Inferring Demographics and Social Networks of Mobile Device Users on Campus From AP-Trajectories

Pinghui Wang[†], Feiyang Sun[†*], Di Wang[†], Jing Tao[†],
Xiaohong Guan[†,‡], Albert Bifet[§]
[†]MOE Key Laboratory for Intelligent Networks and Network Security, Xi'an Jiaotong University, China
[‡]Department of Automation and NLIST Lab, Tsinghua University, Beijing, China
[§]LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France
{phwang, jtao, xhguan}@mail.xjtu.edu.cn,
{fysun, diwang}@sei.xjtu.edu.cn, albert.bifet@telecom-paristech.fr

ABSTRACT

Exploring demographics and social networks of Internet users are widely used for many applications such as recommendation systems. The popularity of mobile devices (e.g., smartphones) and location-based Internet services (e.g., Google Maps) facilitates the collection of users' locations over time. Despite recent efforts to predict users' attributes (e.g., age and gender) and social networks based on utilizing the rich location context knowledge (e.g., name, type, and description) of places of interest (e.g., restaurants and hotels) they checked-in on location-based online social networks such as Foursqure and Gowalla, little attention has been given to inferring attributes and social networks of mobile device users based on their spatiotemporal trajectories with less/no location context knowledge. In this paper we collect logs of thousands of mobile devices' network connections to wireless access points (APs) of two campuses, and investigate whether one can infer mobile device users' demographic attributes and social networks solely from their spatiotemporal AP-trajectories. We develop a tensor factorization based method *Dinfer* to infer mobile device users' demographic attributes from their AP-trajectories by leveraging prior knowledge, such as users' social networks. We also propose a novel method *Sinfer* to learn social networks between mobile device users by exploring patterns of their AP-trajectories, such as fine-grained co-occurrence events (e.g., co-coming, co-leaving, and co-presenting duration). Experimental results on real-word datasets demonstrate the effectiveness of our methods.

Keywords

user profiling; spatiotemporal trajectories; social network

1. INTRODUCTION

In today's digital big data era, human activities on a variety of domains have been collected and monitored. For example, people's webpage browsing, shopping, video viewing, and music listening behaviors on the Internet can be easily collected by providers of web services such as webpage search engines, e-commerce, video, and music websites. These collected user activities facilitate service providers to better understand/profile their users, which is crucial for many applications such as target advertising, business recommendation, and counter-criminal/terrorists. It has revealed that service providers can infer their users' demographic attributes such as age and gender from the users' Internet browsing history [22], linguistics writing [11], mobile call/message records [8,9], music listening history [18], purchase data [27], and available demographic attributes of their friends on online social networks (OSNs) such as Facebook and LinkedIn [21], even though the users do not intend to reveal their attributes to the service providers.

Recently, mining people's locations over time attracts a lot of attention due to the popularity of mobile devices (e.g., smartphones) and location-based Internet services such as navigation applications (e.g., Google Maps and Uber) and location-based online social networks (in short, location-based OSNs, e.g., Instagram and Foursquare). For example, mobile devices such as smartphones and smartwatches can be precisely located outside/inside buildings by current/near-future positioning techniques, which precisely infer the locations of mobile devices using multiple information such as built-in sensors (e.g., GPS and barometers) of mobile devices and the location and signal strength of connected wireless access points (APs)/cell-phone towers. This leads to mobile device users' spatiotemporal trajectories explicitly or implicitly exposed to third parties. For example, people may enable map services to retrieve their GPS locations automatically, and may also explicitly distribute geo-tagged posts (e.g., tweets and photos) or check-in places of interest (POIs) on OSNs. In addition, users' trajectories may also be exposed implicitly even though users do not intend to reveal their locations. For example, banks and mobile phone service providers can learn users' spatiotemporal trajectories from collected credit-card and telecommunication transactions respectively. The collection of a large number of users' spatiotemporal trajectories facilitates the study of research topics such as urban planning, congestion prediction, and point-of-interest recommendation in smart cities.

In this paper, we are interested in testing whether one can infer mobile device users' demographic attributes and social networks from their spatiotemporal trajectories. Existing work [30] learns users' attributes from their check-in POIs on location-based OSNs such as Facebook, Foursquare, Yelp. The method relies on POIs' rich semantic features such as categories, user reviews, and descriptions. However, in practice the semantic features of locations may not be publicly available. For example, daily activities of most col-

*P.W. and F.S. contributed equally to this work.

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054140

lege students and faculties may be centered in campuses. Third parties now or in the near future may be able to collect fine-grained ***indoor*** spatiotemporal trajectories of persons on campus but fail to get the fine-grained context (e.g., library and gym) of a place on campus, because a building may have different functional areas (e.g., research lab and restaurant) and the context of each building/place on campus may not be publicly available on the Internet. In addition to demographic attributes, we also study the problem of predicting mobile device users' close friends from their spatiotemporal trajectories. Existing methods [4–6,24,26,29] mine friendships and infer social strengths mainly from co-coming events between users, e.g., check-in records on OSNs, because most location-based OSNs record the time of users' check-ins but not check-outs (In practice, users check-in at POIs on location-based OSNs but never check-out).

To the best of our knowledge, our work is the first attempt to infer users' attributes by utilizing AP-trajectories without location context knowledge. Moreover, we study three types of fine-grained co-occurrence events: co-coming, co-leaving, and co-presenting duration, and observe that these fine-grained events are effective and complementary to each other for predicting friendships between mobile device users. Our contributions are summarized as:

• We conduct an in-depth measurement study on spatiotemporal trajectories of 52 thousand mobile devices on two campuses, which are revealed by their network connections to wireless APs.

• We develop a tensor factorization based learning method *Dinfer* to infer mobile device users' attributes from their AP-trajectories by leveraging user social networks, which could be learned from users' spatiotemporal trajectories.

• We use pointwise mutual information (PMI) to evaluate whether a fine-grained co-occurrence happens by chance or it is a social event, and propose an effective method *Sinfer* to learn social networks of mobile device users by exploring their fine-grained co-occurrence events.

The rest of this paper is organized as follows. Section 2 describes our datasets. Section 3 presents our measurement study of micro- and macro-statistics of mobile devices' AP-trajectories. Sections 4 and 5 present our methods Dinfer and Sinfer for learning demographic attributes and social networks of mobile device users from their spatiotemporal AP-trajectories. Section 6 presents the performance evaluation and testing results. Section 7 summarizes related work. Concluding remarks then follow.

2. DATASETS

In this section, we introduce the datasets used in this paper. Our work strictly follows the ethical guidelines. To avoid ethical concerns, we have anonymized mobile users' identity information such as netIDs and mobile devices' MAC (Media Access Control) addresses in the datasets. We will make the datasets publicly available.

2.1 Spatiotemporal Data: AP-Trajectories

From May 03 to July 15, 2015, we poll all APs in Campuses A and B regularly (every 5 minute) and collect information (e.g., MAC addresses) of their connected mobile devices via SNMP (Simple Network Management Protocol). Based on this dataset, we can determine whether a mobile device connected to an AP during a time interval. Campus A with a gross floor area of 0.4 km^2 and Campus B with a gross floor area of 0.07 km^2 have 2,283 and 362 APs respectively. For both campuses, each AP covers an area of about 100 m^2 in buildings. During the 74 days, there exist 29,618 and 22,285 active mobile devices on the Internet for Campuses A and B respectively. On average, a mobile was 21 (20) days active, and connected to 54 (38) APs in Campus A (B) respectively.

2.2 Ground Truth I: Device-NetID Networks

In our dataset, we observe that a NetID (user's token) may appear on multiple mobile devices, and also more than one NetID may appear on a mobile device. This happens because a person may have more than one mobile device, and may also release its NetID to its close friends for sharing its network usage quota. We observe that 40% and 37% of NetIDs are used in more than one mobile device, and 42% and 39% of mobile devices have more than one NetID used. In this paper, **we define two NetIDs are "ID_sharing_friends" when they both appeared on at least one same mobile device**. We use ID_sharing_friends as a ground truth to evaluate the performance of our method (i.e., Sinfer in Section 4) for close friendship inference.

2.3 Ground Truth II: Demographics

We obtain a dataset of NetIDs' genders and social roles (i.e., undergraduate student, graduate student, and faculty) from the information center of Campuses A and B, and use the dataset as the ground truth to evaluate the accuracy of our method for inferring demographic attributes of mobile network users. When more than one NetID appeared on a mobile device, we select the NetID most frequently used on the mobile device, and then retrieve the NetID's demographic attributes as the mobile device's user attributes. A summary of persons' demographics is shown in Table 1.

Table 1: Statistics of persons on Campus A (B).

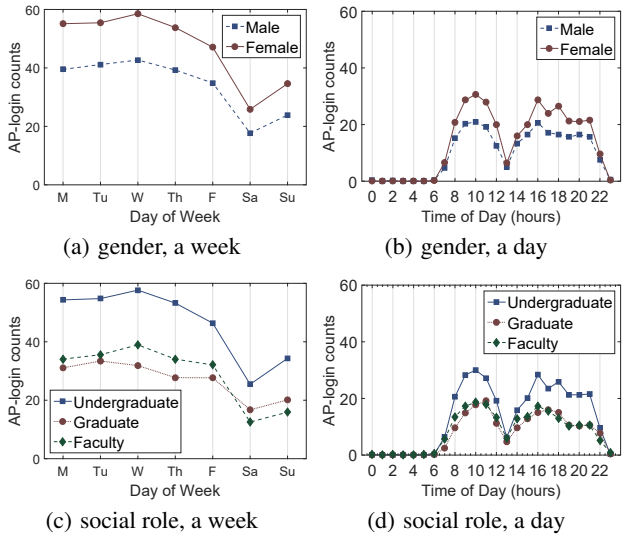
	#devices	#NetIDs
Male	9,364 (5,095)	4,797 (3,212)
Female	20,254 (17,190)	12,038 (10,307)
Undergraduate	21,136 (18,663)	12,982 (11,328)
Graduate	3,630 (1,237)	2,409 (987)
Faculty	4,852 (2,385)	1,444 (1,214)
Total	29,618 (22,285)	16,835 (13,529)

2.4 Ground Truth III: Affiliation Information

Persons on campus with following relationships tend to have strong social connections: 1) students in the same class; 2) students enrolled at school at the same year and affiliated to the same department; 3) faculties in the same department. Based on these observations, we collect the affiliation information of all NetIDs appeared in our datasets. This data is used as another ground truth for evaluating the accuracy of our method (i.e., Sinfer in Section 4) for learning social networks of mobile device users. Mobile devices collected on Campus A belong to users (or, NetIDs) from 21 departments and 812 undergraduate classes. On average, a department of Campus A has 618 undergraduate students, 115 graduate students, and 69 faculties, and an undergraduate class has 16 students. Mobile devices collected on Campus B belong to users from 21 departments and 804 undergraduate classes. On average, a department of Campus B has 539 undergraduate students, 47 graduate students, and 58 faculties, and an undergraduate class has 14 students.

3. MEASUREMENT STUDIES

In this section, we study macro- and micro-statistics of mobile devices' AP-trajectories. Intuitively, persons on campus with the same demographic attributes (such as gender, major, and social role) tend to exhibit similar macroscopic events such as visiting



(a) gender, a week (b) gender, a day

(c) social role, a week (d) social role, a day

Figure 1: (Campus A) A single user's average login counts for each day/hour in the week/day.

places in common. In addition, microscopic behaviors such as co-coming, co-leaving, and co-presenting at a place are strong indicators of close relationships between people. Next, we describe our observations in detail.

3.1 Macroscopic observations

Observation 1. **Macroscopic features are useful for demographic inference**. To some extent, a user's login count (i.e, the number of logins) reflects whether it moves frequently or not in campus, because Campuses A and B allow mobile devices to automatically connect and login to APs near to their current locations. Fig. 1 plots the average login counts of users on Campus A for each day in the week and each hour in the day. We observe: 1) woman moves more frequently than man; 2) undergraduate students move more frequently than graduate students and faculties; 3) compared to graduate students, faculties move more frequently on weekdays but less frequently on weekends. Here we omit the similar results for Campus B. This observation indicates that macro-statistics of AP-trajectories are useful features for predicting user attributes.

Observation 2. **ID_sharing_friends tend to have the same demographics**. We study the homophily between ID_sharing_friends' attributes such as gender, social role, department, grade, and class, where we say two friends are homophily when they have the same attribute of interest. Among all friend pairs of campus A (B), we observe that 81% (96%) are gender homophily, 91% (95%) are social role homophily, 84% (85%) are department homophily, 77% (85%) are grade homophily, and 57% (69%) are class homophily. This observation indicates that social networks of mobile device users are helpful for predicting user attributes.

Observation 3. **Macroscopic features of AP trajectories are not sufficient for identifying friendship relationships.** We test whether it is accurate to identify friendship relationships by directly comparing similarities between macroscopic features of mobile devices' AP-trajectories. We define a macro-trajectory matrix, where the matrix's element (i, j) records the average number of times the device connected the j-th AP during the i-th hour in the week, where $1 \leq i \leq 7 \times 24$. We concatenate the elements in each mobile device's macro-trajectory matrix into a one-dimensional feature vector. To evaluate the similarity between two mobile devices' macro-trajectories, we use two similarity metrics: *cosine similarity* and *Jaccard similarity*, which are defined as $\frac{\mathbf{v_1} \cdot \mathbf{v_2}}{\|\mathbf{v_1}\| \|\mathbf{v_2}\|}$ and $\frac{\mathbf{v_1} \cdot \mathbf{v_2}}{\|\mathbf{v_1}\|^2 + \|\mathbf{v_2}\|^2 - \mathbf{v_1} \cdot \mathbf{v_2}}$ respectively. For mobile devices belonging to undergraduate students, we compute their Top-20 most similar mobile devices. Table 2 shows the distribution of these Top-20 mobile device users. On average, two similarity metric based methods identify only 18.52% of ID_sharing_friends and 26.33% of classmates for each user. In contrary, our microscopic features based method Sinfer, which will be introduced in Section 5, is more effective for close friendship inference.

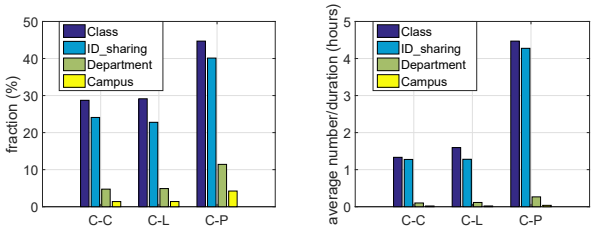
Table 2: (Campus A, undergraduate students) Fraction of friends among Top-20 most similar mobile device users identified based on macroscopic features. Sinfer is our method based on microscopic features, which will be introduced in Section 5.

feathures&method	ID_sharing_friends	classmates
macroscopic&cosine	19.35%	29.07%
macroscopic&Jaccard	17.69%	23.59%
microscopic&Sinfer	**27.48%**	**58.93%**

3.2 Microscopic observations

In this subsection, we study AP trajectories' microscopic features, including co-coming, co-leaving, and co-presenting events, and observe that these fine-grained co-occurrence events are effective for identifying close relationships between mobile device users. Existing studies [5,24,26] focus on inferring friendships between users based on their check-in records on location-based OSNs. For example, a check-in record (u, x, t) indicates that a user u checked in at POI x at time t. However, the exact leaving time of u at POI x is not clear for location-based OSNs, which provides important information for friendship inference.

Observation 4. **Co-coming, co-leaving, and co-presenting duration are all strong close relationship indicators.** For birthday parties, as an example, friends may not come to or leave prearranged places simultaneously. For this case, all of co-coming, co-leaving, and co-presenting events are strong close relationship indicators. Fig. 2 plots the fraction of user pairs having at least one co-occurrence event and the average number of co-occurrence events. We observe that: 1) the pairs of users with stronger relationships (e.g., ID_sharing_friends and classmates) tend to have more occurrence events. For example, ID_sharing_friends and classmates have more occurrence events than pairs of users selected randomly. 2) all



(a) fraction of pairs having co-occurrence events (b) average number/duration of co-occurrence events

Figure 2: (Campus A) Statistics of co-occurrences for pairs of users selected from different sets (e.g., ID_sharing_friends, classmates, and persons in the same department), where abbreviations C-C, C-L, and C-P refer to co-coming, co-leaving, and co-presenting respectively.

of co-coming, co-leaving, and co-presenting duration are helpful for predicting users' close relationships.

Observation 5. **Co-coming, co-leaving, and co-presenting are complementary to each other for indicating close relationships between mobile device users.** Among 153 thousand friend pairs (including ID_sharing_friends and classmates) having co-presenting events in Campus A, 64.3% have co-coming events, 65.2% have co-leaving events, and 23.8% have neither co-coming nor co-leaving events. Among friend pairs having co-coming events, 17.2% have no co-leaving events. In contrary, among friend pairs having co-leaving events, 18.4% have no co-coming events. The results of Campus B are similar. The above results indicate that co-coming, co-leaving, and co-presenting are complementary to each other for friendship inference.

4. DINFER: LEARN DEMOGRAPHICS

Formally, we formulate the **demographic inference problem** as: *Given AP-trajectories of all mobile devices* $u_1, \ldots, u_m$ *and user demographics (such as gender and social role) of a set of training mobile devices* $\{u_i : i \in \Omega_{train}\}$, *our goal is to automatically infer user demographics of remaining mobile devices* $\{u_i : i \in \{1, \ldots, m\} \setminus \Omega_{train}\}$. In this section, we introduce a method Dinfer to solve this problem. The framework of Dinfer is shown in Fig. 3. The center part of Fig. 3 represents the modeling of macro-statistics of AP-trajectories, where we use a tensor $\mathcal{X}$ to model spatial and temporal distributions of all mobile devices' AP-trajectories, and factorize tensor $\mathcal{X}$ into three matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$, which are low-rank latent representations of mobile devices, APs, and time slots respectively. As shown in the lower part of Fig. 3, we pose a constraint on $\mathbf{U}$, which is a Laplacian regularization $\mathbf{L}$ that is learned from social networks G_{MD} of mobile devices. G_{MD} is learned from micro-statistics of mobile devices' AP-trajectories by Sinfer in Section 5, where two mobile devices belong to close friends or the same person are connected in G_{MD}. Finally, mobile device users' demographics are learned from their latent representations $\mathbf{U}$ via a supervised classifier.

4.1 Modeling Information of Locations

We use a tensor $\mathcal{X} \in \mathbb{R}^{m \times n \times h}$ to denote the information of users' AP-trajectories, where tensor element $\mathcal{X}(i, j, k)$ is defined as the total time of the i-th mobile device connected to the j-th AP during the k-th time slot, $1 \le i \le m$, $1 \le j \le n$, and $1 \le k \le h$. As alluded, $\mathcal{X}$ is very sparse. Inspired by previous work on topic modeling [1], we observe that a mobile user may focus on a few topics, which results in $\mathcal{X}$ very sparse and low-rank. To solve this problem, we model mobile devices' trajectory information from the latent topic level. Inspired by the non-negative matrix factorization (NMF) model [3,14], we propose a non-negative tensor factorization (NTF) model to factor $\mathcal{X}$ into three lower-dimension matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$ in order to get a more compact but accurate latent representation of mobile devices by solving the following optimization problem:

$$\min_{\mathbf{U},\mathbf{V},\mathbf{T} \ge 0} \mathbb{O}_{\text{TRA}} = \|\mathcal{X} - \hat{\mathcal{X}}\|^2 \tag{1}$$

with $\hat{\mathcal{X}} = [\![\mathbf{U}, \mathbf{V}, \mathbf{T}]\!] \equiv \sum_{j=1}^{r} \mathbf{u}_{:j} \circ \mathbf{v}_{:j} \circ \mathbf{t}_{:j}$, where $\mathbf{U} \in \mathbb{R}^{m \times r}$, $\mathbf{V} \in \mathbb{R}^{n \times r}$, and $\mathbf{T} \in \mathbb{R}^{h \times r}$ are non-negative factor matrices to be learned, $[\![\cdot]\!]$ is a short-hand notation of the sum of rank-one tensors, $\circ$ represents the vector outer product, and $\mathbf{u}_{:j}$, $\mathbf{v}_{:j}$, and $\mathbf{t}_{:j}$ are the j-th columns of matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$ respectively. $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$ are representations of mobile devices, APs, and time slots in a latent space. The flexibility of NTF model allows us to incorporate prior knowledge such as social networks of mobile device users into this model, which will be introduced in later subsections.

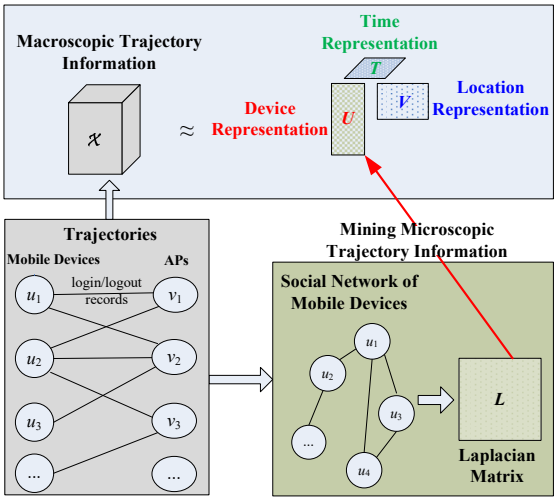


Figure 3: Overview of our method for learning social networks and demographics of mobile users. Social networks are learned by our method Sinfer in Section 5.

4.2 Modeling Social Network Information

We build a graph G_{MD} to model social networks of mobile devices $u_1, \ldots, u_m$. Each node in G_{MD} represents a mobile device, and there exists an edge (u_i, u_j) in G_{MD} when two different mobile devices u_i and u_j are friends, $1 \le i, j \le m$. Let $\mathbf{A} \in \mathbb{R}^{m \times m}$ denote the adjacency matrix of G_{MD}, i.e., $\mathbf{A}(i, j) = 1$ when u_i and u_j are connected, and $\mathbf{A}(i, j) = 0$ otherwise. We let $\mathbf{A}(i, i) = 0$. Denote $\mathbf{u}_i$ (i.e., the i-th row of matrix $\mathbf{U}$) as the representation of u_i, $1 \le i \le m$. In reality, close friends tend to have same attributes such as ages and genders. It indicates that connected nodes in G_{MD} also tend to have the same labels (e.g., gender and social role) of interest. To model this knowledge, we use graph Laplacian, which produces similar representations $\mathbf{u}_i$ and $\mathbf{u}_j$ for two connected mobile devices u_i and u_j. Formally, we formulate it as minimizing the following loss function:

$$\mathbb{O}_{\text{GMD}} = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \|\mathbf{u}_i - \mathbf{u}_j\|^2 \mathbf{A}(i, j). \tag{2}$$

The above loss function $\mathbb{O}_{\text{GMD}}$ incurs a penalty $\|\mathbf{u}_i - \mathbf{u}_j\|^2$ when two connected u_i and u_j have different representations $\mathbf{u}_i$ and $\mathbf{u}_j$.

Let $\mathbf{D} \in \mathbb{R}^{m \times m}$ denote a diagonal matrix, and its diagonal element is the degree of a mobile device in the adjacency matrix $\mathbf{A}$, i.e., $\mathbf{D}(i, i) = \sum_{j=1}^{m} \mathbf{A}(i, j)$. Let $\mathbf{L} = \mathbf{D} - \mathbf{A}$, and then we rewrite eq. (2) as

$$\mathbb{O}_{\text{GMD}} = \sum_{i=1}^{m} \sum_{j=1}^{m} (\mathbf{u}_i \mathbf{A}(i, j) \mathbf{u}_i^{\mathsf{T}} - \mathbf{u}_i \mathbf{A}(i, j) \mathbf{u}_j^{\mathsf{T}}) = \text{Tr}(\mathbf{U}^{\mathsf{T}} \mathbf{L} \mathbf{U}).$$

4.3 Learning User Representation

By considering both of the above two types of information, we formulate the task of learning latent representations of mobile device users as the following optimization problem:

$$\min_{\mathbf{U},\mathbf{V},\mathbf{T} \ge 0} \mathbb{O} = \underbrace{\|\mathcal{X} - [\![\mathbf{U}, \mathbf{V}, \mathbf{T}]\!]\|^2}_{\mathbb{O}_{\text{TRA}}} + \alpha \underbrace{\text{Tr}(\mathbf{U}^{\mathsf{T}} \mathbf{L} \mathbf{U})}_{\mathbb{O}_{\text{GMD}}} + \gamma \underbrace{(\|\mathbf{U}\|^2 + \|\mathbf{V}\|^2)}_{\text{regularization}}, \tag{3}$$

where the first term is to introduce AP-trajectories' macro-statistics, the second term is to consider social networks of mobile device users, and the third term is for regularization to avoid overfitting. α and γ are the trade-off parameters that control the effects of social networks and regularization terms respectively.

There is no close-form solution for the optimization problem (3). To solve this problem, we propose a multiplicative update algorithm to find optimal solutions for variables $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$ based on Oja's iterative learning rule [23,28]. The basic idea behind the multiplicative update algorithm is to optimize the objective $\mathbb{O}$ with respective to one variable while fixing the other. We keep updating the variables $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$ until convergence or reaching the number of maximum iterations. Next, we introduce the algorithm in detail. Let $\mathbf{\Phi} \in \mathbb{R}^{m\times r}$, $\mathbf{\Psi} \in \mathbb{R}^{n\times r}$, and $\mathbf{\Upsilon} \in \mathbb{R}^{h\times r}$ be the Lagrange multipliers for constraints $\mathbf{U} \geq 0$, $\mathbf{V} \geq 0$, and $\mathbf{T} \geq 0$ respectively. The Lagrange function $\mathbb{J}$ is defined as

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V},\mathbf{T}\geq 0} \mathbb{J} =& \|\mathcal{X} - [\![\mathbf{U},\mathbf{V},\mathbf{T}]\!]\|^2 + \alpha \text{Tr}(\mathbf{U}^\top \mathbf{L}\mathbf{U}) \\ &+ \gamma(\|\mathbf{U}\|^2 + \|\mathbf{V}\|^2 + \|\mathbf{T}\|^2) \\ &- \text{Tr}(\mathbf{\Phi}\mathbf{U}^\top) - \text{Tr}(\mathbf{\Psi}\mathbf{V}^\top) - \text{Tr}(\mathbf{\Upsilon}\mathbf{T}^\top).\end{aligned}$$

1) **Computation of $\mathbf{U}$**. For a third-order tensor $\mathcal{X}$ of size $m \times n \times h$, the mode-1, mode-2, and mode-3 matricizations turn $\mathcal{X}$ into matrices $\mathbf{X}_{(1)}$ of size $m\times(nh)$, $\mathbf{X}_{(2)}$ of size $n\times(mh)$, and $\mathbf{X}_{(3)}$ of size $h\times(mn)$, by mapping tensor element $\mathcal{X}(i_1,i_2,i_3)$ to elements (i_1,j_1), (i_2,j_2), and (i_3,j_3) in $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}$, and $\mathbf{X}_{(3)}$ respectively, where

$$j_1 = 1 + (i_3 - 1)n, j_2 = 1 + (i_3 - 1)m, j_3 = 1 + (i_2 - 1)m.$$

When $\mathbf{V}$ and $\mathbf{T}$ are fixed, we have

$$\min_{\mathbf{U}\geq 0} \|\mathcal{X} - [\![\mathbf{U},\mathbf{V},\mathbf{T}]\!]\|^2 = \min_{\mathbf{U}\geq 0} \|\mathbf{X}_{(1)} - \mathbf{U}(\mathbf{T}\odot\mathbf{V})^\top\|^2,$$

where $\odot$ represents Khatri-Rao product. Then, the derivative $\frac{\partial \mathbb{J}}{\partial \mathbf{U}}$ is computed as

$$\begin{aligned}\frac{\partial \mathbb{J}}{\partial \mathbf{U}} =& -2\mathbf{X}_{(1)}(\mathbf{T}\odot\mathbf{V}) + 2\mathbf{U}(\mathbf{T}\odot\mathbf{V})^\top(\mathbf{T}\odot\mathbf{V}) \\ &+ 2\alpha\mathbf{L}\mathbf{U} + 2\gamma\mathbf{U} - \mathbf{\Phi}.\end{aligned}$$

Based on the properties of Khatri-Rao product [25], we have $(\mathbf{T}\odot\mathbf{V})^\top(\mathbf{T}\odot\mathbf{V}) = \mathbf{T}^\top\mathbf{T} * \mathbf{V}^\top\mathbf{V}$, where $*$ represents Hamamard product. By setting the derivative $\frac{\partial \mathbb{J}}{\partial \mathbf{U}} = 0$, we then have

$$\mathbf{\Phi} = -2\mathbf{X}_{(1)}(\mathbf{T}\odot\mathbf{V}) + 2\mathbf{U}(\mathbf{T}^\top\mathbf{T} * \mathbf{V}^\top\mathbf{V}) + 2\alpha\mathbf{L}\mathbf{U} + 2\gamma\mathbf{U}.$$

Based on the Karush-Kuhn-Tucker complementary condition [2] for the nonnegativity constraint of $\mathbf{U}$, we obtain

$$\mathbf{\Phi}(i,j)\mathbf{U}(i,j) = 0, \quad 1 \leq i \leq m, \quad 1 \leq j \leq z.$$

Thus, we have

$$[-\mathbf{X}_{(1)}(\mathbf{T}\odot\mathbf{V}) + \mathbf{U}(\mathbf{T}^\top\mathbf{T} * \mathbf{V}^\top\mathbf{V}) + \alpha\mathbf{L}\mathbf{U} + \gamma\mathbf{U}](i,j)\mathbf{U}(i,j) = 0.$$

The Laplacian matrix $\mathbf{L}$ may take any signs, so we decompose $\mathbf{L}$ into the positive part $\mathbf{L}^+$ and the negative part $\mathbf{L}^-$, i.e., $\mathbf{L} = \mathbf{L}^+ - \mathbf{L}^-$. For any $1 \leq i,j \leq m$, $\mathbf{L}^+(i,j)$ equals $\mathbf{L}(i,j)$ when $\mathbf{L}(i,j) > 0$, and 0 otherwise; $\mathbf{L}^-(i,j)$ equals $-\mathbf{L}(i,j)$ when $\mathbf{L}(i,j) < 0$, and 0 otherwise. By the definition of $\mathbf{L} = \mathbf{D} - \mathbf{A}$, we easily obtain $\mathbf{L}^+ = \mathbf{D}$ and $\mathbf{L}^- = \mathbf{A}$. Similar to [7], we have the following multiplicative updating rule of $\mathbf{U}$:

$$\mathbf{U}(i,j) \leftarrow \mathbf{U}(i,j)\sqrt{\frac{[\mathbf{X}_{(1)}(\mathbf{T}\odot\mathbf{V}) + \alpha A\mathbf{U}](i,j)}{[\mathbf{U}(\mathbf{T}^\top\mathbf{T} * \mathbf{V}^\top\mathbf{V}) + \alpha D\mathbf{U} + \gamma\mathbf{U}](i,j)}}.$$

2) **Computation of $\mathbf{V}$ and $\mathbf{T}$**. Similarly, we have the following multiplicative updating rules of $\mathbf{V}$ and $\mathbf{T}$:

$$\mathbf{V}(i,j) \leftarrow \mathbf{V}(i,j)\sqrt{\frac{[\mathbf{X}_{(2)}(\mathbf{T}\odot\mathbf{U})](i,j)}{[\mathbf{V}(\mathbf{T}^\top\mathbf{T} * \mathbf{U}^\top\mathbf{U}) + \gamma\mathbf{V}](i,j)}},$$

$$\mathbf{T}(i,j) \leftarrow \mathbf{T}(i,j)\sqrt{\frac{[\mathbf{X}_{(3)}(\mathbf{V}\odot\mathbf{U})](i,j)}{[\mathbf{T}(\mathbf{V}^\top\mathbf{V} * \mathbf{U}^\top\mathbf{U}) + \gamma\mathbf{T}](i,j)}}.$$

We can easily find that the above three multiplicative update rules maintain the nonnegativity of $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$ for initial nonnegative matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$. Similar to [7], we observe: (I) $\mathbf{U}(i,j)$ increases when $[\mathbf{X}_{(1)}(\mathbf{T}\odot\mathbf{V}) + \alpha A\mathbf{U}](i,j) > [\mathbf{U}(\mathbf{T}^\top\mathbf{T} * \mathbf{V}^\top\mathbf{V}) + \alpha D\mathbf{U} + \gamma\mathbf{U}](i,j)$, i.e., $\frac{\partial \mathbb{O}}{\partial \mathbf{U}}(i,j) < 0$, and decreases otherwise; (II) $\mathbf{V}(i,j)$ increases when $\frac{\partial \mathbb{O}}{\partial \mathbf{V}}(i,j) < 0$, and decreases otherwise; (III) $\mathbf{T}(i,j)$ increases when $\frac{\partial \mathbb{O}}{\partial \mathbf{T}}(i,j) < 0$, and decreases otherwise. Therefore, there exist two kinds of stationary points in the iterative use of the multiplicative updating rules of $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$: One satisfies $\frac{\partial \mathbb{O}}{\partial \mathbf{U}} = 0$, $\frac{\partial \mathbb{O}}{\partial \mathbf{V}} = 0$, and $\frac{\partial \mathbb{O}}{\partial \mathbf{T}} = 0$, which are the stationary points of the objective function $\mathbb{O}$; The other is $\mathbf{U}(i,j) \to 0$, $\mathbf{V}(i,j) \to 0$, and $\mathbf{T}(i,j) \to 0$, which yields sparsity in $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$ respectively. Formally, the correctness and convergence of the multiplicative updating rules of $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{T}$ can be proven with the standard auxiliary function approach [7,12,15]. The computational complexity of Dinfer can be shown to be $O(mnhr)$ per iteration. One can use distributed computing systems such as GraphLab [19] to accelerate the three multiplicative updating rules' matrix multiplication operations. At last, we learn a supervised model (e.g., SVM or LR) based on labeled mobile device users and their latent features learned, and use this model to predict unlabeled mobile device users' attributes.

5. SINFER: LEARN SOCIAL NETWORKS

In this section, we introduce a Pointwise Mutual Information (PMI) based method Sinfer for learning social networks of mobile devices. PMI has been widely used for measuring the semantic similarity between words. For example, word2vector [20], an effective tool for embedding words into a low-dimensional space and evaluating the similarity between words, is known to be equivalent to factorizing a word-word PMI matrix [16]. Inspired by this, we define the following three new PMI metrics to evaluate the closeness between mobile devices from their co-coming, co-leaving, and co-presenting events:

PMI of Co-Coming Events. To formally describe our method, we first introduce some notations. Let $f_v^{(c)}(u)$ be the number of times a device u logins to AP v, and $f_v^{(c)}(u_i,u_j)$ be the number of times devices u_i and u_j co-come at AP v, The probability that a randomly picked co-coming event at AP v belongs to device u is $p_v^{(c)}(u) = \frac{f_v^{(c)}(u)}{\sum_{x\in U} f_v^{(c)}(x)}$, and the probability that a randomly picked co-coming event at AP v belongs to the pair of devices u_i and u_j is $p_v^{(c)}(u_i,u_j) = \frac{2f_v^{(c)}(u_i,u_j)}{\sum_{x,y\in U} f_v^{(c)}(x,y)}$. To evaluate the closeness between devices u_i and u_j, we compute the PMI of their co-coming events at AP v as $\text{pmi}_v^{(c)}(u_i,u_j) = \log \frac{p_v^{(c)}(u_i,u_j)}{p_v^{(c)}(u_i)p_v^{(c)}(u_j)}$. The value of $\text{pmi}_v^{(c)}(u_i,u_j)$ reflects a co-coming event of u_i and u_j at AP v happens by chance or it is a social event. For example, when u_i and u_j both frequently appear at place v but have only one co-coming event at place v, then $\text{pmi}_v^{(c)}(u_i,u_j)$ is small and it indicates that a co-coming event of them at place v is probably a coincidence. In contrary, when u_i and u_j seldom appear at place v but have a number of co-coming event at place v, then $\text{pmi}_v^{(c)}(u_i,u_j)$ is large and

it indicates that a co-coming event of them at place v is probably a social event.

PMI of Co-Leaving Events. Similarly, let $f_v^{(\mathrm{l})}(u_i, u_j)$ be the number of times devices u_i and u_j co-leave at AP v, and $f_v^{(\mathrm{l})}(u)$ be the number of times a device u logouts of AP v. Denote by $p_v^{(\mathrm{l})}(u_i, u_j) = \frac{2f_v^{(\mathrm{l})}(u_i,u_j)}{\sum_{x,y\in U} f_v^{(\mathrm{l})}(x,y)}$ and $p_v^{(\mathrm{l})}(u) = \frac{f_v^{(\mathrm{l})}(u)}{\sum_{x\in U} f_v^{(\mathrm{l})}(x)}$. We define the PMI of co-leaving events of devices u_i and u_j at AP v as $\mathrm{pmi}_v^{(\mathrm{l})}(u_i, u_j) = \log \frac{p_v^{(\mathrm{l})}(u_i,u_j)}{p_v^{(\mathrm{l})}(u_i)p_v^{(\mathrm{l})}(u_j)}$.

PMI of Co-Presenting Events. Let $t_v^{(\mathrm{p})}(u_i, u_j)$ and $f_v^{(\mathrm{p})}(u_i, u_j)$ be the total duration and the number of times two devices u_i and u_j co-present at AP v, and $t_v^{(\mathrm{p})}(u)$ be the total duration of a device u presents at AP v. At a randomly picked time, the probability that device u presents at AP v is $p_v^{(\mathrm{p})}(u) = \frac{t_v^{(\mathrm{p})}(u)}{\sum_{x\in U} t_v^{(\mathrm{p})}(x)}$, and the probability that devices u_i and u_j co-present at AP v is $p_v^{(\mathrm{p})}(u_i, u_j) = \frac{2t_v^{(\mathrm{p})}(u_i,u_j)}{\sum_{x,y\in U} t_v^{(\mathrm{p})}(x,y)}$. We define the PMI of co-presenting events of devices u_i and u_j at AP v as $\mathrm{pmi}_v^{(\mathrm{p})}(u_i, u_j) = \log \frac{p_v^{(\mathrm{p})}(u_i,u_j)}{p_v^{(\mathrm{p})}(u_i)p_v^{(\mathrm{p})}(u_j)}$.

Based on the above PMI metrics, we use all three kinds of co-coming, co-leaving, and co-presenting events, and compute the closeness between mobile devices u_i and u_j as

$$\begin{aligned} closeness(u_i, u_j) = & \sum_{f_v^{(\mathrm{c})}(u_i,u_j)>0, v\in V} f_v^{(\mathrm{c})}(u_i, u_j)\mathrm{pmi}_v^{(\mathrm{c})}(u_i, u_j) \\ & + \sum_{f_v^{(\mathrm{l})}(u_i,u_j)>0, v\in V} f_v^{(\mathrm{l})}(u_i, u_j)\mathrm{pmi}_v^{(\mathrm{l})}(u_i, u_j) \\ & + \sum_{t_v^{(\mathrm{p})}(u_i,u_j)>0, v\in V} f_v^{(\mathrm{p})}(u_i, u_j)\mathrm{pmi}_v^{(\mathrm{p})}(u_i, u_j). \end{aligned}$$

6. EXPERIMENTS

6.1 Experimental Settings

Data and Evaluation. We use the dataset of AP-trajectories of mobile devices on Campuses A and B introduced in Section 2 to infer mobile device users' close friends and demographic attributes including gender (i.e., male/female) and social roles (i.e., undergraduate/graduate/faculty). To infer mobile device users' social networks and demographic attributes effectively and accurately, we only consider *active* mobile devices who have at least fifty AP login records in the 74 days we collected. Campuses A and B have 21,207 and 13,353 active mobile devices respectively. For learning mobile device users' demographic attributes, similar to [9], we consider each class in genders or social roles is as important as each other, and use *weighted Precision*, *Recall*, and *macro-F1* (in short, *wPrecision*, *wRecall*, and *wF1*) as evaluation metrics. We repeat the experiment 100 times and report the average performance in terms of wPrecision, wRecall, and wF1.

Other State-of-The-Art Algorithms. For inferring mobile device users' social networks, we compare our method Sinfer with the state-of-the-art method:

• **EBM [24]** is an entropy-based method for learning social strength between users from check-in records on location-based OSNs. The design of EBM is mainly inspired by two observations: 1) people are more likely be friends when their co-occurrences happened at diverse places. 2) co-occurrences at popular places happen more likely by chance than those at private places.

For predicting mobile device users' demographics, we compare the performance of our method Dinfer with the state-of-the-art methods S and ST [30], which are described as follows:

• **Spatiality-based method (S)** only considers spatial information, i.e., a mobile device's features consist of the log of logins at different APs. Similar to Dinfer, S learns supervised classifiers using extracted AP-trajectory features to predict unknown mobile device users' demographic attributes.

• **Spatiality and Temporality-based method (ST)** goes one step further than S, i.e., it concatenates both spatial and temporal features for prediction, where a mobile device's temporal features consist of the log of logins during different hours in the week.

Moreover, we also compare our method Dinfer with the following two straightforward methods that learn mobile device representations $\mathbf{U}$ from the tensor of spatiotemporal AP-trajectories $\mathcal{X}$ in a direct manner:

• **NTF** only considers $\mathcal{X}$ but not social networks of mobile device users in comparison with Dinfer, i.e., NTF learns $\mathbf{U}$ by optimizing $\mathbb{O}_{\mathrm{TRA}}$ in eq. (1).

• **NMF** learns $\mathbf{U}$ by applying the non-negative matrix factorization method [3,14] to the mode-1 matricization $X_{(1)}$ of $\mathcal{X}$.

By default, in our experiments we set the training/test ratio as 9:1, the number of latent dimensions as 512 (i.e., $r = 512$) for methods Dinfer, NTF, and NMF, and set the social network factor of Dinfer as $\alpha = 2$.

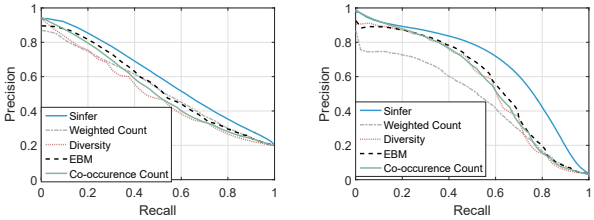
6.2 Accuracy of Learning Social Networks

Results of friendship inference. Table 3 shows the distribution of Top-10 and Top-20 closeness friends learned by our method Sinfer. Among Top-10 and Top-20 closeness friends, 16%-27%, 16%-18%, and 9%-16% are ID_sharing_friends for undergraduate students, graduate students, and faculties respectively. Among Top-10 (Top-20) closeness friends of undergraduate students on Campus A, on average, 76.40% (76.18%) are undergraduate students who enrolled the same department in the same year, 61.47% (58.93%) are their classmates, 5.20% (5.44%) are faculties in the same department, and 4.40% (4.70%) are undergraduate students who enrolled different departments in the same year. Among Top-10 (Top-20) closeness friends of undergraduate students on Campus B, on average, 89.23% (87.96%) are also undergraduate students who enrolled the same department in the same year, 83.87% (80.28%) are their classmates, and 3.38% (3.87%) are undergraduate students who enrolled different departments in the same year. Similarly, among Top-10 and Top-20 closeness friends of graduate students, 73%-77% come from the same department, 16%-21%, 34%-40%, and 16%-22% are undergraduate students, graduate students, and faculties in the same department respectively. Among Top-10 and Top-20 closeness friends of faculties, 73%-86% are affiliated with the same department. From the above results, we observe a strong homophily appeared in close friends learned by Sinfer. These results are consistent to common sense: e.g., classmates and undergraduate students in the same department tend to be close friends. Unlike Campus A, we notice that most learned top closeness friends of faculties on Campus B are not faculties but undergraduate students in the same department. We further investigate this phenomenon and find that faculties on Campus B spend much more time on teaching and co-present with undergraduates more frequently than faculties on Campus A.

Sinfer vs. Prior Art. We conduct experiments to evaluate the performance of our method Sinfer in comparison with the state-of-the-art method EBM [24]. The EBM uses Renyi entropy to evaluate the diversity of two users' co-occurrences at different places. Similarly, it uses Shannon entropy to evaluate the popularity of a place, and then uses this location entropy to reweigh co-occurrence counts at the place. EBM is a linear combination of the co-occurrence diversity (in short, Diversity) and the weighted co-occurrence count

Table 3: Distribution of Top-10 and Top-20 closeness friends learned by Sinfer. The number in the parentheses is the fraction of friends enrolled at school at the same year.

social role	Campus	Top-K	distribution of learned friends (%)							
			ID_sharing_ friends	same department				different department		
				undergraduate		graduate	faculty	undergraduate	graduate	faculty
				classmate	not classmate					
under-graduate	A	Top-10	25.41	**61.47**	16.67 (14.93)	2.27	5.20	10.40 (4.40)	1.07	2.93
		Top-20	27.48	**58.93**	19.32 (17.25)	2.13	5.44	10.47 (4.70)	1.10	2.61
	B	Top-10	15.98	**83.87**	6.43 (5.36)	1.74	0.21	7.45 (3.38)	0.57	0.94
		Top-20	21.28	**80.28**	8.85 (7.68)	2.84	0.18	8.45 (3.87)	0.58	0.90
graduate	A	Top-10	16.54	16.47 (2.94)		**34.71** (12.94)	22.65	13.82 (0.59)	7.06 (4.12)	5.29
		Top-20	16.87	16.71 (3.19)		**34.15** (12.95)	22.34	14.79 (0.25)	7.60 (4.09)	4.41
	B	Top-10	17.43	21.63 (7.86)		**40.15** (28.53)	17.71	14.99 (1.09)	1.83 (1.24)	3.69
		Top-20	18.25	21.54 (7.78)		**38.59** (28.04)	16.83	17.33 (1.19)	1.94 (1.13)	3.78
faculty	A	Top-10	14.53	12.41		4.22	**56.51**	12.96	4.69	9.22
		Top-20	15.70	12.61		4.48	**55.79**	13.04	4.65	9.45
	B	Top-10	9.65	61.22		3.46	21.14	9.59	0.76	3.82
		Top-20	10.91	60.37		3.09	21.16	10.48	0.97	3.93



(a) Campus A (b) Campus B

Figure 4: Precision-Recall curve comparison of our friendship inference method Sinfer, state-of-the-art method EBM, and methods using Diversity, Weighted Count, and Co-occurrence Count as metrics to evaluate the closeness between mobile device users.

(in short, Weighted Count). We set its parameters as used in the original paper. Clearly, it is not easy to obtain the ground truth of top closeness friends especially for graduate students and faculties in our dataset. In our experiment, we focus on undergraduate students, and use their classmates and ID_sharing_friends as the ground truth of their close friends. Fig. 4 shows the precision-recall curves of Sinfer for learning the closeness friends in comparison with EBM and methods using Diversity, Weighted Count, and Co-occurrence Count as metrics to evaluate the closeness between mobile devices in a direct manner. We can see that our method Sinfer significantly outperforms other methods especially for Campus B. We notice that the precision ratio cannot reach 100% in precision-recall curves. This is because in practice, undergraduates' top closeness friends may not be their ID_sharing_friends or classmates, which are wrongly recognized as false friends by our experimental settings.

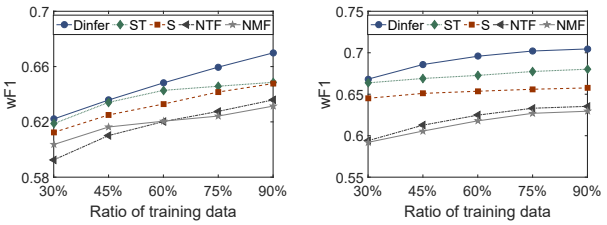
6.3 Accuracy of Learning Demographics

Predictive performance. Tables 4 shows the performance of our method Dinfer in comparison with state-of-the-art methods for learning mobile device users' genders and social roles. Overall, Dinfer outperforms the other methods in terms of wPrecision, wRecall, and wF1. For gender inference, the wF1 of Dinfer is 0.69 and 0.72 for Campuses A and B respectively, and is up to 5% higher than the other methods. For social role inference, the wF1 of Dinfer is 0.63 and 0.61 for Campuses A and B respectively, and is up to 10% higher than the other methods. Compared to NTF, Dinfer uses social networks of mobile device users to learn more effective latent representations from AP-trajectories, and raises the wF1 by 6%-11%.

Training/test ratio. We study how the performance of demographic inference methods changes with the increase of the training/test ratio. As shown in Fig. 5, the performance of Dinfer and other methods improves as the percentage of labeled mobile device users in training set increases. Moreover, we can see that Dinfer significantly outperforms the other methods for gender inference when more than 50% data are used for training. We omit similar results for social role inference.

Classifier sensitivity. We study the performance of Dinfer with different classifiers. Fig. 6 presents the wF1 of three popular classifiers SVM (linear kernel), logistic regression (LR), and linear regression (LSR) for gender inference. We observe that LR and SVM slightly outperform LSR when using Dinfer and other methods for learning mobile device users' latent representations. We omit similar results for social role inference.

Parameter sensitivity. Next, we discuss the parameter sensitivity of Dinfer. Fig. 7 shows the impact of different choices of the number of latent dimensions r and the social network factor α on gender inference. Fig. 7(a) shows how the performance of Dinfer changes with the increase of r. In general, large r provides a better performance. However, the rate of the performance increase becomes slower as we enlarge $r \geq 512$. Considering the increas-

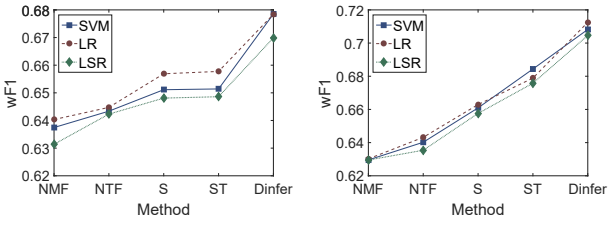


(a) Campus A (b) Campus B

Figure 5: (Gender inference) Performance w.r.t. ratio of training data for Dinfer and state-of-the-art methods.

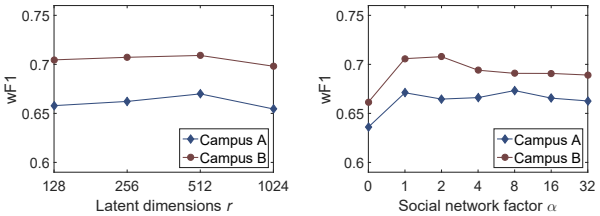
Table 4: Performance of learning demographics of mobile devices on Campus A | B.

	gender inference			social role inference		
	wPrecision	wRecall	wF1	wPrecision	wRecall	wF1
S	.645 | .655	.638 | .665	.641 | .660	.639 | .589	.644 | .549	.605 | .515
ST	.645 | .667	.644 | .688	.644 | .676	.641 | .598	.645 | .552	.612 | .517
NMF	.626 | .621	.625 | .653	.626 | .630	.612 | .531	.609 | .552	.565 | .471
NTF	.627 | .635	.628 | .636	.628 | .635	.627 | .603	.617 | .551	.578 | .502
Dinfer	**.694 | .712**	**.690 | .722**	**.692 | .717**	**.654 | .621**	**.661 | .600**	**.633 | .614**



(a) Campus A (b) Campus B

Figure 6: (Gender inference) wF1 of different classifiers combined with different methods of learning mobile device users' latent representations.



(a) Campus A (b) Campus B

Figure 7: (Gender inference) Parameter sensitivities w.r.t. the number of latent dimensions r and the social network factor α.

ing computational complexity as r increases, we suggest $r = 512$ as a compromise. Fig. 7(b) shows how the performance of Dinfer changes with the increase of α. We can see that the wF1 is low when $\alpha = 0$, i.e., when social networks of mobile device users are not used. When $\alpha > 0$, the performance significantly increases, and fluctuates in a small range (less than 1.5%) for $\alpha = 1, \ldots, 32$. We omit similar results for social role inference.

7. RELATED WORK

Our problem is closely related to research on learning user demographics and social networks from spatiotemporal trajectories. More specially,

Learning user demographics. Given demographic attributes (e.g., gender and affiliation) for a fraction of users on an OSN, Mislove et al. [21] proposed a method to infer the attributes of the remaining users when the OSN has significant homophily, which refers to the tendency of people to connect to others with common attributes. Similarly, [8,9] developed methods to learn users' attributes from mobile social networks, which are generated based on call and SMS records. Zhong et al. [30] observed that users' demographics can also be inferred from their check-in POIs on location-based OSNs. However, their method relies on POIs' rich semantic features such as categories, user reviews, and descriptions, which are not available for the problem in this paper.

Learning users' social networks. There has been a considerable effort to predict social relationships from human movement records collected from GPS-enabled devices [6,10,17] and location-based OSNs [4,5,24,26,29]. Li et al. [17] proposed a framework to mine user similarities based on their spatial trajectories, which is useful for friend and interest recommendation. Similarly, Zhang and Pang [29] proposed a method to predict friendships based on computing distances between users' frequently movement areas. Compared to these coarse-grained metrics, Eagle et al. [10] observed that a fine-grained metric *co-occurrence* more strongly indicates friendships, especially when it happens at non-working time and locations. Crandall et al. [5] developed a probabilistic model of spatiotemporal co-occurrences to evaluate whether a co-occurrence happens by chance or it is a social event. However, their method neglects the impact of the locations of co-occurrences on friendship prediction. In reality, co-occurrences between strangers are more likely to happen in crowded public places than in small private places. Thus, the probability of friendships is strongly related with co-occurrence places. To utilize this knowledge, Cranshaw et al. [6] defined a metric *location entropy* to characterize the location popularity. Furthermore, in addition to the friendship prediction, Pham et al. [24] proposed another entropy-based model to estimate the strength of friendships from co-occurrence events. In addition to the location popularity, Wang et al. [26] proposed a model also considering other factors such as users' personal preferences to each place and time gaps between their co-occurrences. Besides co-occurrences, Cheng et al. [4] observed that the time interval between two users visiting the same place (not necessary co-occurrences) is also useful for predicting their relationship. Jayarajah et al. [13] utilize AP-trajectories to estimate the strength of social events, but they neglect to solve the events happened by chance. The above methods studied only one kind of co-occurrence events, i.e., co-coming events. To the best of our knowledge, we are the first to further study fine-grained co-occurrence events: co-coming, co-leaving, and co-presenting duration. We observe these three kinds of events are complementary to each other for predicting social networks of mobile devices. and develop a method to predict friendships by utilizing all co-occurrence events.

8. CONCLUSIONS

In this paper we investigate whether one can learn mobile device users' demographic attributes and social networks from their spatiotemporal AP-trajectories. We propose a method Sinfer to learn friendships between mobile device users by exploring fine-grained co-occurrence events of AP-trajectories, such as co-coming, co-leaving, and co-presenting events. Moreover, we develop a tensor factorization based learning method Dinfer to infer mobile device users' attributes from their AP-trajectories by leveraging user social networks learned by Sinfer. Experimental results on thousands of

mobile device users demonstrate the effectiveness of our methods Sinfer and Dinfer.

9. ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (61603290, 61103240, 61103241, 61221063, 91118005, 61221063, U1301254), Ministry of Education & China Mobile Joint Research Fund Program (MCM20150506, MCM20160311), Shenzhen Basic Research Grant (JCYJ20160229195940462), 863 High Tech Development Plan (2012AA011003), 111 International Collaboration Program of China, and Application Foundation Research Program of SuZhou (SYG201311).

10. REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. In *NIPS'01*, pages 601–608, 2001.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] D. Cai, X. He, J. Han, and T. S. Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8):1548–1560, 2011.

[4] R. Cheng, J. Pang, and Y. Zhang. Inferring friendship from check-in data of location-based social networks. In *ASONAM'15*, pages 1284–1291, 2015.

[5] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *PNAS*, 107(52):22436–22441, 2010.

[6] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh. Bridging the gap between physical location and online social networks. In *UbiComp'10*, pages 119–128, 2010.

[7] C. H. Q. Ding, T. Li, and M. I. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(1):45–55, 2010.

[8] Y. Dong, F. Pinelli, Y. Gkoufas, Z. Nabi, F. Calabrese, and N. V. Chawla. Inferring unusual crowd events from mobile phone call detail records. In *ECML-PKDD'15*, pages 474–492, 2015.

[9] Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla. Inferring user demographics and social strategies in mobile social networks. In *KDD'14*, pages 15–24, 2014.

[10] N. Eagle, A. S. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *PNAS*, 106(36):15274–15278, 2009.

[11] P. Eckert. Gender and sociolinguistic variation. *Readings in Language and Gender*, 1997.

[12] Q. Gu, J. Zhou, and C. H. Q. Ding. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs. In *SDM'10*, pages 199–210, 2010.

[13] K. Jayarajah, A. Misra, X.-W. Ruan, and E.-P. Lim. Event detection: Exploiting socio-physical interactions in physical spaces. In *ASONAM'15*, pages 508–513, 2015.

[14] D. Lee and H. Seung. Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401:788–791, 1999.

[15] D. D. Lee and H. S. Seung. Algorithms for Non-negative Matrix Factorization. In *NIPS'01*, pages 556–562, 2001.

[16] O. Levy and Y. Goldberg. Neural word embedding as implicit matrix factorization. In *NIPS'14*, pages 2177–2185, 2014.

[17] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma. Mining user similarity based on location history. In *SIGSPATIAL GIS'08*, page 34, 2008.

[18] J.-Y. Liu and Y.-H. Yang. Inferring personal traits from music listening history. In *MIRUM'12*, pages 31–36, 2012.

[19] Y. Low, J. Gonzalez, A. Kyrola, D. Bickson, C. Guestrin, and J. M. Hellerstein. Distributed graphlab: A framework for machine learning in the cloud. *PVLDB'12*, 5(8):716–727, 2012.

[20] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *NIPS'13*, pages 3111–3119, 2013.

[21] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. You are who you know: inferring user profiles in online social networks. In *WSDM'10*, pages 251–260, 2010.

[22] D. Murray and K. Durrel. Inferring Demographic Attributes of Anonymous Internet Users. In *WEBKDD'99*, 1999.

[23] E. Oja. Principal components, minor components, and linear neural networks. *Neural Networks*, 5(6):927–935, 1992.

[24] H. Pham, C. Shahabi, and Y. Liu. Ebm: an entropy-based model to infer social strength from spatiotemporal data. In *SIGMOD'13*, pages 265–276, 2013.

[25] A. Smilde, R. Bro, and P. Geladi. *Multi-way analysis with applications in the chemical sciences*. Wiley, 2004.

[26] H. Wang, Z. Li, and W.-C. Lee. Pgt: Measuring mobility relationship using personal, global and temporal factors. In *ICDM'14*, pages 570–579, 2014.

[27] P. Wang, J. Guo, Y. Lan, J. Xu, and X. Cheng. Your cart tells you: Inferring demographic attributes from purchase data. In *WSDM'16*, 2016.

[28] Z. Yang and J. Laaksonen. Multiplicative updates for non-negative projections. *Neurocomput.*, 71(1-3):363–373, 2007.

[29] Y. Zhang and J. Pang. Distance and friendship: A distance-based model for link prediction in social networks. In *APWeb'15*, pages 55–66, 2015.

[30] Y. Zhong, N. J. Yuan, W. Zhong, F. Zhang, and X. Xie. You are where you go: Inferring demographic attributes from location check-ins. In *WSDM'15*, pages 295–304, 2015.