Exploring the Significance of Human Mobility Patterns in Social Link Prediction

Basma Alharbi
King Abdullah University of Science and Technology (KAUST)
P.O. Box 1639
Thuwal 23955-6900, KSA
basma.harbi@kaust.edu.sa

Xiangliang Zhang
King Abdullah University of Science and Technology (KAUST)
P.O. Box 2925
Thuwal 23955-6900, KSA
Xiangliang.Zhang@kaust.edu.sa

ABSTRACT

Link prediction is a fundamental task in social networks. Recently, emphasis has been placed on forecasting new social ties using user mobility patterns, e.g., investigating physical and semantic co-locations for new proximity measure. This paper explores the effect of in-depth mobility patterns. Specifically, we study individuals' movement behavior, and quantify mobility on the basis of trip frequency, travel purpose and transportation mode. Our hybrid link prediction model is composed of two modules. The first module extracts mobility patterns, including travel purpose and mode, from raw trajectory data. The second module employs the extracted patterns for link prediction. We evaluate our method on two real data sets, GeoLife [15] and RealityMining [5]. Experimental results show that our hybrid model significantly improves the accuracy of social link prediction, when comparing to primary topology-based solutions.

Keywords

Mobility Pattern, Transportation Mode, Link Prediction, Social Ties, Social Networks, Call Detail Records

1. INTRODUCTION

Social networks, where nodes represent individuals and edges symbolize social ties and connections, are very dynamic in nature. Understanding how these networks evolve over time, and what factors induce such changes, is a fundamental research question in the study of human behavior. In general, networks evolve with the addition/deletion of new nodes and/or new edges (links). In this paper, we focus on the specific case of what factors influence new edge formation, and how it can be foreseen, i.e., *link prediction*.

With an aim to infer new links using only topology-extracted features, many proximity measures from graph theory were adopted in social networks [8]. In this paper, we place more emphasis on non-topology based measures, particularly movement-based features. Our goal is to investigate the correlation between new edge formation and high-level individuals' movement behavior. We strive to answer the question, how much of new ties can be predicted with the help of movement patterns? Unlike the previous studies of mobility patterns in link prediction, such as [9, 10, 12, 13, 11] where physical and semantic co-locations were investigated for new similarity/proximity measure, we explore high-level descriptive mobility-based features in link prediction, including both working and leisure patterns as well as travel modes. We extract features from individuals based on the analysis of where they go, how often they visit these places, and what mode of transportation they frequently use.

As smartphones become ubiquitous, the availability of Call Detail Records (CDR) and other location relevant usage records enables the study of individuals' movement behavior [7]. Mobility patterns derived from the movement behaviors can thus provide valuable insights on the human social lifestyles. To explore the value of mobility patterns in link prediction, we propose a hybrid model based on both topology and mobility measures. To the best of our knowledge, this work presents the first attempt to predict social ties from detailed mobility patterns. Precisely, it is the first attempt to explore the impact of both travel patterns and transportation mode in social link prediction.

The proposed framework is composed of two modules, *mobility pattern inference* and *social link prediction*. The module of *mobility pattern inference* firstly formalizes the problem of identifying users' point-of-interests (POIs) by clustering and the problem of transportation mode by classification, and then infers users' mobility patterns. The *link prediction* module focuses on employing the movement behaviors for predicting potential links among users. We evaluate our modules on two real data sets, GeoLife [15] and RealityMining [5]. The experimental results show that our link prediction model performs significantly better than the baseline method where only topology features are utilized.

The rest of the paper is organized as follows. Section 2 discusses related work in link prediction and mobility pattern inference. Section 3 introduces our proposed framework. Section 4 reports our experimental results and findings. Section 5 concludes the paper with perspectives.

2. RELATED WORK

The related work presented in this section is grouped into two categories. The first group introduces research con-

SAC'14 March 24-28, 2014, Gyeongju, Korea.

http://dx.doi.org/10.1145/2554850.2554918

ducted in inferring mobility patterns from location traces, and the second category summarizes work in exploring mobility patterns in link prediction.

Inferring Mobility Patterns from Location History: Studies of mobility pattern inference differ based on the data under investigation and/or the desired granularity of the results. For raw GPS data, Ashbrook and Starner proposed a model for point-of-interests (POIs) identification [2]. POIs are the key places visited by users, e.g., working places, home, and shopping mall. The proposed model is a threshold-based algorithm, which distinguishes stay and movement points by the evaluation of staying time longer than a specified threshold. The detected stay points are then grouped to form main clusters using k-means. Each cluster is then appointed as a point of interest. Zheng et al. proposed a travel mode inference framework for GPS trajectories [14]. The principal design of the model is based on extracting motion descriptive features from trip segments, followed by learning a classification model on top of the extracted features. Due to the comparable characteristics of some transportation modes (e.g., car and bus), a post-processing graph-based algorithm is employed to enhance the achieved results. In our approach, we adopt the same framework for travel mode detection and point of interest identification, yet we explore the impact of new features in differentiating similar transportation modes.

Exploring Mobility Patterns in Link Prediction: Link prediction using location history has gained rapid attention in the past few years [9, 10]. One stream of research has focused on introducing new similarity measures based on actual physical locations [4, 10, 3, 9]. The basic intuition behind these studies is: new links are more likely to form between people who have visited the same places. Scellato et al. examined the correlation between users visiting the same places and link formation in an online social network [9]. The presented approach was able to reduce the prediction space using the findings of correlation. Wang et al. explored the relationship between network-based user similarity measures and actual spatial and temporal physical proximity [10]. The results achieved validate the logical intuition behind the adopted measure. In this study, we adopt the same concept of physical and temporal co-location. However, we extract features based on most-frequent patterns as well as examine the effect of another dimension of mobility pattern, *transportation mode*. In addition, we examine the effect of co-location at a lower level of granularity, i.e., users who share the same cell-tower IDs at a given time are considered co-located.

The other stream of research has investigated the impact of semantic co-locations [12, 13, 11]. The hypothesis behind this direction is: people are more likely to form new links if they share similar interests and/or hobbies. Xiao et al. proposed a user similarity measure which takes into account the semantics of the visited locations, rather than the actual physical location [12]. That is, trajectories of users do not have to overlap in the actual physical space in order for users to be considered similar. In our approach, the semantic co-location notion is considered by measuring the similarity of social activities.

3. INFERRING SOCIAL TIES FROM HUMAN MOBILITY PATTERNS

Human trajectory traces and communication logs are often recorded by mobile service providers for billing purposes. For every user, trajectory traces comprise a set of location information recorded at given timestamps. Communication logs constitute all the phone calls and text messages to and from a given user. Our proposed method aims to extract mobility measures from trajectory traces and explore them in social link prediction, along with topology-based features from communication logs. We will firstly give the definitions and then present the two modules of our model in section 3.2 and 3.3.

3.1 Definitions and Notations

Let $G = (V, E)$ be an undirected graph, where V is the set of vertices/users and E is the set of edges representing communication logs between pair of users. An edge $e = (x, y)$ associated with a timestamp $T(e)$ presents the existence of a communication, e.g., phone call or short text message, between user x and y at time $T(e)$. Each user $x \in V$ is described by a trajectory of locations and related timestamps, such that $Tr(x) = \langle L(x), T(x)\rangle$, where $L(x) = \{l(x)\}$ is the set of movement traces for user x, and $T(x)$ is the timestamp associated with $L(x)$. $l_i(x) \in L(x)$ is either a geographical location (i.e., Latitude and Longitude), or a unique location identifier (e.g., cell tower ID). A *trip* is defined as a user's movement from point a to b, given that a user stays for a significant amount of time at both a and b. *Trips* of a user x are denoted by $trips(x) = \langle \hat{L}(x), \hat{T}(x)\rangle$, where $\hat{L}(x) \subset L(x)$ and $\hat{T}(x) \subset T(x)$.

To apply link prediction methods, we create two sub-networks from G, G_{known} and $G_{predict}$. $G_{known} = (V, E_t)$ such that $E_t = \{e|e \in E, T(e) \leq t\}$, where t is a specified timestamp. That is, G_{known} has the same set of nodes in G, but only includes edges with $T(e) < t$. $G_{predict} = (V, E - E_t)$ also has the same set of nodes, but only includes new edges which appeared after t, i.e., $E_{predict} = \{e|e \in E, e \notin E_t\}$. The same concept of splitting applies to the movement traces Tr for each user. When predicting links, we only consider location traces that occurred before the given timestamp. Given the above, our objective is to infer new edges in $G_{predict}$, using measures extracted from G_{known} and Tr_{known}. $FreqU$ is the set of frequent users in G whose degree is larger than a threshold d, and $FreqU_{1/2}$ is the first half of $FreqU$.

3.2 Mobility Pattern Inference

In the *mobility pattern inference* module, we seek to infer individual's travel patterns from low-level trajectory data. In precise, the aim is to infer travel purpose, i.e., user's main Points-of-Interests (POIs), as well as transportation mode. Individuals' trip patterns are extracted from their trips to/from main POIs.

3.2.1 *Inferring Main Point-of-Interests*

Main points-of-interests are the key places where each user visits frequently and stays for a significant amount of time. In order to differentiate main POIs from infrequent stay points, we adopt a clustering and rule-based approach. Given a set of trips for user x, $trips(x)$, we use the stay points of every trip (a, b) as the initial POIs. Then, we apply Density-Based Clustering (DBSCAN) [6] on the collection of all initial POIs. Each of the formed clusters is considered as one main POI. To distinguish POI clusters of *work* from *home*, we use a rule-based approach, and label each cluster by the most frequent visiting time. A cluster, whose most visits are during working hours, is considered as a *work place*. Similarly, a cluster where staying hours are late at night during weekdays, is labeled as *home*. Other main clusters, where visit times are after working hours but before midnight, are labeled as *leisure*.

3.2.2 Inferring Transportation Mode

Mode of transportation defines the way people commute from place a to b. In general, most frequent travel modes are: *walk*, *bike*, *bus*, and *car*. Travel mode inference can be regarded as a classification problem, where features (extracted from trips) are used as input to a classification model that learns to distinguish modes. Thus, the main task is to extract representative motion descriptive features.

Feature Extraction. For every trip, we extract three types of motion descriptive features: *global* features, *local* features, and *FFT* features. Global features are measured using the first and last points in a trip, while local features are measured on each consecutive pair of points in a trip. Table 1 shows a list of the local and global features adopted in this study.

Given a trip for user x with n points, the *global duration* is measured as $t_n(x) - t_1(x)$, while the j-th *local duration* is measured as $dr_j = t_{j+1}(x) - t_j(x)$, $j = 1, ..., n-1$. From the $n-1$ local duration values, three local features of duration are defined: mean of dr_j, standard deviation of dr_j and maximum of dr_j. The same global and local measures are applied to *Displacement*, *Heading*, *Velocity*, and *Acceleration*. Three other global features evaluate the rate of *Heading Change*, of *Stop* and of *Velocity Change*.

In order to capture main frequencies of velocity changes in different transportation modes, we apply Fast Fourier Transformation (FFT) on the sequence of local *velocity* feature, $[v_1, ..., v_j, ..., v_{n-1}]$. The motivation is that trips of a bus have similar velocity to trips of a car, but a bus stops more frequently than a car. The FFT of each trips results in a 5-dim vector representing the magnitude of velocity changes in frequency like every 30 mins, 10 mins, 5 mins, the variance of magnitudes in frequency domain less than 2 mins and the mean of the four largest magnitudes.

Classification. We utilized three popular classification algorithms, K-Nearest Neighbor (KNN), Random Forests (RF) and Support Vector Machines (SVM). The performance of them on (1) only local features, (2) only global features, (3) only FFT of velocity, and (4) all features is reported in section 4.1.

3.3 Social Link Prediction

Node *proximity* in social networks is often regarded as the go-to approach for link prediction. The idea behind this concept is simple, users (i.e., nodes) are more likely to communicate (i.e., form edges) in the future if they are *close* or *similar* to each other in the present. This similarity, a.k.a., node *proximity*, can be measured by quantities such as common neighbors and shortest path. When additional data is accessible, node proximity can be measured using other means such as physical proximity, co-location rate and/or trajectory overlap. In this study, we propose a link prediction module that exploits both mobility patterns and topology-based measures.

3.3.1 Topology-Based Measures

Two widely used topology-based measures, k-hop common neighbors and Adamic-Adar, are employed as baselines for computing node proximity.

k-hop Common Neighbor measures the number of neighbors in common between user x and y in the k-th degree neighborhood.

$$S^k(x, y) = |\Gamma^k(x) \cap \Gamma^k(y)| \tag{1}$$

where $|\cdot|$ gives the cardinality of a set, and $\Gamma^k(\cdot)$ is the set of k-hop neighbors of a user, i.e., $\Gamma^1(x) = \{y|y \in V, (x, y) \in E\}$, $\Gamma^2(x) = \{y, z|z \in \Gamma^1(x), (y, z) \in E\}$ and $\Gamma^k(x) = \{y, z|z \in \Gamma^{k-1}(x), (y, z) \in E\}$.

k-hop Adamic-Adar, introduced by Adamic and Adar [1], is a modified *common neighbor* measure, which uses the logarithm of node degree, rather than the even weight 1, in the summation over the common neighbors.

$$AA^k(x, y) = \sum_{z \in \Gamma^k(x) \cap \Gamma^k(y)} \frac{1}{log(|\Gamma(z)|)} \tag{2}$$

3.3.2 Mobility-Based Measures

Our objective is to employ high-level mobility-based features in user similarity measures. For this reason, three mobility patterns are defined as follows:

Mobility in Working. We use POIs labeled for working to characterize the users' mobility in working. For one user, we describe him/her by a vector of visit frequency to the POIs during working hours. Users working/studying at same places have close distribution of their working time at the POIs. We thus define $MP\_W(x, y)$, their difference in mobility patterns of working, by the Euclidean distance of their corresponding vectors. Users with small values of $MP\_W(x, y)$ are expected to be colleagues/classmates and are potential to have social links.

Mobility in Evening/Leisure. We define the mobility patterns of evening/leisure in an analogous manner. POIs labeled as evening/leisure are used to characterize user behaviors after working. The measure $MP\_E(x, y)$ quantifies how similar user x and y spend their leisure time at the recognized leisure places. Please note that the POIs of *home* are excluded in link prediction module as work and leisure places are more common than home for people to build their social ties.

Transportation Mode. Public transportation modes, e.g., bus, metro and tram, are of our interest for link prediction, as people have high chance to meet when taking the same way for going work or home. $MP\_T(x, y)$ is assigned with 1 if user x and y have the same public transportation mode, otherwise $MP\_T(x, y) = 0$.

3.3.3 Hybrid Social Link Prediction

In our proposed hybrid model of link prediction, mobility patterns are employed to enrich topology-based measure. Algorithm 1 shows the pseudo code of our approach when predicting the social link of user x and y. Given the mobility patterns of two users, their relationship is evaluated in the following three hypotheses: i) they are similar in working patterns regardless of leisure and travel mode patterns (line 5 and 6); ii) they are similar in both working and leisure patterns regardless of travel mode (line 1 and 2); and iii) they

Table 1: Motion Descriptive Features for a trip

Measure	Definition	Global	Local
Duration	Difference in time between points	$DR = t_n(x) - t_1(x)$	$mean(dr_j), std(dr_j), max(dr_j)$ where $dr_j = t_{j+1}(x) - t_j(x)$, $j = 1, ..., n-1$
Displacement	Geographical distance between points	D	$mean(d_j), std(d_j), max(d_j)$
Heading	Angle degree measured from the true north	-	$mean(h_j), std(h_j), max(h_j)$
Velocity	Displacement/Duration	V	$mean(v_j), std(v_j), max(v_j)$
Acceleration	Change in velocity/Duration	-	$mean(a_j), std(a_j), max(a_j)$
Heading Change Rate	n_CD/Duration n_CD: the N. of locations where the user changed direction Change in direction occurs when $\|h_{j+1} - h_j\| > threshold$	HCR	-
Stop Rate	n_SP/Displacement n_SP: the N. of stop points, where $v_j < threshold$	SR	-
Velocity Change Rate	n_PV/Displacement n_PV: the N. of points where the user changed velocity Change in velocity occurs when $\|v_{j+1} - v_j\| > threshold$	VCR	-

share similar patterns in working, leisure and travel mode (line 3 and 4).

For users holding the 2nd and 3rd relationship, their topology-based measure $S(x, y)$ (or $AA(x, y)$) are enhanced by connecting to all frequent users whose degrees are large enough, $FreqU$. Intuitively, frequent users (hub users) are potentially to be connected with most of other users. Increasing the proximity of users who have strong mobility correlation in such a way cannot only solve the sparsity problem in social graph, but also obey the reality principle. The enriched proximity measure on common neighbors is defined as:

$$S^k(x, y|FreqU) = |(\Gamma^k(x) \cup FreqU) \cap (\Gamma^k(y) \cup FreqU)| \quad (3)$$

The enriched Adamic-Adar measure is

$$AA^k(x, y|FreqU) = \sum_{\substack{z \in (\Gamma^k(x) \cup FreqU) \\ \cap (\Gamma^k(y) \cup FreqU)}} \frac{1}{log|\Gamma(z)|} \quad (4)$$

The 1st relationship is weaker than the other two, users holding it have a boosted neighborhood by including less number of frequent users, $FreqU_{1/2}$. It is worth noting that $FreqU$ is often a small set as social graphs have power-law degree distribution.

Algorithm 1 Hybrid Link Prediction Model

Input: $S^k(x, y)$, $S^k(x, y|FreqU)$, $MP\_W(x, y)$, $MP\_E(x, y)$, $MP\_T(x, y)$, ϵ

1: **if** $MP\_W(x, y) < 0.3$ & $MP\_E(x, y) < 0.15$ **then**
2: x and y are linked if $S^k(x, y|FreqU) > \epsilon$ //work-leisure
3: **else if** $0.5 > MP\_W(x, y) > 0.3$ & $MP\_E(x, y) > 0.15$ & $MP\_T(x, y) = 1$ **then**
4: x and y are linked if $S^k(x, y|FreqU) > \epsilon$ //work-leisure-mode
5: **else if** $MP\_W(x, y) < 0.5$ **then**
6: x and y are linked if $S^k(x, y|FreqU_{1/2}) > \epsilon$ //work
7: **else**
8: x and y are linked if $S^k(x, y) > \epsilon$
9: **end if**

4. EXPERIMENTAL EVALUATION

We evaluate the module of mobility pattern on GeoLife data [15] as it has diverse enough travel modes and easy-to-interpret POIs. We evaluate link prediction on the RealityMining [5] data[1], rather than GeoLife due to its lack of ground truth in social ties.

4.1 Mobility Pattern Inference

Datasets: GeoLife [15] is a GPS trajectory dataset collected by *Microsoft Research Asia* from 2007 to 2012. It contains raw GPS traces for 182 users, including timestamp, latitude, longitude and altitude for each point in the trajectory. We used the data of 72 users, where each of their trips is labeled with a transportation mode, such as walk, bike, bus, car (and taxi), train and airplane.

Experimental Design: Since the dataset provides no ground-truth for users' main POIs, we manually verify the extracted POIs in Google Maps, for a couple of randomly selected users. The clustered POIs, which are not transit points such as bus stops or train stations, indeed have reasonable labels that coincide with the nearby labels in Google Maps.

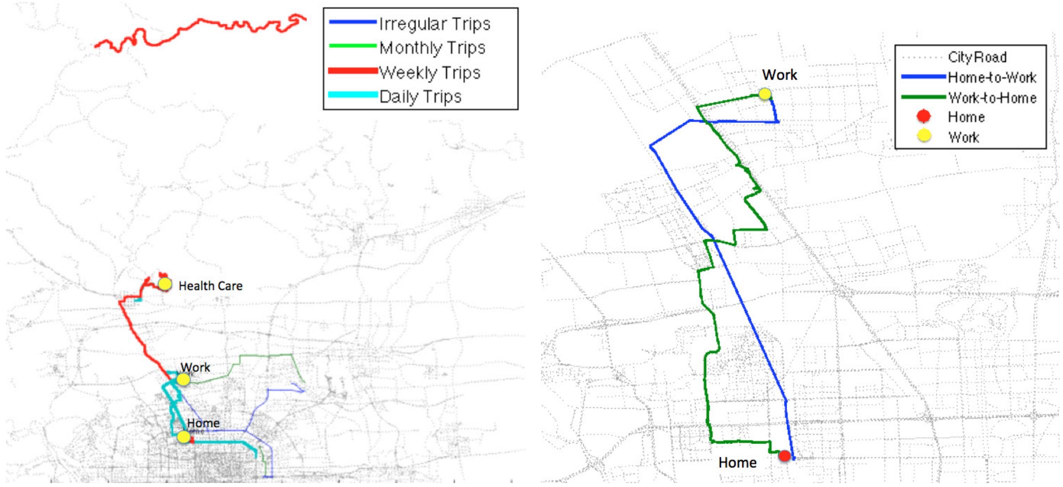
As for evaluating travel mode, we use classification accuracy, defined as $\frac{\sum_{i=1}^{m} C_i}{N}$ where C_i is the correctly classified trips in mode i, m is the number of travel modes, and N is the total number of trips. 5-fold cross validation is applied to report the average accuracy of the model.

Experimental Results: We visualize the travel pattern of a user in Fig. 1 (not for all users due to space constraints). The main POIs labeled as either *home*, *work* or *other* are attached with trips with different frequency, categorized as either *daily*, *weekly*, *monthly*, or *irregular*. It is worth noting that the trips of home-to-work are different from that of work-to-home.

Fig. 2 shows the accuracy of travel mode classification, when using four variations of the extracted features: global, local, FFT, and all. In general, local features perform well in predicting travel modes. The full feature set including all global, local and FFT features improves a little on the accuracy stability. Random forests provide better predictions than SVM and k-NN. The good performance on travel mode classification ensures the proper extraction of mobility patterns, which are used next for link prediction.

4.2 Social Link Prediction

[1]RealityMining is not used for mobility pattern evaluation because most of the users' activities are around the campus.



(a) Daily, weekly, & monthly trip of one user (b) Home-to-work trip, & work-to-home trip

Figure 1: Point-of-interests visualization of mobility patterns for one user. Home and work places are marked in red and yellow respectively (in (b)). Trips with unmarked origins and/or destinations represent other places of interests such as entertainment or leisure. Lines between main places represent trips, where the width of each line shows the frequency of trips, e.g., daily, weekly, monthly, and irregular ones.

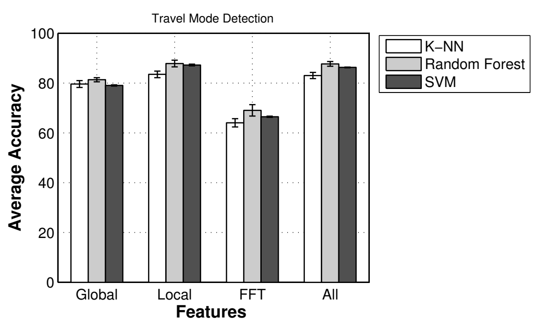


Figure 2: Average accuracy of travel mode classification when employing different sets of features; error bars show standard deviation.

Datasets: RealityMining [5] is a cell-phone dataset collected by *MIT Media Laboratory* from 2004 to 2005. It contains the following for 94 users: call logs, cell tower IDs, and survey data. This dataset is suitable for evaluating our link prediction approach as the ground truth of social ties is available in call logs and user mobility can be explored from mobile records in terms of cell IDs.

Experimental Design: As introduced in section 3.1, we split the graph into G_{known} and $G_{predict}$ by the day of Feb. 1, 2005. 84 users with nonempty travel records are selected. There exist 115 links among them in G_{known} and 36 new links are created in $G_{predict}$. Mobility patterns are extracted from movement recorders before Feb. 1, 2005.

The used evaluation criteria are the ROC curves and the Area Under the Curve (AUC). We compare the performance of our hybrid model and the baseline approach without mobility patterns on these criteria through tuning the threshold ϵ in Algorithm 1. The frequent user set $FreqU$ includes 11 users who are linked with more than 5 others, while the $FreqU_{1/2}$ includes 5 users whose degree is larger than 6.

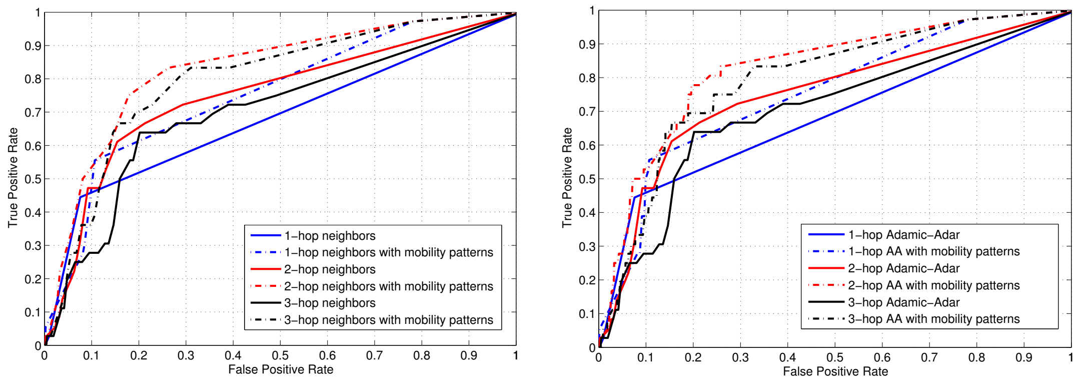
Table 2: AUC - Social Link Prediction without mobility patterns (C2), with only working patterns (C3), with only working and leisure patterns (C4), with working and leisure patterns and modes (C5).

Neighbors	C2	C3	C4	C5
1-hop	0.6836	0.7330	0.7529	0.7652
2-hop	0.7517	0.8100	0.8294	**0.8397**
3-hop	0.7054	0.7787	0.8006	0.8096

Experimental Results: Fig. 3 shows the ROC curves of our hybrid model and the baseline method, when using k-hop common neighbors (a) and k-hop Adamic-Adar (b). We can observe that our hybrid model (dot dash lines) achieves better performance when compared to using only topology-based features (solid lines). This improvement is recognizable for the two adopted topology-based measures that have similar performance, as well as for the 1- to 3-hop variations of the measures. In addition, the 2-hop measures always outperform 1-hop and 3-hop measures. The 1-hop and 3-hop introduce too few or too many neighbors, which result in inaccurate link prediction.

For comparing the improvement made by different mobility patterns, Table 2 shows the AUC values of link prediction approaches when not using any mobility patterns (column 2), using only working patterns (column 3), using only work and leisure patterns (column 4), and using work, leisure patterns and travel mode (column 5). The topology-based measure in this comparison is the 1- to 3-hop common neighbors. That is to say, column 2 and 5 show the AUC of curves in Fig. 3 (a). The continuing increase in AUC values from column 2 to 5 in all rows shows that using more mobility patterns can achieve better link prediction results. The employment of mobility pattern for improving link prediction is thus validated.

5. CONCLUSION



(a) k-hop common neighbors and mobility patterns

(b) k-hop Adamic-Adar and mobility patterns

Figure 3: Social Link Prediction: ROC curves of approaches using only topology-based measures and of hybrid model with mobility patterns.

In this work, we present a hybrid framework for predicting new social ties in social networks. Our key contribution lies in employing descriptive mobility patterns in social link prediction. We demonstrate the validity of these patterns, when paired with low-level node proximity measures, using real world mobile data. Following this work, we will investigate refined individual and group mobility patterns, which are useful for link prediction and community detection in large social networks.

References

[1] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social networks*, 25(3):211–230, 2003.

[2] D. Ashbrook and T. Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286, 2003.

[3] D. J. Crandall, L. Backstrom, D. Cosley, and et al. Inferring social ties from geographic coincidences. *Proc. of the National Academy of Sciences*, 107(52):22436–22441, 2010.

[4] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh. Bridging the gap between physical location and online social networks. In *Proc. of ACM Intl. Conf. on Ubiquitous computing*, pages 119–128, 2010.

[5] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

[6] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *SIGKDD*, pages 226–231, 1996.

[7] S. Isaacman, R. Becker, R. Cáceres, and et al. Identifying important places in people's lives from cellular network data. *Pervasive Computing*, pages 133–151, 2011.

[8] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7): 1019–1031, 2007.

[9] S. Scellato, A. Noulas, and C. Mascolo. Exploiting place features in link prediction on location-based social networks. In *SIGKDD*, pages 1046–1054, 2011.

[10] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi. Human mobility, social ties, and link prediction. In *SIGKDD*, pages 1100–1108, 2011.

[11] X. Xiao, Y. Zheng, Q. Luo, and X. Xie. Finding similar users using category-based location history. In *SIGSPATIAL*, pages 442–445, 2010.

[12] X. Xiao, Y. Zheng, Q. Luo, and X. Xie. Inferring social ties between users with human location history. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–17, 2012.

[13] J. J.-C. Ying, E. H.-C. Lu, W.-C. Lee, T.-C. Weng, and V. S. Tseng. Mining user similarity from semantic trajectories. In *SIGSPATIAL Workshop on Location Based Social Networks*, pages 19–26, 2010.

[14] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma. Understanding transportation modes based on gps data for web applications. *ACM Transactions on the Web (TWEB)*, 4(1):1, 2010.

[15] Y. Zheng, X. Xie, and W.-Y. Ma. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33(2):32–39, 2010.