Received April 7, 2018, accepted May 14, 2018, date of publication May 28, 2018, date of current version July 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2841007

Mining Mobile Internet Lifestyles in Distinct Urban Areas: Tales of Two Cities

YI ZHANG[1], LINTAO YANG[2], HAO JIANG[1], SHUWEN YI[1], ZHIYI HU[1], AND XING LIU[3], (Member, IEEE)

[1]School of Electronic Information, Wuhan University, Wuhan 430072, China
[2]College of Physical Science and Technology, Central China Normal University, Wuhan 430079, China
[3]School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China

Corresponding author: Hao Jiang (jh@whu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702387, in part by the National Key Research and Development Program under Grant 2017YFB0504103 and Grant 2017YFC0503801, in part by the Development Program of China (863 Program) under Grant 2014AA01A707, and in part by the Natural Science Foundation of Hubei Province of China under Grant 2017CFB302.

ABSTRACT The information and communication technology industry has developed rapidly, and data are an important link between society and the Internet. The study of the regional usage features of mobile Internet users provides a guide for network supervision and service distribution. In this paper, we propose an approach to identify the typical mobile Internet patterns to account for the content of the user's Internet access, which is used to discover the online lifestyle of people living in cities of different sizes. We believe that the user Internet behavior is a weighted sum of many typical underlying features. Therefore, a single usage content tensor is constructed to record the sequence of content that the people access on the Internet. The mobile Internet lifestyles based on user context access content were obtained by comparing the composition of a user's typical online behavior under regional conditions. The method is applied to the case of two representative cities of different sizes in mainland China. We have extracted key features for the mobile Internet context for a large number of users in each city from the usage detail records obtained from cellular networks. This approach uncovers interesting features of human behavior on the mobile Internet at fine granularity, some of which allow us to quantitatively compare the online lives of people living in areas with different urban compositions. The identified regional typical user lifestyle can be used to guide urban planning, network supervision, service resource allocation, and urban dynamics analysis.

INDEX TERMS Lifestyles, urban computing, mobile Internet.

I. INTRODUCTION

China has become the largest market in the world for mobile communications [1]. (As of June 2017, people spend an average of 26.5 hours a week online, and 96.3% of these users access the Internet via mobile phones [2].) Online data are gradually replacing traditional social surveys and interviews to understand the living habits of individuals in society. At the mobile Internet technology level, the ever-increasing global coverage of mobile cellular networks and huge amount of generated data have provided support for the prosperity of the mobile Internet. Due to the combined effect of the above factors, although the PC Internet has become increasingly saturated, the mobile Internet has experienced tremendous development. The coverage and usage of large-scale mobile networks provide a new motivation for recording the online life and social activities of urban people. Hence, understanding the underlying mobile Internet lifestyle is crucial to understanding the personal behavior and social dynamics in the Big Data era. The analysis of user's online behavior has great significance to social computing for simulation, forecasting and control. In addition, the hotspot areas where collective behavior exhibits great differences compared to the ordinary behavior patterns can be identified, which is helpful for anomaly detection and network supervision. Besides, the discovery of these lifestyles is also an effective way for city planners to better understand the dynamics of modern cities, including cultural boundaries, Internet security and even economic conditions.

Due to cultural and socio-economic conditions and other constraints, the populations in different regions have differing lifestyles. Current research is based primarily on telephone interviews or survey data, which are always accompanied by relatively large time and financial costs. In addition, due to the limited time and space dimensions, these data do not accurately cover the range of human Internet activities.

The traditional notion of individual lifestyles in cybersociety is that people in metropolitan cities are active on the Internet, preferring social networks and shopping, whereas in small cities, people are quiet and have limited in Internet activity. To validate a wide range of urban understanding, in this work, we consider the space in which the users are located and the Internet content they visit. We propose a solution to this ''user-geo-Internet content'' problem by extracting the typical features of users based on high-dimensional mobile Internet contextual access content.

We believe that an individual's Internet lifestyle is related to the Internet content they visit. In our current work, an individual's style accessing content on the Internet is considered as a weighted combination of multiple qualitative lifestyles. The method uses mobile Internet UDRs to infer people's network lifestyles. Because data usage behavior is common in daily life and is spontaneously generated by users accessing base stations (BSs), it accurately reflects the human spatial location information and network access behavior. After constructing the contextual information tensors for users to access mobile Internet content, we use the improved high-order singular value decomposition (HOSVD) method to determine the sequence features of these potential access contents. The extracted features provide a definition the user's mobile internet associated with a particular lifestyle, and they represent typical access features of a particular group of people, such as entertainment game players. We chose Beijing as a representative big city in mainland China and Jinhua as a representative small city. We determined the Internet lifestyle in different sized cities and regions. In addition, in combination with the spatial and geographical factors, we extracted the reliable spatial structure of mobile users 'online behavior in the two places. This method provides effective guidance for extracting complex sequence patterns in high-dimensional space.

The rest of this article is structured as follows. In Section 2, we review the literature on social computing and mobile Internet mining. Section 3 introduces the framework and process of this study and the data used. Section 4 describes the experimental results and discusses the social implications. Section 5 describes the contributions and limitations of the study.

II. RELATED WORK

Research results have greatly enriched the study of human behavior dynamics. Anonymous call detail records (CDR) are commonly used to capture urban dynamics. In the case of Milan, Parwez *et al.* [7] analyzed abnormal mobile wireless network behavior and compared the detected anomalies with real information to verify the accuracy. Pulselli *et al.* [8] demonstrated the possibility of using data to study urban activities; they use images to illustrate the intensity of urban activity and evolution. Reades *et al.* [9] studied the dramatic changes in the activity of locals and tourists on weekdays and weekends in six different locations in Rome and proposed algorithms for clustering similar geographical locations. Calabrese *et al.* [10] speculated on the origins of people attending a special event in Boston. The literature has also examined behavioral differences between tourists and locals in New York [11]. Because service providers regularly collect service business, planning and billing policies, the resources needed to analyze these data are small. However, CDRs have some limitations. First, they are sparse in time because records are generated only when conversations occur, and second, CDRs do not record the data of users who do not participate in conversations, making the source of regional survey data incomplete.

In mining data from mobile phones, we focus on the extraction of typical features from internet users are as following:

In the traditional method of feature mining, the focus is on the extraction and analysis of single-dimension user features such as spatial feature. Luo *et al.* [12] designed a spherical frame by combining the structure holes of social network models and proved that there is a strong correlation between the features of human network life patterns and social status. Moreover, a user's online behavior has also been shown to have an important connection with the individual's consumer performance [13]. They found a strong connection between user network behavior and consumption and market behavior. Some studies have considered the use of models to discover user features. Park *et al.* [14] proposed an eigen model and found that the eigenvectors of the user's transfer matrix can provide detailed information about the user's mobile mode. Cole *et al.* [15] designed a method based on Markov chains to discover behavior patterns based on individual webpage access sequences. The method can be used to distinguish and represent different tasks. Zhang *et al.* [16] used a hidden Markov model and multi-state model infer the usage patterns of mobile Internet apps . Based on the features of individual mobility, location and travel behavior, Qiao *et al.* [17] proposed a prediction framework for studying the behavior of individuals and groups of people. Kawazu *et al.* [18] used a hidden Markov model to analyze various potential Internet surfing states of users. The advantages and disadvantages of the above methods are related to the statistical learning models used. Matrix and tensor decomposition starts with individual behaviors. On the basis of the founding theme of the composition of online behavior, Shafiq *et al.* [19] proposed a user trust and correlation matrix and through this extraction method, found that the relevant information in the user network is strongly correlated with the relevant information in the social network. Jiang *et al.* [20] proposed a new framework of tensor decomposition and predicted the multi-directional behavior of time and the behavior patterns in MAS and Weibo data. Wu *et al.* [21] used the statistical features of user network access as a basis for optimizing network deployment and found that the radius of user activation in space obeys a lognormal distribution. Furthermore, the literature has proposed a framework for using family and work-related contexts to identify user interests, thereby enriching traditional user behavior [22]. However, to the best of our knowledge, this article is the first to discuss typical user online sequential features from the perspective of high-dimensional individual behavior under the conditions of regional differentiation.

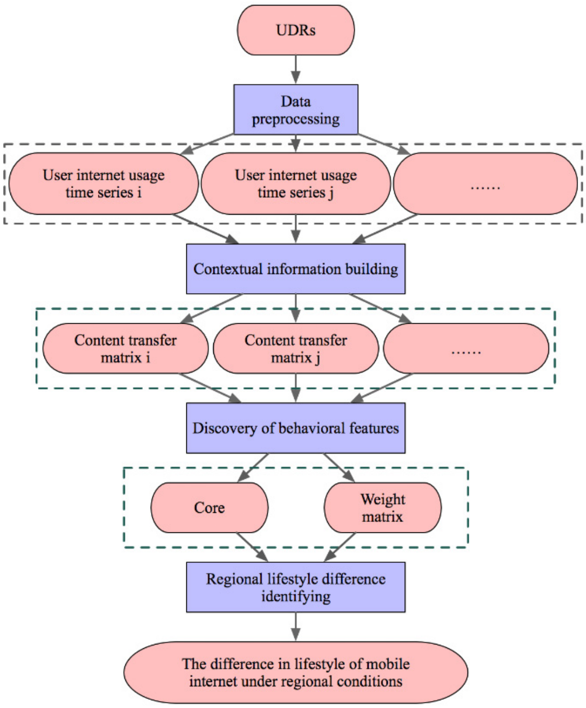


FIGURE 1. Study framework.

III. STUDY FRAMEWORK

This section introduces the overall research process of user mobile Internet behavior patterns, including the three steps shown in Fig. 1. The first step is data preprocessing to organize user access time series by session time, user ID, Internet session time and user access app types. Then, we construct the transfer probability matrix based on user access content. After fitting with multi-state models, the content time sequence forms a context-based transfer distribution. Finally, we can extract the key features and weights through the improved HOSVD and obtain the composition of user mobile Internet lifestyle in cities of different scale through mobile Internet user pattern discovery.

A. DATA PREPROCESSING

In this study, we use UDRs from mobile cellular networks that contain fields for user ID, content information, session time, etc. All user IDs are protected by encryption. The dataset was collected from Beijing and Jinhua, China, including 2,788,384 and 1,519,661 individual users, respectively, and spanning 23 days in 2015. Based on the dataset, we randomly selected 7,000 users from each city for quantitative research. To ensure the sample data were reliable and representative, we screened for mobile users who have used mobile phones for more than a year because we assumed that behaviors of these users were more stable and representative. The field structure of our UDRs is shown in TABLE 1.

TABLE 1. Data structure before preprocessing.

Fields	Description
uid	An encrypted telephone number uniquely indicating a mobile user
stime	The time that a user begins to access the BS for data usage
URL	Uniform resource locator that a mobile user accesses at the corresponding start time

We assumed that a normal mobile visits only one target application during a conversation. To ensure that no duplicate information was included in the sampled users, we filtered out records that did not coincide with the start and end times of the user's conversations, ensuring that users accessed only one application or site at a time.

Due to the rapid growth of smart phones, competition in the service market is becoming increasingly fierce. Therefore, users can select applications provided by different service providers to achieve the same purpose. We divided the services extracted from URL into 12 groups based on Apple Inc.'s classification of mobile applications in the App Store and gave each group a tag, as shown in TABLE 2.

TABLE 2. Data structure before preprocessing.

Service tags	Description
social	Instant communication, video voice service, broadcast online communication service
shopping	Online shopping service
portal & news	Portal and news access
downloads & clouds	Application store downloads or P2P downloading services
search	Internet search service
travel	Travel and tourism resources acquisition
entertainment	Games, live entertainment, online video
forum	Online forum services
life	Service to meet and assist with needs related to life, weather, cooking guide
mail	Mailbox service for mobile devices
ad	Mobile phone ads
fiction	Online Reading

We then sorted users' access information and constructed a time series of users' application types. The structure of each record in the series is shown in Table 3.

The above process was implemented in Apache Spark, which is a distributed computing framework based on Hadoop that has obvious advantages in large-scale data processing [23].

TABLE 3. Time series structure of discrete value categories.

User ID	Start time	App label

B. CONTEXTUAL INFORMATION BUILDING

To construct context information about a user's online behavior, this paper uses a multi-state model to obtain the transfer matrix of user behavior. This model is a continuous time stochastic process tool that allows individuals to move between a limited number of discrete states. Assume that at time t an individual state belongs to $S(t) \in \{1, \ldots, N\}$. In this article, the state corresponds to the category that the individual belongs to, and the discretized interclass transfer strength can be defined as follows:

$$q_{rs}(t, z(t)) = \lim_{\Delta t \to 0} P(S(t+\Delta t)=s \mid S(t) = r)/\Delta t \qquad (1)$$

The transition matrix Q of app categories, which is used to describe the transformation intensity between the two groups, can be obtained by calculating the edge transfer time series. The row sum of the matrix is 0. At $r \neq s$, in line r of the matrix, the element in column s is q_{rs}, which represents the state transfer intensity of an individual transferred from app r to app s. The state transfer intensity represents the frequency of the transition from app r to app s in a single time series. According to the state transfer strength matrix Q, the state transition probability matrix P can be fitted to describe the individual switching between several states. After fitting, matrix P, row r column s elements p_{rs} provide the following information: when $r = s$, $p_{rs} = -q_{rs}/q_{rr}$, which represents the transition probability from app r to app s. The state transfer probability matrix P obtained from the context information construction serves as a further research basis and provides input.

We applied the msm package in R, which was published by Jackson, to train a multi-state model. During training, the exact times is set to TRUE, which means that the start time can be assumed to represent the exact time course conversion. *obstype* is set to 2, which means that the conversion time of the process is the exact conversion time. That is, the user maintains the previous observation from a certain point in time until the next observation, preventing a change of state in the middle of the observation interval, which is in line with our expectations.

We obtain the state transition distribution of each user from the training results of the multi-state model. As mentioned above, the probability of state transition is more meaningful than the original transition intensity. Given the average length of stay and transition probability matrix used in training the multi-state models, we chose the probability matrix as the basis for further research and provided input, because we are mostly concerned with user information about the relationships among the services.

C. DISCOVERY OF BEHAVIORAL FEATURES

For the state transfer probability matrix P, which contains the user's behavior context information in the Internet, it is a two-dimensional vector, which can clearly indicate the preference of the individual to the content of the Internet in a certain period. However, such a state transfer probability matrix cannot reflect the tendency of the group to the content in the process of collective behavior analysis, and some unsupervised learning algorithms, such as clustering, have the problem of lack of representativeness. Therefore, we introduce the high-dimensional tensor of " User-Transition Matrix " to characterize collective user internet behavior.

To discover potential features of user mobile Internet context behavior, we believe that human activity can be represented by a 3-dimensional tensor Y called the "User-Transition Matrix" structure. N is the number of mobile Internet app tags. For target users:

$$\begin{aligned} Y &\cong \mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)} \\ &= \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} A_p^{(1)} \cdot A_q^{(2)} \cdot A_r^{(3)} \end{aligned} \qquad (2)$$

At the same time:

$$\mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)} = \left[\mathcal{G}; A^{(1)}, A^{(2)}, A^{(3)}\right] \qquad (3)$$

$\mathcal{G}$ is the tensor of K types of transition matrices of the N types of apps, which describe K types of potential features of the individual's mobile Internet lifestyle. To balance the interpretability and accuracy of the results, we set K to 4. $A^{(n)}$ represents the decomposition result of the tensor along the N direction. On the basis of the objective function, we can obtain the following:

$$\begin{aligned} &\left\| Y - \left[\mathcal{G}; A^{(1)}, A^{(2)}, A^{(3)}\right] \right\| \\ &\quad = \left\| vec(Y) - \left(A^{(3)} \otimes A^{(2)} \otimes A^{(1)}\right) vec(\mathcal{G}) \right\| \end{aligned} \qquad (4)$$

Thus, $\mathcal{G}$ is subject to:

$$\mathcal{G} = Y \times_1 A^{(1)T} \times_2 A^{(2)T} \times_3 A^{(3)T} \qquad (5)$$

Therefore, the square of the objective function is converted to:

$$\begin{aligned} &\left\| Y - [\mathcal{G}; A^{(1)}, A^{(2)}, A^{(3)}] \right\|^2 \\ &= \|Y\|^2 - 2\langle Y, [\mathcal{G}; A^{(1)}, A^{(2)}, A^{(3)}] \rangle \\ &\quad + \left\| \left[\mathcal{G}; A^{(1)}, A^{(2)}, A^{(3)}\right] \right\|^2 \\ &= \|Y\|^2 - 2\langle Y \times_1 A^{(1)T} \times_2 A^{(2)T} \times_3 A^{(3)T}, \mathcal{G} \rangle \\ &\quad + \|\mathcal{G}\|^2 = \|Y\|^2 - 2\langle \mathcal{G}, \mathcal{G} \rangle + \|\mathcal{G}\|^2 \\ &= \|Y\|^2 - \left\| Y \times_1 A^{(1)T} \times_2 A^{(2)T} \times_3 A^{(3)T} \right\|^2 \end{aligned} \qquad (6)$$

Then, the problem is transformed to:

$$\begin{aligned} &min \left\| Y - [\mathcal{G}; A^{(1)}, A^{(2)}, A^{(3)}] \right\| \\ &\qquad \Leftrightarrow max \left\| Y \times_1 A^{(1)T} \times_2 A^{(2)T} \times_3 A^{(3)T} \right\| \end{aligned} \qquad (7)$$

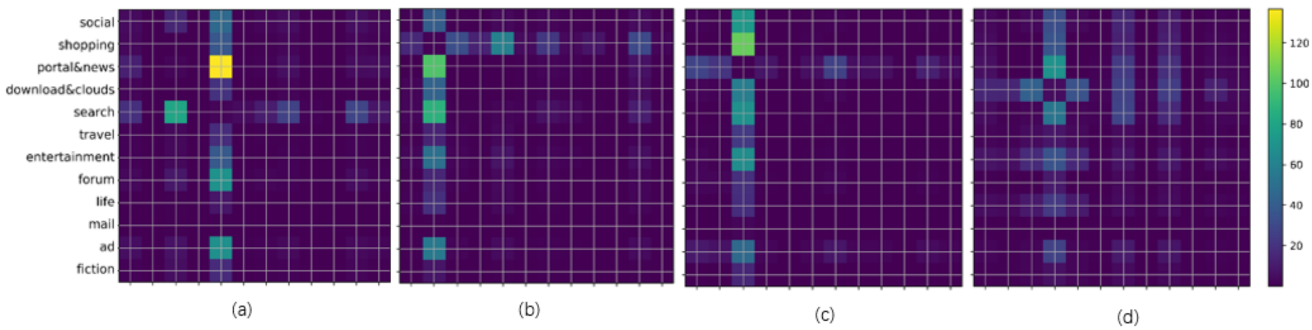


FIGURE 2. Context features of users accessing mobile Internet in Jinhua.

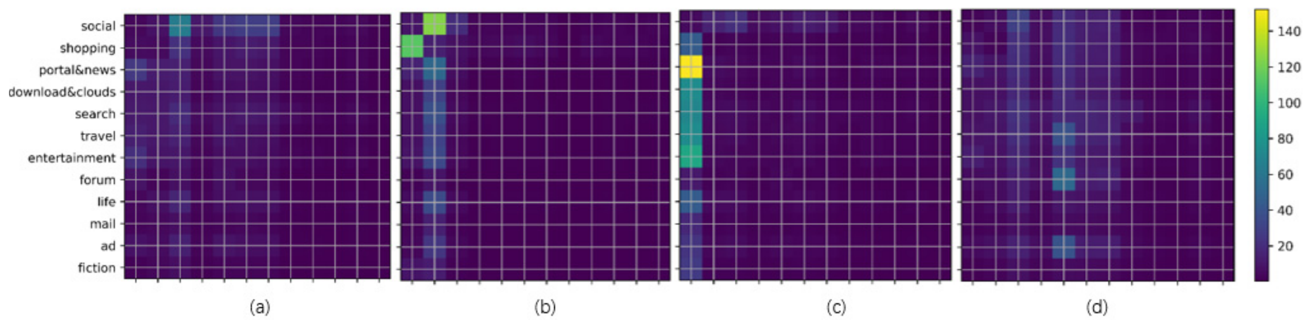


FIGURE 3. Context features of users accessing mobile Internet in Beijing.

The problem can be transformed into a subproblem by using the idea of an alternating solution:

$$max \left\| A^{(n)T} W \right\| \tag{8}$$

This solution is subject to $W = Y_{(n)}(A^{(3)} \otimes A^{(2)} \otimes A^{(1)})$. We adopt high-order orthogonal iteration (HOOI) to solve the tensor decomposition optimization problem. This method was first proposed by Lathauwer *et al.* [24], and the HOOI-Tucker decomposition was performed using the *tensorly toolkit*.[1] The aggregation of kernel tensors along the three different decomposition modes transforms the HOSVD into nonnegative matrix decomposition derived form, that is:

$$Y \cong \mathcal{G} \times \mathscr{w} \tag{9}$$

$\mathscr{w}$ is a weight vector that represents the user's preference for each lifestyle and records the number of individuals under a single typical feature. To determine and compare the typical lifestyles of different cities, we first set the vector of activities of the residents to the single matrix of each city. We define:

$$Y = [p_1, p_2, \ldots, p_n]^T, \quad n \in \{1, 2, ..., N\} \tag{10}$$

p_n is the transfer probability set of app n, in the form $p_n = [p_{1n}, p_{2n}, \ldots, p_{nn}]$. In the decomposed results, $\mathcal{G}$ represents the contextual features of app access behavior, which combined constitute the user's mobile Internet lifestyle. The weight represents the occupancy of the different features of the individual under different regional conditions, that is, the user's preferred mobile Internet lifestyle.

[1] https://pypi.python.org/pypi/tensorly/0.2.0

IV. RESULTS ANALYSIS AND DISCUSSION

This section presents a detailed explanation of the mobile Internet lifestyles (MILS) of regional users based on the results of the tensor decomposition method. By decomposing the context-aware Internet usage tensor, MILS is described by the core tensor $\mathcal{G}$. The weighted combination of multiple MILS can form the mobile Internet behavior of a specific user in a period. MILS is a basic unit of human mobile Internet behavior. It records the potential contextual preferences of users using the mobile Internet.

A. DISCOVERY OF BEHAVIORAL FEATURES

In the analysis of large-scale data, individual behavior is considered to be formed through the combination of user features. Therefore, we attempt to use tensor decomposition to study the individual features of users and to analyze user MILS.

In this paper, improved HOSVD is used to extract the features of user behavior. From the experiment, we obtain the typical user features that differ among users in Beijing and Jinhua, as shown in Fig. 2 and Fig. 3.

Firstly, we decompose the tensor of users in Jinhua. By decomposing the above tensors, we find four basic types of user preference behavior features. In Fig. 2, the values of users transfer behavior are depicted by different colors:

darker yellow represents a higher value and darker blue represents a lower value. We find the following four features.

The first feature indicate that users prefer to use search-type content. The figure shows that users prefer to flow from portal news, forums and advertisement types to search categories and that they tend to link to portal news content. Additionally, the figure shows that users follow a ''portal news - search'' cycle, that is, users switch back and forth between portal news and searching, with access to other types of mobile Internet content during the handover process.

The second feature show the characteristics of users' access behavior related shopping content, where users link to shopping-related mobile Internet content from portals, search, entertainment, and advertising. In addition, users tend to link content from shopping category to search category content.

The third feature show the characteristics of users' access behavior related to portal news content, that is, users tend to access portal news content from shopping, downloading, searching, entertainment and other content.

The fourth feature represent other user characteristics, e.G., the user accesses a small amount of mobile Internet content, such as downloading, entertainment, living, and advertisement. Downloading service is used as the main traffic port for this feature.

We also analyzed the results of users in Beijing. We found that users in Beijing have a notable feature in switching between Internet content; that is, users have a strong dependence on social content. We decompose the user's transfer probability tensor and obtain four typical features of user access.

The first type of user feature represents users with low interest in Internet content. The most common behavior in this case is access from the social category to the portal news category.

The second type of user feature represents user access behavior related to shopping content, that is, the user tends to shift from other content to the shopping category and back and forth between social and shopping category content.

The third type describes the scenario in which a user accesses social content, where users prefers to migrate between mainly portrayal and entertainment content and social content. However, these users have low tendency to roll out.

The fourth feature represents users that have similar preferences among all Internet content, such as news portal, search, travel and other Internet services. This phenomenon shows a preference for moving from search, forums, and advertising to search content.

Within the small allowed deviation, we can define the MILS for the core tensors of the two cities in the above figure. The four elements related to users in Jinhua are as follows:

MILS 1: Searcher, the inflow and outflow of search category apps is much larger than those of other types among these users, so these individuals are keen to use a variety of search software to understand the world.

MILS 2: Shopaholic, users who show a special preference for shopping apps.

MILS 3: Current affairs observers, users who are passionate about current events and have a high interest in news. This group probably corresponds to middle-aged and elderly people in the region.

MILS 4: Casual Player, users who are less dependent on the network and occasionally use downloading and entertainment due to traffic expenses.

By contrast, the results for Beijing are as follows:

MILS 1: Busy bee, users who have low overall access intensity, only for occasional news and search content, which is similar to the behavior of white-collar workers.

MILS 2: Gregarious bird, users who are the spokespersons of social networks. On the basis of the intensity of the columns, it is clear that social apps serve as their entryway to all types of content. Additionally, their online life cannot be separated from social applications such as WeChat or Weibo.

MILS 3: Sociable shopper, the shopping and social software access behavior of these users forms a circle, and they are more inclined to switch between these two types of software.

MILS 4: Businessman, users who are not interested in Internet usage. They only search on the Internet to understand other content, such as forums and travel.

Some meaningful comparisons between the two core tensors exist. In Jinhua, as a developing city, the composition of user Internet access is relatively simple, and it is difficult to identify an obvious circle. However, a combination of various potential features can be found due to the specific online behavior of users. Based on the results of the feature discovery, service providers can formulate special recommendations in the region, for example, placing a product recommendation in a news application or providing attractive game links under a search engine. These recommendations will result in more traffic. In Beijing, as a well-developed city, user behavior is composed of the mixed usage of multiple apps, obvious circle structures, such as social-shopping loops, can be found. The presence of the busy bee and businessman makes it even more challenging for service owners to seize market opportunities. We found that users in Beijing have more concentrated Internet surfing behavior than do users in Jinhua; therefore, we recommend that urban administrators take social applications as the main entrance to Internet supervision to reduce the expenditure of other applications.

Taking into account the accuracy of the results, we resampled the current data and applied the feature extraction method to the MILS results. The resampling was performed randomly. The number of samples was the same as that of the previous experiment. The results in Fig. 2 and Fig. 3 are based on the Jensen–Shannon (JS) similarity of divergence. The method was proposed by Shannon and later published in the literature [25]. The evaluating indicator is symmetrical, and the distance is between 0 and 1, with a larger value indicating a larger similarity. The similarity among the sampling results of the two cities are shown in Fig. 4. As shown in Fig. 4,

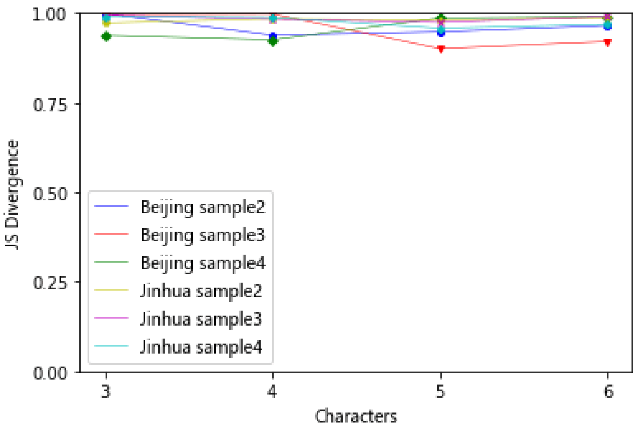


FIGURE 4. Comparison of the sampling similarity in two cities based on JS divergence.

the decomposition results of different samples are roughly similar when the number of MILS is 3-6, which statistically represents the composition of the mobile Internet surfing features that exist in the two regions.

B. REGIONAL INTERNET LIFESTYLES

Based on the weighted results of the two cities, we need to divide the distribution of MILS weights more granularly into urban areas. Analyzing the spatial pattern of human activities can provide a good understanding of the patterns of daily activities. In addition, analysis of human initiative may give rise to an awareness of the patterns that improve their daily activities. In cellular networks, users always contact the nearest base station, so the location data accuracy is approximately 300-500 *m* [26]. Therefore, each user's location can be approximated by the sequence of locations of BSs. Exploration of the geographical distribution of people with a small range of daily activities (R) is interesting [27]. Considering the mobility of users, we study the radius of users' activity. The activity radius of user u in space r is defined as follows:

$$r_g^u = \sqrt{\frac{1}{n_c^u} \sum\nolimits_{i=1} (r_i^u - r_{cm}^u)^2} \tag{11}$$

r_{cm} represents the centroid position of the movable range of u during the observation period, r_i is the ith position of the user's activity track, and n_c represents the number of user track points. Fig. 5 shows the cumulative density function (CDF) and probability density function (PDF) of mobile user gyration in both regions. Mobile users have limited spatial activity regions in both Beijing and Jinhua. In particular, approximately 55% of users in Beijing move less than 1 km, and 62% of users in Jinhua and 90% of users in Beijing move less than 10 *km*. On the double logarithmic axis, the PDF of user gyration in Beijing and Jinhua shows two distinct distributions. In Beijing, for the user groups whose gyration radius is more than 10 *km*, we find that the probability density shows a straight line. For the users whose gyration radius is less than 10 *km*, the probability density shows an exponential curve. In Jinhua, the areas on both sides of the boundary show a linear distribution, but the slope shows a significant difference.

Due to the local policies, local customs and cultural influences, cities have natural cultural and economic differences [28]. The finer-grained urban individual users' online lifestyle can use the administrative area as an entry point. In both locations, the gyration radius of more than 90% of mobile users is within 10 km. The average administrative area of Jinhua is 1215.8 km^2 and that of Beijing is 1025.7 km^2, which cover most users' activity radii. Therefore, a hypothesis about the region can be proposed, that is, the majority of mobile users do not leave their own administrative area. According to this hypothesis, we found the distribution features of the weights of MILS in the region. Fig. 6 shows the MILS representative statistics of the two locations. If we use the largest proportion in the region as the representative of the entire administrative region, as shown in Fig. 6, it is clear that the most important proportion of Jinhua is MILS 3, corresponding to "Casual Player", which means that in the entire city, the mobile Internet users in all regions are mainly reflected by this network personality. However, different results are observed in Beijing, where the distribution is polarized. For example, in Changping District, the main user group is busy bee, whose overall rate of Internet access is not very high. In the five urban districts and Changping District, users showed more enthusiasm for social networking and higher network activity. These two results are consistent with the conclusion of Vincente and Lopez [29] for users in the European Union-27. In the country survey, users in urban areas or more developed regions were found to have a higher degree of Internet access. However, we found that in relatively underdeveloped areas, Internet users are not represented by these features. The ratios of MILS types in the two regions are shown in Fig. 7 and Fig. 8.

Fig. 8 shows the specific distribution of MILS. GDP per capita and population density are important indicators of regional development. Therefore, we obtained the per capita GDP and population density of Beijing in 2015 [30]. The 16 districts are divided into two categories, one includes Dongcheng, Xicheng, Chaoyang, Haidian, Fengtai, Shijingshan, Shunyi, and Daxing, representing the economically developed regions, and the other one includes Yanqing, Miyun, Pinggu, Mentougou, Fangshan, Huairou, Tongzhou, and Changping, representing the less developed areas. As shown by the distribution diagram, the top five regions of MILS 1 are Yanqing, Pinggu, Mentougou, Fangshan and Daxing, while the areas with the highest number of MILS 2 are Chaoyang, Xicheng, Haidian, Dongcheng and Shijingshan. Therefore, users in MILS 1 (who are not enthusiastic about the Internet) mostly live in more developed areas, and users in MILS 2 (who are keen on surfing the Internet and love social issues) mostly live in underdeveloped areas. This further validates that those with social preferences are more likely to be affluent, which is in line with the conclusion

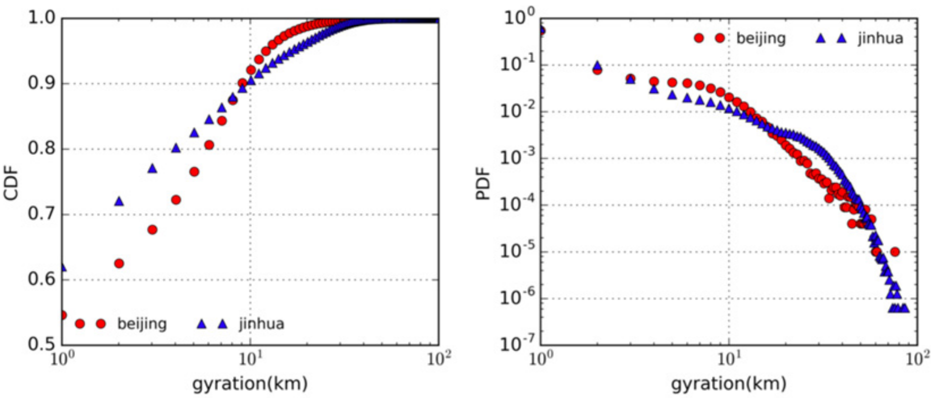


FIGURE 5. Statistics of the radius of the users in two cities. Note the cumulative density function diagram of the active radius on the left and the probability density function diagram on the right.

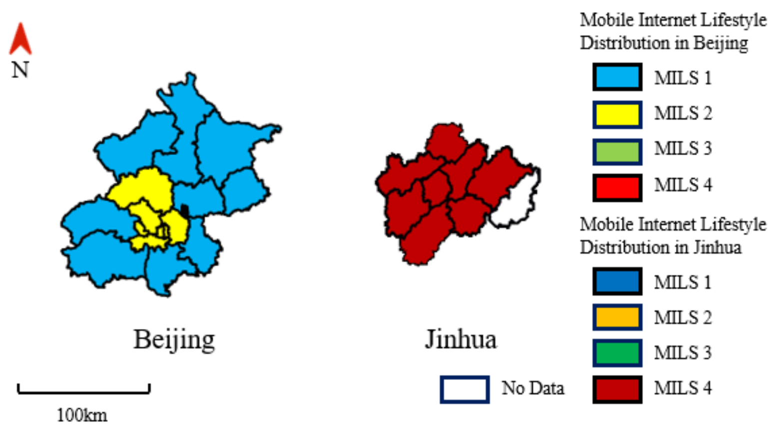


FIGURE 6. Two most representative MILS map based on the administrative division of the city. Note that the map on the left represents Beijing, and the map on the right represents Jinhua.

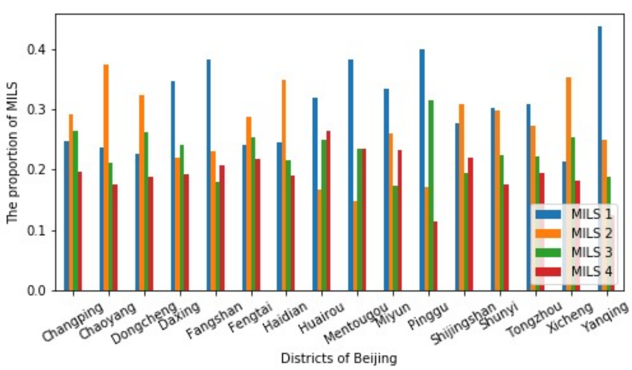


FIGURE 7. Intra-regional occupation ratios in Beijing's MILS.

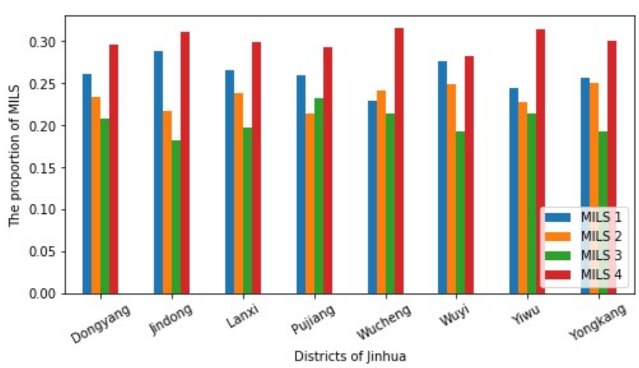


FIGURE 8. Intra-regional occupation ratios in Jinhua's MILS.

reached by Luo *et al*. [12] that the socio-economic status of people with a large social circle is high.

Two interesting areas are observed in the image, namely, Daxing and Changping Districts. In these two regions only, the population features are opposite to the statistics. However, the social behavior indicates that there is a larger number of non-resident residents than in the other regions (1,963,000 and 1,562,000, respectively). In addition, the network character is related to the potential for development in these two regions (both have been incorporated into the Beijing Urban Development Area).

On the other hand, Jinhua has become the largest commodity distribution center in the world due to the existence of Yiwu District (with a GDP of 104.6 billion yuan in 2015 [31]). In recent years, the development of e-commerce has provided a new source for the development of small commodities, thus leading to a high presence of MILS 2 (Shopaholic) in Jinhua City. However, as a developing city (GDP of 340.65 billion yuan in 2015 compared to 2.301559 trillion yuan in Beijing), different urban development processes (developed metropolitan and developing commercial cities) and urban economic development process led to a clear pattern of MILS distribution in Jinhua (MILS 4 > MILS 1 > MILS 2 > MILS 3). Information apps (search, current affairs), message carriers (shopping apps and product launching websites) and entertainment are most common for self-employed individuals in the small-commodity market. The ''Yiwu Mode'' is widely used by people on the Internet, a manifestation of the influence of highly developed regions in ordinary cities. One more effort is required, namely, to find the factors significantly influencing the representation in a city, that is, the ''Yiwu Mode'' in many developing regions, which portrays the city's unique demographics and provides an important guide for mobile operators, app service providers and policy designers.

The distribution of derivative weights is a hallmark of the everyday activity patterns of individual networks, enabling a systematic comparison of the behavioral features of human network users in different geographic regions. The results of the analysis reveal that the impact of human activity in both cities is worth discussing. First, a considerable part of the network life in these two cities is limited. This unique mode of activity may reflect a number of social issues that are mainly related to low-income migrant communities.

For most people in both cities, small-scale network activity is usually sufficient to satisfy daily Internet needs, which is in line with the government and operators' goals because small-scale access can make network regulation and market policies more accurate.

However, under the conditions of economic development and the differences between North and South, the typical features of the network users in two cities and their geographical distribution differ. This context-based universal feature not only provides an exciting market channel for operators and service providers but also a network regulatory understanding of the ability to guide the apps for the government. In the context of the weak ties and structural holes in social networks [12], it is important to discover the network dominance of mobile applications.

V. CONCLUSION

The development of the mobile Internet is accompanied by clear regional differences. The community has benefited from mobile Internet traffic. Operators and service providers have also enjoyed tremendous business opportunities and profits from the user-centered Internet. However, the expansion of the market is also accompanied by problems, namely, higher load and more complex service. A reasonable understanding of the key features of mobile Internet access behavior in different regions of the world can help us to overcome these problems. In this study, we found a way to explain the regional user lifestyles of mobile Internet users based on an access sequence model. From MILS, we can easily acquire typical APP transfer sequence patterns and understand key user profiles. They can not only provide operators with abundant business information, but also derive rational allocation of resources and support the creation of intelligent cities. On the other hand, we have also investigated the potential connotation of MILS combined with geographical information and socio-economic factors, which broadens our horizons of mobile Internet supervision. Meanwhile, this method is suitable for large scale data sets for its low complexity.

The main contributions of this paper include the following: (1) an MILS integrated mining theory based on improved HOSVD for mobile Internet users, (2) fine-grained extraction and comparison of typical network lifestyles in two urban areas with different economic development levels, and (3) analysis of the relationship between various mobile Internet services from the perspective of app category transfer. These results objectively confirm the stereotypes of the network activities in cities of different sizes and provide operators and service providers with service suggestions for different regions. Even though this paper produces similar results after multiple sampling target analyses, the dataset might be not complete. For example, under certain conditions, individuals in the network may go online through other means, such as wireless networks, and this type of access is not considered. In this paper, we compare and contrast the lifestyles of individuals in the Internet in terms of the transfer process. However, Internet lifestyles can be an imperfect concept that covers many aspects of behavior on the Internet. In future work, we could investigate which types of cyber-lifestyles can be classified culturally across regions. Despite these limitations, we believe that this research provides an opportunity for network supervision and is a powerful tool for social computing in the Big Data era.

REFERENCES

[1] M. J. Haught, R. Wei, Y. Xuerui, and J. Zhang, ''Understanding the psychology of mobile phone use and mobile shopping of the 1990s cohort in China: A lifestyle approach,'' in *Mobile Commerce: Concepts, Methodologies, Tools, and Applications*. Hershey, PA, USA: IGI Global, 2018, pp. 88–105.

[2] *The 40th China Statistical Report on Internet Development*, CNNIC, Beijing, China, 2017.

[3] Y. Lee, J. Kim, I. Lee, and H. Kim, ''A cross-cultural study on the value structure of mobile Internet usage: Comparison between Korea and Japan,'' *J. Electron. Commerce Res.*, vol. 3, no. 4, pp. 227–239, 2002.

[4] A. Pasagada, P. Reddy, B. Rohit, and V. Suneetha, ''Lifestyle comparison between india vs germany and Telangana vs Chennai-a brief study,'' *Res. J. Pharmaceutical Biol. Chem. Sci.*, vol. 7, no. 3, pp. 1651–1659, 2016.

[5] X. Chen and K. Siu, ''Smart communication in the context of lifestyle transformation: A case study of smartphone user behaviour among Hong Kong's young people,'' *Int. J. Interdiscipl. Studiesin Commun.*, vol. 9, no. 2, pp. 13–28, 2015.

[6] S. Jiang, J. Ferreira, and M. C. Gonzalez, ''Analyzing household lifestyles mobility and activity profiles: A case study of Singapore,'' 2012.

[7] M. S. Parwez, D. B. Rawat, and M. Garuba, ''Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network,'' *IEEE Trans. Ind. Inform.*, vol. 13, no. 4, pp. 2058–2065, Aug. 2017.

[8] R. Pulselli, P. Ramono, C. Ratti, and E. Tiezzi, "Computing urban mobile landscapes through monitoring population density based on cellphone chatting," *Int. J. Des. Nature Ecodyn.*, vol. 3, no. 2, pp. 121–134, 2008.

[9] J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti, "Cellular census: Explorations in urban data collection," *IEEE Pervasive Comput.*, vol. 6, no. 3, pp. 30–38, Jul./Sep. 2007.

[10] 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 *Proc. Int. Conf. Pervasive Comput.* Berlin, Germany: Springer, 2010, pp. 22–37.

[11] F. Girardin, A. Vaccari, A. Gerber, A. Biderman, and C. Ratti, "Towards estimating the presence of visitors from the aggregate mobile phone network activity they generate," in *Proc. Int. Conf. Comput. Urban Planning Urban Manage.*, Jun. 2009, pp. 1–12.

[12] S. Luo, F. Morone, C. Sarraute, M. Travizano, and H. A. Makse, "Inferring personal economic status from social network location," *Nature Commun.*, vol. 8, May 2017, Art. no. 15227.

[13] A. Decuyper *et al.* (2014). "Estimating food consumption and poverty indices with mobile phone data." [Online]. Available: https://arxiv.org/abs/1412.2595

[14] J. Park, D.-S. Lee, and M. C. González, "The eigenmode analysis of human motion," *J. Stat. Mech., Theory Exp.*, vol. 2010, no. 11, p. 11021, 2010.

[15] M. J. Cole, C. Hendahewa, N. J. Belkin, and C. Shah, "User activity patterns during information search," *ACM Trans. Inf. Syst.*, vol. 33, no. 1, 2015, Art. no. 1.

[16] X. Zhang, C. Wang, Z. Li, J. Zhu, W. Shi, and Q. Wang, "Exploring the sequential usage patterns of mobile Internet services based on Markov models" *Electron. Commerce Res. Appl.*, vol. 17, pp. 1–11, May/Jun. 2016.

[17] Y. Qiao, X. Zhao, J. Yang, and J. Liu, "Mobile big-data-driven rating framework: Measuring the relationship between human mobility and app usage behavior," *IEEE Network*, vol. 30, no. 3, pp. 14–21, May/Jun. 2016.

[18] H. Kawazu, F. Toriumi, M. Takano, K. Wada, and I. Fukuda, "Analytical method of Web user behavior using hidden Markov model," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 2518–2524.

[19] O. Shafiq, R. Alhajj, and J. G. Rokne, "On personalizing Web search using social network analysis," *Inf. Sci.*, vol. 314, pp. 55–76, Sep. 2015.

[20] M. Jiang, P. Cui, F. Wang, X. Xu, W. Zhu, and S. Yang, "FEMA: Flexible evolutionary multi-faceted analysis for dynamic behavioral pattern discovery" in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 1186–1195.

[21] L. Wu, Y. Li, C. Zhou, H. Jiang, and X. Qian, "Statistic analysis of data access behavior in the mobile Internet," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2013, pp. 89–93.

[22] C. Zhou, H. Jiang, Y. Chen, L. Wu, and S. Yi, "User interest acquisition by adding home and work related contexts on mobile big data analysis," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2016, pp. 201–206.

[23] J. G. Shanahan and L. Dai, "Large scale distributed data science using apache spark," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 2323–2324.

[24] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and Rank-(R_1 ,R_2 ,. . .,R_N) approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, 2000.

[25] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.

[26] Y. Yuan, M. Raubal, and Y. Liu, "Correlating mobile phone usage and travel behavior—A case study of Harbin, China," *Comput., Environ. Urban Syst.*, vol. 36, no. 2, pp. 118–130, 2012.

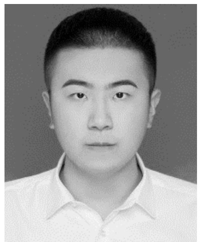
[27] Y. Xu *et al.*, "Another tale of two cities: Understanding human activity space using actively tracked cellphone location data," *Ann. Amer. Assoc. Geographers*, vol. 106, no. 2, pp. 489–502, 2016.

[28] L. Wu, H. Leung, H. Jiang, H. Zheng, and L. Ma, "Incorporating human movement behavior into the analysis of spatially distributed infrastructure," *PLoS ONE*, vol. 11, no. 1, p. e0147216, 2016.

[29] M. R. Vicente and A. J. López, "Assessing the regional digital divide across the European Union-27," *Telecommun. Policy*, vol. 35, no. 3, pp. 220–237, 2011.

[30] *Statistical Communiqué on the National Economy and Social Development of Beijing in 2015*, Beijing Municipal Bureau Statist., Beijing, China, 2016.

[31] *Statistical Communiqué on the National Economy and Social Development of Jinhua in 2015*, Jinhua Municipal Bureau Statist., Jinhua, China, 2016.



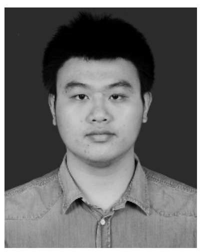
YI ZHANG was born in Lanzhou, Gansu, China, in 1994. He received the B.S. degree in communication engineering from Lanzhou University, Lanzhou, China, in 2016.

He is currently pursuing the M.S. degree with the Department of Information and Communication Engineering, School of Electronic Information, Wuhan University. His research interests include data mining, social computing, machine learning, and statistical analysis.

LINTAO YANG received the M.Eng. and Ph.D. degrees in communication networks from Wuhan University, Wuhan, China, in 2005 and 2011, respectively. He is currently a Lecturer with the College of Physical Science and Technology, Central China Normal University. His research interests include complex networks and wireless ad hoc networks.

HAO JIANG received the B.Eng. degree in communication engineering and the M.Eng. and Ph.D. degrees in communication and information systems from Wuhan University, China, in 1999, 2001, and 2004, respectively. From 2004 to 2005, he undertook his post-doctoral research work at LIMOS, Clermont-Ferrand, France. He was a Visiting Professor with the University of Calgary, Canada, ISIMA, and Blaise Pascal University, France. He is currently a Professor with Wuhan University. He has authored over 60 papers in different journals and conferences. His research interest includes mobile ad hoc networks and mobile big data.

SHUWEN YI received the B.S. degree in communications engineering from Wuhan University, China, in 2014, where he is currently pursuing the Ph.D. degree with the Electronic Information School. His current research interests include wireless communication networks, machine learning, data mining, and big data analytics.



ZHIYI HU was born in Shaoyang, Hunan, China, in 1994. He received the bachelor's degree from the College of Communications Engineering, Chongqing University, Chongqing, China, in 2016. He is currently pursuing the M.S. degree with the Department of Information and Communication Engineering, School of Electronic Information, Wuhan University. His research interests include data mining, machine learning, and distributed computing.

XING LIU (M'16) received the Ph.D. degree from Wuhan University in 2014. He is currently an Associate Professor with the School of Computer Science and Technology, Wuhan University of Technology. He has authored over 20 publications and book chapters. His research interests include the Internet of Things, the big data, the cloud computing, and the embedded systems. He is a member of the China Computer Federation. He received the Best Paper Award from the IEEE International Conference PIC in 2014.