Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2018.DOI

Discovering Individual Life Style from Anonymized WiFi Scan Lists on Smartphones

SHA ZHAO[1], ZHE ZHAO[1], RUNHE HUANG[2], ZHILING LUO[1], SHIJIAN LI[1], JIANRONG TAO[1], SHIWEI CHENG[3], JING FAN[3], GANG PAN[1], (Member, IEEE)

[1]Zhejiang University, Hangzhou, China (e-mail: (szhao, zhezhao, luozhiling, shijianli, taojianrong0, gpan)@zju.edu.cn)

[2] Hosei University, Tokyo, Japan (e-mail: rhuang@hosei.ac.jp)

[3]Zhejiang University of Technology, Hangzhou, China (e-mail: (swc, fanjing)@zjut.edu.cn)

Corresponding author: (e-mail: shijianli@zju.edu.cn).

This work was supported by NSF of China (No. 61802342 and No.61772460), China Postdoctoral Science Foundation under Grant No. 2017M620246 and No. 2018T110591, National Key Research and Development Plan under Grant No. 2016YFB1001203, and NSF of China (No. 61772468)

ABSTRACT The prevalence of smartphones equipped with various sensors enables pervasive capture of users' location data. WiFi scan lists on one smartphone, i.e., scan results of network in a range, can roughly indicate the physical location of the phone in a time period. Considering the close relationship between location and daily life, users' life style can be inferred from their WiFi scan lists. Given the issue of user privacy, in this paper, we explore anonymized WiFi scan lists to discover users' life style. Individual life style about mobility and important places of home and workplaces is discovered, respectively, based on the stay places extracted from anonymized WiFi scan lists and the reconstructed mobility trajectories. We first learn the life style about mobility by detecting activity areas from mobility trajectories and introducing two metrics of activeness and diversity to measure individual mobility. Then, we discover the life style about the home and workplaces identified from anonymized WiFi scan lists, such as stay duration at home, activeness of going outside at night, and working hours on weekdays and weekends. Experiments were conducted on a real-world large-scale dataset, which contains records of smart phone usage of more than 17,000 volunteering participants. Our work is a promising step towards automatically discover people's life style from anonymized smartphone data.

INDEX TERMS Anonymized WiFi scan lists, life style, mobility, smartphones.

I. INTRODUCTION

NOWADAYS, more than 2 billion people worldwide have been covered by smartphones [1], which are becoming people's essential belongings [1], [2]. Smartphones are going almost everywhere with their owners, recording whereabouts of users. The location of smartphones can be captured as with the rapid development in location-acquisition technologies, such as GPS, GSM network, and WiFi MAC AP (Access Point). Moreover, some specific applications, such as Lausanne data collection campaign (LDCC) for Mobile Data Challenge [3] and Device Analyzer [4], are developed to collect the captured location data. A continuous collection of individual history location data for a long duration provides detailed records on user mobility in daily lives. The ubiquitous and large-scale individual history data derived from smartphones brings us opportunities as well as challenges to discover valuable knowledge.

In our daily lives, most of human activities are linked directly or indirectly to geographic location. Considering the close relationship between everyday life and geographic location, it is claimed that one's general life style can be discovered from his/her history location. Compared with history location, life style represents a higher level of knowledge, reflecting the way a person or a group live. For example, does one user visit diverse places or tend to return to frequently visited places? Is he/she a homebody who stays at home for long hours every day? Does he/she often go outside at late

[1]https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

night? How many working hours in one day? Does he/she often work extra hours on weekends and holidays?

Discovering users' life style enable new and targeted service opportunities based on their location and other location-specific variables such as time of day, which can lead to enhanced user experience, improved life quality, potential business benefits, etc. More specifically, knowing users' life style from their geographic location has a manifold of application scenarios, such as smart recommendation system, context-based computing system, smart environment, and personalized services. For example, popular tourist attractions can be recommended to one user who often visits new and diverse places, economical tickets or hotels as well. For the users who prefer to stay at home for relatively long hours every day, we could suggest some popular movies, TV shows, or books to them for entertainment and killing time at home. For the users who frequently go outside in the evening or at late night, navigation or taxi services could be recommended to them at the right time. Meanwhile, the discovered knowledge about life style can be used for improving human models, such as SmartShadow [5] and CyberI [6].

There have been several studies on discovering users' life style from historical location captured by smartphones. They analyzed the smartphone location data captured through GPS (Global Positioning System), CDR (Call Detail Record), WiFi, and GSM (Global System for Mobile Communication), to discover important places in daily life, mine individual mobility patterns, and infer daily routines and social ties. For example, Jiang *et al.* identified users' daily mobility networks and extracted mobility patterns from raw CDRs [7]. They discovered daily motifs consisting of one to four places, and analyzed patterns of tours and trip-chaining behaviors in daily mobility networks. Diao *et al.* inferred activity patterns from mobile phone traces, such as working at home, eating meal outside of home, outdoor recreation, and routine shopping [8]. Zhao *et al.* discovered individual's life regularity from anonymized WiFi logs, such as the visiting orders of different places [9]. Nguyen *et al.* measured the social similarity among mobile phone users by analyzing their cell tower logs and Bluetooth proximity traces, and determined social groups among individuals in human society [10].

Many studies have learned users' life style using history location data derived from GPS and GSM network. However, there are some limitations of GPS and GSM network data. To be specific, it was found that the continuous operation of GPS at a fine temporal and spatial scale can shorten battery lifetime of mobile phones significantly [11]. Although GSM saves cell phone battery, it cannot provide precise localization information. Compared with GPS and GSM network, WiFi logs of smartphones are relatively easily to be collected, such as WiFi AP MAC addresses. Taking advantage of the rapid growth in recent years of wireless access points in urban areas, localization using WiFi has been becoming popular. In this work, we attempt to discover individual life style from *WiFi scan lists*. WiFi scan list refers to a list of WiFi APs that is scanned on smartphones when one user visits a place. The WiFi APs are periodically scanned and change as one user moves among different places, and the WiFi scan list changes accordingly. Thus, the WiFi scan lists can roughly indicate the physical location of the phone in a certain time period, providing us an opportunity to discover users' life style.

Given the issue of user privacy, we explore anonymized WiFi scan lists. Although the anonymized WiFi scan lists make it more challenging to discover users' life style, the users' privacy is protected. The location data used in some current studies are sensitive to users' privacy, such as GPS data with detailed longitude and latitude. Mobility data is among the most sensitive data currently being collected [12]. Blumberg *et al.* pointed out that, a list of potentially professional and personal information could be inferred about an individual knowing only his mobility trace [13]. Four spatio-temporal points are enough to uniquely identify 95% of individuals in a dataset, where the location of the individual is specified hourly [12].

In this paper, we address the discovery of people's life style from anonymized WiFi scan lists. To this end, we first extract stay places from anonymized WiFi scan lists and reconstruct one's mobility trajectories by building a mobility graph. Then, we learn the life style related to mobility and the important places of home and workplaces. A large-scale real-world dataset of WiFi scan lists collected by Device Analyzer [4] was used for evaluation, containing over 17,000 participants from more than 150 countries in the world.

The contributions of this work are three-fold:

1) We discover people's life style from anonymized WiFi scan lists. Although the anonymous data makes it more challenging, we can still discover individual life style. Our work is a promising step towards automatically mining people's life style from anonymized mobility data captured by smartphones.
2) We measure users' mobility in terms of activeness and diversity based on the activity areas that are detected from the mobility graphs through the method of community detection.
3) We learn the life style related two important places in our daily lives: home and workplaces, which are identified from anonymized WiFi scan lists.

The remainder of the paper is organized as follows. In Section II, the related work is reviewed. An overview about the data used in this study is given in Section III. Section IV describes the approach to discovering life style from anonymized WiFi scan lists. Section V describes how to extract users' stay places and reconstruct mobility trajectories. The life style about mobility is learned in Section VI. In Section VII, we discover the life style about home and workplaces. Our conclusion and discussion on the future work are given in Section VIII.

II. RELATED WORK

The increasing availability of individual location history data has brought about many relevant studies in the last decades, such as extracting significant places, mining trajectories or mobility, detecting social events, inferring social relationship between users, and discovering daily routines.

The studies in [14]–[17] identified significant places from smartphone traces. For example, Isaacman *et al.* identified generally important locations and discerned semantically meaningful locations, such as home and workplaces from anonymized cellular network [14]. Do *et al.* addressed the problem of automatic place labeling based on co-occurrence of WiFi AP and GPS data [15]. The frequently visited places were recognized reliably (over 80%) while it was much more challenging to recognize infrequent places. Montoliu *et al.* discovered places-of-interest from location information integrating GPS, WiFi, GSM and accelerometer sensors [16]. Dousse *et al.* inferred significant places directly from a set of raw WiFi fingerprints by a density based clustering approach [17].

Among [18]–[22], users' mobility or movement was mined from mobile phone data. Mobility patterns exhibit time-of-day periodicity and strong location preference [18]. It was discovered that people tend to follow simple reproducible patterns. Alhasoun *et al.* inferred individual home/ work locations by analyzing users' CDRs, then investigated the formation of segregated communities based on users' home and work locations, and estimated people flows within the city within a day time scale [20]. Bayir *et al.* used cellular networks of real-world cell phone data to analyze human mobility in city-wide level. They found that a total of 15% of a cell phone user's time is spent on average in locations that each appears with less than 1% of total time [21].

More recently, studies have focused on the macroscopic structure of mobile devices such as communities of nodes that meet each other frequently. Such community structure is generally assumed because of the social nature of human mobility. For example, Hossmann *et al.* found that mobility shows typical small world characteristics by representing a mobility scenario as a weighted contact graph, and analyzing the structure of a scenario using tools from complex network analysis and graph theory [23]. Liang *et al.* studied the impact factors that may affect the regularity and variability of human mobility patterns using social network analysis [24]. They showed that lots of factors such as environmental, temporal and age factors contribute to the shape of human mobility patterns. Many studies have analyzed the location data to detect social events and infer social relationship [23]–[27]. For example, Traag *et al.* introduced a Bayesian location inference framework to detect social events in massive mobile phone data [25]. Calabrese *et al.* showed that there is a strong correlation in that: people who live close to an event are preferentially attracted by it; events of the same type show similar spatial distribution of origins [26]. Zignani *et al.* deduced social relationships from traces by projections of the entire system node, geo-community on nodes [27].

It has been shown that users' daily routines were discovered from location data [9], [28]–[30]. For example, Farrahi *et al.* automatically discovered location-driven routines from the day in the life of a person without any supervision [28]. The discovered daily routines include "going to work at 10am", "leaving work at night", "being at home in the mornings and evenings while being out in the afternoon" and so on. Yang *et al.* mined individual life pattern based on GPS data collected by GPS devices [29], and focused on significant places of individual life. Ying *et al.* discovered trajectory patterns of users, namely GTS (Geographic-Temporal-Semantic) patterns, to capture frequent movements by taking into account users' geographic-triggered intentions, temporal-triggered intentions and semantic-triggered intentions [30]. Zhao *et al.* discovered people's life patterns, capturing individual's life regularity from anonymous WiFi scan lists [9].

To summarize, there have been many studies to learn users' mobility, life patterns and social relationship from history location data. However, there are some limitations. First, the scale of some location data was limited. To be specific, some data was collected from a specific population, in limited regions, and in a relatively short duration. Second, some mobility data used in aforementioned works was sensitive to users' privacy, such as GPS data with longitude and latitude. In this paper, we explore anonymized WiFi scan lists to discover people's life pattern, and conduct experiments on a large-scale real-world dataset that was collected from over 17,000 volunteers from all over the world.

TABLE 1: Sample of anonymous WiFi scan lists in the dataset.

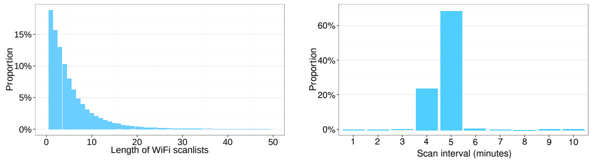
	Time	
	2013-01-25T12:20:55.576	ScanComplete: **5**
1	2013-01-25T12:20:55.576	Anonymized AP: 4724809ed90f2825cf0cec381c
2	2013-01-25T12:20:55.579	Anonymized AP: f5d92bf14c73c58d6a2e3706ad
3	2013-01-25T12:20:55.580	Anonymized AP: 6ed268b401af7995197a7aad57
4	2013-01-25T12:20:55.580	Anonymized AP: 7b6d307fe4021bf0d13b933af1
5	2013-01-25T12:20:55.580	Anonymized AP: a1272298ee210dacd5adf639aa

III. DATA OVERVIEW AND PREPROCESSING

A. DATA OVERVIEW

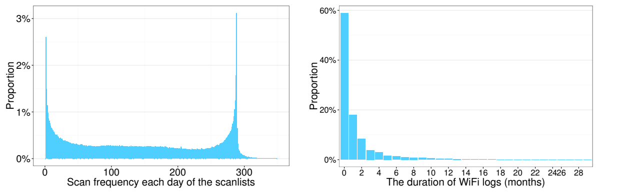
The dataset used in this work is called Device Analyzer dataset, a collection of smartphone usage of more than 17,000 volunteering participants from over 150 countries. The volunteers have installed a copy of Device Analyzer from the Android market and consented to their data being collected [4]. Device Analyzer collects a rich, highly detailed time-series log of approximately 300 different events, and there are 37 attributes of phone data in total, including CDR, app usage, alarm, battery, WiFi, etc.[2]. The dataset has been made publicly available for the first time in the context of the UbiComp/ISWC 2014 Programming Competition. Here, we focus on the anonymized WiFi scan lists, each of which consists of the scanned time and anonymized AP MAC addresses. A sample of the WiFi scan lists is shown in Tab. 1. 'ScanComplete' is a marker that indicates that a WiFi scan finished. The value contains the number of visible APs,

[2]http://deviceanalyzer.cl.cam.ac.uk/keyValuePairs.htm



(a) (b)

FIGURE 1: The length and scan interval of the WiFi scan lists. a) the length, b) the scan interval.



(a) (b)

FIGURE 2: a) The scan frequency of the scan lists and b) the duration of the WiFi logs of all the users.

which follows immediately after this marker. In the sample, the scan list consists of 5 anonymized scanned APs.

According to our observation, there are more than one visible AP in most scan lists. We define the number of the APs in each WiFi scan list as the *length* of the WiFi scan list. Fig. 1(a) shows the the distribution of the WiFi scan lists' length of all the users in their duration. The x-axis is the length of WiFi scan lists, and the y-axis is the proportion of the scan lists with the exact length. As we can see, about 19% of the scan lists have only one AP, and about 16% consist of 2 APs. Almost 85% of the scan lists consist of less than 10 APs, while only about 3% consist of more than 20 APs. We also investigated the scan interval, shown in Fig 1(b). Most scan lists are scanned less than every 10 minutes. As we can see, about 85% of scan lists are scanned every five and four minutes.

B. PREPROCESSING

We first performed preprocessing work and filtering. Fig. 2(a) shows the scan frequency of each day for the selected users, , i.e., how many WiFi scan lists are scanned in one day. The horizontal axis is the scan frequency in each day, and the vertical axis is the proportion of the days with the exact scan frequency. As we can see, there are two obvious peaks, which are 1 and 288, respectively. The scan frequency in the range from 20 to 270 distributes evenly. Few days are with scan frequency more than 300. The scan frequency of each day depends on the on-off state of Device Analyzer and WiFi subsystems. Some users may occasionally launch Device Analyzer someday and quit it immediately, and their WiFi logs are scanned only once on many days. For the users who always keep both of the WiFi subsystem and Device Analyzer on, the WiFi logs are scanned 288 times (one scan every 5 minutes, and 12 scan times in one hour). According to our calculation, there are about 80% of the days with scan frequency higher than 40. Considering low scan frequency makes the data sparse, we remove the days on which the number of scan lists is less than 40 for each user.

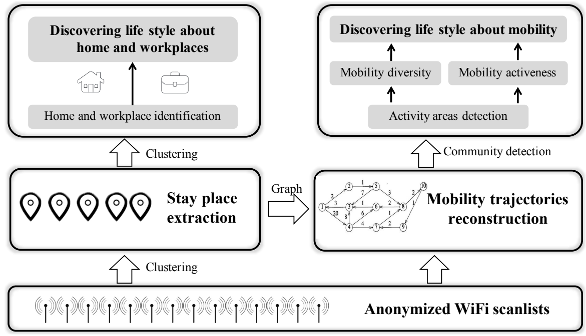


FIGURE 3: Framework to discovering life style from anonymized WiFi scan lists.

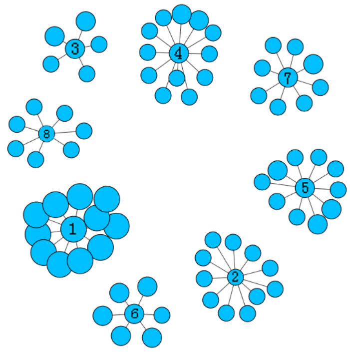
According to our observation, the duration of WiFi logs of different participants varies from a day to nearly three years, since they join the data collection program at different time. As shown in Fig. 2(b), around 90% of the users have contributed data less than six months. Here, we consider users whose duration is more than six months, and there are 746 users in total, which will be used in the following experiments. In this work, we focus on the WiFi logs on weekdays, since the patterns of workdays are more representative [29].

IV. APPROACH TO DISCOVERING LIFE STYLE FROM ANONYMIZED WIFI SCAN LISTS

In order to discover individual life style, it is necessary to first extract stay places for each user from his/her anonymized WiFi scan lists. Then, incorporating the time stamp of each stay place, we reconstruct one user's trajectories to depict his/her mobility by building a graph. Based on the extracted stay places and the trajectories on the mobility graph, the life style related to mobility and the important places of home and workplaces is learned, respectively. To learn the life style related to mobility, we detect each users' activity areas from each user's mobility graph through the idea of community detection. In each activity area, one user frequently moves among the places for specific requirements. We then measure users' mobility by introducing two metrics of mobility diversity and activeness based on the activity areas. For the discovery of life style related to home and workplaces, we first identify home and workplaces from the extracted stay places by the method of clustering, and then analyze the life style, such as stay duration at home every day, activeness of going outside at night, and working hours on workdays and weekends. Fig. 3 shows the framework to discovering individual life style from anonymized WiFi scan lists.

V. RECONSTRUCTING MOBILITY TRAJECTORIES

A WiFi scan list is recorded periodically. It roughly indicates the physical location of the phone at the time. In this section, we first extract stay places from WiFi scan lists, and then depict the mobility trajectories by building a graph.



(a)

Beginning	Ending	Stay place ID
2013/2/11 0:02	2013/2/11 08:12	1
2013/2/11 08:32	2013/2/11 10:02	4
2013/2/11 10:22	2013/2/11 10:47	4
2013/2/11 11:48	2013/2/11 12:13	2
2013/2/11 13:29	2013/2/11 14:04	4
2013/2/11 14:09	2013/2/11 15:40	5
2013/2/11 16:04	2013/2/11 17:28	2
2013/2/11 18:02	2013/2/11 19:27	2
2013/2/11 19:46	2013/2/11 23:57	1

(b)

FIGURE 4: The results of stay place extraction from User A's anonymized WiFi scan lists. a) The 8 selected stay places, b) A sample of User A's mobility records.

A. EXTRACTING STAY PLACES

In the real-world, some places are frequently visited while some are visited less often or only fleetingly. For stay place detection, a big challenge is how to deal with the intermittent APs during users' moving. We take one's stay places that has been visited for a significant amount of time into consideration and ignore those places where the user just passes by. Here, we remove the WiFi APs that appear less than 10 minutes. A stay-place p_i can be detected when the same AP MAC addresses appear in the consecutive WiFi scan lists. In a long duration at the same place, one's WiFi scan lists are similar, consisting of the APs around the place. The scan lists can be taken as coordinates of the physical place. Mapping continuous coordinates onto a discrete set of places is nothing else than a clustering exercise [17]. Scan lists in the same cluster represent a stay place p_i and each scan list belongs to only one cluster. Here, the UIM clustering algorithm [31] is employed.

By the UIM algorithm, firstly, the scan lists that frequently appear together are selected and defined as a good set of scan lists. Sequentially, the similarity between the scan lists in the good set is computed and a similarity graph G_s is constructed. Then, a candidate cluster set C_c is obtained by clustering the vertex of the G_s. Finally, the candidate clusters are merged based on the similarity measures to obtain the set C_F of final clusters. Each cluster in C_F represents a stay place.

1) Results of stay place extraction

We took one user sample as an example, named User A, to show the results of stay place extraction. There are 95 stay places extracted from his/her anonymized WiFi scan lists, in each of which he/she stays more than 10 minutes. Here, we selected the top 8 stay places to analyze, each of which appears more than 10% of days, since 90% of one's locations appears less than 10% of total time [21]. Fig. 4(a) shows the 8 selected stay places. Each cluster represents one stay place, in which each node represents a WiFi scan list, and the node size means the average stay duration. The numbers in the figure represent the ID of each stay place. As we can see from Fig. 4(a), User A stays at the Place 1 with the biggest node size for the longest duration in average, while he/she stays at the Place 4 and Place 8 for relatively short duration.

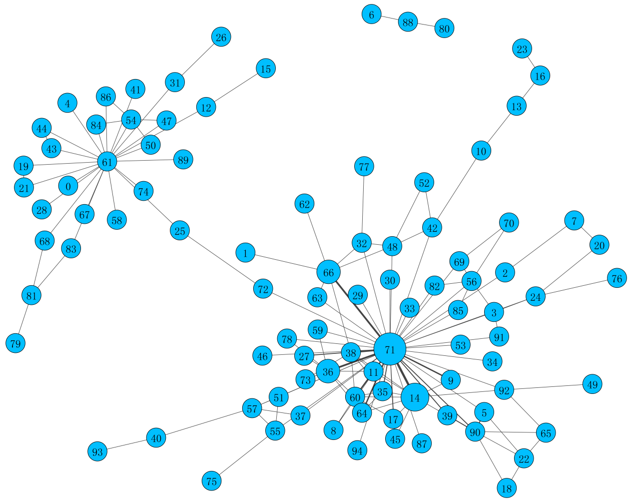


FIGURE 5: User A's mobility graph.

With the time stamp, the User A's mobility records are obtained, indicating he/she visits which place at what time. Fig. 4(b) shows his/her mobility records on Feb 11th, 2013. He/she stays in Place 1 overnight (from 0am to 8am), and in the morning leaves for Place 4. At noon, he/she stays at Place 2, and then visits Place 4, Place 5, and Place 2 sequentially. At the end of the day, he/she comes back to Place 4. By comparing two consecutive mobility records, we can calculate the stay duration at each stay place. For example, User A spends the most time at night at Place 1, while he/she stays at Place 4 for a long time at daytime.

B. BUILDING MOBILITY GRAPHS

Human mobility involves moving to and from a set of places, such as comings and goings between any two places. As a kind of visualizer, graph provides a fairly succinct way to analyze the relationship between objects in nature [32]. Users' transition among multiple places can be depicted through a weighted undirected graph, which is called mobility graph.

From one user's consecutive mobility records, how he/she moves among different places is known. If comings or goings happen between two places of p_i and p_{i+1}, the two places are connected by an arc, and the value of the arc is the number of the comings and goings between p_i and p_{i+1}. In this way, the mobility graph is built for each user sample. In each mobility graph, a stay place corresponds to a node v, and an edge e between node v to node u is established if there exists transition, with the corresponding transition times being the weight w of the edge. The degree d of a stay place v is the number of edges incident to v.

We also took User A as an example to show the mobility graph, shown in Fig. 5. Each node represents a stay place, the node size means the average stay duration at the stay place, and the numbers in the figure represent the ID of each stay place. The width of edge connecting two stay places indicates the weight of the edge, e.g. the number of the comings and goings. There are 95 stay places in total on User A's mobility graph. From Fig. 5, it can be seen there are very few places with big size and more places with smaller size, indicating User A visits many places but stays for a short duration. The size of Place 71 and Place 14 is much bigger than that of others, which means User A usually spends a great amount of time at the two places. There are not any isolated places, and each place is connected with the other places by edges. In particular, there are lots of edges directly connected to Place 71, indicating User A visits Place 71 frequently. The edge between Place 71 and Place 66 and the edge between Place 71 and Place 14 are much wider than other edges, implying User A frequently moves among these places. Considering the long stay duration and the high visit frequency, it could be inferred that Place 71 is important for User A.

VI. DISCOVERING LIFE STYLE ABOUT MOBILITY

In the real world, individuals display significant regularity on their movement, and they usually return to a few highly frequented locations [18]. The locations involved in the regular movements form an "area". Imagine a typical day in one's daily life, she leaves home for the work office at 8 am, and stays in the work office until 12 pm. Then she goes outside for lunch in a frequently visited restaurant near her workplace. After lunch she comes back to the work office and stays there until 5pm. At last, she leaves for home. The places of home, work office, and the launch restaurant are frequently visited, and the route repeats almost every workday. These places form a "living area", in which the user transfers frequently among the places for residence, food and work. Similarly, some places compose other areas, such as "business area". Most of the time we travel only over short distances, between home and work, whereas occasionally we take longer trips [18], such as business trip in a different city, and the places around the accommodation that are visited form a "business area".

In our daily life, individuals conduct different activities (e.g. living and traveling) in different areas according to their needs. The areas are defined as *activity areas*, which consist of multiple places and users frequently move among the places for specific needs. Thus, activity areas indicate the types of activities, and reflect one user's need to a certain degree. For example, the number of one user's activity areas reflect how many kinds of activities he/she conducts in daily life. The transition among different places in each activity area shows the activeness in mobility. We try to detect one user's activity areas in the daily life to understand the related life style.

A. DETECTING ACTIVITY AREAS

Intuitively, people move frequently among the stay places in the same activity area while rarely in different activity areas. It indicates that there is a higher density of edges within activity areas than between them on the mobility graph. Similarly, the weight of the edges among the places within activity areas is much bigger. The activity area can be seen as a community of nodes in the mobility graph, in which there is a higher density and bigger weight of edges. Thus, the problem of activity area discovery can be addressed by community detection on a graph.

When it comes to community detection, there exist a lot of algorithms [33]–[36]. Here, the algorithm in [37] was employed, which uses edge betweenness as a metric to find the boundaries of communities. The extra information contained in the edge weights does indeed help us enormously to discern the community structure [38]. This algorithm uses the modularity Q as the metric to measure the result of community detection, which is defined as Equation 1. Higher values of Q correspond to better divisions of a mobility graph into communities. In practice, it is found that a value above 0.3 is a good indicator of significant community structure.

$$Q = \frac{1}{2m}\sum_{vw}[A_{vw} - \frac{k_v k_w}{2m}]\delta(c_v, c_w) \tag{1}$$

where v and w are the vertices of the mobility graph, m is the number of edges, k_v and k_w are the degree of vertex v and w, respectively. The probability of an edge existing between vertices v and w if connections are made at random but respecting vertex degrees is $k_v K_w/2m$. c_v and c_w are the communities which v and w belong to, respectively. The $\delta(i, j)$ is 1 if $i = j$ and 0 otherwise. A_{vw} is an element of the adjacency matrix of the mobility graph.

A set of activity areas of the user u_i can be denoted by C_{u_i}. If there are n activity areas for u_i, C_{u_i} can be defined by: $C_{u_i} = \{c_1, c_2, \cdots c_i, \cdots, c_n\}$, where $c_i = \{\}$represents for an activity area.

1) Results of activity areas detection

By applying the algorithm [37], we discovered users' activity areas from their mobility graphs. For one user's activity areas, we focus on three aspects: the total number of the activity areas, the size of each activity area, and the number of visits to each activity area. An activity area's size refers to the number of the stay places in it. The number of visits means the number of days on which the activity area is visited. Here, we take the union of the days on which the stay places are visited, as the visit days of one activity area.

We also took User A's activity areas as an example, and $Q = 0.37$, shown in Fig. 6. There are 7 different activity areas in total, with 7 different colorful background. Each

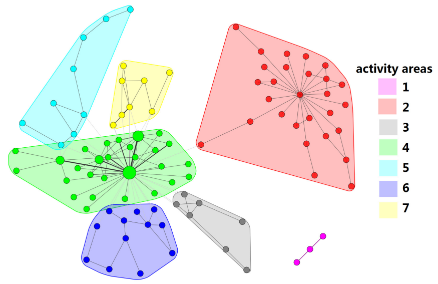
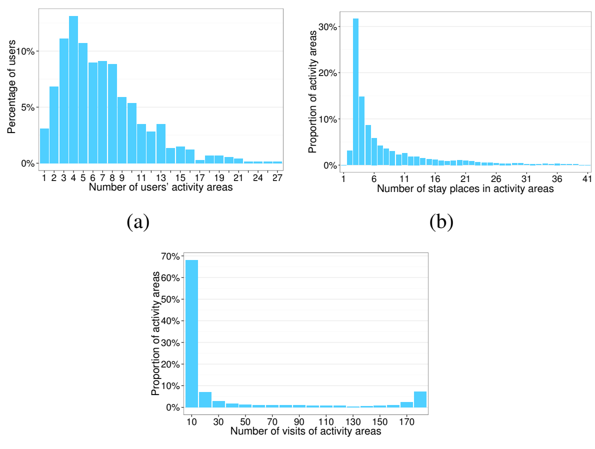


FIGURE 6: User A's activity areas discovered from his/her mobility graph.

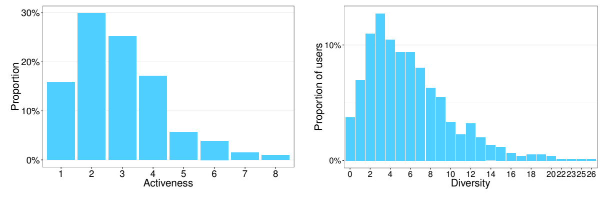


(c)

FIGURE 7: (a) The number of users' activity areas. (b) The number of places in activity areas. (c) The number of visits of activity areas.

node represents a stay place, and an edge represents the comings and goings between two stay places. The node size means the number of days on which the corresponding stay place is visited. Each stay place belongs to one activity area, and there is no overlap between any two different activity areas. It's obvious that the nodes are tightly connected in the same one activity area, while there are a few edges between two different activity areas. It indicates that User A moves frequently among the places in the same activity areas, while he/she moves rarely between the two places in two different activity areas.

Among the 7 activity areas, activity area 2 and 4 have the most stay places in them, both of which have 28 places, respectively. The activity area 1 has the fewest places, and there are only 3 places. There exists higher density of edges in activity area 2 and 4, which means User A moves frequently in the two activity areas. Activity area 2 is visited the most frequently, whose number of visits is 172 and the size of most nodes is relatively big. The activity area 1 is visited for only once. It is worth noting that the activity area 1 is the only one that is isolated from the others. It may be because the activity area is occasionally visited by the User A, e.g. a business area or travel area.



(a) (b)

FIGURE 8: The mobility activeness and diversity of users. a) the mobility activeness, b) the mobility diversity.

For the 746 user samples, there are 5,100 activity areas in total. There are differences in activity areas across users, and we investigated the number of activity areas, the size, and the number of visits, shown in 7. Fig. 7(a) shows the frequency of user samples in term of the number of activity areas. It can be seen that, there are 98 users with 4 activity areas, accounting for 13%. Most users have 3 to 8 activity areas, accounting for 62%, and only a few users have more than 20 activity areas.

Figure 7(b) shows the frequency of activity areas in term of the number of stay places. As we can see, the activity areas with 3 stay places are the most, accounting for about 30%. Around 3% of the activity areas consist of two stay places, and about 7.5% consist of more than 20 stay places. Fig. 7(c) shows the number of visits of the activity areas. About 70% of the activity areas are visited in less than 10 days. The activity areas visited in less than 90 days account for 85%, and 7% of the activity areas are visited in 180 days, which means these activity areas are the most frequently visited.

B. LIFE STYLE DISCOVERED FROM ACTIVITY AREAS

As mentioned above, the activity areas can reflect users' mobility style. For example, the weight of the edges between any two nodes in one activity area manifest the transitions between the two stay places, indicating how frequently the user moves among multiple stay places. The number of visits to one activity area implies how important it is to the user. Here, we define two metrics from the users' activity areas to measure their mobility, activeness and diversity, respectively.

1) Mobility activeness

In the real world, some people prefer to stay in the same place and rarely visit the other places, while some regularly visit a few fixed places almost every day and move frequently among them. And some often visit some new places every a few days because of various of needs. Intuitively, the latter ones are more active. In order to measure how active in mobility one user is, we introduce the indicator of activeness. We use one's average transitions between different places every day to measure his/her activeness. The activeness of one user u_i can be denoted by:

$$A_{u_i} = \frac{1}{n}\sum_{k=1}^{n} Tr_{ik} \tag{2}$$

where Tr_{ik} means the transitions of u_i on the kth day, computed by summing the weight of all the edges in and between all of the activity areas.

We investigated the mobility activeness of users, shown in Fig. 8(a). It can be seen that about 70% of users' activeness mainly lies in the range from 2 to 4, and very few users' activeness is higher than 6, accounting for 6%. It indicates most users transfer 2 to 4 times among multiple places every day, while few users transfer over 6 times. Many users transfer twice every day, accounting for 30%. According to our observation, the users with two transitions usually transfer from the stay place p_1 to p_2, and then come back to p_1 ($p_1 \rightarrow p_2 \rightarrow p_1$).

2) Mobility diversity

In the daily life, individuals conduct different activities (e.g. live and travel) in different areas [18]. The variety of activity areas implies people's diversity in mobility. According to our observation, the users have at least one significant activity area, which is for individual living, such as residence and work. Intuitively, these activity areas are not so helpful to show the mobility diversity. On the contrary, the other activity areas used for other activities, such as travel, entertainment, and shopping, are useful to measure the diversity. Thus, we used the number of the insignificant activity areas to measure one user's diversity in mobility. The number of visits to each activity area can reflect the importance of the activity area to users. As we can see from Fig. 7(c), about 85% of activity areas are visited in less than 90 days. Here, we defined the activity areas visited in less than 90 days as insignificant activity areas.

Users' diversity in mobility is shown in Fig. 8(b). The users' diversity mainly varies from 2 to 7, which means that 61% of users go to 2 to 7 activity areas for different activities except for living. About 14% of users' diversity is very low (0 and 1), indicating they are used to stay at the significant places, such as home or workplace, and rarely visit any other places. On the contrary, the diversity of very few users, accounting for less than 1% (0.93%), is very high (higher than 20), indicating the users usually go to many activity areas for different activities.

VII. DISCOVERING LIFE STYLE ABOUT HOME AND WORKPLACES

People spend most of their time at a few key locations, such as home and workplaces. A geographic location where a person spends a significant amount of time and/or she visits frequently, is defined as an important place. Examples of important places include: home, workplace, gym, grocery store, and a house of worship [14]. Focusing on these important places is helpful for understanding users' life style. Here, we focused on the important places of home and workplaces. In this section, we show how to identify home and workplace in people's lives and discover the related patterns from anonymous WiFi scan lists.

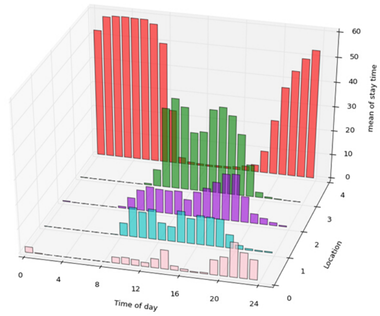


FIGURE 9: The top 5 important places of User A.

A. IDENTIFYING HOME AND WORKPLACES

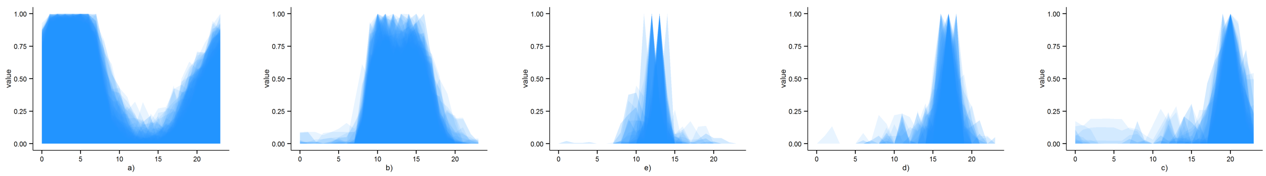
We identified home and workplaces from top important stay places rather than all of the stay places, so as to improve computation efficiency. Our approach for identifying home and workplaces consists of two steps. We first detected important places from all of the stay places, and represented each important place as a vector with 24 dimensions, indicating the average stay duration at the important place at each hour slot. Secondly, we temporally clustered the important places and then identified which clusters are home or workplaces.

1) Important places detection and representation

In order to identify home and workplaces, it is necessary to first detect important places from all of the places. In the real world, one important place is visited frequently by users, and usually taken as a hub for connecting other places. After mapping users' movements to mobility graphs, we can identify the important places by detecting the important nodes on the mobility graphs. The importance of nodes on a graph is mainly evaluated by the connection between its adjacent nodes [39], such as degree of node, shortest paths and weight. We exploited the degree of one node and betweenness to measure its importance. Degree of one node is the number of edges directly connected to it, and the betweenness is the number of shortest paths from all vertices to all others that pass through the node [40]. One place's degree is higher, and it is visited more frequently. One place's betweenness is bigger, and the probability as a hub for connecting other places is higher. The importance I_{p_i} of the stay place p_i was computed by:

$$I_{p_i} = 0.5 \times degree_{p_i} + 0.5 \times betweenness_{p_i} \qquad (3)$$

For each user sample, we selected the top 5 important places and there are 3,730 important places in total. In order to cluster important places for identifying home and workplaces, it is necessary to represent them. Intuitively, people stay at different places in different time periods. For example, people spend most time of night at home, and tend to spend most of day time at workplaces on weekdays. That is to say, there is a close relationship between the type of the place and the time periods when one user spends lots of time at the place. Thus, we represented each important place as a



(a) Cluster A (b) Cluster B (c) Cluster C (d) Cluster D (e) Cluster E

FIGURE 10: An illustration of the clustering results, 5 types of places.

vector of 24 dimensions, each of which is the average stay duration in the corresponding time slot. The stay duration was obtained by comparing two neighbor mobility records.

Taking User A as an example, we illustrated the top 5 important places and showed the average stay duration (minutes) in each place in Fig. 9. As we can see, User A spends most time at Place 4 at night (from 7pm to 7am), while he/she spends very little time here at daytime (from 8am to 6pm). On the contrary, User A spends most time at Place 3 at daytime (from 9am to 5pm), and spends very little time at this place at night and early morning (from 5pm to 7am). Place 4 is most likely User A's home, where he/she spends a significant account of time at night, and Place 3 may be his/her workplace since he/she stays here for lots of time at daytime. Similar with Place 3, the stay duration of both Place 1 and Place 2 mainly covers day time (from 9am to 5pm). But, the stay duration is a little shorter than that of Place 4. For Place 0, User A stays here mainly from 8pm to 11pm, and the average stay duration is short.

2) Clustering important places

We exploited the method of k-means to cluster the 3,730 important places to identify home and workplaces. In order to well characterize an important place, given that the value of each dimension among users varies in a wide range, each value is brought into the range [0,1] by unity-based normalization: $v_i' = \dfrac{v_i - min(v_i)}{max(v_i) - min(v_i)}$. Cosine similarity was used to measure the similarity between any two places, and we obtained 5 clusters for which the cosine distance between the points in the same cluster was the smallest while the cosine distance between the points in different clusters was the biggest. There are 707, 844, 1187, 370, and 622 stay places in each cluster, respectively. To well understand the clustering results, for each cluster, we chose 30 places to analyze which are the nearest to the centroid. In each subfigure, we drew 30 curves for the 30 selected places, and the value of each point on the curve represents for the normalized stay duration at the specific hour at the place, shown in Fig. 10. The color at the specific hours is darker, and the stay places are visited more frequently at the corresponding hours.

For Cluster C, users visit these places around noon. Maybe the stay places are some restaurants where people have lunch. The stay places in Cluster D are usually visited around 6pm. These stay places may be grocery stores, and people may go to stores to buy some groceries or food. The stay places in the Cluster E may be some places of recreation for entertainment, such as bars and theaters, which are usually visited after dinner time, at around 8pm. For Cluster A, users spend a long duration (almost 60 minutes in each hour slot) at the stay places from 9pm to 6am. Intuitively, people tend to stay at home at evening and night. It was also found that people usually stay at home from 7pm to 7am [14]. Thus, it can be inferred the stay places in Cluster A are home. For Cluster B, users spend a long duration at the stay places from 9am to 6pm. In the real world, people usually go to work during these hours [14]. Thus, the stay places in Cluster B can be taken as workplaces.

There are 707 home places and 844 workplaces for the 746 user samples. According to our observation, we cannot discover any home places for 59 users. This may be brought about by the missing data, for example, some users may turn off the WiFi subsystem or even shut down their phones after they go to bed. On the contrary, we found two home places for 20 users. For each of them, the stay place where the user spends the most time from 9pm to 6am was selected as the unique home place. For 35 user samples, their workplaces cannot be identified, while we discovered more than one workplaces for 106 users. Similarly, we took the stay place where the users spend the most time from 9am to 6pm as the unique workplace. In this way, one user has one home place and workplace at most. There are 687 home places and 711 workplaces in total.

B. LIFE STYLE ANALYSIS ABOUT HOME

Intuitively, home is very important to people, since they spend a great amount of time at home almost every day. Understanding individual patterns at home is crucial to understand users, such as how much time in average one stays at home every day, how regularly he/she stays at home, and how often he/she stays out at night.

1) Average stay duration at home

In the real-world, users spend different amount of time at home. For each user, we computed his/her average stay duration at home, formulated by:

$$E(H) = \frac{1}{n}\sum_{i=1}^{n} h_i \tag{4}$$

where h_i means one's stay duration (hours) at home at ith day, n is the number of days on which the user stays at home.

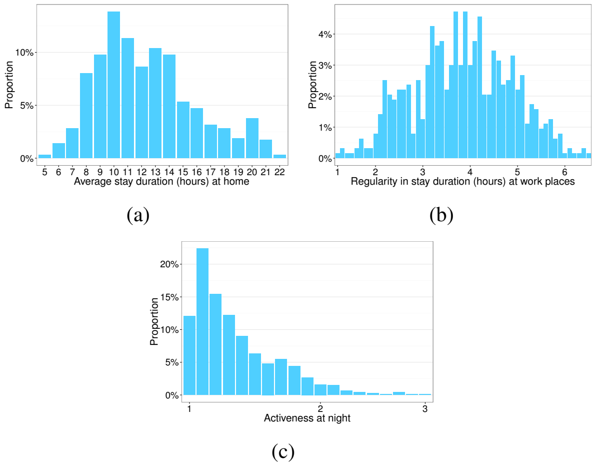


FIGURE 11: a) Users' average stay duration (hours) at home every day; b) Users' regularity in hours of staying at home; c) Users' activeness at night.

Fig. 11(a) shows the users' average stay duration at home every day. It can be seen that most users stay at home for about 8 to 14 hours every day in average, accounting for 72%. Some users stay at home for less than 6 hours every day (1.7%), who may work extra hours at workplaces, and leave for home late. On the contrary, some stay at home for over 20 hours, accounting for 5.8%, who may work at home, and rarely visit other places.

2) Regularity in stay duration at home

Some users stay at home for a long time on some days, while they spend a very short time at home on some other days. Here, we used the standard deviation $D(H)$ to measure one's regularity in hours of staying at home, formulated by

$$D(H) = \sqrt{E(H^2) - (E(H))^2} \tag{5}$$

Fig. 11(b) shows users' regularity in hours of staying at home. Most users' difference in hours of staying at home is from 2 to 6 hours, accounting for 81.7%.

3) Activeness at night

According to our observation, some users stay at home overnight while some visit other places at night. The number of the places visited in the evening or at late night reflect users' activeness at night. We took the average number of places visited by one user from 7pm to 0am as his/her activeness at night. For each user, we extracted the places he/she visits from 7pm to 0am every day based on his/her mobility graph. Fig. 11(c) shows the users' activeness at night. As shown, most users' activeness at night is from 1 to 2, accounting for 82%. It means most users visit 1 or 2 places in average at night. 12% of users visit only one stay place at night, and this place is their home. Few users visit more than 2 places at night, accounting for 5%. Only one user visits more than 3 places in average.

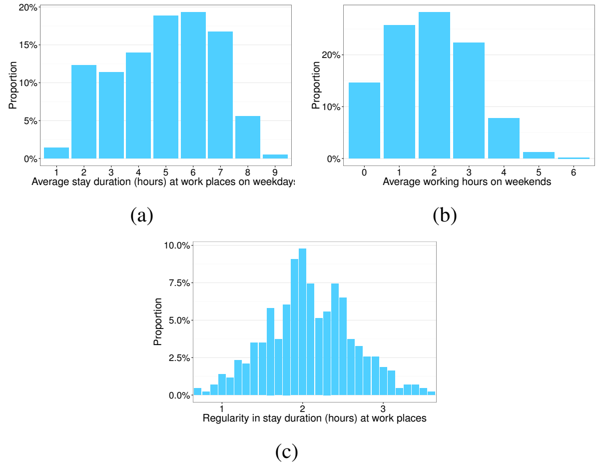


FIGURE 12: a) Users' average stay duration (hours) at their workplaces on weekdays; b) Users' average working hours on weekends; c) Users' regularity in stay duration (hours) at workplaces.

C. LIFE STYLE ANALYSIS ABOUT WORKPLACES

Workplace is also very important in our daily life. We try to understand individual patterns about workplaces in terms of how many hours users usually stay at their workplaces on weekdays and weekends, respectively, and how regularly users stay at the workplaces.

1) Working hours on weekdays

We first investigated the working hours on weekdays. It was assumed that the hours of one user staying at his/her workplace are his/her working hours. We took the time slots from 9am to 6pm as working time. For each user, his/her average stay duration at workplace on weekdays $E(W)$ was computed by $E(W) = \frac{1}{n}\sum_{i=1}^{n} w_i$, where w_i is one's stay duration at his/her workplace on the ith day, and n is the total number of workdays on which he/she visits the workplace. Fig. 12(a) shows users' average stay duration at their workplaces on weekdays. Most users spend 4 to 7 hours in average at their workplaces on weekdays, accounting for about 70%. According to our observations, these users stay at their stay places during the time periods from 10am-12pm, and 1pm-6pm. 13.6% of users spend less than 2 hours in average at their workplaces every day. These users usually move frequently among multiple places, and stay not so long at the discovered workplaces. 6% of users spend over 8 hours at their workplaces, who tend to stay at the workplaces from 9am to 6pm and rarely visit any other places.

2) Working hours on weekends

Generally, people go to their workplaces on weekdays. Sometimes, people need to work extra hours on weekends. Similarly, we investigated each user's average stay duration at workplace on weekends $E(W^*)$: $E(W^*) = \frac{1}{n}\sum_{i=1}^{n} w_i^*$, where n is the number of the weekends, and w_i^* means each user's average stay duration at their workplaces on the ith weekend. Fig. 12(b) shows users' average working hours on weekends. Lots of users stay at workplaces for 1 to 3 hours on weekends, accounting for 76%. About 15% of users don't visit their workplaces on weekends. Only 1% of users spend over 5 hours at their workplaces on weekends.

3) **Regularity in stay duration at workplaces**

We also explored the regularity in hours of staying at the workplaces on weekdays. It was measured by the standard deviation $D(W)$ of the stay duration at workplaces: $D(W) = \sqrt{E(W^2) - (E(W))^2}$. Fig. 12(c) shows the distribution of all users' regularity at workplaces. Most users' difference in hours of staying at workplaces is 2 hours, accounting for 65%. Comparing with users' regularity in stay duration at home shown in Fig. 11(b), we found that users' stay duration at workplaces is more regular.

VIII. CONCLUSION

WiFi scan list can roughly indicate the physical location of the phone in a certain time period. One's life style can be inferred from his/her WiFi scan lists, since there is a close relationship between location and our daily life. Considering the concern of user privacy issues, we explored WiFi scan lists that are anonymized, to discover one's life style about his/her mobility and the important places of home and workplaces. We first extracted stay places from anonymized WiFi scan lists for each user, and reconstructed the mobility trajectories by building mobility graphs. Based on the mobility graph, we then detected users' activity areas through the idea of community detection, and measured users' mobility by two metrics of activeness and diversity. We also learned the life style about the identified home and workplaces for users, such as how many hours in average one stays at home, activeness of going outside at night, and average working hours on weekdays and weekends. Although the WiFi scan lists are anonymized, the experiments conducted on the large-scale dataset of 17,000 users from over 150 countries showed that we can still discover the users' life style.

We emphasize, however, that due to the absence of ground truth, all our conclusions are, at best, educated guess which are based on real-world data. We believe such results are meaningful and insightful for a wide range of people such as smart service providers, and advertisers. In the future work, we will make our efforts to collect more individual information to verify our results and discover more insights.

REFERENCES

[1] S. Zhao, G. Pan, Y. Zhao, J. Tao, J. Chen, S. Li, and Z. Wu, "Mining user attributes using large-scale app lists of smartphones," IEEE Systems Journal, vol. 11, no. 1, pp. 315–323, 2017.

[2] S. Zhao, J. Ramos, J. Tao, Z. Jiang, S. Li, Z. Wu, G. Pan, and A. K. Dey, "Discovering different kinds of smartphone users through their application usage behaviors," in UbiComp '16. ACM, 2016, pp. 498–509.

[3] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila, "Towards rich mobile phone datasets: Lausanne data collection campaign," in ICPS '10. Berlin: ACM, 2010.

[4] D. T. Wang, A. Rice, and A. R. Beresford, "Device analyzer: Understanding smartphone usage," in MOBIQUITOUS '13. ACM, 2013.

[5] Z. Wu and G. Pan, SmartShadow: Models and Methods for Pervasive Computing. Springer, 2013.

[6] J. Ma, J. Wen, R. Huang, and B. Huang, "Cyber-individual meets brain informatics," IEEE Intelligent Systems, no. 5, pp. 30–37, 2011.

[7] S. Jiang, J. Ferreira, and M. C. Gonzalez, "Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore," IEEE Transactions on Big Data, vol. 3, no. 2, pp. 208–219, 2017.

[8] M. Diao, Y. Zhu, J. Ferreira Jr, and C. Ratti, "Inferring individual daily activities from mobile phone traces: A boston example," Environment and Planning B: Planning and Design, vol. 43, no. 5, pp. 920–940, 2016.

[9] S. Zhao, Z. Zhao, Y. Zhao, R. Huang, S. Li, and G. Pan, "Discovering people's life patterns from anonymized wifi scanlists," in UIC '14. IEEE, 2014, pp. 276–283.

[10] C.-B. Nguyen, S. Yoon, and J. Kim, "Discovering social community structures based on human mobility traces," Mobile Information Systems, vol. 2017, 2017.

[11] I. Constandache, S. Gaonkar, M. Sayler, R. R. Choudhury, and L. Cox, "Enloc: Energy-efficient localization for mobile phones," in IEEE INFOCOM 2009. IEEE, 2009, pp. 2716–2720.

[12] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," Scientific reports, vol. 3, 2013.

[13] A. J. Blumberg and P. Eckersley, "On locational privacy, and how to avoid losing it forever," Electronic frontier foundation, vol. 10, no. 11, 2009.

[14] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky, "Identifying important places in peopleâĂŹs lives from cellular network data," in PerCom '11. Springer, 2011, pp. 133–151.

[15] T. M. T. Do and D. Gatica-Perez, "The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data," IEEE Transactions on Mobile Computing, vol. 13, no. 3, pp. 638–648, 2014.

[16] R. Montoliu, J. Blom, and D. Gatica-Perez, "Discovering places of interest in everyday life from smartphone data," Multimedia Tools and Applications, vol. 62, no. 1, pp. 179–207, 2013.

[17] O. Dousse, J. Eberle, and M. Mertens, "Place learning via direct wifi fingerprint clustering," in MDM '12. IEEE, 2012, pp. 282–287.

[18] M. C. GonzÃąlez, C. A. Hidalgo, and A.-L. BarabÃąsi, "Understanding individual human mobility patterns," Nature, vol. 453, no. 7196, pp. 779–782, 2008.

[19] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti, "Estimating origin-destination flows using mobile phone location data," IEEE Pervasive Computing, vol. 10, no. 4, pp. 36–44, 2011.

[20] F. Alhasoun, A. Almaatouq, K. Greco, R. Campari, A. Alfaris, and C. Ratti, "The city browser: Utilizing massive call data to infer city mobility dynamics," in The 3rd International Workshop on Urban Computing, 2014.

[21] M. A. Bayir, M. Demirbas, and N. Eagle, "Mobility profiler: A framework for discovering mobility profiles of cell phone users," Pervasive and Mobile Computing, vol. 6, no. 4, pp. 435–454, 2010.

[22] H. Du, Z. Yu, F. Yi, Z. Wang, Q. Han, and B. Guo, "Recognition of group mobility level and group structure with mobile devices," IEEE Transactions on Mobile Computing, vol. 17, no. 4, pp. 884–897, 2018.

[23] T. Hossmann, T. Spyropoulos, and F. Legendre, "A complex network analysis of human mobility," in 2011 IEEE Conference on Computer Communications Workshops. IEEE, 2011, pp. 876–881.

[24] Y. Liang, X. Zhou, B. Guo, and Z. Yu, "Understanding the regularity and variability of human mobility from geo-trajectory," in WI-IAT '12, vol. 1. IEEE, 2012, pp. 409–414.

[25] V. A. Traag, A. Browet, F. Calabrese, and F. Morlot, "Social event detection in massive mobile phone data using probabilistic location inference," in PASSAT '11 and SocialCom '11. IEEE, 2011, pp. 625–628.

[26] F. Calabrese, F. C. Pereira, G. Di Lorenzo, L. Liu, and C. Ratti, "The geography of taste: analyzing cell-phone mobility and social events," in PerCom '10. Springer, 2010, pp. 22–37.

[27] M. Zignani, "Human mobility model based on time-varying bipartite graph," in WoWMoM '11. IEEE, 2011, pp. 1–4.

[28] K. Farrahi and D. Gatica-Perez, "What did you do today?: discovering daily routines from large-scale mobile data," in MM '08. ACM, 2008, pp. 849–852.

[29] y. ye, y. zheng, y. chen, j. feng, and x. xie, "Mining individual life pattern based on location history," MDM '09, pp. 1–10, 2009.

[30] J. J.-C. Ying, W.-C. Lee, and V. S. Tseng, "Mining geographic-temporal-semantic patterns in trajectories for location prediction," TIST, vol. 5, no. 1, p. 2, 2013.

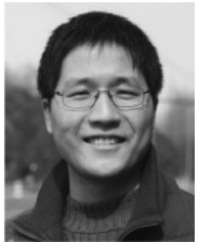
[31] L. Vu, Q. Do, and K. Nahrstedt, "Jyotish: A novel framework for constructing predictive model of people movement from joint wifi/bluetooth trace," in PerCom '11. IEEE, 2011, pp. 54–62.
[32] A. Sadri, Y. Ren, and F. D. Salim, "Summarizing movement graph for mobility pattern analysis," in Proceedings of the Knowledge Capture Conference. ACM, 2017, p. 41.
[33] L. Wan, J. Liao, and X. Zhu, "Cdpm: Finding and evaluating community structure in social networks," in ADMA '08. Chengdu: Springer Berlin Heidelberg, 2008, pp. 620–627.
[34] J. Chen, O. R. Zaiane, and R. Goebel, "Detecting communities in social networks using max-min modularity," in Proceedings of the 2009 SIAM International Conference on Data Mining, 2009, pp. 20–24.
[35] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: The state-of-the-art and comparative study," ACM Computing Surveys, vol. 45, no. 43, 2013.
[36] F. D. Malliarosa and M. Vazirgiannisa, "Clustering and community detection in directed networks: A survey," Physics Reports, vol. 533, no. 4, pp. 95–142, 2013.
[37] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," Physical Review E, vol. 70, no. 6, p. 066111, 2004.
[38] M. E. Newman, "Analysis of weighted networks," Physical Review E, vol. 70, no. 5, p. 056131, 2004.
[39] R. S. Burt, M. J. Minor, and R. D. Alba, Applied network analysis: A methodological introduction. Sage Publications Beverly Hills, CA, 1983.
[40] L. C. Freeman, "A set of measures of centrality based on betweenness," Sociometry, pp. 35–41, 1977.

SHA ZHAO is currently a Postdoctoral Research Fellow of the College of Computer Science and Technology, Zhejiang University. She won the Best Paper Award of ACM UbiComp'16. Zhao visited the Human-Computer Interaction Institute at Carnegie Mellon University as a visiting PhD student from September 2015 to September 2016. She received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2017. Her research interests include pervasive computing, mobile sensing, data mining, and machine learning.

ZHE ZHAO received his B.Sc. in computer science from Zhengzhou University, Zhengzhou, Chian, 2013, and M.S in computer science from Zhejiang University, Hangzhou, China, in 2016. His research interests include machine learning and data mining.

RUNHE HUANG received the Ph.D. degree in computer science and mathematics from the university of the West of England, Bristol, U.K., in 1993. She is a Full Professor withe the Faculty of Computer and Information Sciences, Hosei University, Tokyo, Japan. Her research fields include artificial intelligence, computational intelligence computing, ubiquitous intelligence computing, and big data, machine intelligence.

ZHILING LUO is an Assistant Professor in Computer Science at Zhejiang University, China. He received his B.S. and Ph.D. degree in Computer Science from Zhejiang University in 2012 and 2017, respectively. He was the visiting scholar of Georgia Institute of Technology, US, in 2016. His research interests include service computing, machine learning and data mining.



SHIJIAN LI received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2006.In 2010, he was a Visiting Scholar with the Institute Telecom SudParis, Evry, France. He is currently with the College of Computer Science and Technology, Zhejiang University. He was published over 40 papers. His research interests include sensor networks, ubiquitous computing, and social computing. Dr. Li serves as an Editor of the International Journal of Distributed Sensor Networks and as Reviewer or PC Member of more than ten conferences.

JIANRONG TAO received his B. Sc. degree in computer science from Huazhong University of Science and Technology, Wuhan, China, in 2014, and received M.S in computer science from Zhejiang University, Hangzhou, China, in 2017. His research interests include machine learning and data mining.

SHIWEI CHENG received B.Sc. degree in industrial design from Northwestern Polytechnical University, Xi'an, China, in 2004; and Ph.D degree in computer science from Zhejiang University, Hangzhou, China, in 2009. He is currently a Associate Professor with the College of Computer Science, Zhejiang University of Technology. He was with the Carnegie Mellon University, Pittsburgh, USA, as a Postdoctoral Research Fellow from 2012 to 2014. His research interests include human-computer interaction, ubiquitous computing and collaborative computing.

JING FAN received B.Sc., M.S, and Ph.D degrees in software engineering and computer science from Zhejiang University, Hangzhou, China, in 1990, 1993 and 2003, respectively. She is currently a Professor with the College of Computer Science and Technology, Zhejiang University of Technology. She was with the St. Francis Xavier University, Canada, and the University of Aberdeen, UK, as a Visiting Scholar from 2008 to 2009 and 2011, respectively. Her research interests include service computing, virtual reality and human-computer interaction.

GANG PAN received B.Sc. and Ph.D. degrees in computer science from Zhejiang University, Hangzhou, China, in 1998 and 2004, respectively. He is currently a Professor with the College of Computer Science and Technology, Zhejiang University. He was with the University of California, Los Angeles, Los Angeles, CA, USA, as a Visiting Scholar, from 2007 to 2008. He has published more than 100 refereed papers. His research interests include pervasive computing, computer vision, and pattern recognition.