SNOW: Detecting Shopping Groups Using WiFi

Jiaxing Shen, Jiannong Cao, *Fellow, IEEE*, Xuefeng Liu, and Shaojie Tang

***Abstract*—Detecting shopping groups is gaining popularity as it enables various applications ranging from marketing to advertising. Existing methods exploit WiFi probe requests to detect shopping groups by identifying co-located customers. However, the probe request is prone to suffer from device heterogeneity which might pose a severe data sparseness problem. More importantly, we find that a certain amount of shopping groups would separate sometimes which makes traditional methods unreliable. In this paper, we propose a shopping group detection system using WiFi (SNOW). Instead of collecting probe requests, SNOW utilizes the WiFi data from smartphones associated with the deployed access points (APs). We could thus obtain data from different devices and even ensure a data granularity of seconds using Arping. Besides, we exploit an effective heuristic extracted from two observations of shopping group dynamics to improve the detection performance. First, the probability of group separation differs in diverse areas. Second, the proportion of group participation and individual engagement differs in different activities of the mall. Therefore, APs under which shopping groups appear more frequently and barely separate should contribute more in measuring customer similarity. Lastly, we represent the measured similarity into a matrix format and apply matrix factorization with a sparsity constraint to derive grouping results directly. According to our experiments in a large shopping mall, SNOW improves the detection performance of baseline approaches by 13.2% on average.**

***Index Terms*—Group detection, shopping group, WiFi.**

I. Introduction

DETECTING shopping groups is not only the foundation of many areas but also an enabler of various applications ranging from marketing to advertising [1]. The insights of shopping groups can help retailers to provide a context-specific incentive to potential customers on the one hand and add more intelligence to their business analytics on the other [2].

A *group* refers to people with similar properties or behaviors like network association histories [3] and mobility patterns [1], [4]. *Detecting shopping groups* is defined as a task to cluster a set of customers into disjoint subsets.

Manuscript received November 11, 2017; revised March 21, 2018 and April 23, 2018; accepted May 18, 2018. Date of publication May 22, 2018; date of current version November 14, 2018. This work was supported in part by the Hong Kong Polytechnic University for the Project of Strategic Importance under Grant 1-ZE26, in part by the RGC General Research Fund under Grant PolyU 152199/17E, in part by the National Natural Science Foundation of China (NSFC) Key Grant under Project 61332004, and in part by the NSFC under Grant 61572218. *(Corresponding author: Jiaxing Shen.)*

J. Shen and J. Cao are with the Department of Computing, Hong Kong Polytechnic University, Hong Kong (e-mail: csjshen@comp.polyu.edu.hk; csjcao@comp.polyu.edu.hk).

X. Liu is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China (e-mail: csxfliu@gmail.com).

S. Tang is with the Naveen Jindal School of Management, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: shaojie.tang@utdallas.edu).

Digital Object Identifier 10.1109/JIOT.2018.2839525

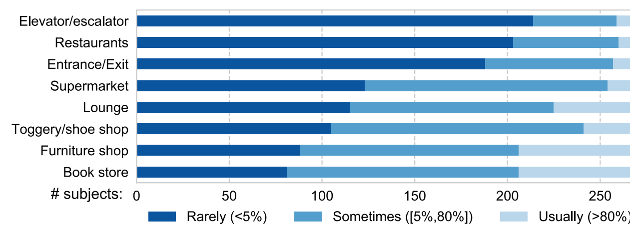


Fig. 1. Answers to the online survey problem: "How often will you get separated with your companion(s) in these regions?"

Some existing works detect co-located people as groups with *probe requests* (or probes) [1], [4]. Those probes are broadcast by smartphones to seek information about nearby access points (APs). Received signal strength indicator (RSSI) contained in probes could be used to represent smartphone users' mobility information. Compared to other approaches, the probe method requires neither high deployment cost (e.g., deploy cameras [5], [6]) nor user intervention (e.g., carry wearable devices or install mobile applications [2], [7], [8]).

However, the probe method might have two difficulties in detecting shopping groups. The first problem results from the probe request itself. Pervasive as the probe is, it suffers from multiple issues like MAC randomization [9], meaningless devices [10], and especially device heterogeneity [9]. The timing of sending probes are mainly determined by user-device interaction and the internal mechanism of the device. Therefore, different devices might generate data with various granularities which makes it difficult to measure customer similarity [9]. The second problem arises from the shopping group. Existing group detection methods assume group members always stick together while shopping groups might sometimes get separated. According to our online survey of 268 subjects, most group customers are often separated with their companions in the mall, especially in bookstores. The detailed answers to the survey problem are shown in Fig. 1. This fact requires group detection methods can not only distinguish strangers who are close to each but also identify groups that might disperse.

We ask the following question: *Can we reliably detect shopping groups using WiFi?* In this paper, we provide an affirmative answer by proposing SNOW. Instead of sniffing controversial probe requests, we collect the *WiFi data* from customers who associate with the deployed APs. WiFi data refers to the information contributed by any captured wireless traffic. According to a survey [11], over 75% people use public WiFi, which indicates the WiFi data is also pervasive enough

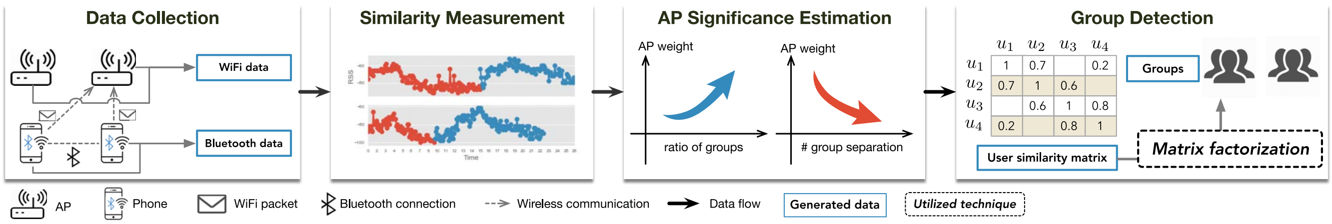


Fig. 2. Overview of SNOW.

for many application scenarios. Due to extra wireless packets, WiFi data could derive more continuous information for different devices. As WiFi data comes from only connected customers, extra efforts to handle MAC randomization and remove meaningless devices could be exempted. To handle the dynamics of shopping groups, we exploit an effective heuristic derived from two key observations.

1) *Observation I:* The probability of group separation differs in diverse areas, which is also reflected in the survey results in Fig. 1. We think this happens might due to different interests of group members.
2) *Observation II:* The proportion of group participation and individual engagement differs in different activities of the mall.

It is reported when considering whether to engage in hedonic and public activities like going to a movie alone, individual consumers anticipate negative inferences from others about their social connectedness that reduce their interests of engaging in such activities [12].

We show a toy example in Fig. 3 to highlight the main idea of SNOW. We could see that less group separation occurs in the cinema (Observation I), indicating there would be less false negative detections (groups are detected as strangers). Besides, the ratio of engaged groups over individuals is higher in the cinema (Observation II), showing the probability of false positive detection (strangers are detected as groups) would be smaller. Therefore, customer similarity measured in the cinema is more important than that of the bookstore. Accordingly, the AP deployed in the cinema (a_1) should bear higher importance than other APs.

The vision of SNOW, however, entails significant challenges when applied to real conditions. First, it is difficult to compare customers' WiFi data directly since the data from different devices are usually defined on different time instants with different lengths. Besides, other issues like packet loss and not sending any packets make it even difficult to measure the customer similarity. Second, the measured similarity could be incomplete and noisy which might lead to false detections like detecting strangers as a group and vice versa.

To address the first challenge, we propose a three-step data preprocessing including time interpolation, noise filtering, and noneffective instant removal. After the preprocessing, issues like packet loss and not sending any packets could be appropriately addressed. Then we could measure pairwise customer similarity based on the WiFi data. For the second challenge, we propose to represent customer similarity into a matrix format and apply matrix factorization (MF) to derive the grouping results. The advantages are twofold. First, MF

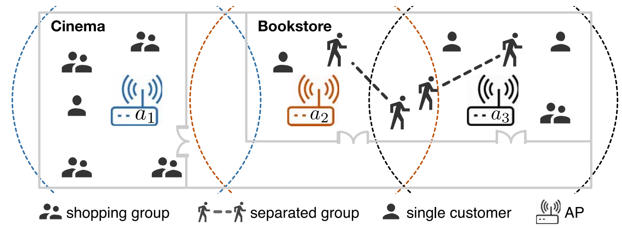


Fig. 3. Toy example for illustrating the main idea of SNOW.

is a popular approach for noise filtering and data completion by decomposing the input matrix into several factor matrices. Second, MF with sparsity constraint is an alternative for clustering. Therefore, we could directly derive the grouping results without extra clustering processes by imposing the sparsity constraint to the decomposition.

According to our experimental evaluation in a large mall, SNOW could achieve robust and reliable performance in detecting shopping groups. Compared to baseline approaches, SNOW improves the performance of detection by 13.8% and 12.6% on labeled and semi-labeled datasets, respectively.

The contributions of this paper are summarized as follows.

1) We extract an effective heuristic from observations of shopping group dynamics that could significantly improve the detection performance.
2) We propose a general three-step preprocessing method for processing the WiFi data.
3) We evaluate the proposed system using data collected in a large shopping mall for three weeks.

The remainder of this paper is organized as follows. We present design details of the proposed system in Section II. Section III demonstrates results of the experimental evaluation. Related work is introduced in Section IV. We provide further discussion on issues that might be unclear in Section V. Finally, we conclude this paper in the last section.

II. System Design

In this section, we elaborate on design details of SNOW. Fig. 2 shows an overview of the proposed system which consists of four main components. First, data collection is to collect the WiFi data from customers who associate with the deployed APs. We also collect the Bluetooth data from volunteers to estimate the significance of different APs. Note the Bluetooth data is not required when the system is in service. Second, we measure pairwise customer similarity based on their WiFi data in similarity measurement. Third, AP significance estimation exploits both WiFi and Bluetooth data to evaluate AP weights. Last, the customer similarity is refined combining both the AP weights and the WiFi data. We further represent the refined similarities in a matrix and apply MF to detect shopping groups.

A. Data Collection

1) WiFi Data: The WiFi data refers to the information contributed by any wireless traffic of connected customers. It would not violate customers' privacy since only nonsensitive information in the packet header is used. Compared to probe requests, the WiFi data has two advantages. First, smartphones use the MAC randomization mechanism to protect user privacy nowadays [9]. Probe requests might also come from passers-by outside the mall. Therefore, it usually requires extra processes for probe methods to handle MAC randomization and filter out passers-by. While these efforts could be exempt for the WiFi data as only consumers would take the initiative to connect to the deployed APs in the mall. Second, the timing of sending probes are mainly determined by user-device interaction and the internal mechanism of devices. Generally, Android devices send more probes than iOS devices and devices with old operating systems send more probes. [9]. Therefore, probe requests might generate sparse data with different data granularity. For the WiFi data, however, it is always available for both iOS and Android devices once connected to the AP.

We exploit off-the-shelf WiFi APs to collect and store the WiFi data and upload them to a server daily. Each AP works under OpenWrt (a GNU/Linux distribution for embedded devices) and uses IW (a tool for managing wireless configuration) to collect WiFi data from connected devices. An example of using IW is "IW *interface* station dump" (*interface* is the wireless interface of the AP). We extract two fields from the output of IW. The first field is *signal* which indicates RSSI. The second field *inactive time* refers to an interval since receiving the last packet. We execute the command every second to extract both fields and record the signal information when it updates.

However, when the smartphone is not used, the granularity of WiFi data could be unsatisfying. To overcome this issue, we use Arping to force the connected device to send packets more frequently. Arping is a tool for discovering an MAC address given an IP address. IP addresses of the connected devices could be found in ARP table of DHCP lease. This act might cause more energy consumption and we further discuss this issue in Section V. According to our experiments under laboratory environment, Arping could on average boost the data granularity by 43% for devices with different systems and ensure a data granularity of seconds.

2) Bluetooth Data: We also collect the Bluetooth data from volunteers' smartphones for significance estimation of different APs. When the system is applied in service, this data is not required for customers.

We develop an Android application to scan and record its associated network information and nearby Bluetooth devices every 30 s. The interval is determined empirically as it takes several seconds to complete the scan process. The app collects

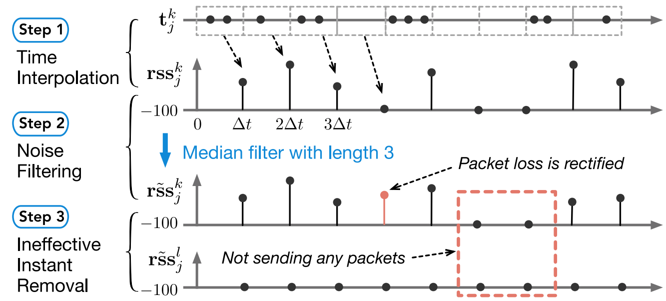


Fig. 4. Three-step preprocessing of the WiFi data.

data entries in the format of [*time*, *local_address*, *associated_MAC*, *scan_info*]. The parameter *local_address* is the Bluetooth address of the device, *associated_MAC* refers to the MAC address of the connected AP, and *scan_info* is a list of scanned device addresses and corresponding RSSI. Given a group of two users, if their Bluetooth RSS is smaller than a threshold for a certain period of time, then it is regarded as a group separation. The settings of the RSS threshold and the period might vary in different scenarios. According to our experiments, the performance peaks when the RSS threshold is set to −90 and the period is set to 2 min.

B. Similarity Measurement

1) Data Preprocessing: The WiFi data from different devices usually have different lengths. They might be defined on different time instants. Also, conditions like packet loss and not sending any packets should be properly handled. To address those issues, we propose a three-step data preprocessing: 1) time interpolation; 2) noise filtering; and 3) noneffective instant removal. An illustration of the preprocessing procedure is shown in Fig. 4.

In step 1, we align the WiFi data by generating RSS vectors on an equally spaced time instants. As shown in Fig. 4, $\mathbf{P}_j^k$ represents all packets received by AP k from smartphone j. We generate an RSS vector out of $\mathbf{P}_j^k$ on the unified time instants $\mathbf{T}_u = [0, \Delta t, \ldots, n\Delta t]$. The RSS vector of packets received by AP k from smartphone j is denoted as $\mathbf{rss}_j^k = [\mathbf{rss}_j^k(0), \mathbf{rss}_j^k(1), \ldots, \mathbf{rss}_j^k(p)]$, where $\mathbf{rss}_j^k(p)$ represents the median RSS value of packets during the period of time $[p\Delta t, (p+1)\Delta t]$. If no packets are received, we take the typical lowest RSS value (−100 dB) as a replacement. We call an RSS value a valid RSS if it is unequal to −100.

In step 2, we filter out two types of noises with median filter for each RSS vector. The filtered RSS vector is represented as $\tilde{\mathbf{rss}}_j^k$ for the given input $\mathbf{rss}_j^k$. As shown in Fig. 4, $\mathbf{rss}_j^k$ might have intermittent −100 which are mainly caused by packet loss. These −100 among RSS are one of the noises. The other type of noise is the isolated RSS that appears among a long sequence of −100 which might be caused by device noise or multipath effect.

The main idea of the median filter is to run through the signal entry by entry, replacing each entry with the median of neighboring entries. For a sequence of x, a median filter with length $2n+1$ will generate a sequence y which is defined as

$$y(i) = \text{median}([x(i-n), \ldots, x(i), \ldots, x(i+n)]). \quad (1)$$

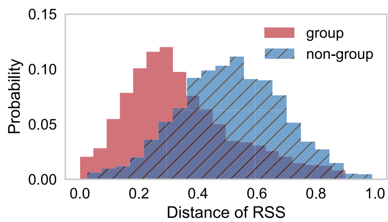


Fig. 5. Distance distributions of group and nongroups using RSS.

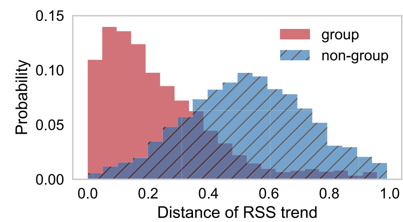


Fig. 6. Distance distributions of group and nongroups using RSS trend.

The length of the median filter is determined by the maximum number (n) of packet loss that is allowed during the period of $(2n+1)\Delta t$. For example, we set $2n+1=3$ in Fig. 4. Given a certain time instant $p\Delta t$, such a median filter could generate a valid $\tilde{\mathbf{rss}}_j^k(p)$ if the number of instants with packet loss is no more than 1 during $[(p-1)\Delta t, (p+1)\Delta t]$.

In step 3, we remove ineffective time instants when a smartphone does not send any packets. We represent the effective time instants of smartphone j as $\mathbf{T}_j$ which is defined in (2). Parameter $\mathbf{K}$ is a set of all APs

$$\mathbf{T}_j = \left\{ t \middle| \exists k \in \mathbf{K}, \tilde{\mathbf{rss}}_j^k(t) \neq -100, t \in \mathbf{T}_u \right\}. \tag{2}$$

Effective times refer to time instants when a smartphone send packet(s) or have a valid RSS. When $\tilde{\mathbf{rss}}_j^k(p) = -100$, it indicates that AP k does not receive any packets from the smartphone j during $[p\Delta t, (p+1)\Delta t])$. If all APs do not receive packets from the devices in a certain instant, it is believed that the smartphone does not send any packets.

2) Similarity Measurement With RSS: According to [13], the RSS difference of customers i and j at AP k could be measured with (3). $\mathbf{T}_{i\cap j} = \mathbf{T}_i \cap \mathbf{T}_j$ represents the intersection of effective time instants from customers i and j

$$\mathbf{d}_k(i,j) = \frac{1}{|\mathbf{T}_{i\cap j}|}\sqrt{\sum_{t\in\mathbf{T}_{i\cap j}}\left[\tilde{\mathbf{rss}}_i^k(t) - \tilde{\mathbf{rss}}_j^k(t)\right]^2}. \tag{3}$$

We could derive an overall RSS distance $\mathbf{d}(i,j)$ by averaging $\mathbf{d}_k(i,j)$ over all APs. However, we should note that the RSS distance at a certain AP would change over time. Instead of calculating a single statistic of $\mathbf{d}_k(i,j)$ for the whole testing period, a better way is to look at the distribution consisting of multiple $\mathbf{d}_k(i,j)$ from different time slots and different APs. Intuitively, if i and j are in the same group, the center of the distribution should be close to 0.

For pairwise customers, we could obtain two distributions $\mathcal{D}_g$ and $\mathcal{D}_{\tilde{g}}$ of RSS difference from shopping groups and nongroups, respectively. Whether RSS is an appropriate feature could be verified with the WiFi data collected from 104 volunteer shopping groups by comparing $\mathcal{D}_g$ and $\mathcal{D}_{\tilde{g}}$. As an illustration, Fig. 5 shows both distributions from our experiments described in Section III. In particular, the unit time instant $\Delta t = 1s$, and the length of the median filter is set to 11. Both parameters are determined empirically and experimentally. Using the testing data involving 104 groups in three weeks, we have generated a pool containing about 9675 data samples from group pairs and another pool containing about 32 410 samples from nongroup pairs.

Although $\mathcal{D}_g$ and $\mathcal{D}_{\tilde{g}}$ are obviously different, the level of difference is still not large enough to derive satisfying group detection performance. Since there is a large overlap between the two distributions, it is difficult to find a threshold to differentiate both distributions. After an investigation, we realize that one of the potential reasons for the large overlap is device heterogeneity. Even though group members stick together all the time, the difference between their smartphones could lead to a large gap. Fig. 8 shows two typical examples. In both cases, group members with different smartphones walk closely with each other, but there still exist certain gaps in their RSS signals which might be caused by hardware difference.

3) Similarity Measurement With RSS Trend: We find through experiments that when two customers are walking together closely, the change of their RSS reveals quite similar patterns, which could be exploited for group detection. From Fig. 8 we also notice that despite the gap in group members' RSS signals, their general trends are similar. This property has been observed in [14] and [15] that although RSS is very unstable, the trend of RSS is relatively stable. RSS values increase or decrease when approaching or leaving an AP. This has been utilized for indoor localization in [15] and [16].

Following the procedures as described in the previous section, we could find out the filtered common RSS vectors $\tilde{\mathbf{rss}}_i^k(\mathbf{T}_{i\cap j})$ and $\tilde{\mathbf{rss}}_j^k(\mathbf{T}_{i\cap j})$ for smartphone i and j. Let $X = \tilde{\mathbf{rss}}_i^k(\mathbf{T}_{i\cap j})$ and $Y = \tilde{\mathbf{rss}}_j^k(\mathbf{T}_{i\cap j})$, we calculate the distance of RSS trend with

$$\mathbf{d}'_k(i,j) = 1 - \rho(X,Y) = 1 - \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{4}$$

where $\rho(X,Y)$ is the Pearson correlation coefficient of sequence X and Y, σ_X is the standard deviation of X, cov is the covariance defined as $\mathrm{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$, and μ_X is the mean of X. Here, Person correlation is chosen for its better simplicity and efficiency compared with other measurements like dynamic time warping.

Fig. 6 shows distributions from group pairs and nongroup pairs, we can find the overlap is smaller than using RSS values. To compare features in Figs. 5 and 6 more objectively, we plot the receiver operating characteristic (ROC) curve of both methods. ROC curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate against the false positive rate at various threshold settings. True positive occurs when we correctly detect a customer pair as a group. False positive occurs when two strangers are improperly detected as a group. From Fig. 7,

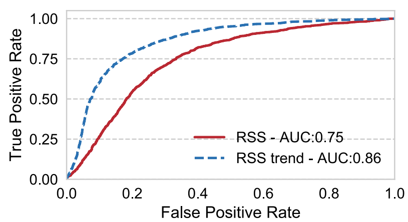
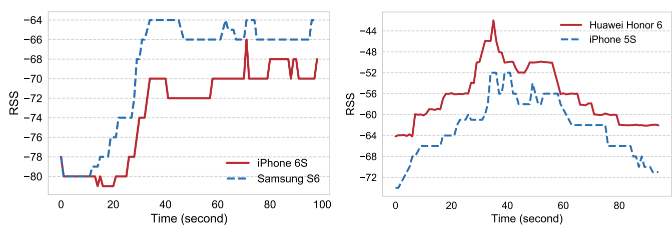


Fig. 7. ROC curves of using RSS and RSS trend.



(a) (b)

Fig. 8. RSS of group members using different smartphones. In both cases, group members stick together all the time, but there exist gaps in their RSS signals. (a) Case I. (b) Case II.

we could see the performance of using RSS trend is better than RSS.

C. AP Significance Estimation

Although we could detect groups by identifying customers with similar RSS trend, it is not good enough as shopping groups might naturally separate sometimes. However, we have two observations that indicate different APs should have different weights in measuring customer similarity. First, groups are more likely to separate in certain areas like bookstores and supermarkets as group members might have different interests. APs in those areas should have smaller weight due to frequent separation. Second, the ratio of customer groups over individuals is higher in public entertainment areas. A study of customer behaviors indicates individual customers are less interested in public entrainment activities since they anticipate negative inferences from others about their social connectedness. Therefore, we propose a probabilistic representation of different AP weights.

To calculate the probability, we combine the WiFi data and the Bluetooth data collected from volunteers. The following information could be extracted from the combined data: the number of individual customers connected to AP k ($\mathbf{N}_i^k$); the number of group customers in AP k ($\mathbf{N}_g^k$); the number of shopping groups in AP k ($\mathbf{M}_g^k$); and the number of group separation in AP k ($\mathbf{M}_s^k$). Fig. 9 illustrates an example with two APs and some individual customers and group customers. For AP a_1, $\mathbf{N}_i^1 = 3$, $\mathbf{N}_g^1 = 4$, $\mathbf{M}_g^1 = 2$, and $\mathbf{M}_s^1 = 1$.

We calculate a posterior probability ($P(G\bar{D}|A)$) as the AP weight in (5). $A = \{a_1, \ldots, a_n\}$ is a variable indicating the target AP, event D represents groups appear in the AP, and event $\bar{D}$ means groups do not disperse within the AP coverage. Therefore, $P(G\bar{D}|A)$ refers to the probability that groups

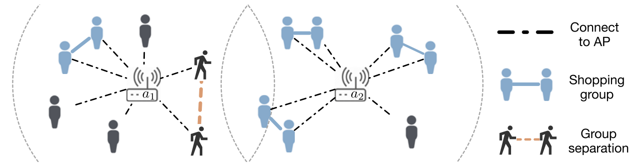


Fig. 9. Simple example for calculating AP weights.

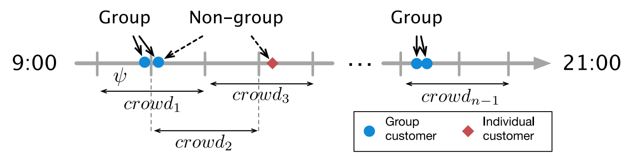


Fig. 10. Illustration of partitioning customers into different crowds with the temporal constraint.

appear and do not separate within the coverage of a certain AP

$$P(G\bar{D}|A) = P(G|A)\cdot P(\bar{D}|A) = \frac{\mathbf{N}_g}{\mathbf{N}_g + \mathbf{N}_i}\cdot\left(1 - \frac{\mathbf{M}_s}{\mathbf{M}_g}\right). \quad (5)$$

In the example of Fig. 9, $w_1 = P(G\bar{D}|a_1) = (4/4+3)\cdot(1-(1/2)) = 0.29$, $w_2 = P(G\bar{D}|a_2) = (6/6+1)\cdot(1-(0/3)) = 0.86$. It is clear that a_2 has a larger weight than a_1 and thus is more important in measuring customer similarity.

D. Group Detection

1) Customer Similarity Matrix: Given the weights of different APs, we could refine the similarity with

$$\mathbf{Sim}(i,j) = \frac{\sum_{k=1}^{\mathbf{K}} w_k \cdot \mathbf{d}_k'(i,j)}{\sum_{k=1}^{\mathbf{K}} w_k} \quad (6)$$

where $\mathbf{Sim}(i,j) \in [0,1]$ is the refined similarity between customers i and j, $\mathbf{d}_k'(i,j)$ is the similarity under AP k measured with RSS trend, and w_k is the weight of AP k.

Instead of detecting shopping groups out of all customers in a whole day, we first utilize the temporal constraint of groups to separate customers into different crowds and then identify groups out of each crowd. We equally partition the business hours of the mall into nonoverlapping fragments using a threshold ψ which is determined by the customers' dwell time. The partition process is depicted in Fig. 10. For each adjacent segment, we measure customers' similarity and construct a similarity matrix. The idea is quite straightforward. If two customers have a large gap in the time domain, they are more likely to be strangers.

2) Group Detection With Matrix Factorization: Group detection is essentially a hard clustering problem which means each user can only belong to a cluster or not. Existing works mostly apply graph clustering methods like Markov clustering to detect groups. Here, we resort to MF for the following two reasons. First, the constructed matrix has some noises, like strangers being regarded as groups and vice versa. MF can help in reducing these noises and preserving the latent group information. Second, MF can directly derive clustering results by imposing a sparseness constraint, which is similar to K-means but the performance is much better.

Given a similarity matrix $\mathcal{A} \in \mathrm{IR}^{m\times m}$ and an integer $k < m$, MF aims to find two factors $\mathcal{W} \in \mathrm{IR}^{m\times k}$ and $\mathcal{H} \in \mathrm{IR}^{m\times k}$ such that $\mathcal{A} \approx \mathcal{W}\mathcal{H}^{\mathsf{T}}$. The solutions can be found by solving the optimization problem with non-negative and sparseness constraints

$$\min_{\mathcal{W},\mathcal{H}} \frac{1}{2}\left[\left\|\mathcal{A} - \mathcal{W}\mathcal{H}^{\mathsf{T}}\right\|_F^2 + \eta\left\|\mathcal{W}\right\|_F^2 + \beta\sum_{i=1}^{m}\left\|\mathcal{H}(i,:)\right\|_1^2\right]$$
$$\text{s.t. } \mathcal{W},\mathcal{H} \geq 0 \tag{7}$$

where $\|\cdot\|_F$ means Frobenius Norm which has a Guassian noise interpretation and the objective function can be easily transformed into a matrix trace version, $\mathcal{H}(i,:)$ is the ith row vector of $\mathcal{H}$. Parameter $\eta > 0$ controls the size of the elements of $\mathcal{W}$. It is usually determined by the largest element of input matrices [17]. Parameter $\beta > 0$ balances the tradeoff between the accuracy of approximation and the sparseness of $\mathcal{H}$. A larger value of β implies stronger sparseness while smaller values of β can achieve better accuracy of approximation. The imposed non-negative constraint is due to physical meanings (similarity of pairwise users) of entries in the original matrix. Positive factors facilitate direct physical connections. The sparseness on the $\mathcal{H}$ factor could directly derive the clustering results with ℓ1-norm regularization.

Although (7) is a nonconvex problem, it is convex separately in each factor, i.e., finding the optimal factor $\mathcal{W}$ corresponding to fixed factors $\mathcal{H}$ reduces to a convex optimization problem. Algorithms based on alternating non-negative least squares are often used for sparse non-negative MF. More details of solving the optimization problem and determining the appropriate value of k can be found in [17].

III. Experimental Evaluation

A. Settings

1) Setup: We conduct experiments in a large shopping mall with four floors covering an area of 4890 m^2. This mall is at the bottom of an office building which is adjacent to a subway station. Most of the shops in the mall are related to food like restaurants and bakeries. There are originally 20 APs installed in the mall for customers to access the Internet. We use those APs to collect the WiFi data. The configurations of the AP are as follows: AR9341 (WLAN chip), 64M (RAM), and 8M (flash memory).

Within three weeks, we conduct 34 experiments at different times of a day. For each experiment, we recruit $2 \sim 4$ volunteer groups with each group containing $2 \sim 4$ customers and record their MAC addresses and grouping information. The majority of experiments last less than 3 h. To ensure authenticity, volunteers are only told to keep the smartphone WiFi function enabled without knowing the purpose of experiments.

2) Dataset: During three weeks, we collect the WiFi data from volunteer customers and other customers. Detailed information about the collected data is illustrated in Fig. 11. For the first two weeks, we record volunteer customers' WiFi data and Bluetooth data to estimate the different significance of the deployed APs. For the last week, we record the WiFi data from volunteer and nonvolunteer customers appeared in the mall to evaluate group detection performance.

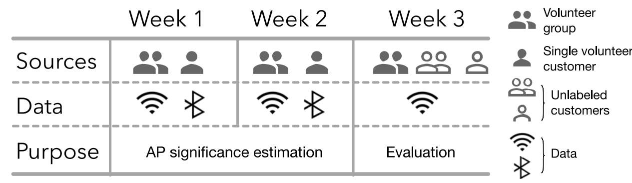


Fig. 11. Detailed information of the collected data in 3 weeks.

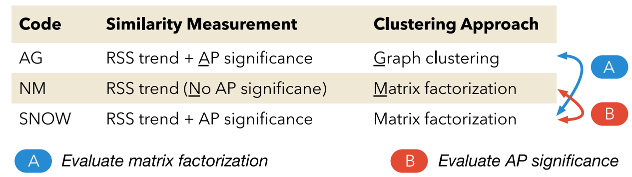


Fig. 12. Illustration of baseline approaches.

The WiFi data comes from 104 volunteer shopping groups during 34 experiments, including 258 group pairs (positive data samples) and 864 nongroup pairs (negative samples). For example, given an experiment with two shopping groups, each group has three customers, then we have $2 \times C_3^2 = 6$ positive samples and $C_3^1 \times C_3^1 = 9$ negative samples. We call the dataset above *labeled dataset* which contains only volunteers' data. In other words, we know the relation of all pairwise customers in the dataset. We also have a *semi-labeled dataset* that includes both volunteers and other customers which we do not know their grouping information. But one thing for sure is that volunteers and other customers must be strangers. Therefore, the semi-labeled dataset has much more negative samples than labeled dataset.

The Bluetooth data are from 58 volunteers during the first two weeks. Combined with the WiFi data, we find that customers groups are more likely to separate in places like restrooms (27.7%) and cosmetics shops (21.5%). One potential reason is that the mall has limited types of shops and most of the shops are related to food. As reported by our survey results, customers are not frequently get separated in restaurants.

B. Evaluation

1) Baseline Approaches: To detect shopping groups using WiFi data, similarity measuring and group clustering are two essential steps. We have different baselines for evaluating different steps. Since it is demonstrated that RSS trend is better than RSS, all baseline approaches are based on RSS trend.

As illustrated in Fig. 12, to evaluate the effectiveness of AP significance and MF, we need two baseline approaches apart from SNOW. The first baseline is called AG which measures customer similarity with estimating AP significance and then constructs a user graph with each node representing a customer and each edge representing the similarity between pairwise customers. Then AG detects groups with the help of Markov cluster algorithm (MCL) [1], [2]. MCL works well when the cluster size is small and it does not require the number of clusters as an input. The second baseline is NM which

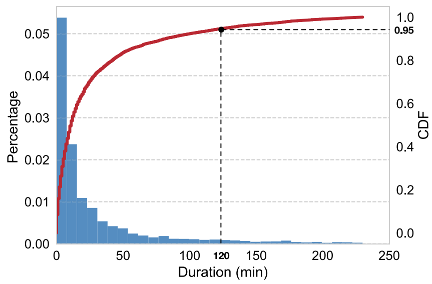


Fig. 13. Distribution and CDF of customers' dwell time in the mall.

TABLE I
PERFORMANCE COMPARISON ON BOTH DATASETS

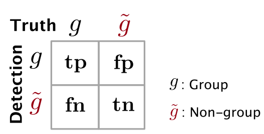
	Labeled Dataset			Semi-labeled Dataset		
	P[1]	R[2]	F[3]	P	R	F
AG	0.863	0.841	0.852	0.782	0.811	0.796
NM	0.750	0.787	0.768	0.691	0.734	0.712
SNOW	0.912	0.927	**0.919**	0.833	0.860	**0.846**

[1] Precision [2] Recall [3] F-score

measures customer similarity without AP significance and then detects groups using MF.

2) Evaluation Metric: As pointed out in [6], there is no consensus on which metrics should be used to evaluate groups detection. Here, we use *precision* and *recall* to measure the performance of group detection, which are defined as

$$\begin{cases} \text{precision} = \frac{\mathbf{tp}}{\mathbf{tp}+\mathbf{fp}} \\ \text{recall} = \frac{\mathbf{tp}}{\mathbf{tp}+\mathbf{fn}} \end{cases}$$



as shown in the confusion matrix, *tp* is the number of cases that positive samples being detected as groups. We also use a combined metric F-score defined in (8) to represent the general performance

$$F\text{-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{8}$$

3) Parameter Selection: We have two important parameters to determine for SNOW. First, parameter ψ represents the maximum duration time of customers. As shown in Fig. 13, over 90% customers stay in the mall for less than 2 h. Therefore, we simply set $\psi = 120$ (min). Second, parameter β balances the tradeoff between accuracy of approximation and sparseness. Even though the performance is not that sensitive to β, too big β is undesirable since that might lead to worse approximation. Therefore, we set $\beta = 0.3$ for all methods.

4) Performance Evaluation: We evaluate the performance of SNOW and baseline approaches. As shown in Table I, SNOW outperforms baseline approaches on both datasets by 6.3% $\sim$ 19.7%. The performance of all three methods on semi-labeled data are slightly worse than that of labeled dataset. This effect is reasonable and could be explained by the following reasons. First, there are much more customers in semi-labeled datasets which bring in more noise for the clustering process. Second, the number of negative samples

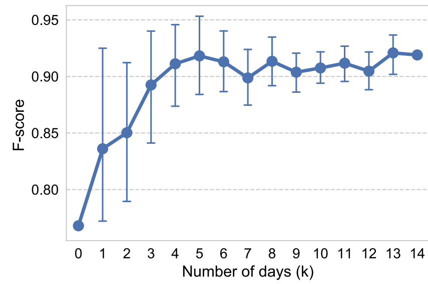


Fig. 14. Impact of the amount of Bluetooth data in estimating AP weights.

increased sharped that may cause more potential false positive detections and decrease precision.

To evaluate the effectiveness of MF, we could compare SNOW and AG. On average, SNOW outperforms AG by 7.1% in F-score which means MF could achieve better results than graph clustering. As explained, one potential reason is that MF could reduce the effect of false positive and false negative detections to some extent. MF is known for removing noises and preserving latent group information.

As one of the core components, AP significance is supposed to capture the dynamics of shopping groups. We could evaluate the effectiveness of AP significance by comparing SNOW and NM. From Table I, on average the F-score of SNOW is 19.3% better than that of NM. This indicates that estimating AP significance could greatly improve the performance of detecting shopping groups. Since a certain number of shopping groups may actually separate from time to time, their similarity in the RSS space can be affected. Straightforward as this method is, it reflects the pattern of most shopping group activities and successfully refines customers' similarity in the signal space.

5) Impact of Bluetooth Data: The Bluetooth data are critical in AP significance estimation (Section II-C). Intuitively, more adequate Bluetooth data can achieve more accurate approximation to group separation in real situations. Fig. 14 shows the detailed performance of SNOW on the label dataset using different numbers of the Bluetooth data. When the number is k, we randomly choose k days' Bluetooth data to calculate the AP significance and refine customer similarity. When k is small, the final performance may not be good enough and there exists a large deviation. With the increase of k, the performance gradually increases and becomes more stable (the deviation gets smaller). In our scenario, we could see that using no less than one week Bluetooth data achieves relatively stable performance.

6) Impact of AP Density: Since different scenarios may have different AP deployments and AP densities, we evaluate the performance of different methods with various AP densities. To derive different AP density, we adopt a sampling method over the semi-labeled dataset. For example, to evaluate the performance under 0.8 AP density, we randomly choose 16 out 20 APs and use the WiFi data of chosen APs for all users. We average the result for 100 times to eliminate the impact of randomness. As illustrated in Fig. 15, the performance drops as AP density decreases. One of the potential reasons is information loss. However, we can also find that

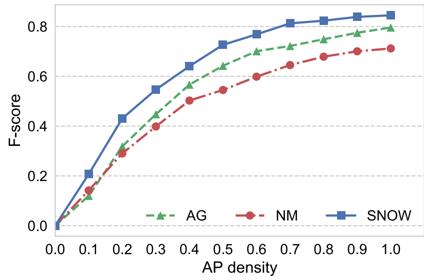


Fig. 15. Performance under different AP density on semi-labeled dataset.

SNOW still outperforms baseline approaches under different AP density.

IV. Related Work

With the development of IoT [1], [5], [7], [18], [19], there exist various group detection systems. However, none of them are particularly designed for shopping groups. These methods detect groups mainly by separating strangers who are close to each other but overlook the fact that shopping groups might separate sometimes. Under this situation, existing methods might generate many false negative detections and thus degrade the usability of the system. Literature methods can be classified as vision-based approaches, sensor-based approaches, and probe-based approaches according to different means.

Vision-based approaches regard group detection as a task of clustering a set of users' trajectories into disjoint subsets [5], [6]. However, this kind of methods have some apparent limitations. First of all, the most significant problem is privacy erosion. Besides, video surveillance suffers from environmental issues such as nonline-of-sight, and low brightness.

Sensor-based approaches use wearable devices or install apps on smartphones to collect users' behavioral data. Groups are detected through correlation analysis of multiple sensor data. For instance, MIT researchers use specially designed wearable devices called "sociometric badges" [8], [20] to measure group behavior through face-to-face interaction and physical proximity. Some research works [2], [7], [21] combine several sensor modalities (WiFi, accelerometer, compass, etc.) to measures users' similarity. However, these methods might be difficult to collect data on a scale, as they require user intervention which would be cumbersome in some scenarios. Besides, engaging multiple sensors drains smartphone battery more quickly.

Probe-based approaches utilize the information contained in probe requests to detect groups. The probe contains significant information like timestamp, smartphone MAC address, RSSI, and service set identifier (SSID), which enables a wide range of applications like passive tracking [15], [22], crowd counting [23], [24], and facility utilization analysis [25]. Compared to other approaches, probe-based approaches do not require high deployment cost or user intervention. SSID and RSSI are two frequently used information to detect groups. Barbera *et al.* [3], Cunche *et al.* [26], and Cheng *et al.* [27] link different smartphones through SSID similarity. However, 80% of the devices reply with empty SSID list [28], approaches that rely on SSID may not work well anymore. Then researchers' focus transfer to RSSI which indicate users' mobility. Kjærgaard *et al.* [4] extracted spatial features, signal-strength features, and pseudo-spatial features from signal strength to detect social groups which they call pedestrian flocks. It is found that the performance of spatial features is unreliable since mapping RSSI into locations is not accurate enough. Besides, the mapping process itself is usually time-consuming and labor-intensive. To avoid the cumbersome mapping process, directly measure the similarity of RSSI fingerprints to detect co-located mobile users. These methods get rid of absolute locations, thus eliminate labor-intensive calibration and protect users' privacy. SocialProbe [1] considers the hardware diversity and uses the normalized RSSI vector to achieve co-location detection. However, the timing of sending probes are mainly determined by user-device interaction and internal mechanism of the device. Different devices might generate various data granularity which makes it hard to compare their similarity [9].

V. Discussion

In this section, we provide further discussions to clarify potential issues that might be confusing and unclear. Issues to be discussed including AP deployment, the energy issue, and the generality of the system.

For AP deployment, we do not have any special requirements since we do not care about where exactly customers visit. Although our observations are related to different areas in the mall, we do not need to locate the customers. Because we assume that when customers are in different areas they would connect to different APs. This assumption stands in most of the practical scenarios because the coverage radius of common APs is tens of meters. Besides, AP deployments in shopping malls are usually conducted by experts which would try to use as less APs as possible while ensuring the network quality.

For the energy issue, it is no doubt that SNOW will increase energy consumption since we exploit Arping to lure smartphones to send more packets. However, we also need to note two points. First, the amount of extra energy consumed is very limited due to the low frequency of sending Arping packets and their small packet size [29]. Second, we notify customers of the potential energy consumption in the agreement when they initially connect to the deployed APs. If they really care about the extra energy to be consumed they would disconnect from those APs by themselves.

Lastly, the proposed system could be generalized to different shopping malls since we do not rely on specific scenarios or any device configurations. As discussed above, we do not have special requirements for the AP deployment. Our observations of group dynamics are also independent of places. They are based on the online survey with over 250 subjects (Fig. 1) and the research result of consumer behaviors [12] from Americans, Chinese, and Indian respondents.

VI. Conclusion

In this paper, we propose a practical SNOW using WiFi. One of our contributions is an effective heuristic that could significantly improve the detection performance of shopping groups. The heuristic indicates APs under which groups appear more frequently and barely separate should have larger weights in measuring customer similarity. Our second contribution is to apply MF to detect groups without extra clustering processes. MF could properly handle data issues in the measured similarity including noise filtering and data completion. Besides, imposing a sparsity constraint to the factorization process could derive the clustering results directly. Finally, we conduct extensive experiments in a large shopping mall to validate the performance of SNOW. Experimental results indicate SNOW can detect over 90% groups with a precision of 91.2%.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for their insightful comments.

REFERENCES

[1] H. Hong, C. Luo, and M. C. Chan, "SocialProbe: Understanding social interaction through passive WiFi monitoring," in *Proc. ACM MobiQuitous*, Hiroshima, Japan, 2016, pp. 94–103.
[2] R. Sen, Y. Lee, K. Jayarajah, A. Misra, and R. K. Balan, "GruMon: Fast and accurate group monitoring for heterogeneous urban spaces," in *Proc. ACM SenSys*, Memphis, TN, USA, 2014, pp. 46–60.
[3] M. V. Barbera, A. Epasto, A. Mei, V. C. Perta, and J. Stefa, "Signals from the crowd: Uncovering social relationships through smartphone probes," in *Proc. Conf. Internet Meas. Conf.*, ACM, 2013, pp. 265–276.
[4] M. B. Kjærgaard, M. Wirz, D. Roggen, and G. Tröster, "Mobile sensing of pedestrian flocks in indoor environments using WiFi signals," in *Proc. IEEE PerCom*, Lugano, Switzerland, 2012, pp. 95–102.
[5] W. Ge, R. T. Collins, and R. B. Ruback, "Vision-based analysis of small groups in pedestrian crowds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 1003–1016, May 2012.
[6] F. Solera, S. Calderara, and R. Cucchiara, "Socially constrained structural learning for groups detection in crowd," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 995–1008, May 2016.
[7] M. B. Kjærgaard, M. Wirz, D. Roggen, and G. Tröster, "Detecting pedestrian flocks by fusion of multi-modal sensors in mobile phones," in *Proc. ACM UbiComp*, Pittsburgh, PA, USA, 2012, pp. 240–249.
[8] D. O. Olguın, P. A. Gloor, and A. S. Pentland, "Capturing individual and group behavior with wearable sensors," in *Proc. AAAI Spring Symp. Human Behav. Model.*, vol. 9, 2009.
[9] J. Freudiger, "How talkative is your mobile device? An experimental study of Wi-Fi probe requests," in *Proc. 8th ACM Conf. Security Privacy Wireless Mobile Netw.*, ACM, 2015, p. 8.
[10] J. Shen, J. Cao, X. Liu, and C. Zhang, "DMAD: Data-driven measuring of Wi-Fi access point deployment in urban spaces," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, p. 11, 2017.
[11] *Public WiFi Usage Survey*. Accessed: Aug. 24, 2017. [Online]. Available: https://www.idtheftcenter.org/images/surveys_studies/PublicWiFiUsageSurvey.pdf
[12] R. K. Ratner and R. W. Hamilton, "Inhibited from bowling alone," *J. Consum. Res.*, vol. 42, no. 2, pp. 266–283, 2015.
[13] G. Vanderhulst, A. Mashhadi, M. Dashti, and F. Kawsar, "Detecting human encounters from WiFi radio signals," in *Proc. ACM MUM*, Linz, Austria, 2015, pp. 97–108.
[14] G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang, "Walkie–Markie: Indoor pathway mapping made easy," in *Proc. 10th USENIX Conf. Netw. Syst. Design Implement.*, USENIX Assoc., 2013, pp. 85–98.
[15] J. Shen, J. Cao, X. Liu, J. Wen, and Y. Chen, "Feature-based room-level localization of unmodified smartphones," in *Smart City 360°*. Cham, Switzerland: Springer, 2016, pp. 125–136.
[16] D. L. Lee and Q. Chen, "A model-based WiFi localization method," in *Proc. 2nd Int. Conf. Scalable Inf. Syst.*, ICST, 2007, p. 40.
[17] J. Kim and H. Park, "Sparse nonnegative matrix factorization for clustering," Georgia Inst. Technol., Atlanta, GA, USA, CSE Tech. Rep., 2008.
[18] X. Liu *et al.*, "Multi-category RFID estimation," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 264–277, Feb. 2017.
[19] X. Liu *et al.*, "Completely pinpointing the missing RFID tags in a time-efficient way," *IEEE Trans. Comput.*, vol. 64, no. 1, pp. 87–96, Jan. 2015.
[20] D. O. Olguín *et al.*, "Sensible organizations: Technology and methodology for automatically measuring organizational behavior," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 43–55, Feb. 2009.
[21] Y. Lee *et al.*, "SocioPhone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion," in *Proc. ACM MobiSys*, 2013, pp. 375–388.
[22] A. Di Luzio, A. Mei, and J. Stefa, "Mind your probes: De-anonymization of large crowds through smartphone WiFi probe requests," in *Proc. IEEE INFOCOM*, San Francisco, CA, USA, 2016, pp. 1–9.
[23] L. Schauer, M. Werner, and P. Marcus, "Estimating crowd densities and pedestrian flows using Wi-Fi and Bluetooth," in *Proc. EAI MOBIQUITOUS*, 2014, pp. 171–177.
[24] W. Xi *et al.*, "Electronic frog eye: Counting crowd using WiFi," in *Proc. IEEE INFOCOM*, Toronto, ON, Canada, 2014, pp. 361–369.
[25] T. S. Prentow, A. J. Ruiz-Ruiz, H. Blunck, A. Stisen, and M. B. Kjærgaard, "Spatio-temporal facility utilization analysis from exhaustive WiFi monitoring," *Pervasive Mobile Comput.*, vol. 16, pp. 305–316, Jan. 2015.
[26] M. Cunche, M. A. Kaafar, and R. Boreli, "I know who you will meet this evening! Linking wireless devices using Wi-Fi probe requests," in *Proc. WoWMoM*, San Francisco, CA, USA, 2012, pp. 1–9.
[27] N. Cheng *et al.*, "Inferring user relationship from hidden information in WLANs," in *Proc. MILCOM*, Orlando, FL, USA, 2012, pp. 1–6.
[28] X. Hu, L. Song, D. Van Bruggen, and A. Striegel, "Is there WiFi yet? How aggressive probe requests deteriorate energy and throughput," in *Proc. ACM IMC*, 2015, pp. 317–323.
[29] L. Sun, R. K. Sheshadri, W. Zheng, and D. Koutsonikolas, "Modeling WiFi active power/energy consumption in smartphones," in *Proc. IEEE 34th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, IEEE, 2014, pp. 41–51.

Jiaxing Shen received the B.S. degree from Jilin University, Changchun, China, in 2014. He is currently pursuing the Ph.D. degree at the Department of Computing, Hong Kong Polytechnic University, Hong Kong.

He was a Visiting Student with the MIT Media Lab, from 2017 to 2018. His current research interests include data mining, mobile sensing, and group dynamics.

Mr. Shen served as a Reviewer of several international conferences.

Jiannong Cao (M'93–SM'05–F'15) received the M.Sc. and Ph.D. degrees in computer science from Washington State University, Pullman, WA, USA, in 1986 and 1990, respectively.

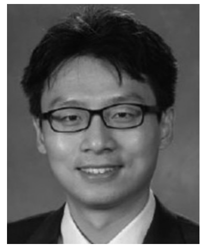
He is currently the Chair Professor with the Department of Computing, Hong Kong Polytechnic University, Hong Kong. His current research interests include parallel and distributed computing, mobile computing, and big data analytics.

Dr. Cao has served as a member of the Editorial Boards of several international journals, a Reviewer for international journals/conference proceedings, and also as an Organizing/Program Committee member for many international conferences. He is a Senior Member of the China Computer Federation.

Xuefeng Liu received the M.S. degree from the Beijing Institute of Technology, Beijing, China, in 2003, and the Ph.D. degree from the University of Bristol, Bristol, U.K., in and 2008.

He is currently an Associate Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China. His current research interests include wireless sensor networks and in-network processing.

Dr. Liu has served as a Reviewer for several international journals/conference proceedings.



Shaojie Tang received the Ph.D. degree in computer science from the Illinois Institute of Technology, Chicago, IL, USA, in 2012.

He is currently an Assistant Professor with the Naveen Jindal School of Management, University of Texas at Dallas, Dallas, TX, USA. His current research interests include social networks, e-business, and optimization.

Dr. Tang served as a Chair and a TPC member for numerous conferences.