Information Processing and Management 54 (2018) 475–489

Contents lists available at ScienceDirect

Information Processing and Management

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

The role of location and social strength for friendship prediction in location-based social networks

Jorge C. Valverde-Rebaza[a], Mathieu Roche[c,d], Pascal Poncelet[b], Alneu de Andrade Lopes[*,a]

[a] Department of Computer Science, ICMC, University of São Paulo, C.P. 668, São Carlos, 13560-970, SP, Brazil
[b] LIRMM, University of Montpellier, 860 rue de Saint Priest F-34095, Montpellier, France
[c] CIRAD, TETIS, F-34398 Montpellier, France
[d] TETIS, University of Montpellier, AgroParisTech, Cirad, CNRS, Irstea, Montpellier, France

ARTICLE INFO

Keywords:
Location-based social networks
Link prediction
Friendship recommendation
Human mobility
User behavior

ABSTRACT

Recent advances in data mining and machine learning techniques are focused on exploiting location data. These advances, combined with the increased availability of location-acquisition technology, have encouraged social networking services to offer to their users different ways to share their location information. These social networks, called location-based social networks (LBSNs), have attracted millions of users and the attention of the research community. One fundamental task in the LBSN context is the friendship prediction due to its role in different applications such as recommendation systems. In the literature exists a variety of friendship prediction methods for LBSNs, but most of them give more importance to the location information of users and disregard the strength of relationships existing between these users. The contributions of this article are threefold, we: 1) carried out a comprehensive survey of methods for friendship prediction in LBSNs and proposed a taxonomy to organize the existing methods; 2) put forward a proposal of five new methods addressing gaps identified in our survey while striving to find a balance between optimizing computational resources and improving the predictive power; and 3) used a comprehensive evaluation to quantify the prediction abilities of ten current methods and our five proposals and selected the top-5 friendship prediction methods for LBSNs. We thus present a general panorama of friendship prediction task in the LBSN domain with balanced depth so as to facilitate research and real-world application design regarding this important issue.

1. Introduction

In the real world, many social, biological, and information systems can be naturally described as complex networks in which nodes denote entities (individuals or organizations) and links represent different interactions between these entities. A social network is a complex network in which nodes represent people or other entities in a social context, whilst links represent any type of relationship among them, like friendship, kinship, collaboration or others (Barabási, 2016).

With the growing use of Internet and mobile devices, different web platforms such as Facebook, Twitter and Foursquare implement social network environments aimed at providing different services to facilitate the connection between individuals with similar interests and behaviors. These platforms, also called as online social networks (OSNs), have become part of the daily life of

* Corresponding author.
E-mail address: alneu@icmc.usp.br (A.d.A. Lopes).

https://doi.org/10.1016/j.ipm.2018.02.004
Received 21 August 2017; Received in revised form 14 January 2018; Accepted 14 February 2018

millions of people around the world who constantly maintain and create new social relationships (Yu et al., 2015; Zheng & Zhou, 2011). OSNs providing location-based services for users to check-in a physical place are called location-based social networks (LBSNs) (Cho, Myers, & Leskovec, 2011; Ozdikis, Ouztzn, & Karagoz, 2016; Valverde-Rebaza, Roche, Pocelet, & Lopes, 2016; Zhu, Chang, Luo, & Li, 2014).

One fundamental problem in social network analysis is link prediction, which aims to estimate the likelihood of the existence of a future or missing link between two disconnected nodes based on the observed network information (Liben-Nowell & Kleinberg, 2007; Lü & Zhou, 2011; Martínez, Berzal, & Cubero, 2016; Wu, Zhang, & Ren, 2017). In the case of LBSNs, the link prediction problem should be dealt with by considering the different kinds of links (Bao, Zheng, Wilkie, & Mokbel, 2015; Li et al., 2008; Zheng & Zhou, 2011). Therefore, it is called *friendship prediction* when the objective is to predict social links, i.e. links connecting users, and *location prediction* when the focus is to predict user-location links, i.e. links connecting users with places (Pálovics et al., 2017; Valverde-Rebaza et al., 2016; Wang, Tan, Zhang, & You, 2016).

Since location information is a natural source in LBSNs, several techniques have been proposed to deal with the location prediction problem (Bao et al., 2015; Zheng & Zhou, 2011). However, to the best of our knowledge no studies have analyzed the performance of friendship prediction methods in the LBSN domain.

In this paper, we review existing friendship prediction methods in the LBSN domain. Moreover, we organize the reviewed methods according to the different information sources used to make their predictions. We also analyze the different gaps between these methods and then propose five new friendship prediction methods which more efficiently explore the combination of the different identified information sources. Finally, we perform extensive experiments on well-known LBSNs and analyze the performance of all the friendship prediction methods studied not only in terms of prediction accuracy, but also regarding the quality of the correctly predicted links. Our experimental results highlight the most suitable friendship prediction methods to be used when real-world factors are considered.

The remainder of this paper is organized as follows. Section 2 briefly describes the formal definition of an LBSN. Section 3 formally explains the link prediction problem and how it is dealt with in the LBSN domain. This section also presents a survey of different friendship prediction methods from the literature. Section 4 presents our proposals with a detailed explanation on how they exploit different information sources to improve the friendship prediction accuracy. Section 5 shows experimental results obtained by comparing the efficiency of existing friendship prediction methods against our proposals. Finally, Section 6 closes with a summary of our main contributions and final remarks.

2. Location-based social networks

A location-based social network (LBSN), also referred to as *geographic social network* or *geo-social network*, is formally defined as a specific type of social networking platform in which geographical services complement traditional social networks. This additional information enables new social dynamics, including those derived from visits of users to the same or similar locations, in addition to knowledge of common interests, activities and behaviors inferred from the set of places visited by a person and the location-tagged data generated during these visits (Allamanis, Scellato, & Mascolo, 2012; Bao et al., 2015; Narayanan & Cherukuri, 2016; Valverde-Rebaza et al., 2016; Zheng & Zhou, 2011).

Formally, we represent an LBSN as an undirected network $G(V, E, \mathscr{L}, \Phi)$, where V is the set of users, E is the set of edges representing social links among users, $\mathscr{L}$ is the set of different places visited by all users, and Φ is the set of check-ins representing connections between users and places. This representation reflects the presence of two types of nodes: users and locations, and two kinds of links: user-user (social links) and user-location (check-ins), which is an indicator of the heterogeneity of LBSNs (Mengshoel, Desail, Chen, & Tran, 2013; Zhang, Kong, & Yu, 2014; Zheng & Zhou, 2011).

Multiple links and self-connections are not allowed in the set E of social links. On the other hand, only self-connections are not allowed in the set Φ of check-ins. Since a user can visit the same place more than once, the presence of multiple links connecting users and places is possible if a temporal factor is considered. Therefore, a check-in is defined as a tuple $\theta = (x, t, \ell)$, where $x \in V$, t is the check-in time, and $\ell \in \mathscr{L}$. Clearly, $\theta \in \Phi$ and $|\Phi|$ defines the total number of check-ins made by all users.

3. Link prediction

In this section, we formally describe the link prediction problem and how this mining task is addressed in the LBSN domain. Moreover, we also review a selected number of friendship prediction methods for LBSNs.

3.1. Problem description

Link prediction is a fundamental problem in complex network analysis (Barabási, 2016; Lü & Zhou, 2011), hence in social network analysis (Liu et al., 2016; Shahmohammadi, Khadangi, & Bagheri, 2016; Valverde-Rebaza & Lopes, 2013; Wu et al., 2017). Formally, the link prediction problem aims at predicting the existence of a future or missing link among all possible pairs of nodes that have not established any connection in the current network structure (Liben-Nowell & Kleinberg, 2007).

Consider as *potential link* any pair of disconnected users $x, y \in V$ such that $(x, y) \notin E$. U denotes the universal set containing all potential links between pairs of nodes in V, i.e. $|U| = \frac{|V| \times (|V| - 1)}{2}$ since G is an undirected network. Also consider a *missing link* as any potential link in the set of nonexistent links $U - E$. The fundamental link prediction task here is thus to detect the missing links in the