Computer Communications 73 (2016) 251–262

Contents lists available at ScienceDirect

Computer Communications

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

Socio-spatial affiliation networks

Konstantinos Pelechrinis*, Prashant Krishnamurthy

University of Pittsburgh, Pittsburgh, PA 15260, USA

ARTICLE INFO

Article history:
Available online 11 June 2015

Keywords:
Location-based social networks
Affiliation networks
Friendship inference

ABSTRACT

Location-based social networks (LBSNs) have recently attracted a lot of attention due to the number of novel services they can offer. Prior work on analysis of LBSNs has mainly focused on the social part of these systems. Even though it is important to know how different the structure of the social graph of an LBSN is as compared to the friendship-based social networks (SNs), it raises the interesting question of what kinds of linkages exist between locations and friendships. The main problem we are investigating is to identify such connections between the social and the spatial planes of an LBSN. In particular, in this paper we focus on answering the following general question "What are the bonds between the social and spatial information in an LBSN and what are the metrics that can reveal them?" In order to tackle this problem, we employ the idea of *affiliation networks*. Analyzing a dataset from a specific LBSN (Gowalla), we make two main interesting observations; (i) the social network exhibits *signs of homophily* with regards to the "places/venues" visited by the users, and (ii) the "nature" of the visited venues that are common to users is powerful and informative in revealing the social/spatial linkages. We further show that the "entropy" of a venue can be used to better connect spatial information with the existing social relations. The entropy records the diversity of a venue and requires only location history of users (it does not need temporal history). Finally, we provide a simple application of our findings for predicting existing friendship relations based on users' historic spatial information. We show that even with simple unsupervised or supervised learning models we can achieve significant improvement in prediction when we consider features that capture the "nature" of the venue as compared to the case where only apparent properties of the location history are used (e.g., number of common visits).

1. Introduction

During the last few years, boosted by advancements in mobile handheld devices (e.g., smartphones), a new class of digital social networks, namely location-based social networks (LBSNs), has emerged. It is now possible to bring into the equation of online social networks (OSNs) another dimension, that of ***location***, due to the significantly improved ability of mobile devices to accurately estimate their position or location. The underlying communities not only have social ties (e.g., friendship) and/or interests in common (e.g., sports), but they are also "connected" with regards to their geographic locations (often mapped into "venues" as described later). In other words, LBSNs bond the online and physical social ties through location information.

This bond can enable a number of novel, convenient, and appealing services making LBSNs popular. People can now track their children's locations. By tracking friends, applications such as better coordination for scheduled meetings can be enabled. Applications can also include exploring new places through a list of venues that are within the proximity of the current location. This list can now be accompanied by tips and recommendations from people/friends that have visited these places. Even simply the number of people that have visited a locale in the past or are present at the moment might be helpful and informative. Other systems can also offer Groupon-like deals, providing additional monetary incentives for someone to adopt their usage. A recent study has also shown that "gaming" aspects of LBSNs form an important motivation for people to start using them [15].

With LBSNs becoming prevalent, it becomes critical to comprehend and discriminate the types of knowledge we can obtain from the bond between locations and social ties. For example, what correlations exist between users' spatial trails and their social behaviors as expressed through their friendships and do the spatial trails provide any information about social ties? Our primary objective in this work is to identify the existing correlations and the metrics that can best capture them. Using the knowledge we obtain from our study we further examine *whether we can use these correlations and metrics to infer social information* **only** *from users' locations*. Going forward this can stimulate our ability to deconstruct the interplay between the social and the spatial information plane and apply it to new applications.

Interactions in an LBSN: An LBSN has two distinct components; a social network and a location log for each member. The social part of

* Corresponding author. Tel.: +1 4126249417.
E-mail address: kpele@pitt.edu (K. Pelechrinis).

http://dx.doi.org/10.1016/j.comcom.2015.06.002

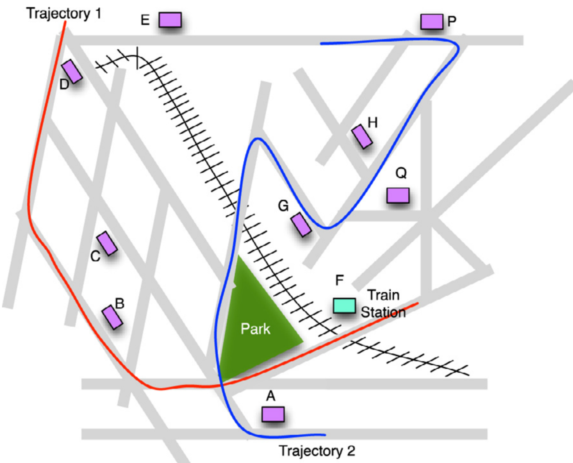

Fig. 1. Trajectory-based LBSN.

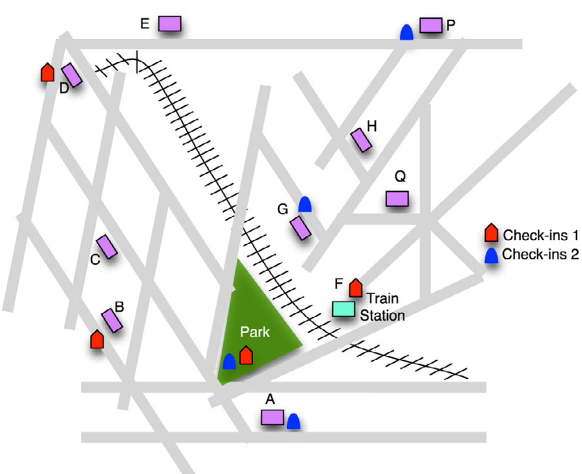

Fig. 2. Check-in based LBSN.

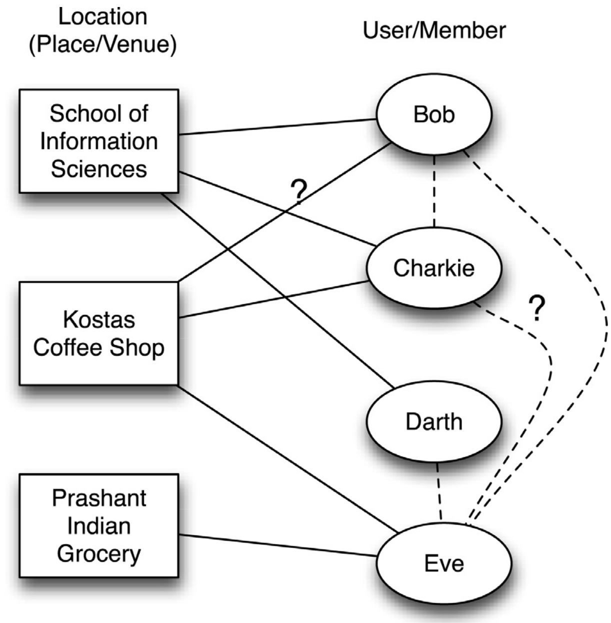

Fig. 3. LBSN as affiliation network.

the system resembles any other existing online social network, where friendships are declared and people can interact with their friends. What differentiates LBSNs from other OSNs are the type of interactions that are feasible between the members of the network. The main feature of this interaction is location sharing. While the "visible" interactions in a traditional OSN are restricted to the virtual world, we can observe interactions within an LBSN in the physical world as well. This is especially important for our study since it can shed light on patterns that are otherwise difficult to identify.

Location sharing can be realized either through continuous tracking, in the form of a temporal latitude/longitude trajectory (e.g., Loopt - see Fig. 1) or via "check-ins", where users announce their presence in a place or venue at their convenience (e.g., Gowalla, Foursquare etc. - see Fig. 2). Clearly, the second approach, where location is tagged with semantic information as compared to a flat geographic trajectory, offers a richer set of information, but with coarse location granularity. All major LBSNs follow this latter approach and consequently, in this work we consider systems in which spatial information is created via check-ins. We note here that using "check-in" history can be challenging since fine grained temporal information is absent (e.g., users do not "check-out" etc.).

Hence, we now have two types of information – the social ties between members and check-ins of members of the LBSN. To analyze socio-spatial interactions within an LBSN, we model it as an "affiliation network", where the members are nodes of one type and venues/places are nodes of the second type (see Fig. 3). Using a dataset from Gowalla [2], we analyze how the *number* and *type* of users' common affiliations (as measured through the number of common locales visited by them) are related to the affinities in the underlying social graph. The main findings of our study can be summarized as follows:

- We identify clear signs of location homophily, that is, members of the LBSN that are friends are more *similar* compared to those that are non-friends. "Similarity" here refers to the **percentage** of visited places that are common between two users (to be formally defined later).
- While simply the number of common places visited by two users does not provide rich social knowledge, the user similarity as well as the "type" of their common venues is a very descriptive feature.

Using the affiliation network model we are able to define the clustering coefficient (cc) of a venue, which can capture the nature of a place in a variety of ways to be elaborated on later. As we will see later, this cc has a strong correlation with the social relations in the graph; exactly what we are looking for! However, its computation utilizes knowledge from the friendship graph, resulting in the problem of circular reasoning. Hence, we examine other metrics, and in particular we show that the entropy of a venue is very informative and helpful for dealing with our problem.

Finally, we investigate the importance of the different features we consider through simple unsupervised and supervised friendship prediction models. In particular, we seek to infer the existing affinity relations using *only* the users' location history. Our evaluations reveal that features that account for the type of a venue, can significantly improve the estimations as compared to features that consider all venues equal.

Scope of our study: We would like to emphasize that our work is a study of the interplay between the social and spatial information present in an LBSN. Even though this connection can enable many new applications, such as location prediction, this study is not focused on any specific one of them. Despite the fact that we examine some simple friendship inference models utilizing our findings, our objective in this study is not to provide a social affinity classifier but to provide insights into the value of the location information present in an LBSN and its ability towards predicting social ties. For instance, the relation between spatial and social data can have significant