Distance and Friendship: A Distance-based Model for Link Prediction in Social Networks

Yang Zhang[1] and Jun Pang[1,2]

[1] Faculty of Science, Technology and Communication
[2] Interdisciplinary Centre for Security, Reliability and Trust
University of Luxembourg
Luxembourg, Luxembourg
firstname.lastname@uni.lu

Abstract. With the emerging of location-based social networks, study on the relationship between human mobility and social relationships becomes quantitatively achievable. Understanding it correctly could result in appealing applications, such as targeted advertising and friends recommendation. In this paper, we focus on mining users' relationship based on their mobility information. More specifically, we propose to use distance between two users to predict whether they are friends. We first demonstrate that distance is a useful metric to separate friends and strangers. By considering location popularity together with distance, the difference between friends and strangers gets even larger. Next, we show that distance can be used to perform an effective link prediction. In addition, we discover that certain periods of the day are more social than others. In the end, we use a machine learning classifier to further improve the prediction performance. Extensive experiments on a Twitter dataset collected by ourselves show that our model outperforms the state-of-the-art solution by 30%.

1 Introduction

Online social networks have gained a huge success during the past decade and play an important role in our daily life. For example, people publish statuses on Facebook, read news through Twitter and share photos on Instagram. During the past five years, with the development and deployment of mobile devices, such as smart phones and tablets, people begin to use social network services more often on mobile devices. For example, Facebook has 703 million daily active mobile users, and 30% of all the Facebook users only use mobile devices for Facebook services.[3] One interesting and important service related to mobile social network applications is location sharing which can be achieved through a number of localization and positioning techniques, including the GPS satellite navigation system, Wi-Fi-based positioning system, and mobile phone tracking. Nowadays, it is very common to see that users publish photos or statuses that are labeled with the corresponding locations. Moreover, a new type of social network services, namely location-based social networks (LBSNs), has emerged, e.g., Foursquare and Yelp. In LBSNs, users may just share their location to participate in some kinds of social games or get coupons and discounts from restaurants and shops.

[3] http://newsroom.fb.com/

Human movement or mobility have been studied for a long time. With more and more people's location information becoming available through social networks, quantitative study on human mobility becomes achievable (e.g., see [1,2,3,4,5,6,7]). Understanding human mobility can result in appealing applications such as targeted advertising and friends recommendation. In a broader context, it also helps us to tackle the challenges that we are facing at the moment, e.g., urban planning, disease spread, pollution control, etc.

In this paper, we focus on mining social relationship between users based on their mobility information (see related work in Section 5). The main idea of our method is to measure the geographical distance between two users and use this distance to predict whether they are friends.

Contributions. Our contributions in this work are summarized as follows.

- We profile each user's mobility using his frequent movement areas and propose several metrics to quantify the distance between two users' frequent movement areas. Through data analysis, we discover that distance is an effective metric to separate friends and strangers.
- We exploit an important property of location, namely location popularity, to further adjust the distance between users. The adjusted distance achieves a better performance in differentiating friends and strangers.
- We directly exploit distance between two users to predict their friendship, and the prediction performance is promising.
- Moreover, we discover that certain periods of the day are more social than others w.r.t. friendship prediction accuracy. We combine all the distances under these social time periods into a machine learning classifier to further improve the performance of our friendship predictor.

Through extensive experiments on a Twitter dataset collected by ourselves, we have demonstrated that our predictor outperforms the state-of-the-art solution by 30%. Note that the experiment code as well as the dataset are available upon request.

Organization. After the introduction, we present the notations as well as the dataset used in the paper in Section 2. The relationship between distance and user friendship in LBSNs is analyzed in Section 3. We present our friendship predictor as well as its experimental evaluation in Section 4. Related works are summarized in Section 5. Section 6 concludes the paper with future work.

2 Preliminaries

We first introduce notations we use in the paper, and then describe the dataset we collect to conduct the experiments.

Notations. Each user is denoted by u and set U contains all the users. If two users are friends, then there exists an edge between these two users. The social network thus forms a undirected graph $G = (U, E)$, where E contains all the edges between users. We use ℓ to represent a location and it corresponds to a point denoted by (lat, lon). For two locations ℓ and ℓ', $d(\ell, \ell')$ represents their Euclidean distance. A user visiting

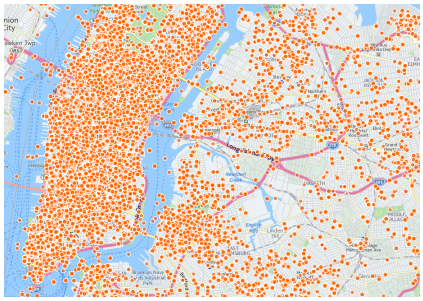

Fig. 1: Check-ins in New York

Starting date	30/11/2014
Ending date	17/02/2015
# of check-ins	2,543,776
# of users	175,566
# of edges	9,490,426
# of active users	4,673
# of edges among active users	37,382

Table 1: Dataset summary

a location, namely check-in, is represented as a tuple $\langle u, t, \ell \rangle$, where t is the time when the check-in happens. Without ambiguity, we use location and check-in interchangeably in the rest of the paper.

Dataset. We exploit Twitter's streaming API[4] to collect geo-tagged tweets in New York city from November 30th, 2014 to February 17th, 2015. When a user shares a geo-tagged tweet at a location, we say that he checks in at that location. In total, we have collected 2,543,776 check-ins from 175,566 users. Each check-in is represented as

$$\langle userID, time, latitude, longitude \rangle.$$

Figure 1 depicts a sample geographical distribution of the check-ins in our dataset. To obtain the social network data, we exploit Twitter's REST API[5] to query each user's followers and followees. Two users are considered friends if they follow each other, this totally gives us a social network with 9,490,426 edges. In the rest of the paper, we only concentrate on users who have at least 50 and no more than 500 check-ins.[6] In the dataset there are totally 4,673 users of this kind, and we refer them as *active users*. The detailed description of the dataset is listed in Table 1.

3 Distance and Friendship

In this section, we first summarize each user's check-ins as his frequent movement areas. Then, based on two users' frequent movement areas, we propose several metrics to quantify the distance between them. In the end, we study the relationship between distance and friendship.

3.1 Frequent movement areas

A user does not visit places randomly in a city. Instead, studies have shown that a user's mobility is centered around several points such as home and office [8]. In Fig-

[4] `https://dev.twitter.com/streaming/overview`

[5] `https://dev.twitter.com/rest/public`

[6] Users with more than 500 check-ins within the time period (2.5 months) are normally public accounts. For instance, `@NewYorkCP` (New York Press) publishes almost 9,000 tweets at the exact same location.

ure 2, the user u mainly visits places around up-east side and Empire state building, while another user u' visits Midtown and Brooklyn often. To measure the geographical distance between two users based on their check-ins, we first need to summarize the places each of them has been, then define the distance based on this summarization. Clustering is the natural solution for this task. In this work, we exploit the centroid-based hierarchical clustering to profile each user's mobility. The cut-off distance of the clustering is set to 500 meters which is a meaningful human movement range. In Section 4, we will further study the sensitivity of this linkage distance. All the places that a user has visited are then grouped into several clusters, we use the central point of a cluster to represent this cluster. Each central point is named as a *frequent movement area* of the user. A user u's mobility is then summarized by all his frequent movement areas denoted by $m(u)$. The frequent movement areas of the two users u and u' are depicted in Figure 2 as well.

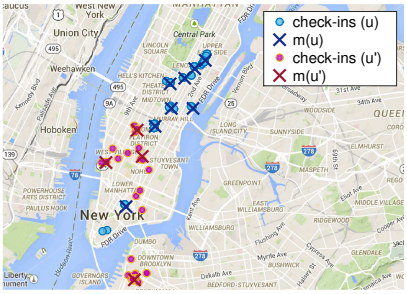

Fig. 2: Frequent movement areas of two users

3.2 Distance between users

After profiling each user's check-ins, the distance between two users u and u' can be measured by the distance between their frequent movement areas, i.e., between the two sets $m(u)$ and $m(u')$. Our goal is to measure the distance between these two sets. There are several metrics proposed to quantify the distance between sets. In this work we adopt the following three ones.

Minimal distance. The minimal distance between u and u' is defined as

$$mind(u, u') = \min(pd(u, u'))$$

where min gives the smallest number in a set of values, and $pd(u, u')$ is the *pairwise distance* between the frequent movement areas of u and u' formally defined as

$$pd(u, u') = \{d(\ell, \ell') \mid \forall\ (\ell, \ell') \in m(u) \times m(u')\}.$$

Here, the minimal distance is simply the distance between two frequent movement areas, one from each user, that are closest to each other.

Average distance. Besides minimal distance, we also exploit the average distance between u and u' as another metric, it is defined as

$$avgd(u, u') = \frac{\sum_{d(\ell,\ell') \in pd(u,u')} d(\ell, \ell')}{\mid pd(u, u') \mid}.$$

Hausdorff distance. Hausdorff distance is another classic metric for quantifying the distance between two sets. Here, we define the Hausdorff distance between u and u' as

$$hausd(u, u') = \max\{\sup_{\ell \in m(u)} \inf_{\ell' \in m(u')} d(\ell, \ell'), \sup_{\ell' \in m(u')} \inf_{\ell \in m(u)} d(\ell, \ell')\},$$

where sup represents the supremum and inf the infimum.

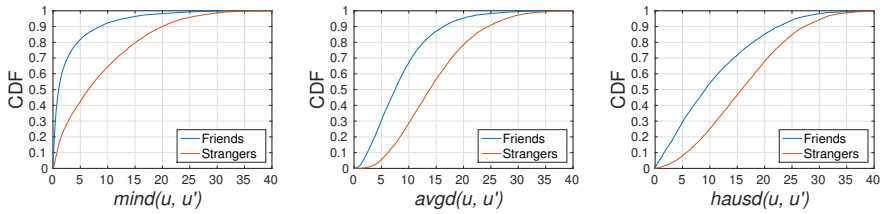

Fig. 3: CDF of distances between friends and strangers

3.3 Distance and friendship

Previous works [9,8,10] have shown that the distance between two users' home locations is related to their relationship. Concretely, they have discovered that friends tend to live closer than strangers. Generally speaking, we would like to check if friends' frequent movement areas are closer than the ones of strangers.

Figure 3 depicts the cumulative distribution functions (CDF) of distances between friends as well as strangers under the three distance metrics. As we can see, the friends and strangers are easily separable by all the three metrics. In particular, more than 80% of friends' minimal distances are less than 5km while only 40% of strangers' minimal distances have the same value. This indicates that the minimal distance between two users is an effective metric to separate friends and strangers. The same happens with the average distance. On the other hand, the difference driven by the Hausdorff distance is relatively smaller compared with the other two metrics. We further study the relationship of social strength (quantified by embeddedness) and the minimal distance between users. As shown in Figure 4, with the increase of $mind(u, u')$, the embeddedness between them drops. This indicates the minimal distance betwen users is correlated with their social strength as well.

3.4 Location popularity

The popularity of locations can potentially affect the distance between friends and strangers. For two users who both have frequent movement areas near a popular place, such as a metro station, the chance that they know each other is low since they may just happen to take the metro everyday. On the other hand, as we show in the previous section the short distance between these two frequent movement areas is a strong indicator that these two users are friends. Therefore, to find a meaningful distance metric to separate friends and strangers, it is necessary to take location popularity into account.

In [11], the authors propose a metric named *location entropy* to quantify a location's popularity. It is defined as follows

$$locent(\ell) = -\sum \frac{|ci(\ell, u)|}{|ci(\ell)|} \log \frac{|ci(\ell, u)|}{|ci(\ell)|},$$

where $ci(\ell)$ represents all the check-ins at location ℓ and $ci(\ell, u)$ contains the check-ins of u at ℓ. Popular places have higher location entropies than unpopular ones. Figure 5

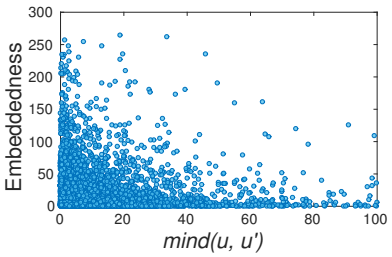

Fig. 4: Social strength as a function of minimal distance

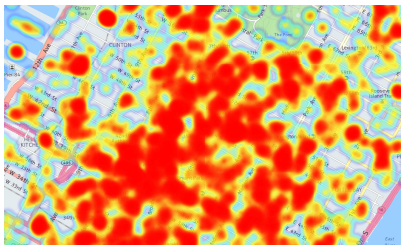

Fig. 5: Heat map of Midtown w.r.t. location entropy

depicts a heatmap of Midtown w.r.t. location entropy. Since we focus on each user's frequent movement areas, we further define the location entropy of a frequent movement area as the average location entropy of all locations in that cluster represented by the frequent movement area. Similarly, more popular a frequent movement area is, its location entropy is also getting higher. Moreover, instead of considering location entropy directly, we use *location diversity* defined as

$$locdiv(\ell) = \exp(locent(\ell))$$

to represent a location's popularity.

After having the distance between two frequent movement areas, we multiply the distance by the maximal location diversity of these two frequent movement areas. The adjusted pairwise distance between u and u', namely $ldpd(u, u')$ is

$$ldpd(u, u') = \{d(\ell, \ell') \cdot \max(locdiv(\ell), locdiv(\ell')) \mid \forall(\ell, \ell') \in m(u) \times m(u')\}.$$

With this adjustment, the distance measure between popular frequent movement areas is increased, while the long ones between unpopular places are reduced. Based on this new pairwise distance, we redefine the minimal and average distances between users accordingly (denoted by $ldmind(u, u')$ and $ldavgd(u, u')$). In addition, the adjusted Hausdorff distance w.r.t. location diversity is defined as

$$ldhausd(u, u') = \max\{ld\_di\_hausd(u, u'), ld\_di\_hau\_d(u', u)\},$$

where $ld\_di\_hausd(u, u')$ is defined as

$$\sup_{\ell \in m(u)} \inf_{\ell' \in m(u')} (d(\ell, \ell') \cdot \max(locdiv(\ell), locdiv(\ell'))).$$

Figure 6 depicts the CDF of the adjusted distances between friends and strangers. As we can see, $ldmind(u, u')$ can better differentiate friends and strangers compared with the other two metrics. Almost 70% of friends' $ldmind(u, u')$ are less than 10 while the value is less than 20% for strangers. Moreover, we notice that this difference is even larger than $mind(u, u')$ depicted in Figure 3. In Section 4, we will demonstrate the effectiveness of this adjusted minimal distance on friendship prediction. On the other hand, the separations driven by $ldavgd(u, u')$ and $ldhausd(u, u')$ get less clear when compared with $avgd(u, u')$ and $hausd(u, u')$.

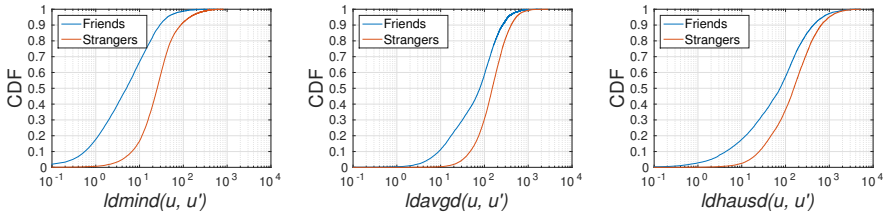

Fig. 6: CDF of adjusted distances between friends and strangers

4 Link Prediction

Link prediction plays an essential part in the context of social networks. For example, recommending friends to a newly joined user is crucial for attracting the user to stay with the social network service. In Section 3, we have shown that friendship is related to distance, i.e., friends' frequent movement areas are closer than strangers. In this section, we aim to predict friendship between users based on their distances.

4.1 Experiment setup and metrics

All our experiments are conducted on a machine with 2.6GHz Intel Core i7 processor and 8Gb memory. We extract all the friends pairs from the active users in New York and randomly sample the same number of stranger pairs to construct a balanced dataset.

We exploit ROC curve and three other standard metrics including *AUC* (area under the ROC curve), *Accuracy* and *F1score* to evaluate our friendship predictor. Let *TP*, *FP*, *TN* and *FN* denote true positive, false positive, true negative and false negative, respectively. Accuracy and F1score are defined as the following:

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |FP| + |FN| + |TN|};$$

$$F1score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \text{ with}$$

$$Precision = \frac{|TP|}{|TP| + |FP|}, \; Recall = \frac{|TP|}{|TP| + |FN|}.$$

4.2 Prediction based on distances

In Section 3, we have proposed 6 different metrics to quantify the distances between two users including $mind(u, u')$, $avgd(u, u')$, $hausd(u, u')$, $ldmind(u, u')$, $ldavgd(u, u')$ and $ldhausd(u, u')$. We start by directly using these distances to predict whether u and u' are friends or not. More precisely, we tune a threshold τ and predict pairs of users whose distances are less than τ to be friends. Figure 7 shows the AUC value of all the different distances for predicting friendships. Among all the six distances, $ldmind(u, u')$ achieves the best performance (AUC = 0.81) followed by $mind(u, u')$. On the other hand, $ldhausd(u, u')$ has the worst performance, with AUC equals to 0.65.

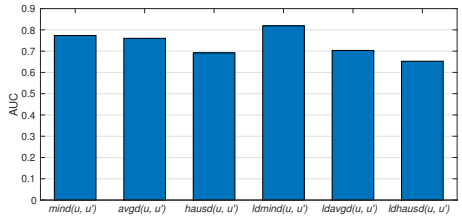

Fig. 7: AUC under different distances

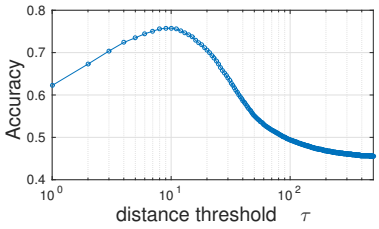

Fig. 8: Accuracy vs. threshold τ

This result is consistent with the friends and strangers separation results in Figure 3 and Figure 6, where $ldmind(u, u')$ achieves the best result on differentiating friends and strangers while $ldhausd(u, u')$ performs the worst. We also notice that adding location diversity to distances only improves the perform of the minimal distance. Meanwhile, $ldavgd(u, u')$'s performance gets even worse than $avgd(u, u')$. This indicates that the adjusted average distance cannot capture the distance between users very well. The same happens with the adjusted Hausdorff distance.

Since the results of all the distance-based link predictors are obtained by tuning the threshold τ, we proceed by studying which is the optimal threshold for predicting friendships. Here, we focus on the best performance distance, i.e., $ldmind(u, u')$ and find τ w.r.t. prediction accuracy. As shown in Figure 8, when the threshold τ falls into the range [8, 11], the accuracy achieves the highest value (0.75). In the following experiments, we set τ to 10 when computing accuracy and F1score for $ldmind(u, u')$.

4.3 Parameter sensitivity

In our whole settings, there is only one parameter to adjust, which is the cut-off distance used in the hierarchical clustering algorithm for finding each user's frequent movement areas (see Section 3.1). For the above evaluation as well as the evaluation in Section 3, we have set it to 500m. Next, we focus on the sensitivity of this cut-off distance.

We have performed friendship predictions through $ldmind(u, u')$ on multiple cut-off distances. As we can see from Figure 9, our predictor achieves the best performance when the cut-off distance falls between [400m, 700m]. Moreover, we also notice that F1score drops when the distance becomes longer (such as 5km and 10km). This is expected considering the human movement range. For example, if the cut-off distance is set to 10km, then a user probably will only have one frequent movement area that covers the whole city. This is too coarse-grained and cannot properly reflect the user's mobility. Based on this analysis, we set the cut-off distance as 500m in our experiments.

4.4 Time and friendship

So far, our distance-based friendship predictor only considers mobility information from users (i.e., frequent movement areas) and locations (i.e., location popularity). On the other hand, time also plays an essential role on users' mobility. For example, a user may check in at places close his office during the working hours while visiting bars and cinemas at his spare time. Intuitively, if two users are close at a social time, the

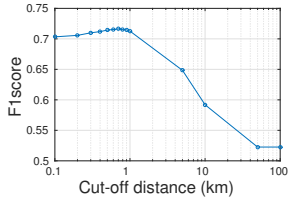

Fig. 9: Parameter vs. F1score

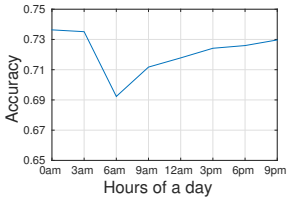

Fig. 10: Hour vs. Accuracy

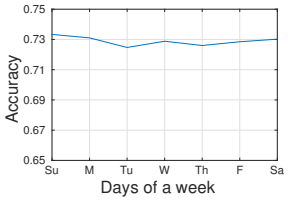

Fig. 11: Day vs. Accuracy

probability for them being friends is high. For example, if u and u' are close (appearing at a same metro station) at 8am in the morning, then we are less confident that they are friends, compared with the case when their frequent movement areas are close during the midnight. Next, we verify whether this hypothesis holds in general.

We consider time on the daily scale as well as the weekly scale, respectively. For the daily scale, we divide a day by eight with each part being a three hour range staring from 0am. We extract each user's check-ins under different time periods and construct the corresponding frequent movement areas. The social level of each time periods is evaluated by the accuracy on predicting friendship through $ldmind(u, u')$. For the weekly scale, we consider users' frequent movement areas on each day (from Sunday to Saturday) and perform an evaluation similar to the daily scale.

As shown in Figure 10, on the daily scale, 0-3am is the most social time of the day followed by 3-6am and 9-12pm. This indicates that if two users' frequent movement areas are close at these hours, the chance that they are friends is high compared with other hours. There is a sudden drop when the period is from 6am to 9am, this means that 6am to 9am is the least social hours of the day, most probably because 6-9am is the commuting hours of the day. On the other hand, the prediction accuracy stays stable on a weekly scale with Sunday and Saturday slightly higher than the weekdays (Figure 11).

4.5 Combining features with machine learning

We have shown that different distance measures between users together with location and time information can capture different aspects of friendship. As a consequence, it is natural to ask whether combining all the information together can further improve the results of link prediction.

Feature description. For two users u and u', we adopt $mind(u, u')$, $avgd(u, u')$ and $hausd(u, u')$ as three features for the classifier. Besides, we take the maximal value and standard deviation of the pairwise distance, i.e., $pd(u, u')$, as another two features as well. We also consider all the above distances' adjusted versions, i.e., taking into account location popularity. This totally gives us 10 features. As some hours are more social than the others on the daily scale, besides the general distance between two users (10 features), we further take their distances at 0-3am and 3-6am into account. Each time period provides 10 features. In total, we collect 30 features for each pair of users.

Learning techniques. Regarding the machine learning classifier, we have tried logistic regression, support vector machine, gradient boosting machine and random forest. It

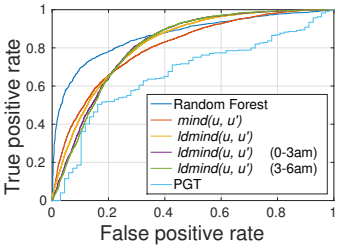

Fig. 12: ROC of Random Forest

	AUC	Accuracy	F1score
Random Forest	0.87	0.79	0.78
$ldmind(u, u')$	0.82	0.75	0.71
$mind(u, u')$	0.77	0.61	0.68
$ldmind(u, u')$ (0-3am)	0.82	0.73	0.65
$ldmind(u, u')$ (3-6am)	0.81	0.74	0.68
PGT	0.66	0.62	0.63

Table 2: Performance comparison with PGT

turns out that random forest outperforms the other algorithms. Thus we adopt random forest as our classifier. We put 70% of user pairs into our training set and the rest 30% are used for testing. In all sets, we perform 10-fold cross validation.

Comparison with the state-of-the-art model. For baseline models, except for some of the distance-based predictors presented before, we also choose a state-of-the-art friendship predictor that exploits the location information, namely the PGT model, proposed in [12]. For each pair of users, PGT first extracts their meeting events. Then, the model considers personal, global and temporal factors based on the meeting events to discover friendship. For the personal factor, PGT proposes a density-based function for each user. If the place of a meeting event is less visited by the user, PGT will value more on this event. For the global factor, PGT adopts the location entropy to adjust the meeting location popularity. In the end, PGT introduces the temporal factor to penalize the meeting event that happens closely with the following events. According to the experiment results in [12], PGT outperforms other location information based link prediction models including [13,11,14]. In our experiments, we follow the parameter settings as specified in [12] for the PGT model including the distance and time for discovering a meeting event and two parameters used in the personal and temporal factors.

Results. The experiment results including the ROC curve and three evaluation metrics are listed in Figure 12 and Table 2. As we can see, our random forest achieves a good prediction result and it outperforms all the other distance-based predictors. Besides, all our models outperform the PGT model. In particular, the machine learning classifier achieves a 30% improvement. We believe this is because PGT only focuses on meeting events between two users, which is too strict since friends can hang out at the same place but they do not have to check in together. In a broader view, according to the homophily theory [15], friends have similar interests. In the context of mobility, this means that friends tend to visit similar places such as same kinds of shops or bars, and this does not have to happen at the same time. Our model considers two users' mobility at a macro level (through their frequent movement areas) which naturally captures the concept of homophily, while PGT and other solutions such as EBM [14] fail to do so.

5 Related Work

With the development of location-based social networks, research on analyzing the social relationship and mobility has attracted a lot of attention. The research can be

roughly partitioned into two groups. One is to use friendship to understand mobility, such as location prediction [8,16] and recommendation [17]; the other is to use mobility information to infer friendship. Our work belongs to the second group.

Authors of [13] propose a probabilistic model to infer friendships from location data shared on Flickr. Their model considers both temporal and spatial information. However, they make a strong assumption that each user only has one friend which is not the case in the real-life applications. Cranshaw et al. [11] propose to use a machine learning classifier to infer the friendship between two users. The features they consider include the ones related to locations as well as the social network structure. Besides, they also propose location entropy to characterize the popularity of a location. The effectiveness of location entropy has been demonstrated in [14,12], we use it to measure the distance between users in this work as well. In [14], the authors propose an entropy-based model, namely EBM. The model first extracts a vector of meeting events between two users, then EBM builds two components based on these meeting events. The first component of EBM is named diversity which is a Rényi entropy formalization on the meeting events vector. More locations two users visit together, better chance they will be friends. The second component is the weighted frequency which exploits location entropy to penalize meeting events at popular locations. Then diversity and weighted frequency are fitted by a linear regression to two users' social strength quantified by Katz score. With the learned model, by tuning a threshold on Katz score, friendship prediction is achieved. More recently, Wang et al. [12] propose the PGT model that we use in this work as the baseline model for friendship prediction.

Advantages of our model. Besides prediction performance, our solution has the following advantages compared with the two recently developed models, i.e., EBM and PGT. First, our model is intuitive and easy to implement. There is only one parameter to adjust in our solution which is the cut-off distance used in the clustering algorithm. On the other hand, both EBM and PGT have several parameters to adjust. Second, both EBM and PGT compute their model components based on the locations two users have visited together, i.e., their meeting events. However, meeting events largely depend on the parameter settings such as the time range and location distance. In PGT, the authors consider two check-ins to be a meeting event if their locations' distance is less than 30m and they happen within one hour. How to set these parameters to find a meeting event is rather unclear. Moreover, as mentioned in Section 4, meeting events cannot capture important social behaviors such as homophily. On the other hand, our model studies two users' distance at a macro level which can naturally capture the concept of homophily.

6 Conclusion

In this work, we have proposed to exploit the distance between two users to predict their friendship. We further integrated location popularity and time information into the prediction. With experiments, we have demonstrated the effectiveness of our approach.

There are several directions we would like to pursue in the future. First, we plan to integrate location and time semantics into the process of summarizing users' frequent movement areas. Based on the new summarized frequent movement areas, we aim to further analyze the relation between friendship and distance. Second, our experiments

only focus on the check-in data in New York, we also want to check whether our discoveries in this paper generally hold in other cities or countries.

References

1. González, M., Hidalgo, C., Barabási, A.L.: Understanding individual human mobility patterns. Nature **453** (2008) 779–782
2. Song, C., Koren, T., Wang, P., Barabási, A.L.: Modelling the scaling properties of human mobility. Nature Physics **6**(10) (2010) 818–823
3. Simini, F., González, M., Maritan, A., Barabási, A.L.: A universal model for mobility and migration patterns. Nature **484** (2012) 96–100
4. Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., Aberer, K.: Semantic trajectories: Mobility data computation and annotation. ACM Transactions on Intelligent Systems and Technology **4**(3) (2013) 49
5. Ying, J.J.C., Lee, W.C., Tseng, V.S.: Mining geographic-temporal-semantic patterns in trajectories for location prediction. ACM Transactions on Intelligent Systems and Technology **5**(1) (2013) 2
6. Chen, X., Pang, J., Xue, R.: Constructing and comparing user mobility profiles for location-based services. In: Proc. 28th ACM Symposium on Applied Computing (SAC), ACM (2013) 261–266
7. Chen, X., Pang, J., Xue, R.: Constructing and comparing user mobility profiles. ACM Transactions on the Web **8**(4) (2014) 21
8. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: Proc. 17th ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM (2011) 1082–1090
9. Backstrom, L., Sun, E., Marlow, C.: Find me if you can: Improving geographical prediction with social and spatial proximity. In: Proc. 19th International Conference on World Wide Web (WWW), ACM (2010) 61–70
10. McGee, J., Caverlee, J., Cheng, Z.: Location prediction in social media based on tie strength. In: Proc. 22nd ACM International Conference on Information & Knowledge Management (CIKM), ACM (2013) 459–468
11. Cranshaw, J., Toch, E., Hone, J., Kittur, A., Sadeh, N.: Bridging the gap between physical location and online social networks. In: Proc. 12th ACM International Conference on Ubiquitous Computing (UbiComp), ACM (2010) 119–128
12. Wang, H., Li, Z., Lee, W.C.: PGT: Measuring mobility relationship using personal, global and temporal factors. In: Proc. 14th IEEE International Conference on Data Mining (ICDM), IEEE (2014) 570–579
13. Crandalla, D.J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., Kleinberg, J.: Inferring social ties from geographic coincidences. Proceedings of the National Academy of Sciences **107**(52) (2010) 22436–22441
14. Pham, H., Shahabi, C., Liu, Y.: EBM: an entropy-based model to infer social strength from spatiotemporal data. In: Proc. 2013 ACM International Conference on Management of Data (SIGMOD), ACM (2013) 265–276
15. Tang, J., Chang, Y., Liu, H.: Mining social media with social theories: a survey. ACM SIGKDD Explorations Newsletter **15**(2) (2014) 20–29
16. Pang, J., Zhang, Y.: Exploring communities for effective location prediction. In: Proc. 24th World Wide Web Conference (Companion Volume) (WWW), ACM (2015) 87–88
17. Gao, H., Tang, J., Hu, X., Liu, H.: Content-aware point of interest recommendation on location-based social networks. In: Proc. 29th AAAI Conference on Artificial Intelligence (AAAI), The AAAI Press (2015)