Exploring Communities for Effective Location Prediction

Jun Pang
University of Luxembourg
jun.pang@uni.lu

Yang Zhang
University of Luxembourg
yang.zhang@uni.lu

ABSTRACT

Humans are social animals, they interact with different communities to conduct different activities. The literature has shown that human mobility is constrained by their social relations. In this work, we investigate the social impact on a user's mobility from his communities in order to conduct location prediction effectively. Through analysis of a real-life dataset, we demonstrate that (1) a user gets more influences from his communities than from all his friends; (2) his mobility is influenced only by a small subset of his communities; (3) influence from communities depends on social contexts. We further exploit a SVM to predict a user's future location based on his community information. Experimental results show that the model based on communities leads to more effective predictions than the one based on friends.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data mining*

1. INTRODUCTION

Humans are social animals, everyone is a part of the society and receives influences from it. For example, our daily behaviors such as what types of music to listen or where to go for lunch are largely dependent on our social relations. Sociology has shown that we can categorize our social relations (or friends) into communities, using different criteria and considerations. In daily life, humans are engaged in various social environments, and interact with different communities depending on the environments. Therefore, for a specific behavior of a user, in most cases social influence comes from one of *his communities*, but not from *all his friends*. For example, one listens to similar music as his close friends; and he has lunch together with his colleagues on weekdays.

With the large amount of location (check-in) and social relation data from location-based social networks (LBSNs) available, studying human mobility and its connection with social relationships become quantitatively possible. Understanding human mobility can lead to compelling applications including location recommendation, urban planning, etc. In this paper, we aim to study the impact from communities on a user's mobility and effectively predict his future locations based on his community information.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742720.

2. COMMUNITY AND FRIENDS

The LBSN dataset we use is collected by the authors of [1]. We focus on the check-in data from four metropolises in US including New York, LA, Bay Area and Dallas. We apply the widely used algorithm *InfoMap* to partition each user's friends into communities. If a user is engaged in a number of communities, then he is considered an active society member. Therefore, his daily behaviors are largely dependent on his social relations. To quantify a user u's social diversity, we introduce the notion of *community entropy*:

$$coment(u) = \frac{1}{1-\alpha} \log \sum_{c(u)\in C(u)} (\frac{|c(u)|}{|f(u)|})^{\alpha}.$$

The community entropy is defined based on Rényi entropy where $C(u)$ is the set of u's communities, $c(u)$ is one of u's community containing a subset of his friends, $f(u)$ is the set of all u's friends, and α is the order of diversity (set to 10).

For a user u, we exploit single linkage clustering algorithm to group his check-ins into his frequent movement areas $fa(u)$ (a set of his clusters' central points). Then, the impact from his communities on his mobility is defined as

$$im(u)^c = \frac{1}{|fa(u)|} \sum_{l\in fa(u)} \min\{d(l, \{fa(c(u))|\forall c(u) \subseteq C(u)\})\},$$

where d represents the distances between $\{fa(c(u))|\forall c(u) \subseteq C(u)\}$ and l. The community impact is the average of all the shortest distances between a user's frequent movement areas and any movement areas from any communities of the user. Moreover, the communities that are used when computing $im(u)^c$ are named u's *influential communities*. We define a user's friends impact on his mobility in a similar way. A smaller distance indicates more impact on mobility. Therefore, if $im(u)^f > im(u)^c$, then u gets more impact on his mobility from his communities than his friends.

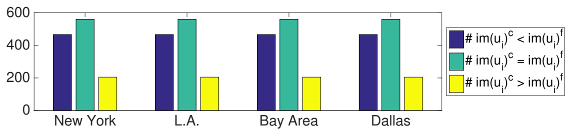


Figure 1: The numbers of users who get more, less or equal impact from their communities than their friends.

As shown in Fig. 1, in general more users get more impact from communities than friends in the four metropolises.

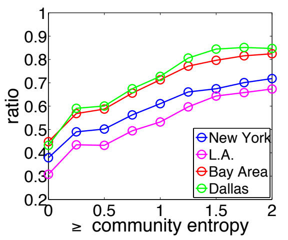


Figure 2: Users having more impact from communities

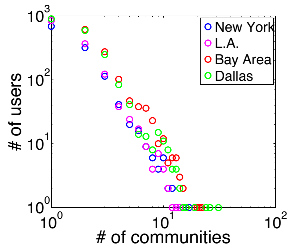
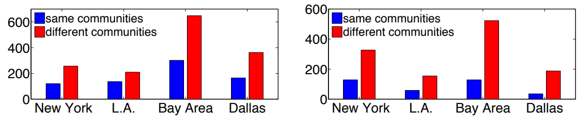


Figure 3: Distribution of # of influential communities



(a) Temporal (b) Spatial

Figure 4: The number of common and distinct influential communities under temporal or spatial contexts.

In New York, Bay Area and Dallas, almost twice users get more impact from communities than friends. Fig. 2 shows that more diverse a user's community (as described by community entropy), more probably he gets more influence from his communities than friends.

3. COMMUNITY AND MOBILITY

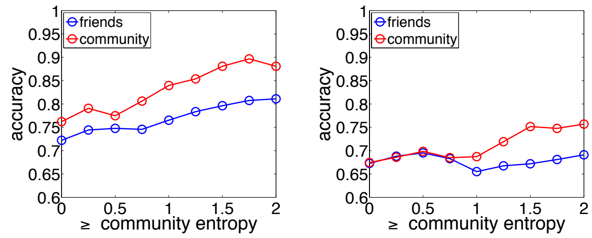
Next, we focus on how many communities can actually influence a user's mobility. We plot the distribution of the number of user's influential communities in Fig. 3. This distribution follows the power law, indicating that most of users are influenced only by a small subset of their communities.

Different communities give influences under different social contexts. For instance, a user has lunch with his colleagues and spends time with his family near his home. To study this, we proceed by considering *temporal* and *spatial* contexts. For temporal context, we choose two time periods including lunch (11am–1pm) and dinner (7pm–9pm) hours on weekdays. We first find users' frequent movement areas during lunch and dinner respectively, and then find the influential communities for users w.r.t. these two time periods. Fig. 4a's result indicates that the influential communities of users during lunch and dinner time are quite different. This simply reflects the fact that the people that users have lunch and dinner with are quite different. For spatial contexts, we pick two disjoint areas in each metropolis. Fig. 4b shows again the number of common and distinct influential communities are quite different, meaning that the influential communities are constrained by spatial contexts as well.

4. LOCATION PREDICTION

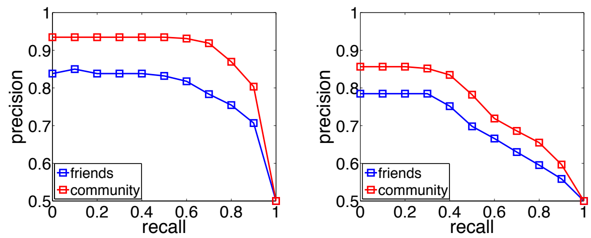
Following the above analysis, we continue to investigate whether it is possible to use community information to effectively predict users' locations, using machine learning classifiers. The question that we want to answer is: *given a user's community information, whether he will check in at a certain place in the future.*

Features. The features we use in the machine learning classifier are from three domains including community, time and location. For the community-related features, we first pick the closest community to the location, and then use six features including the community's distance to the location, the community size, the number of its frequent movement areas, the number of its check-ins and its density. For the time-related features, we use the number of check-ins at that day and hour. For the location-related feature, we adopt location entropy [2] to capture the location's popularity.



(a) New York (b) Bay Area

Figure 5: Prediction accuracy



(a) New York (b) Bay Area

Figure 6: Precision-recall for users with $coment(u) \geq 1$.

Experiments. We split the dataset from [1] into two: one from 2011/03/01 to 2011/09/25 for training and the other from 2011/09/26 to 2011/10/23 testing. We partition the four metropolises into 0.01×0.01 degree latitude and longitude cells, a user is said to be in a cell if he checked in at the cell. We construct a balanced dataset for each metropolis. A support vector machines (SVM) with Gaussian kernel is exploited as our classifier. For the baseline model, we build the same set of features out of a user's all friends.

Result. As shown in Fig. 5, the prediction accuracy is fairly reasonable (around 70%). With the increase of community entropy, the accuracy grows faster for the community-based model which means the predictor works better for users with high community entropies. This validates our observation that users with high social diversities get more impacts from their communities than friends. Fig. 6 summarizes the precision-recall results for users with community entropies ≥ 1. We can conclude that community information can be explored to achieve promising location predictions, especially for those users with high community entropies.

5. CONCLUSION AND FUTURE WORK

We have studied influence from communities on user's mobility (see [3] for details). In the future, we will investigate community impact on other social behaviors.

6. REFERENCES

[1] CHO, E., MYERS, S. A., AND LESKOVEC, J. Friendship and mobility: user movement in location-based social networks. In *KDD 2011.*

[2] CRANSHAW, J., TOCH, E., HONE, J., KITTUR, A., AND SADEH, N. Bridging the gap between physical location and online social networks. In *UbiComp 2010.*

[3] PANG, J., AND ZHANG, Y. Location Prediction: communities speak louder than friends. *CoRR abs/1408.1228, 2014.*