E-mail address: cole.xu@gmail.com (X. Zhang).

http://dx.doi.org/10.1016/j.neucom.2015.09.070
0925-2312/

friendship, kinship and ties among the participants. Zheleva investigated the predictive power of overlaying friendships and family ties on three interesting social networks in the real world [17]. Users can push and show their outdoor activities including touring and cheering with other users in LBSNs, and they can conveniently manage their footprints and also share them with friends. Ref. [18] studied the correlation between friendship and users' attributes, such as their mobility features, social graph properties, and users' profiles in a commercial LBSNs. A potential friendship links among individuals can be predicted via their common friends, interests and visited places in mobile social network. Social structures tend to be highly dynamic objects because they develop and vary rapidly over time via adding new nodes or edges, so the friendship link prediction problem become more challenging.

In this paper, we have presented a measurement study of mining community detection and inferring friendship on two different datasets generated by users' cellphone in mobile social network. The cellphone communication dataset is a virtual campus mobile network (V-Net) provided by a major mobile communication service corporation in China. Users in the V-Net are students or faculties who are from a college or university, and they could dial with each other more conveniently and cheaply via virtual short phone number. We extract individuals' call detail records (CDR) during 3 months along with their registration information. The second dataset is an online location-based social network, namely Gowalla, which is a location-based service platform launched in 2007 and shut down in 2012. Users are able to check in at venues in their local proximity via a special application on mobile devices.

Our main contributions can be summarized as follows:

(1) We propose an algorithm for finding community and evaluating modularity of the virtual campus mobile network (V-Net). We find that community structure is influenced by important nodes, edge density and internal degree (see Section 3). The number of communities in the network vary the combination entropy threshold (see Section 5).

(2) We investigate potential friend links by analyzing not only users' friends of friends, but also places visited by users' friends in Gowalla (see Section 4). We describe the probability distributions of friendships per number of friends, number of check-ins and number of places of users respectively, those follow log-normal distribution (see Section 5).

(3) We study the friendship prediction by taking into account place properties and user similarity (see Section 4). The average opportunity for a couple of individuals who have visited at the common positions may be friends drops while the entropy of the position increases. There exits strong correlative between friendship probability and user similarity according to the growing trend of the curvy line (see Section 5).

Our findings uncover that users reveal different social activity in the two mobile social network, e.g., in the V-Net some key users tend to promote friendship and arise community, however, in Gowalla community members are likely to visit the common positions and form friendship around the places.

2. Related worked

Social networks such as Facebook, Foursquare, Weibo, and Momo have aroused interest of millions of users, some of these social platforms have grew rapidly more than three quarters in the past year. Recent statistics reports present that social networking services have overcome search engines in terms of application. eMarketer [19] addresses that it expects 4.55 billion people worldwide to use a mobile phone in 2014, and mobile phone penetration will rise from 61.1% to 69.4% of the global population from 2013 to 2017 according to a new eMarketer report, namely, Worldwide Mobile Phone Users: H1 2014 Forecast and Comparative Estimates. This reveals how the mobile phone users have immerged social networks into their daily activities. Many mobile applications have already been developed and are available to users via their mobile phones and wirelessly connected the Internet. Mobile phones or smartphones are extremely fast becoming the principal information platform as well as communication device in human beings lives.

Recently location-based social networks (LBSNs) have provided explicit mechanisms to encourage users to share location information with activities. So users can see where their friends are and what they do, to generate or inquire location-tagged information within their social network, and to meet up with other individuals in close proximity who may have the common favourites. The principal mobile social networks are Foursquare, Brightkite and Gowalla, while other hugely popular social networking platforms such as Facebook, Twitter and Wechat also introduce location-based services. Mobile location-based services are generally flourishing, and provide an extraordinary chance to gather fine grained spatiotemporal information about the spots visited by users.

The study of community structure mainly focus on mining and analysis tasks from sociological data. Understanding the structure and dynamics of social community is a principal target for social network analysis because such subgroups are likely to be combined with larger communities [20,21]. A community develops mainly via the aggressive recruitment of friends by other friends, which would appear as a subgraph branching out quickly over time along connects in the network. Newman addresses a hierarchical agglomeration algorithm to find meaningful communities from Amazon purchasing network, and to mine large-scale patterns from the purchasing behaviors data of consumers [22]. Kimura discusses the combinatorial optimization problem of finding the most influential vertices to hold community together in a large-scale social network [23]. An algorithm ComTector (Community DeTector) based on the nature of overlapping communities is proposed for community detection in real life networks, such as Zachary Karate Club, American College Football, Scientific Collaboration, and Telecommunications Call networks [24]. Lancichinetti does experiment on several methods compare to a recently introduced benchmark graphs with heterogeneous distributions of degree and community size and represent a comparative analysis report [25]. Yang presents an overlapping community detection algorithm BIGCLAM (Cluster Affiliation Model for Big Networks), which scales to large networks of millions of vertices and links. They design a model-based community detection method which can discover densely overlapping, hierarchically nested and nonoverlapping communities in massive networks [26]. Backstrom studies how to evaluate the influence of community social properties including the structure, graph density and evolution trend in the underlying social networks. They discover that both the tendency of members to join community and the propensity of community to develop rapidly depend in delicate ways on the underlying network structure [27].

Friendship link prediction has become an interesting research issue in recent years after the seminal study of Liben-Nowell and Kleinberg [28]. The link prediction issue is also referred to the problem of inferring missing links from an observed network in a variety of fields, which constructs a network of interactions built upon visible data and then tries to infer additional links that are not directly observable but likely to exist. Existing approaches have focused on defining various proximity measurements on network topology, and it also is likely to consider particular attributes of the vertices in the network rather than to evaluate