Information Sciences 463–464 (2018) 110–128

Contents lists available at ScienceDirect

Information Sciences

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

Predicting attributes and friends of mobile users from AP-Trajectories☆

Pinghui Wang[a,b], Feiyang Sun[a,1,*], Di Wang[a], Jing Tao[a,c,*], Xiaohong Guan[b,a,d,*], Albert Bifet[e]

[a] NSKEYLAB, Xi'an Jiaotong University, China
[b] Shenzhen Research School, Xi'an Jiaotong University, China
[c] Zhejiang Research Institute, Xi'an Jiaotong University, China
[d] Department of Automation and NLIST Lab, Tsinghua University, China
[e] LTCI, Télécom ParisTech, Université Paris-Saclay, France

ARTICLE INFO

Article history:
Received 3 July 2017
Revised 11 June 2018
Accepted 12 June 2018
Available online 14 June 2018

Keywords:
User profiling
Spatiotemporal trajectories
Social network

ABSTRACT

Exploring the demographic attributes and social networks of Internet users is widely employed by many applications, such as recommendation systems. The popularity of mobile devices (notably smartphones) and location-based Internet services (e.g., Google Maps) facilitates the collection of users' locations over time. There have been recent efforts to predict users' attributes (e.g., age and gender) from this data, and location-based social networks such as Foursquare and Gowalla are based on using the rich location context knowledge of points of interest (e.g., the name, type and description of restaurants and hotels) where users check-in online. However, little attention has been paid to inferring the attributes and social networks of mobile device users based on their spatiotemporal trajectories where there is little or 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' attributes from their AP-trajectories by leveraging prior knowledge. Compared with our previous work, which only considered users' social networks, Dinfer further utilizes AP spatial information and achieves a 2% improvement. 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-world datasets demonstrate the effectiveness and efficiency of our methods.

☆ An earlier conference version appeared as a WWW'17 companion paper [27]. In this journal version, we extend the proposed Dinfer method to incorporate AP spatial closeness information to improve its accuracy. Moreover, we conduct more experiments to validate our proposed methods, include user department inference, grade inference, and class inference, and compare Dinfer with more classic methods, e.g., PCA and HOSVD.

* Corresponding authors.
E-mail addresses: fysun@sei.xjtu.edu.cn (F. Sun), xhguan@sei.xjtu.edu.cn (X. Guan).

[1] P. Wang and F. Sun contributed equally to this work.

https://doi.org/10.1016/j.ins.2018.06.029
0020-0255/

1. Introduction

In today's era of digital big data, information on human activities has been collected and monitored in a variety of domains. For example, people's webpage browsing, shopping, video viewing and music listening behaviors on the Internet can easily be collected by providers of web services such as search engines and e-commerce, video and music websites. This data helps service providers to understand/profile their users, which is crucial for many applications such as targeted advertising and business recommendations, as well as for counter-criminal/terrorist strategies. Utilization of this data has revealed that service providers can infer their users' demographic attributes, including age and gender, even though users do not intend to reveal these attributes. The information is derived by the service providers from users' Internet browsing history [21], linguistics writing [11], mobile call/message records [8,9], music listening history [18], purchase data [26], and available demographic attributes of their friends on online social networks (OSNs) such as Facebook and LinkedIn [20].

Recently, mining people's locations over time has attracted a lot of attention due to the popularity of mobile devices along with location-based Internet services that include navigation applications (e.g., Google Maps and Uber) and location-based online social networks (location-based OSNs, e.g., Instagram and Foursquare). Smartphones, smartwatches and other mobile devices can be precisely located outside or inside buildings by current and near-future positioning techniques, which precisely infer the locations of these devices using multiple information sources that include built-in sensors (e.g., GPS and barometers) and the location and signal strength of connected wireless access points (APs) and cell-phone towers. These techniques lead to mobile device users' spatiotemporal trajectories being explicitly or implicitly exposed to third parties. Explicit exposure occurs when people enable map services to retrieve their GPS locations automatically, distribute geo-tagged posts (e.g., tweets and photos) or check-in at points of interest (POIs) on OSNs. Implicit exposure of users' trajectories can take place 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 POI recommendations 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] derives users' attributes from their check-in POIs on location-based OSNs such as Facebook, Foursquare and 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, the daily activities of most college students and faculty staff may be concentrated 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 or 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 or place on campus may not be publicly available on the Internet. In addition to demographic attributes, we also studied the problem of predicting who mobile device users' close friends are from their spatiotemporal trajectories. Existing methods [4–6,23,25,29] 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 with AP spatial information instead of location context knowledge. Compared with our previous work [27], introducing locations' closeness makes the model more accurate. We studied three types of fine-grained co-occurrence events: co-coming, co-leaving, and co-presenting duration, and observed 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 conducted an in-depth measurement study on spatiotemporal trajectories of 52,000 mobile devices on two campuses, which were revealed by their network connections to wireless APs. To facilitate further research in this field, we made these trajectories publicly available[2].
- We developed a tensor factorization-based learning method *Dinfer* to infer mobile device users' attributes from their AP-trajectories by leveraging AP spatial information and user social networks that are learned from users' spatiotemporal trajectories.
- We used pointwise mutual information (PMI) to evaluate whether a fine-grained co-occurrence happens by chance or 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, consisting of AP-trajectories of mobile devices on campus and the ground truth of mobile device users' social networks and demographic attributes. Section 3 presents our measurement study of micro- and macro-statistics of mobile devices' AP-trajectories. Section 4 formulates the problem and the overview of our methods. Sections 5 and 6 present our methods Dinfer and Sinfer for learning demographic attributes and social networks of mobile device users from their spatiotemporal AP-trajectories respectively.

[2] http://nskeylab.xjtu.edu.cn/dataset/phwang/data/APtraj.zip.

Table 1
Statistics of persons on Campuses A and B.

Attributes	Campus A		Campus B	
	#devices	#NetIDs	#devices	#NetIDs
Male	9364	4797	5095	3,212
Female	20,254	12,038	17,190	10,307
Undergraduate	21,136	12,982	18,663	11,328
Graduate	3630	2409	1237	987
Faculty	4852	1444	2385	1,214
Total	29,618	16,835	22,285	13,529

Section 7 presents the performance evaluation and testing results. Section 8 summarizes related work. Concluding remarks then follow.

2. Datasets

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

2.1. Spatiotemporal Data: AP-Trajectories

From May 03 to July 15, 2015, we polled all APs in Campuses A and B regularly (every 5 minutes) and collected information (e.g., MAC addresses) on their connected mobile devices via SNMP (Simple Network Management Protocol). Based on this dataset, we could determine whether a mobile device was 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 1921 and 362 APs respectively. For both campuses, each AP covers an area of about 100 m^2 in the buildings. During the 74 days, there were 29,618 and 22,285 active mobile devices on the Internet for Campuses A and B respectively. On average, a mobile device was active for 21 days in Campus A and 20 days in Campus B, and connected to 54 APs in Campus A and 38 APs in Campus B.

2.2. Ground Truth I: Device-NetID Networks

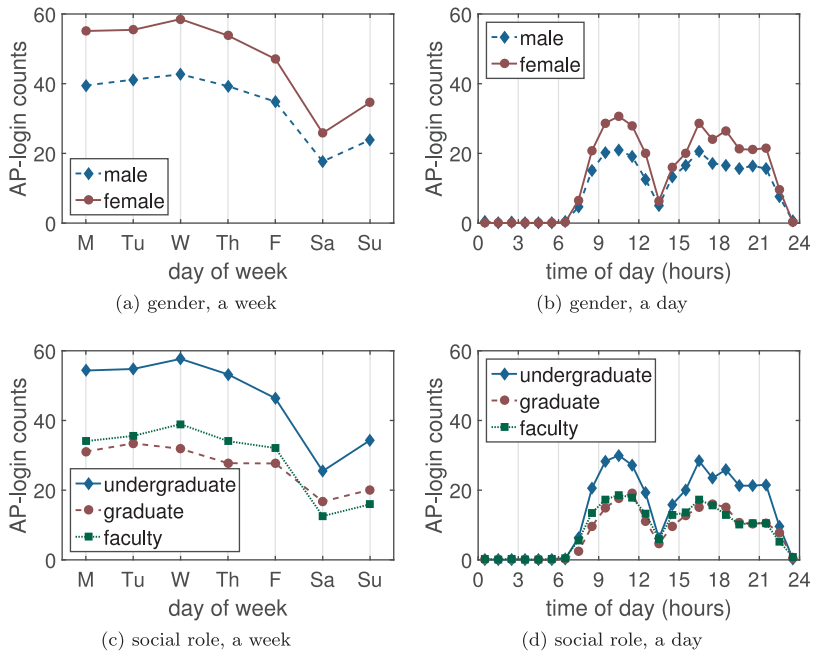
In our dataset, we observed that a NetID (i.e., user's token) may appear on multiple mobile devices, and also that more than one NetID may appear on a mobile device. This happens because users may have more than one mobile device, and may also release their NetID to their close friends for sharing a network usage quota. We observed that 40% and 37% of NetIDs were used on more than one mobile device, and 42% and 39% of mobile devices had more than one NetID used on them, for Campuses A and B respectively. In this paper, **we define two NetIDs as "ID_sharing_friends" when they both appear on at least one same mobile device**. We used ID_sharing_friends as one of the ground truths to evaluate the performance of our method (i.e., Sinfer in Section 6) for close friendship inference.

2.3. Ground Truth II: Demographics

We obtained a dataset of NetIDs' genders, social roles (i.e., undergraduate student, graduate student, and faculty staff), departments, enrolled years and classes from the information center of Campuses A and B, and used the dataset as the ground truth to evaluate the accuracy of our method (i.e., Dinfer in Section 5) for inferring the demographic attributes of mobile network users. When more than one NetID appeared on a mobile device, we selected the NetID most frequently used on the mobile device and then retrieved the NetID's demographic attributes as the mobile device users' attributes. As shown in Table 1, on average, a NetID appeared on 1.79 and 1.67 mobile devices for Campuses A and B respectively. Among the mobile network users on Campus A, 32% were male and 68% female, and 71.5% were undergraduate students, 12% graduate students, and 16.5% faculty staff. Among mobile network users on Campus B, 23% were male and 77% female, and 83.5% were undergraduate students, 6% graduate students, and 10.5% faculty staff.

2.4. Ground Truth III: Affiliation Information

People on campus with the following relationships tend to have strong social connections: 1) students in the same class; 2) students enrolled at school in the same year and affiliated to the same department; 3) faculty staff in the same department. Based on these observations, we collected the affiliation information of all NetIDs that appeared in our datasets. This data was used as another ground truth for evaluating the accuracy of our method (i.e., Sinfer in Section 6) for learning the social networks of mobile device users. The data indicated that mobile devices on Campus A belonged to users (or NetIDs) from 21 departments and 812 undergraduate classes. On average, a department of Campus A has 618 undergraduate



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

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

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

students, 115 graduate students and 69 faculty staff, and an undergraduate class has 16 students. Mobile devices on Campus B belonged 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 faculty staff, and an undergraduate class has 14 students.

3. Measurement studies

Intuitively, people on campus with the same demographic attributes (such as gender, major or social role) would be expected to be found at similar macroscopic events, such as visiting places in common. Similarly, microscopic behaviors such as co-coming, co-leaving and co-presenting at a place would be expected to be strong indicators of close friendships between people. In this section, we study the macro- and micro-statistics of mobile devices' AP-trajectories against the background of these expectations and describe our observations in detail.

3.1. Macroscopic observations

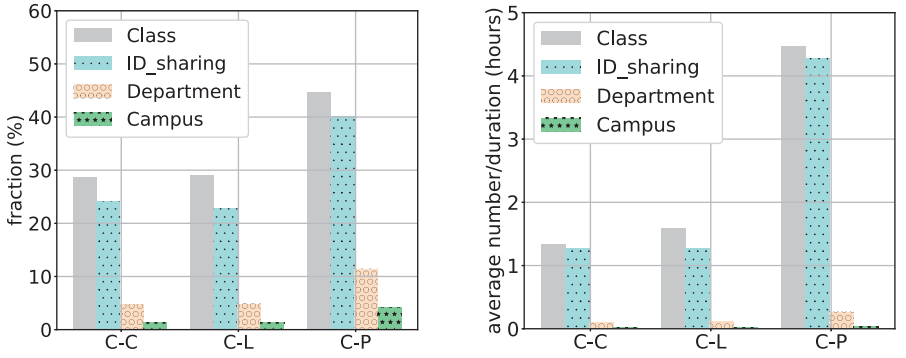
Observation 1: **Macroscopic features are useful for demographic inference**. To some extent, users' login counts (i.e, the number of logins) reflect whether or not they move frequently on campus, because Campuses A and B allow mobile devices to connect automatically and login to APs near to their current locations. Fig. 1 plots the average login counts of users on Campus A for each day of the week and each hour in the day. We observe: 1) women move more frequently than men; 2) undergraduate students move more frequently than graduate students and faculty staff; 3) compared to graduate students, faculty staff 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 demographic attributes**. We studied the homophily between ID_sharing_friends' attributes such as gender, social role, department, grade and class, where we considered two friends as exhibiting homophily when they had the same attributes of interest. Among all friend-pairs of Campus A (resp. Campus B), we observe that 81% (resp. 96%) exhibit gender homophily, 91% (resp. 95%) exhibit social role homophily, 84% (resp. 85%) exhibit department homophily, 77% (resp. 85%) exhibit grade homophily, and 57% (resp. 69%) exhibit 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 social relationships.** We tested whether social relationships can be accurately identified by directly comparing similarities between macroscopic features of mobile devices' AP-trajectories. We defined a macro-trajectory matrix, where the matrix's element (i, j) records the average number of times the device connected to the j-th AP during the i-th hour in the week, where $1 \le i \le 7 \times 24$. We concatenated 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 used 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

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 6.

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



(a) fraction of pairs having co-occurrence events

(b) average number/duration of co-occurrence events

Fig. 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.

students, we computed 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 identified only 18.52% of ID_sharing_friends and 26.33% of classmates for each user. In contrast, our microscopic features-based method Sinfer, which will be introduced in Section 6, is more effective for close friendship inference.

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 friendships between mobile device users. Existing studies [5,23,25] 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; this information would provide important information for friendship inference.

Observation 4: **Co-coming, co-leaving, and co-presenting duration are all strong friendship indicators.** For birthday parties (as an example) friends may not arrive at or leave prearranged places simultaneously. In this case, all of the co-coming, co-leaving and co-presenting events are strong friendship 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 social relationships (e.g., ID_sharing_friends and classmates) tend to have more co-occurrence events. For example, ID_sharing_friends and classmates have more co-occurrence events than pairs of users selected randomly. 2) all of the durations of co-coming, co-leaving and co-presenting are helpful for predicting users' friendships.

Observation 5: **Co-coming, co-leaving, and co-presenting are complementary to each other for indicating friendship between mobile device users.** Among 153,000 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 contrast, 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. Overview

In this section, we formally define the problem of learning mobile device users' demographic attributes, present the basic idea behind our methods for learning attributes and social networks of mobile device users, and introduce the notation and basic operations used throughout the paper.

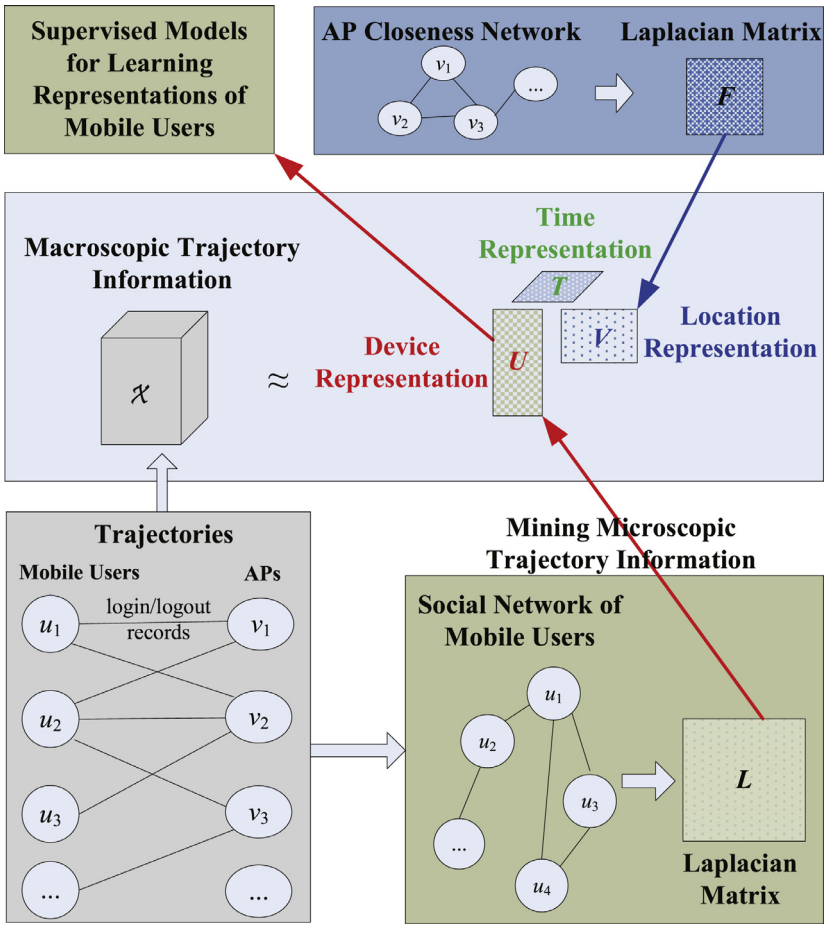


Fig. 3. Overview of our methods for learning social networks and demographic attributes of mobile users. Social networks are learned by our method Sinfer in Section 6. AP closeness network is generated according to the spatial closeness between APs.

4.1. Problem formulation

We formulate the **demographic inference problem** as: *Given AP-trajectories of all mobile devices* $u_1, \ldots, u_m$ *and user demographic attributes (such as gender and social role) of a set of training mobile devices* $\{u_i : i \in \Omega_{\text{train}}\}$, *our goal is to infer automatically user attributes of remaining mobile devices* $\{u_i : i \in \{1, \ldots, m\} \setminus \Omega_{\text{train}}\}$.

4.2. Basic idea behind our methods

We introduce our method Dinfer for identifying mobile device users' demographic attributes from macro- and micro-statistics of mobile devices' AP-trajectories. The framework of our method Dinfer is shown in Fig. 3. The details of Dinfer will be introduced in Section 5. 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 **U, V**, and **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 **U**, which is a Laplacian regularization **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 our method Sinfer, which will be introduced in Section 6, where two mobile devices belonging to close friends are connected in G_{MD}. The second constraint on **V** is a Laplacian regularization **F** that is learned from an AP closeness network G_{AP}. G_{AP} is built based on the spatial closeness between APs, where two APs are among the top-K nearest APs to each other and are connected in G_{AP}. The usage of AP spatial closeness information is the major technical difference between this paper and our previous work [27]. Finally, mobile device users' attributes are learned from their latent representations **U** via a supervised classifier.

4.3. Notation and preliminaries of vector, matrix and tensor operations

Before we describe our methods formally, we first present notation and introduce several vector, matrix and tensor operations, which will be used in our methods discussed later.

Table 3
Table of notation.

m	the number of mobile devices
n	the number of APs
h	the number of time slots
r	the number of latent semantic topics
$\mathcal{X} \in \mathbb{R}^{m\times n\times h}$	the tensor representation of macro-statistics of mobile devices' trajectories
$\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}$, $\mathbf{X}_{(3)}$	the mode-1, mode-2, and mode-3 matricizations of tensor $\mathcal{X}$
graph G_{MD}	mobile devices u_i and u_j in graph G_{MD} are connected when the users of u_i and u_j are close friends
$\mathbf{A} \in \mathbb{R}^{m\times m}$	the adjacency matrix of G_{MD}
$\mathbf{D} \in \mathbb{R}^{m\times m}$	a diagonal matrix, where $\mathbf{D}(i,i) = \sum_{j=1}^{m} \mathbf{A}(i,j)$
$\mathbf{L} = \mathbf{D} - \mathbf{A}$	Laplacian matrix of graph G_{MD}
graph G_{AP}	APs v_i and v_j in graph G_{AP} are connected when the distance between v_i and v_j is close
$\mathbf{S} \in \mathbb{R}^{n\times n}$	the adjacency matrix of G_{AP}
$\mathbf{Q} \in \mathbb{R}^{n\times n}$	a diagonal matrix, where $\mathbf{Q}(i,i) = \sum_{j=1}^{n} \mathbf{S}(i,j)$
$\mathbf{F} = \mathbf{Q} - \mathbf{S}$	Laplacian matrix of graph G_{AP}

Notation. Lowercase letters (e.g., a) denote scalars, lowercase bold letters (e.g., $\mathbf{a}$) denote column vectors, and upper-case bold letters (e.g., $\mathbf{A}$) denote matrices. Third-order tensors are denoted by upper-case calligraphic script letters, e.g., $\mathcal{A}$. Let $\mathbf{a}(i)$ denote the i-th element of vector $\mathbf{a}$ and $\mathbf{A}(i,j)$ denote the element (i,j) of matrix $\mathbf{A}$, i.e., the entry at the i-th row and j-th column of $\mathbf{A}$. Similarly, let $\mathcal{A}(i,j,k)$ denote the element (i,j,k) of $\mathcal{A}$. Denote by $\mathbf{a}_i$ and $\mathbf{a}_{:j}$ the i-th row and j-th column of $\mathbf{A}$ respectively. Let $\mathbf{A}^{\mathsf{T}}$ and $\text{Tr}(\mathbf{A})$ denote the transpose and trace of $\mathbf{A}$ respectively. Let $\|\mathbf{A}\|$ denote the Frobenius norm of $\mathbf{A}$, that is, $\|\mathbf{A}\| = \sqrt{\sum_{i=1}^{m}\sum_{i=1}^{n}\mathbf{A}(i,j)^2}$, where m and n are the numbers of rows and columns of $\mathbf{A}$ respectively. Let m denote the number of mobile devices, n denote the number of APs, and h denote the number of time slots. For ease of reference, we list notations used throughout the paper in Table 3.

Vector outer product $\circ$**.** The outer product of three vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{t}$ is a third-order tensor defined as $\mathcal{X} = \mathbf{u} \circ \mathbf{v} \circ \mathbf{t}$, where $\mathcal{X}(i,j,k) = \mathbf{u}(i)\mathbf{v}(j)\mathbf{t}(k)$.

Kronecker product $\otimes$**.** The Kronecker product of two matrices $\mathbf{U}$ and $\mathbf{U}$ of sizes $m \times r$ and a matrix $n \times s$ is a matrix of size $(mn) \times (rs)$ defined as:

$$\mathbf{U} \otimes \mathbf{V} = \begin{bmatrix} u_{1,1}\mathbf{V} & \dots & u_{1,r}\mathbf{V} \\ \vdots & \ddots & \vdots \\ u_{m,1}\mathbf{V} & \dots & u_{m,r}\mathbf{V} \end{bmatrix}.$$

Khatri-Rao product $\odot$**.** The Khatri-Rao product of two matrices $\mathbf{U}$ and $\mathbf{U}$ of sizes $m \times r$ and a matrix $n \times r$ is a matrix of size $(mn) \times r$ defined as:

$$\mathbf{U} \odot \mathbf{V} = [\mathbf{u}_1 \otimes \mathbf{v}_1, \dots, \mathbf{u}_r \otimes \mathbf{v}_r].$$

Hadamard product ***.** The Hadamard product of two matrices $\mathbf{U}$ and $\mathbf{V}$ both of size $m \times r$ is a matrix of size $m \times r$ defined as:

$$\mathbf{U} * \mathbf{V} = \begin{bmatrix} u_{1,1}v_{1,1} & \dots & u_{1,r}v_{1,r} \\ \vdots & \ddots & \vdots \\ u_{m,1}v_{m,1} & \dots & u_{m,r}v_{m,r} \end{bmatrix}.$$

Matricizations $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}$, and $\mathbf{X}_{(3)}$ of a third-order tensor $\mathcal{X}$. Matricization (also known as unfolding or flattening) is an operation that transforms a tensor into a matrix. 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:

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

5. Dinfer: Learn demographics

In this section, we discuss each part of our method Dinfer in detail. We first discuss how to model macro-statistics of AP-trajectories and present our method to learn mobile devices' latent representations from mobile device users' social networks and AP closeness networks computed based on AP-trajectories and APs' geometric locations respectively.

5.1. Modeling AP-Trajectories

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 we mentioned, $\mathcal{X}$ 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 $\mathcal{X}$ being very sparse and low-rank. To solve this problem, we model mobile devices' AP-trajectories 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 **U, V**, and **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. **U, V**, and **T** are representations of mobile devices, APs, and time slots in a latent space. The flexibility of the NTF model allows us to incorporate prior knowledge such as social networks of mobile device users and AP closeness networks into this model, which will be introduced in later subsections.

5.2. Modeling social networks

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 the same attributes such as age and gender. 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 to produce 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:

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

5.3. Leveraging AP closeness networks

Besides the social network information, additional spatial information may be available for APs. For example, buildings are divided into different functional areas and close places tend to provide similar functionality. This point indicates that spatially close APs have similar semantics, so their latent representations should be close. Formally, we define an AP closeness network G_{AP} to model this information, where a node in G_{AP} represents an AP, (v_i, v_j) in G_{AP} when v_i and v_j are on the same floor in the same building, and one of them is among the top-K nearest APs to the other. Let $\mathbf{S} \in \mathbb{R}^{n \times n}$ denote the adjacency matrix of G_{AP}, Let $\mathbf{v}_i$ (i.e., the i-th row of matrix $\mathbf{V}$) be the representation of the i-th AP v_i, $1 \le i \le n$. To model this knowledge, similarly, we use graph Laplacian to produce similar representations $\mathbf{v}_i$ and $\mathbf{v}_j$ when v_i and v_j are similar. Formally, we formulate it as minimizing the following loss function:

$$\mathbb{O}_{\text{GAP}} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{v}_i - \mathbf{v}_j\|^2 \mathbf{S}(i, j). \tag{3}$$

The above loss function $\mathbb{O}_{\text{GAP}}$ incurs a penalty $\|\mathbf{v}_i - \mathbf{v}_j\|^2$ when two APs v_i and v_j have different latent representations $\mathbf{v}_i$ and $\mathbf{v}_j$ but they are close to each other.

Let $\mathbf{Q} \in \mathbb{R}^{n\times n}$ denote a diagonal matrix, and its diagonal element $\mathbf{Q}(i,i)$ is defined as $\mathbf{Q}(i,i) = \sum_{j=1}^{n} \mathbf{S}(i,j)$. Let $\mathbf{F} = \mathbf{Q} - \mathbf{S}$. We can rewrite Eq. (3) as:

$$\begin{aligned}\mathbb{O}_{\text{GAP}} &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{v}_i \mathbf{S}(i,j)\mathbf{v}_i^{\top} - \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{v}_i \mathbf{S}(i,j)\mathbf{v}_j^{\top} \\ &= \sum_{i=1}^{n} \mathbf{v}_i \mathbf{Q}(i,i)\mathbf{v}_i^{\top} - \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{v}_i \mathbf{S}(i,j)\mathbf{v}_j^{\top} \\ &= \text{Tr}(\mathbf{V}^{\top}(\mathbf{Q}-\mathbf{S})\mathbf{V}) \\ &= \text{Tr}(\mathbf{V}^{\top}\mathbf{F}\mathbf{V}).\end{aligned}$$

We have introduced the modeling of macro-statistics of mobile devices' AP-trajectories, social networks between mobile device users, and AP closeness networks above. Next, we propose a model to incorporate all of these three types of information in a general framework.

5.4. Learning user representations

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

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V},\mathbf{T}\geq 0} \mathbb{O} = &\underbrace{\|\mathcal{X} - [\![\mathbf{U},\mathbf{V},\mathbf{T}]\!]\|^2}_{\mathbb{O}_{\text{TRA}}} + \alpha \underbrace{\text{Tr}(\mathbf{U}^{\top}\mathbf{L}\mathbf{U})}_{\mathbb{O}_{\text{GMD}}} + \beta \underbrace{\text{Tr}(\mathbf{V}^{\top}\mathbf{F}\mathbf{V})}_{\mathbb{O}_{\text{GAP}}} \\ &+ \gamma \underbrace{(\|\mathbf{U}\|^2 + \|\mathbf{V}\|^2 + \|\mathbf{T}\|^2)}_{\text{regularization}},\end{aligned} \tag{4}$$

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

There is no close-form solution for the optimization problem (4). To solve this problem, we propose a multiplicative update algorithm to find optimal solutions for variables **U, V**, and **T** based on Oja's iterative learning rule [22,28]. The basic idea behind the multiplicative update algorithm is to optimize the objective $\mathbb{O}$ with respect to one variable while fixing the other. We keep updating the variables **U, V**, and **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:

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

1) Computation of U. When **V** and **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 $\mathbf{X}_{(1)}$ is the mode-1 matricization of $\mathcal{X}$. Then, the derivative $\frac{\partial \mathbb{J}}{\partial \mathbf{U}}$ is computed as:

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

Based on the properties of Khatri–Rao product [24], we have $(\mathbf{T}\odot\mathbf{V})^{\top}(\mathbf{T}\odot\mathbf{V}) = \mathbf{T}^{\top}\mathbf{T} * \mathbf{V}^{\top}\mathbf{V}$. 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 **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 **L** may take any signs, so we decompose **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 **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 V and T. Similarly, we have the following multiplicative updating rules of **V** and **T**:

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

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

where $\mathbf{X}_{(2)}$ and $\mathbf{X}_{(3)}$ are the mode-2 and mode-3 matricizations of $\mathcal{X}$ respectively.

We can easily find that the above three multiplicative update rules maintain the nonnegativity of **U, V**, and **T** for initial nonnegative matrices **U, V**, and **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 **U, V**, and **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 **U, V**, and **T** respectively. Formally, the correctness and convergence of the multiplicative updating rules of **U, V**, and **T** can be proven with the standard auxiliary function approach [7,12,15]. Last, we learn a supervised model (e.g., LDA, 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.

6. Sinfer: Learn social networks

In this section, we introduce a Pointwise Mutual Information-based (PMI) method Sinfer for learning social networks of mobile device users. PMI has been widely used for measuring the semantic similarity between words. For example, word2vector [19] is an effective tool for embedding words into a low-dimensional space and evaluating the similarity between words, and 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 notation. Let $f_v^{(\mathrm{c})}(u)$ be the number of times a device u logins to AP v, and $f_v^{(\mathrm{c})}(u_i,u_j)$ 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 $p_v^{(\mathrm{c})}(u)=\frac{f_v^{(\mathrm{c})}(u)}{\sum_{x\in U} f_v^{(\mathrm{c})}(x)}$, 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 $p_v^{(\mathrm{c})}(u_i,u_j)=\frac{2f_v^{(\mathrm{c})}(u_i,u_j)}{\sum_{x,y\in U} f_v^{(\mathrm{c})}(x,y)}$. To evaluate the closeness between devices u_i and u_j, we compute the PMI of their co-coming events at AP v as $\mathrm{pmi}_v^{(\mathrm{c})}(u_i,u_j)=\log\frac{p_v^{(\mathrm{c})}(u_i,u_j)}{p_v^{(\mathrm{c})}(u_i)p_v^{(\mathrm{c})}(u_j)}$. The value of $\mathrm{pmi}_v^{(\mathrm{c})}(u_i,u_j)$ 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 $\mathrm{pmi}_v^{(\mathrm{c})}(u_i,u_j)$ is small and it indicates that a co-coming event of them at place v is probably a coincidence. In contrast, when u_i and u_j seldom appear at place v but have a number of co-coming events at place v, then $\mathrm{pmi}_v^{(\mathrm{c})}(u_i,u_j)$ 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 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

$$closeness(u_i,u_j)=\sum_{f_v^{(\mathrm{c})}(u_i,u_j)>0,v\in V} f_v^{(\mathrm{c})}(u_i,u_j)\mathrm{pmi}_v^{(\mathrm{c})}(u_i,u_j)$$

$$+ \sum_{f_v^{(\mathrm{l})}(u_i,u_j)>0, v\in V} f_v^{(\mathrm{l})}(u_i,u_j)\mathrm{pmi}_v^{(\mathrm{l})}(u_i,u_j)$$
$$+ \sum_{t_v^{(\mathrm{p})}(u_i,u_j)>0, v\in V} f_v^{(\mathrm{p})}(u_i,u_j)\mathrm{pmi}_v^{(\mathrm{p})}(u_i,u_j).$$

7. Experiments

In this section, we evaluate the performance of Dinfer and Sinfer. The experiments were conducted on a computer with Intel Core i7 3.5 Ghz CPU and 32 GB memory.

7.1. Experimental settings

Data and Evaluation. We used 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), social role (i.e., undergraduate/graduate/faculty), department, grade (i.e., enrolled year), and class. To infer mobile device users' social networks and attributes effectively and accurately, we only considered *active* mobile devices which had at least 50 AP login records in the 74 days we collected. Campuses A and B had 21,207 and 13,353 active mobile devices respectively. For learning mobile device users' attributes, similar to [9], we considered each class in gender, social role and other attributes as important as each other and used *weighted Precision, Recall*, and *macro-F1* (in short, *wPrecision, wRecall*, and *wF1*) as evaluation metrics.

Comparing State-of-the-Art Algorithms: For inferring mobile device users' social networks, we compared our method Sinfer with the state-of-the-art method:

- **EBM** [23] 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 to be friends when their co-occurrences happened at diverse places; 2) co-occurrences at popular places are more likely to happen by chance than those at private places.

For predicting mobile device users' demographic attributes, we compared 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 number of logins at different APs. Similar to Dinfer, S learns supervised classifiers using extracted AP-trajectory features to predict unknown mobile device users' 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 number of logins during different hours in the week.

We also compared our method Dinfer with the following classic methods that learn mobile device representations **U** from the tensor of spatiotemporal AP-trajectories $\mathcal{X}$ in a direct manner:

- **PCA** learns **U** by applying the principal component analysis method to the mode-1 matricization $X_{(1)}$ of $\mathcal{X}$.
- **NMF** learns **U** by applying the non-negative matrix factorization method [3,14] to the mode-1 matricization $X_{(1)}$ of $\mathcal{X}$.
- **HOSVD** learns **U** by applying high order singular value decomposition method to tensor $\mathcal{X}$.
- **NTF** only considers $\mathcal{X}$ but not social networks of mobile device users or AP closeness information in comparison with Dinfer, i.e., NTF learns **U** by optimizing $\mathbb{O}_{\mathrm{TRA}}$ in eq. (1).
- **DinferSN** [27] only considers $\mathcal{X}$ and social networks of mobile device users but not AP closeness information in comparison with Dinfer, i.e., DinferSN learns **U** by optimizing $\mathbb{O}_{\mathrm{TRA}} + \mathbb{O}_{\mathrm{GMD}}$.

Parameter Settings: Dinfer has four major parameters: (1) the number of latent dimension r; (2) the *closeness* between connected APs in G_{AP}; (3) the trade-off parameter of user social networks α; and (4) the trade-off parameter of AP closeness networks β. We study the effects of different parameters as follows: (1) $r = 100, 300, 500, 1000, 1500, 2000$; (2) select the *closeness* from "none" (no APs are connected), "top-2", "top-4", "top-6", "same floor" (APs on the same floor are connected), and "same building"; (3) $\alpha = 0, 0.01, 0.1, 1, 10$; (4) $\beta = 0, 0.01, 0.1, 1, 10$.

For each campus, we conducted a 10-fold cross-validation for inferring each demographic attribute. By default, in our experiments, we set the training/test ratio as 9:1; set the number of latent dimensions as 1000 in Campus A and 500 in Campus B (i.e., $r = 1000, 500$) for methods NTF, DinferSN, and Dinfer; let one AP connect to the other top-4 nearest APs in G_{AP}; set the social network factor of DinferSN and Dinfer as $\alpha = 0.1$; set the AP closeness factor of Dinfer as $\beta = 0.1$; and set the regularization factor of NTF, DinferSN, and Dinfer as $\gamma = 0.1$.

7.2. Accuracy of learning social networks

Results of Friendship Inference. Tables 4, 5, 6 show the distribution of Top-10 and Top-20 close friends learned by our method Sinfer. Among Top-10 and Top-20 close friends, 16%-27%, 16%-18%, and 9%-16% were ID_sharing_friends for undergraduate students, graduate students, and faculty staff respectively. Among the Top-10 (Top-20) close friends of undergraduate students on Campus A, on average 76.40% (76.18%) were undergraduate students who enrolled in the same department

Table 4
Distribution of Top-10 and Top-20 close friends learned by Sinfer for undergraduate students. The number in the parentheses is the fraction of friends enrolled in school at the same year.

distribution of learned friends for undergraduate students (%)		A		B	
		Top-10	Top-20	Top-10	Top-20
ID_sharing_friends		25.41	27.48	15.98	21.28
same department	undergraduate, classmate	**61.47**	**58.93**	**83.87**	**80.28**
	undergraduate, not classmate	16.67 (14.93)	19.32 (17.25)	6.43 (5.36)	8.85 (7.68)
	graduate	2.27	2.13	1.74	2.84
	faculty	5.20	5.44	0.21	0.18
different department	undergraduate	10.40 (4.40)	10.47 (4.70)	7.45 (3.38)	8.45 (3.87)
	graduate	1.07	1.10	0.57	0.58
	faculty	2.93	2.61	0.94	0.90

Table 5
Distribution of Top-10 and Top-20 close friends learned by Sinfer for graduate students. The number in the parentheses is the fraction of friends enrolled in school at the same year.

Distribution of learned friends for graduate students (%)		A		B	
		Top-10	Top-20	Top-10	Top-20
ID_sharing_friends		16.54	16.87	17.43	18.25
same department	undergraduate	16.47 (2.94)	16.71 (3.19)	21.63 (7.86)	21.54 (7.78)
	graduate	**34.71** (12.94)	**34.15** (12.95)	**40.15** (28.53)	**38.59** (28.04)
	faculty	22.65	22.34	17.71	16.83
different department	undergraduate	13.82 (0.59)	14.79 (0.25)	14.99 (1.09)	17.33 (1.19)
	graduate	7.06	7.60	1.83	1.94
	faculty	5.29	4.41	3.69	3.78

Table 6
Distribution of Top-10 and Top-20 close friends learned by Sinfer for faculties.

distribution of learned friends for faculties (%)		A		B	
		Top-10	Top-20	Top-10	Top-20
ID_sharing_friends		14.53	15.70	9.65	10.91
same department	undergraduate	12.41	12.61	61.22	60.37
	graduate	4.22	4.48	3.46	3.09
	faculty	**56.51**	**55.79**	21.14	21.16
different department	undergraduate	12.96	13.04	9.59	10.48
	graduate	4.69	4.65	0.76	0.97
	faculty	9.22	9.45	3.82	3.93

in the same year, 61.47% (58.93%) were their classmates, 5.20% (5.44%) were faculty staff in the same department, and 4.40% (4.70%) were undergraduate students who enrolled in different departments in the same year. Among the Top-10 (Top-20) close friends of undergraduate students on Campus B, on average, 89.23% (87.96%) were also undergraduate students who enrolled in the same department in the same year, 83.87% (80.28%) were their classmates, and 3.38% (3.87%) were undergraduate students who enrolled in different departments in the same year. Similarly, among the Top-10 and Top-20 close friends of graduate students, 73%-77% came from the same department, 16%-21%, 34%-40%, and 16%-22% were undergraduate students, graduate students, and faculty staff in the same department respectively. Among the Top-10 and Top-20 close friends of faculty staff, 73%-86% were affiliated with the same department. From the above results, we observed that a strong homophily appeared in close friends learned by Sinfer. These results are consistent with common sense; e.g., classmates and undergraduate students in the same department tend to be close friends. Unlike Campus A, we noticed that most learned top close friends of faculty staff on Campus B were not faculty staff but undergraduate students in the same department. We further investigated this phenomenon and found that faculty staff members on Campus B spent much more time on teaching and they co-present with undergraduates more frequently than staff on Campus A.

Sinfer vs. Prior Art. We conducted experiments to evaluate the performance of our method Sinfer in comparison with the state-of-the-art method EBM [23]. 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)

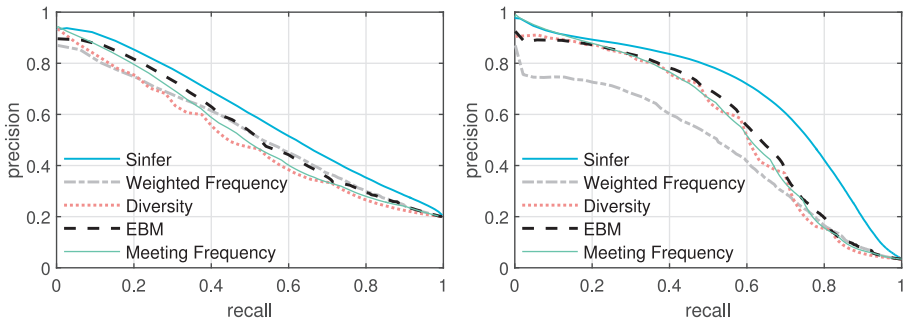


Fig. 4. Precision-Recall curve comparison of our friendship inference method Sinfer, state-of-the-art method EBM, and methods using Diversity, Weighted Count, and Co-occurrence Count as metrics to evaluate the closeness between mobile device users.

Table 7
Performance of learning genders of mobile users.

Methods	Gender inference					
	Campus A			Campus B		
	wPrecision	wRecall	wF1	wPrecision	wRecall	wF1
S	0.645	0.638	0.641	0.672	0.636	0.649
ST	0.645	0.644	0.644	0.670	0.644	0.654
PCA	0.620	0.634	0.621	0.610	0.602	0.605
NMF	0.626	0.625	0.626	0.605	0.597	0.600
HOSVD	0.618	0.621	0.619	0.626	0.598	0.607
NTF	0.627	0.628	0.628	0.635	0.636	0.635
DinferSN	**0.694**	**0.690**	**0.692**	**0.712**	**0.722**	**0.717**
Dinfer	**0.705**	**0.698**	**0.701**	**0.722**	**0.726**	**0.724**

Table 8
Performance of learning social roles of mobile users.

Methods	Social role inference					
	Campus A			Campus B		
	wPrecision	wRecall	wF1	wPrecision	wRecall	wF1
S	0.639	0.616	0.605	0.589	0.549	0.515
ST	0.641	0.621	0.612	0.598	0.552	0.517
PCA	0.636	0.550	0.578	0.644	0.462	0.494
NMF	0.624	0.568	0.589	0.599	0.487	0.517
HOSVD	0.644	0.561	0.588	0.639	0.514	0.545
NTF	0.652	0.584	0.609	0.610	0.528	0.551
DinferSN	**0.698**	**0.635**	**0.659**	**0.686**	**0.616**	**0.642**
Dinfer	**0.713**	**0.665**	**0.685**	**0.701**	**0.621**	**0.651**

and the weighted co-occurrence count (in short, Weighted Count). We set its parameters as used in the original paper. Clearly, it was not easy to obtain the ground truth of top close friends especially for graduate students and faculty staff in our dataset. In our experiment, we focused on undergraduate students, and used 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 close 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 noticed that the precision ratio could not reach 100% in precision-recall curves; this is because, in practice, undergraduates' top close friends may not be their ID_sharing_friends or classmates, which are wrongly recognized as friends by our experimental settings.

7.3. Accuracy of learning demographics

7.3.1. Predictive performance

Tables 7, 8, 9, 10, 11 show the performance of our method Dinfer in comparison with state-of-the-art methods for learning mobile device users' demographic attributes, including genders, social roles, departments, classes, and grades. Overall, Dinfer outperforms the other methods in terms of wPrecision, wRecall, and wF1.

Table 9
Performance of learning departments of mobile users.

Methods	Department inference					
	Campus A			Campus B		
	wPrecision	wRecall	wF1	wPrecision	wRecall	wF1
S	0.642	0.603	0.613	0.525	0.530	0.518
ST	0.649	0.604	0.615	0.535	0.544	0.533
PCA	0.652	0.606	0.610	0.881	0.526	0.627
NMF	0.660	0.604	0.619	0.706	0.618	0.646
HOSVD	0.657	0.617	0.626	0.726	0.616	0.650
NTF	0.680	0.603	0.625	0.692	0.640	0.657
DinferSN	**0.729**	**0.680**	**0.689**	**0.773**	**0.672**	**0.703**
Dinfer	**0.744**	**0.692**	**0.702**	**0.788**	**0.683**	**0.715**

Table 10
Performance of learning grades of mobile users.

Methods	Grade inference					
	Campus A			Campus B		
	wPrecision	wRecall	wF1	wPrecision	wRecall	wF1
S	0.474	0.461	0.463	0.487	0.455	0.462
ST	0.480	0.477	0.474	0.502	0.473	0.481
PCA	0.503	0.488	0.492	0.616	0.457	0.494
NMF	0.532	0.497	0.509	0.554	0.524	0.526
HOSVD	0.527	0.511	0.516	0.568	0.539	0.540
NTF	0.528	0.501	0.510	0.552	0.547	0.539
DinferSN	**0.588**	**0.562**	**0.571**	**0.587**	**0.578**	**0.576**
Dinfer	**0.611**	**0.588**	**0.597**	**0.620**	**0.606**	**0.603**

Table 11
Performance of learning classes of mobile users.

Methods	Class inference					
	Campus A			Campus B		
	wPrecision	wRecall	wF1	wPrecision	wRecall	wF1
S	0.504	0.418	0.431	0.569	0.563	0.544
ST	0.505	0.423	0.436	0.571	0.579	0.555
PCA	0.498	0.465	0.455	0.929	0.571	0.687
NMF	0.532	0.439	0.452	0.811	0.701	0.734
HOSVD	0.555	0.485	0.495	0.814	0.695	0.730
NTF	0.564	0.457	0.484	0.824	0.693	0.735
DinferSN	**0.710**	**0.585**	**0.616**	**0.862**	**0.708**	**0.765**
Dinfer	**0.726**	**0.613**	**0.641**	**0.897**	**0.738**	**0.789**

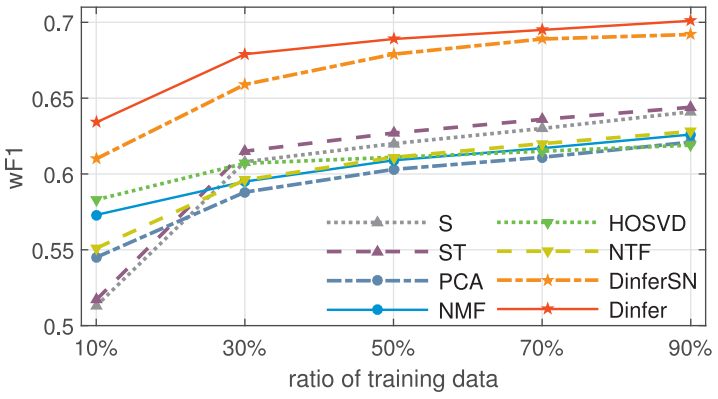
Gender Inference. For gender inference (Table 7), the wF1 of Dinfer is above 0.7 for both Campuses A and B, which significantly outperforms competing methods. In particular, the Dinfer method, which considers user social networks and AP closeness information, gains more than 5% improvement over the ST-based method and 7% improvement over other dimension-reduction methods. This finding indicates that the user social networks and AP closeness information contribute substantial predictive power to Dinfer.

Social Role Inference. Table 8 presents the performance for inferring social role. The wF1 of Dinfer is 0.685 and 0.651 for Campuses A and B respectively. Although the inference performance is reduced as a 3-class classification task, Dinfer still improves the wF1 by more than 7% and 10% for Campuses A and B respectively.

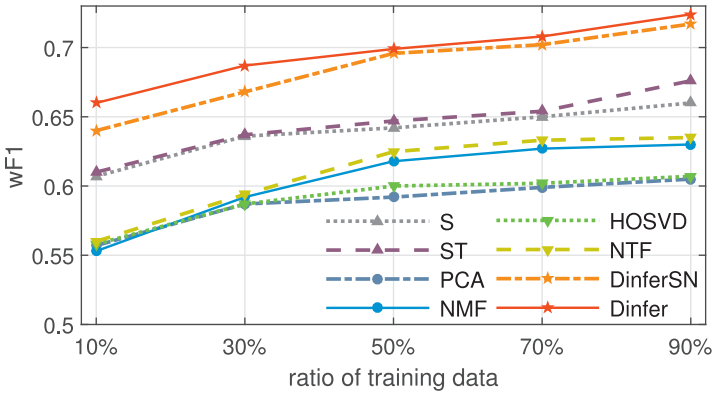
Department, Grade, and Class Inferences. Similar to the above tasks, Dinfer achieves 5%-15% improvement over competing methods. For department inference, the wF1 of Dinfer is 0.702 and 0.715 for Campuses A and B respectively (Table 9). For grade inference, the wF1 of Dinfer is around 0.6 for both campuses (Table 10). For class inference, Dinfer significantly increases the wF1 for more than 15% by Campus A when other methods' wF1 is less than 50%, and the wF1 of Dinfer achieves 0.789 for Campus B (Table 11).

In summary, DinferSN uses social networks of mobile device users, and Dinfer further uses AP closeness networks to learn more effective latent representations from AP-trajectories, and raises the wF1 by 5%-15%. This demonstrates that social networks of mobile device users and AP closeness networks effectively improve the performance of Dinfer.

Training/Test Ratio. We studied 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 improve as the number of labeled



(a) Campus A



(b) Campus B

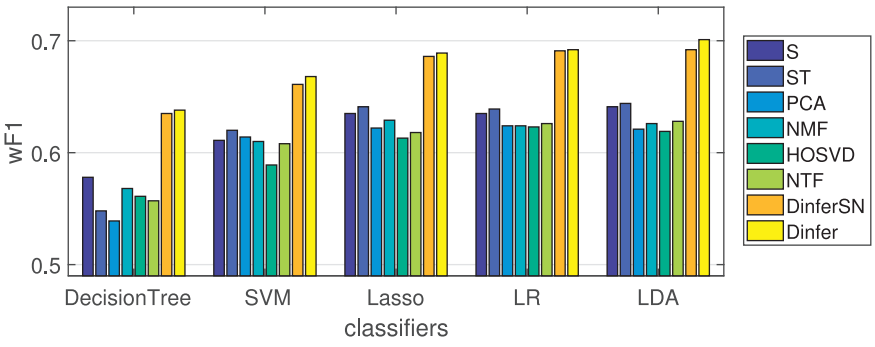
Fig. 5. (Gender inference) Performance w.r.t. ratio of training data for Dinfer and state-of-the-art methods.

mobile device users in the training set increases. Moreover, we see that Dinfer and DinferSN significantly outperform the other methods for gender inference when not less than 10% of data is used for training. We omit similar results of learning other demographic attributes.

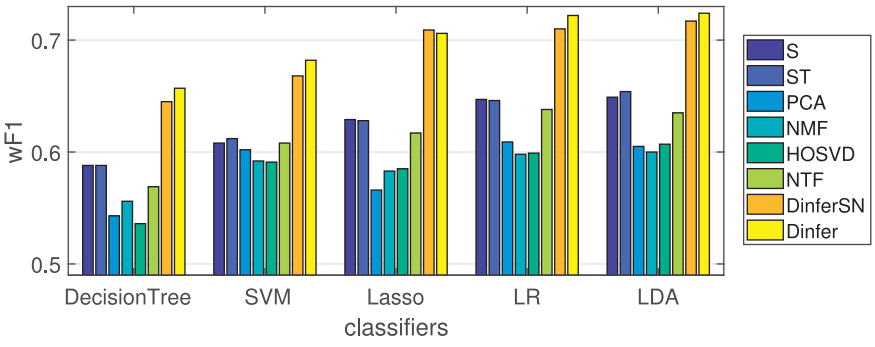
Classifier Sensitivity. We studied the performance of Dinfer with different classifiers. Fig. 6 reports the wF1 of five popular classifiers: support vector machine (linear kernel, SVM), linear discriminant analysis (LDA), logistic regression (LR), decision tree (DecisionTree), and linear regression (Lasso) for gender inference. We observed that LDA, LR, and Lasso slightly outperform SVM and DecisionTree when using Dinfer and other methods for learning mobile device users' latent representations. Dinfer outperforms other methods for all these classifiers. Considering accuracy and efficiency, we suggest LDA and LR classifiers for inference tasks. We omit similar results of learning other user attributes.

7.3.2. Parameter selection

Next, we discuss the parameter sensitivity of Dinfer. As mentioned earlier, there are four major parameters in Dinfer: *r, closeness,* α, and β. After tuning these parameters on Campuses A and B for five user attribute inference tasks, we find that the trends are very similar on these tasks for all the four parameters. Fig. 7 shows the impact of different choices of *r, closeness,* α, and β on school class inference. Fig. 7(a) shows the performance of Dinfer for different *r*. In general, large *r* provides a better performance. However, the rate of the performance increase becomes slower as we enlarge $r \geq 1000$ for Campus A and enlarge $r \geq 500$ for Campus B. The optimal value of *r* is different for Campuses A and B because of the different number of APs in these two campuses. In general, the larger number of APs (i.e., number of rows in matrix $\mathbf{V}$) requires a larger *r* to obtain effective user representations. Considering the increasing computational complexity as *r* increases, we suggest $r = 1000$ and $r = 500$ for Campuses A and B respectively as a compromise. Fig. 7(b) shows the performance of Dinfer for different AP closeness networks. We find that when we connect one AP to its top-4 nearest APs, Dinfer performs best, then the wF1 decreases as we enlarge the distance between connected APs in G_{AP}. Fig. 7(c) shows how the performance of Dinfer changes with the increase of α and β in Campus A. We can see that the wF1 is low when $\alpha = 0$ and $\beta = 0$, i.e. when social networks of mobile device users and AP closeness networks are not used. When $\alpha > 0$, the performance significantly increases, and fluctuates in a small range (less than 2%) for $\alpha = 0.01, \ldots, 10$. When $\beta > 0$, the performance increases slightly (around 1%) for $\beta = 0.01, \ldots, 10$. We omit similar results for the inference of other user attributes.



(a) Campus A



(b) Campus B

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

7.3.3. Efficiency study

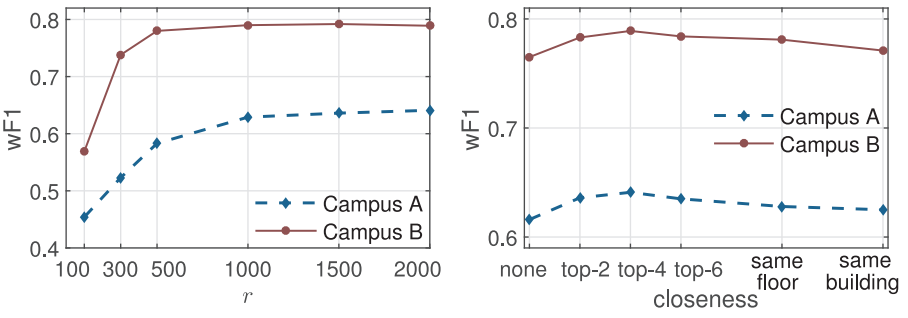
In the final set of experiments, we validated the efficiency of Dinfer. We found that the running time changed with the number of latent dimension r and the scale of the tensor $\mathcal{X}$. We set the number of latent dimensions $r = 100, \ldots, 1000$, and tested the running time for the two campuses. Fig. 8 shows the running time of Dinfer. Obviously, the time complexity of Dinfer is linear in latent dimension r. As for the scale of $\mathcal{X}$, the ratio between the scale of Campus A ($m = 21207, n = 1921, h = 168$) and Campus B ($m = 13353,\ n = 362,\ h = 168$) is 8.43: 1, but the ratio of the running time is only 4.12:1 (when $r = 1000$). It validates that the time complexity of Dinfer is sub-linear in the scale of $\mathcal{X}$.

7.4. Discussion

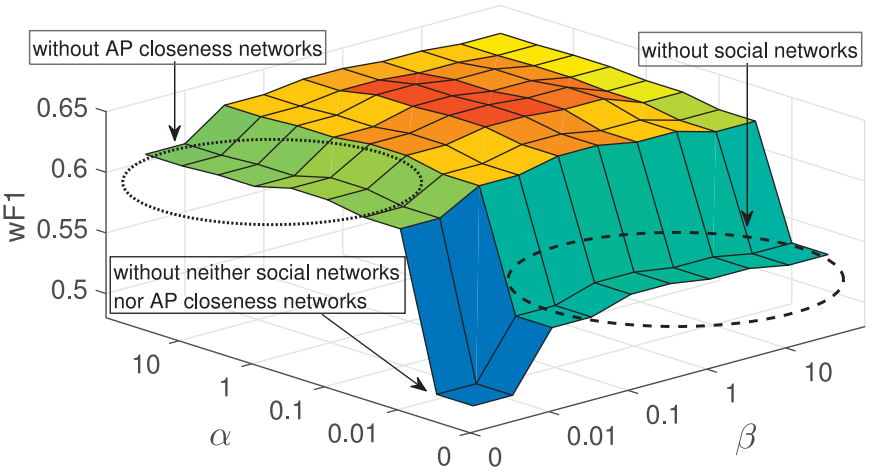
We evaluated our methods on two real-world campus datasets but no city-wide or wider-space trajectory datasets due to such datasets not being publicly available. We believe our methods are also effective for city-wide or wider-space datasets. In contrast to wider-space trajectory datasets, one feature of campus datasets is that many users have similar moving trajectories, which makes it difficult to distinguish different users. Our method Dinfer utilizes social relationship and location diversity to overcome this challenge, and such an approach can also be effective for wider-space trajectory datasets by adjusting users' trajectory tensor $\mathcal{X}$, social network G_{MD}, and AP closeness network G_{AP}. For wider-space datasets, data sparsity is a challenge. We used tensor factorization-based methods to reduce dimension and relieve sparsity. Fig. 9 shows the performance of our methods for sparse datasets. Our dataset is sparse since each user in tensor $\mathcal{X}$ has only 218 non-zero values (i.e., user check-in records) among 322,560 dimensions in Campus A on average. Similarly, in Campus B, there are only 205 non-zero values from 60,816 dimensions on average. When we randomly reduced the fraction of non-zero elements to 10%, many users had less than 5 non-zero elements. In such a situation, Dinfer still achieves the wF1 at around 0.5, which is acceptable for a multi-class classification problem. Therefore, our method Dinfer is effective for wider-space and sparser datasets. For Sinfer, the confidence of co-occurrence events improves in wider space (in small space, co-occurrences are more likely to be coincidences), which leads to better performance for learning social networks between mobile device users.

8. Related work

Our problem is closely related to research on learning user attributes and social networks from spatiotemporal trajectories. More especially:



(a) latent dimensions r (b) *closeness* in G_{AP}



(c) trade-off parameters α and β

Fig. 7. (Class inference) Parameter sensitivities w.r.t. the number of latent dimensions r, the *closeness* in G_{AP}, the social network factor α, and the AP closeness network factor β.

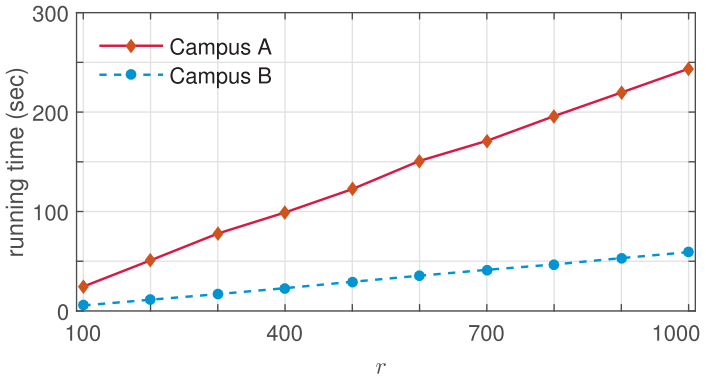


Fig. 8. running time w.r.t. r.

Learning user demographic attributes: Given demographic attributes (e.g., gender and affiliation) for a fraction of users on an OSN, Mislove et al. [20] 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 considerable effort to predict social relationships from human movement records collected from GPS-enabled devices [6,10,17] and location-based OSNs [4,5,23,25,29]. Li et al. [17] proposed a

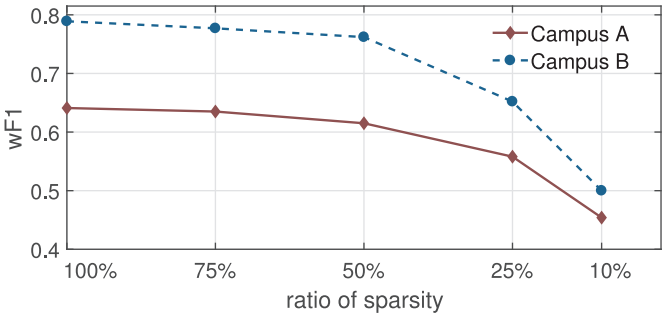


Fig. 9. (Class inference) performance w.r.t. sparser dataset.

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' frequent 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 relates to 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 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 location popularity. Furthermore, in addition to the friendship prediction, Pham et al. [23] proposed another entropy-based model to estimate the strength of friendships from co-occurrence events. In addition to the location popularity, Wang et al. [25] proposed a model that also considered 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 social relationship. Jayarajah et al. [13] utilized AP-trajectories to estimate the strength of social events, but they neglected to determine whether 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 observed that these three kinds of events are complementary to each other for predicting social networks of mobile devices and developed a method to predict friendships by utilizing all co-occurrence events.

9. 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 their AP-trajectories, such as co-coming, co-leaving and co-presenting events. Moreover, we developed 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 and efficiency of our methods Dinfer and Sinfer.

Acknowledgments

The research presented in this paper is supported in part by National Natural Science Foundation of China (61603290, 61602371, U1301254), the Ministry of Education - China Mobile Research Fund (MCM20160311), Shenzhen Basic Research Grant (JCYJ20160229195940462), SZSTI (JCYJ20170816100819428), China Postdoctoral Science Foundation(2015M582663), Natural Science Basic Research Plan in Shaanxi Province of China (2016JQ6034, 2017JM6095), Natural Science Basic Research Plan in ZheJiang Province of China (LGG18F020016).

References

[1] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, in: NIPS'01, 2001, pp. 601–608.
[2] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[3] D. Cai, X. He, J. Han, T.S. Huang, Graph regularized nonnegative matrix factorization for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1548–1560.
[4] R. Cheng, J. Pang, Y. Zhang, Inferring friendship from check-in data of location-based social networks, in: ASONAM'15, 2015, pp. 1284–1291.
[5] D.J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, J. Kleinberg, Inferring social ties from geographic coincidences, PNAS 107 (52) (2010) 22436–22441.
[6] J. Cranshaw, E. Toch, J. Hong, A. Kittur, N. Sadeh, Bridging the gap between physical location and online social networks, in: UbiComp'10, 2010, pp. 119–128.
[7] C.H.Q. Ding, T. Li, M.I. Jordan, Convex and semi-nonnegative matrix factorizations, IEEE Trans. Pattern Anal. Mach. Intell. 32 (1) (2010) 45–55.

[8] Y. Dong, F. Pinelli, Y. Gkoufas, Z. Nabi, F. Calabrese, N.V. Chawla, Inferring unusual crowd events from mobile phone call detail records, in: ECML-PKDD'15, 2015, pp. 474–492.
[9] Y. Dong, Y. Yang, J. Tang, Y. Yang, N.V. Chawla, Inferring user demographics and social strategies in mobile social networks, in: KDD'14, 2014, pp. 15–24.
[10] N. Eagle, A.S. Pentland, D. Lazer, Inferring friendship network structure by using mobile phone data, PNAS 106 (36) (2009) 15274–15278.
[11] P. Eckert, Gender and sociolinguistic variation, Readings in Language and Gender.
[12] Q. Gu, J. Zhou, C.H.Q. Ding, Collaborative Filtering: Weighted Nonnegative Matrix Factorization Incorporating User and Item Graphs, in: SDM'10, 2010, pp. 199–210.
[13] K. Jayarajah, A. Misra, X.W. Ruan, E.P. Lim, Event detection: Exploiting socio-physical interactions in physical spaces, in: ASONAM'15, 2015, pp. 508–513.
[14] D. Lee, H. Seung, Learning the parts of objects by nonnegative matrix factorization, Nature 401 (1999) 788–791.
[15] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: NIPS'01, 2001, pp. 556–562.
[16] O. Levy, Y. Goldberg, Neural word embedding as implicit matrix factorization, in: NIPS'14, 2014, pp. 2177–2185.
[17] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, W.Y. Ma, Mining user similarity based on location history, in: SIGSPATIAL GIS'08, 2008, p. 34.
[18] J.Y. Liu, Y.H. Yang, Inferring personal traits from music listening history, in: MIRUM'12, 2012, pp. 31–36.
[19] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: NIPS'13, 2013, pp. 3111–3119.
[20] A. Mislove, B. Viswanath, K.P. Gummadi, P. Druschel, You are who you know: inferring user profiles in online social networks, in: WSDM'10, 2010, pp. 251–260.
[21] D. Murray, K. Durrel, Inferring demographic attributes of anonymous internet users, in: WEBKDD'99, 1999.
[22] E. Oja, Principal components, minor components, and linear neural networks, Neural Networks 5 (6) (1992) 927–935.
[23] H. Pham, C. Shahabi, Y. Liu, Ebm: an entropy-based model to infer social strength from spatiotemporal data, in: SIGMOD'13, 2013, pp. 265–276.
[24] A. Smilde, R. Bro, P. Geladi, Multi-way Analysis with Applications in the Chemical Sciences, Wiley, 2004.
[25] H. Wang, Z. Li, W.C. Lee, Pgt: Measuring mobility relationship using personal, global and temporal factors, in: ICDM'14, 2014, pp. 570–579.
[26] P. Wang, J. Guo, Y. Lan, J. Xu, X. Cheng, Your cart tells you: Inferring demographic attributes from purchase data, in: WSDM'16, 2016, pp. 173–182.
[27] P. Wang, F. Sun, D. Wang, J. Tao, X. Guan, A. Bifet, Inferring demographics and social networks of mobile device users on campus from AP-trajectories, in: WWW '17 Companion, 2017, pp. 139–147.
[28] Z. Yang, J. Laaksonen, Multiplicative updates for non-negative projections, Neurocomput. 71 (1-3) (2007) 363–373.
[29] Y. Zhang, J. Pang, Distance and friendship: A distance-based model for link prediction in social networks, in: APWeb'15, 2015, pp. 55–66.
[30] Y. Zhong, N.J. Yuan, W. Zhong, F. Zhang, X. Xie, You are where you go: Inferring demographic attributes from location check-ins, in: WSDM'15, 2015, pp. 295–304.