Exploring Human Mobility Patterns in Urban Scenarios: A Trajectory Data Perspective

Feng Xia, Jinzhong Wang, Xiangjie Kong, Zhibo Wang, Jianxin Li, and Chengfei Liu

The authors propose an integrated computing method to rescale heterogeneous traffic trajectory data, which leverages MLE and BIC. Their analysis is based on two real datasets generated by subway smart card transactions and taxi GPS trajectories from Shanghai, China, which contain more than 451 million trading records by 14 subway lines and 34 billion GPS records by 13,695 taxis.

Abstract

Smart cities have been recognized as a promising research focus around the world. To realize smart cities, computation and utilization of big data are key factors. More specifically, exploring the patterns of human mobility based on large amounts of multi-source data plays an important role in analyzing the formation of social-economic phenomena in smart cities. However, our acquired knowledge is still very limited for smart cities. In this article, we propose an integrated computing method to rescale heterogeneous traffic trajectory data, which leverages MLE and BIC. Our analysis is based on two real datasets generated by subway smart card transactions and taxi GPS trajectories from Shanghai, China, which contain more than 451 million trading records by 14 subway lines and 34 billion GPS records by 13,695 taxis. Specifically, we quantitatively explore the patterns of human mobility on weekends and weekdays. Through logarithmic binning and data fitness, we calculate the Bayesian weights to select the best fitting distributions. In addition, we leverage three metrics to analyze the patterns of human mobility in two datasets: trip displacement, trip duration, and trip interval. We obtain several important human mobility patterns and discover quite a few interesting phenomena, which lay a solid foundation for future research.

Introduction

In recent years, with the rapid development of information communication technology and the continuous improvement of people's living standards, smart cities have been recognized as a promising research focus around the world, and have emerged as a novel paradigm to deal with existing problems and circumvent potential issues in modern cities, such as social computing, traffic congestion, energy consumption, and environment pollution. In particular, numerous cities have given priority to constructing smart cities and made remarkable achievements, for example, Vienna, Toronto, and Paris.

In smart cities, people usually use modern transportation facilities for their social activities (working, shopping, and traveling), which leads to a novel communication network referred to as the vehicular social network [1]. As a a new typical cyber-physical social system, it faces some challenging problems including mobility pattern analysis, data dissemination, node privacy, and trust management [2]. In this work, we mainly focus on and solve the first of these challenge.

Nowadays, multi-source traffic data are more easily collected than before, benefiting from the pervasive application of intelligent transportation equipment such as GPS devices, traffic cameras, smart cards, and road deceleration devices. Analyzing and mining the patterns of human mobility hidden in traffic big data have been a hot research field associated with transportation management, urban planning, epidemic control, mobile platform application, and so on [3, 4]. Meanwhile, there are a variety of traffic trajectory data including mobile phone calls, GPS, bank notes traces, and social media check-ins, which contribute to exploring the patterns of human mobility more deeply.

To our knowledge, researchers have proposed quite a few spatio-temporal patterns of human mobility, such as the Lévy flight model [5] and power law distribution [6]. Although the findings mentioned above provide a beneficial reference on mining the scaling laws in human mobility, they mostly focus on a single dataset, which is not enough to analyze the patterns of human mobility very well. Furthermore, the evaluation standards are different based on different research data. Zhao *et al.* [7] explain the power law distribution in detail through integrating four transportation modes. However, the datasets only contain 182 participants in Beijing, and the transportation mode information is manually logged, which is not sufficient to exploit the patterns of human mobility. Therefore, our research, based on real large-scale traffic trajectory datasets, is extremely indispensable and gains valuable insight into human mobility patterns.

In this article, we use two traffic trajectory datasets: subway smart transit data from 14 subway lines and Qiangsheng taxi data in Shanghai with 13,965 taxis. Specifically, we introduce three metrics (trip displacement, trip duration, and trip interval) to analyze the patterns of human mobility on weekends and weekdays.

First, we propose an integrated computing method which leverages maximum likelihood estimation (MLE) and the Bayesian information

Feng Xia, Jinzhong Wang, and Xiangjie Kong (corresponding author) are with Dalian University of Technology and the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province; Jinzhong Wang is also with Shenyang Sport University; Zhibo Wang is with Wuhan University; Jianxin Li is with the University of Western Australia; Chengfei Liu is with Swinburne University of Technology.

Digital Object Identifier: 10.1109/MCOM.2018.1700242

criterion (BIC) to acquire the best fitting distribution and the suitable parameters; second, we find that human mobility has its own characteristics different from existing statistical results in Shanghai, China, which contributes to inferring human mobility patterns more deeply; third, we discover that the patterns of trip displacement by subway and taxi are exactly the same and follow log-normal distribution rather than exponential distribution, which has been proved, and the scaling laws in trip duration are fitted to Weibull distribution for subway and log-normal distribution for taxi. Finally, for trip interval, we conclude that Weibull distribution can fit the probability curve by taxi rather than log-normal distribution. However, there is a two-regime pattern of human mobility in riding the subway.

The remainder of the article is organized as follows. In the following section, we briefly review the related work from two aspects. The section following that mainly introduces methods used in this work and illustrates two traffic trajectory datasets in detail. Then we quantitatively analyze human mobility patterns between subway and taxi on weekends and weekdays. The work is concluded in the final section.

Related Work

We comparatively analyze quite a few relevant works in the literature and give a brief review on the related work in the following two subsections.

Trajectory Data Processing

Recently, there is a bunch of research focusing on processing human mobility trajectory data more efficiently, which includes clustering algorithms to partition and group traces into clusters [8], tree-based ensemble classifiers to identify different transportation modes [9], and a neural Turing machine to learn human mobility patterns [10]. In addition, the algorithm for density-based spatial clustering of applications with noise is introduced to cluster pickup and dropoff locations, and uncovers human mobility from taxi trajectories [11].

Unlike the above-mentioned research, we focus on the characteristics of human travel behaviors, and leverage MLE and BIC to analyze human mobility patterns.

Patterns of Human Mobility

At first, a random walk illustrates a stochastic process of particles and waves traveling, which is a solid foundation for the following proposed Lévy flight model with the distance following the heavy-tailed distribution and the direction being isotropic [5]. In [6], the authors propose a power law distribution fitting human travel by studying the circulation of bank notes in the United States. Song *et al.* [12] find that 93 percent of human mobility patterns can be predicted by analyzing mobile phone users' entropy and are not dependent on their trip distances.

To explain power law distribution, Zhao *et al.* [7] divide human transportation mode into four types, and verify that log-normal distribution approximates single movement, while power law distribution approximates a mixture of different transportation modes. In [13], credit card records are analyzed according to gender, age, and occupation; then the authors emphasize the importance of socio-demographic characteristics. Jiang *et al.* [14] introduce the activity-based approach to extract individual mobility patterns from mobile phone call detailed record data. Wang *et al.* [15] propose exponential distribution fitting the trip displacement rather than power distribution.

The results mentioned above are mainly based on single-source trajectory data. To capture the spatio-temporal laws in human mobility more efficiently, we utilize multi-source trajectory data and obtain some travel distributions differing from the existing patterns. For example, log-normal distribution fits trip displacement better than power distribution and exponential distribution.

An Integrated Computing Method

Overview

In our daily life, people usually go out for different purposes, which consist of multi-source traffic data such as mobile phone data, GPS data, transit data, and social web check-in data. To our knowledge, researchers have discovered quite a few human mobility patterns, including the power law and the exponential law. However, multi-source traffic data are seldom focused on such patterns. Furthermore, it is well known that the patterns of human mobility on weekends are different from those on weekdays. We quantitatively analyze the travel regularities between weekends and weekdays.

Definition 1: A human mobility network (HMN) $G = (V, E)$ is a directed graph, where V denotes the set of the origins or the destinations ($V \neq 0$), and E represents the set of edges that connect an origin location to a destination location ($\|E\| \geq 1$).

As shown in Fig. 1, we first collect the traffic data from the Shanghai SODA competition website. Then we leverage data manipulation language to clean the acquired datasets. According to graph theory, we construct an HMN. In particular, we introduce three evaluation metrics, trip displacement, trip duration, and trip interval, and analyze two datasets quantitatively and comparatively. Using MLE and BIC, we finally explore the patterns of human mobility on weekends and weekdays.

Dataset Preprocessing

At present, multi-source heterogeneous traffic data make analysis of human mobility patterns more precise. In this work, we use two datasets generated by subway smart card transactions and Qiangsheng taxi GPS trajectories from Shanghai, China. The datasets contain more than 451 million trading records by 14 subway lines and 34 billion GPS records by 13,695 taxis, respectively.

The two datasets are collected from April 1, 2015 to April 30, 2015. The subway data contains 346 subway stations and demonstrates the spatio-temporal laws in urban human mobility by the Shanghai subway system. In the taxi data, each transaction record consists of eight fields, in which the longitude and latitude fields denote a subway station's geographical location. The data for taxis is provided by Shanghai Qiangsheng Co., Ltd., and contains about 25 percent of the taxis in Shanghai, and the status field holds 4 numeric values of 0, 1, 2, and 3. 0 means that the taxi is occupied, 1 indicates the taxi is vacant, and 2 and

To our knowledge, researchers have discovered quite a few human mobility patterns, including the power law and the exponential law. However, multi-source traffic data are seldom focused on. Furthermore, it is well known that the patterns of human mobility on weekends are different from those on weekdays.

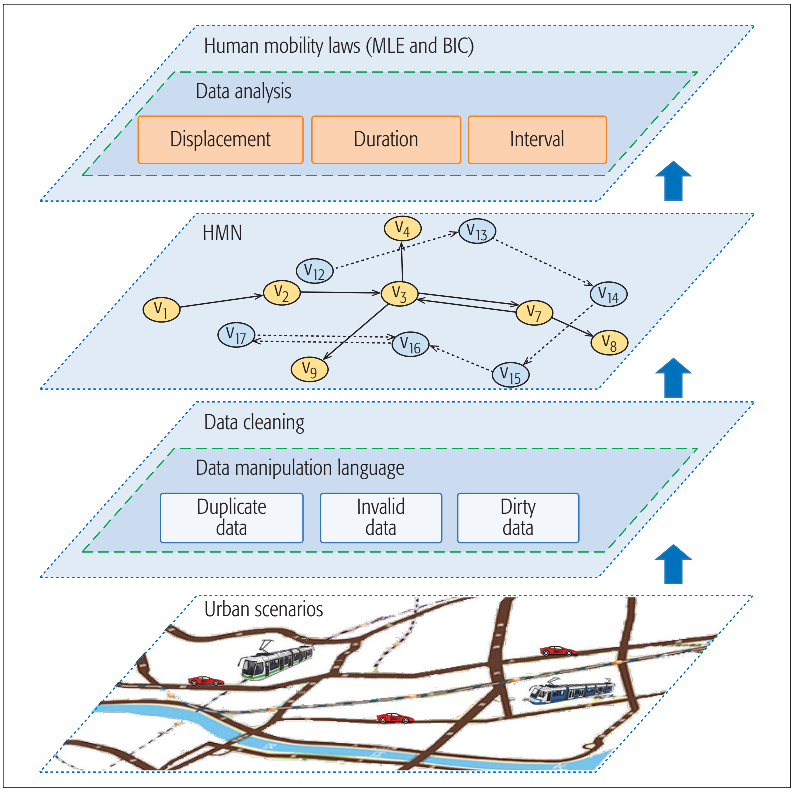

Figure 1. Overview of exploring the patterns of human mobility.

3 represent the non-service state. The status is triggered to change as the taxi picks up and drops off passengers. Each taxi GPS record is composed of nine fields. The average sampling rate is 6 s. To more conveniently describe the data, D_{s1} and D_{s2} denote the dataset of taking the subway on weekends and weekdays, respectively, and D_{t1} and D_{t2} correspondingly represent the dataset by taxi on weekends and weekdays.

To improve the quality of data, we clean duplicate data and invalid data caused by device fault. Leveraging the data manipulation language of SQL, we modify a slice of dirty data with a percentage of 1.85 percent according to the context. Then we extract urban folks' trips by subway and taxi, and discard the data whose duration is less than 1 min or larger than 3 h because passengers seldom spend a few seconds or a long time in their daily routines. Additionally, we also exclude trip displacement less than 200 m or greater than 120 km based on the urban terrain. Specifically, we focus on a certain urban area in Shanghai, which is located at 120.52° E to 122.12° E and 30.40° N to 31.53° N.

Logarithmic Binning and Data Fitness

Logarithmic binning is an important method to analyze the scaling laws of data with noise. Furthermore, logarithmic binning is intensely suitable for diagnosis of Lévy flight motions. We introduce it to unveil the patterns of urban human mobility in Shanghai, and determine whether the spatio-temporal pattern follows power law distribution, exponential law distribution, or any other distributions. To decrease the statistical errors in data fitting, we first normalize the dataset, show the probability distribution in a log-log coordinate system, and acquire quite a few relatively good results. Subsequently, we plot a scatter diagram and make the samples collapse into bins of exponentially growing widths.

This article leverages MLE and BIC to select the best fitting distribution. First of all, we use MLE to estimate the parameters in every distribution. Then we continue to calculate the value of BIC as follows:

$$BIC_i = -2\log(L_i(\hat{\theta} \mid data)) + K_i \log(n), \tag{1}$$

where BIC_i denotes the value of BIC in the ith model, $L_i(\cdot)$ represents the likelihood function, $\hat{\theta}$ indicates the fitted model parameters acquired by MLE, K_i is the number of parameters in the estimating model, and n is the number of observations. According to the BIC value of every distribution, we go a step further to obtain the Bayesian weights W_i to determine which is the best law. More specifically, W_i is generated by normalizing the BIC_i and denoted by

$$W_i = \frac{\exp(-\Delta_i 2)}{\sum_{i=1}^{n} \exp(-\Delta_i 2)}, \tag{2}$$

where n denotes the number of fitting models, and Δ_i is the difference between BIC values and defined by

$$\Delta_i = BIC_i - BIC_{min}.$$

Overall Analysis of Human Mobility Patterns

To intuitively show the patterns of human mobility on weekdays and weekends, we comparatively analyze trajectory data and divide a day into 12 time slots with an interval of 2 h. Then we select two peak hours and one off-peak hour of a day such as 7:00–9:00, 11:00–13:00, and 17:00–19:00, which represent morning rush hours, lunch breaks, and evening rush hours.

The Spatio-Temporal Patterns by Subway

As shown in Figs. 2a–2c, we notice that there are several popular subway stations and relatively more passenger metro trips. More specifically, Shanghai Railway Station and Hongqiao Railway Station stay busy from 7:00 to 19:00 (yellow), which is consistent with our experience. For Zhongshan Park and People Square, the passenger flow in the morning rush hours is obviously less than other time slots, which implies that individuals prefer to leave from or go to the two stations later on weekends. In addition, Xujiahui Station is located in the downtown area, which contributes to more volume in evening rush hours than others due to the downtown area location. We also note that the hottest trip is from Tonghe Xincun Station to Shanghai Railway Station in the morning rush hours, which is closely related to the urban region function. Likewise, more characters travel between Xinzhuang Station and Xujiahui Station in the evening rush hours for their social activities. If you want to know more about Shanghai subway lines, go to https://exploreshanghai.com/metro/.

As shown in Figs. 2d–2f, the passenger flow has its own characteristics, that is, much more in the rush hours than during lunch breaks. In particular,

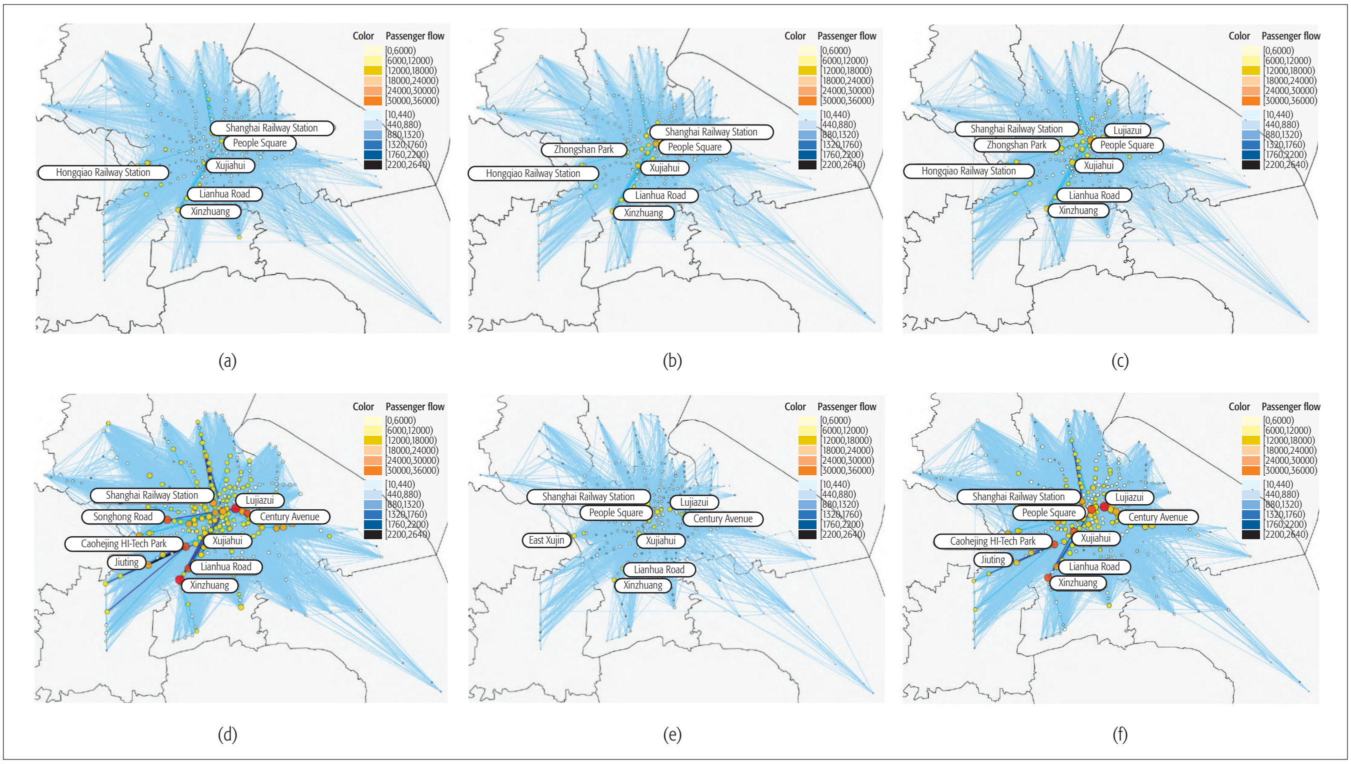

Figure 2. The spatio-temporal evolution of passenger flow by subway: a) 7:00–9:00 on weekends; b) 11:00–13:00 on weekends; c) 17:00–19:00 on weekends; d) 7:00–9:00 on weekdays; e) 11:00–13:00 on weekdays; f) 17:00–19:00 on weekdays.

folks who travel by subway in the morning rush hours are more than those in the evening rush hours, which suggests that a slice of individuals probably work later than 19:00. Moreover, there is the hottest trip between Jiuting and Caohejing Hi-Tech Park on subway line 9 in the rush hours, which is a busy commuter route in Shanghai.

Comparing the passenger volume on weekends and weekdays, we discover that the flow of the latter is four times the volume of the former. Through trip origins and destinations, we go a step further to analyze human mobility patterns. Most individuals prefer to have social activities on the weekend and go to work on weekdays. In addition, subway line 9 is intensely busy on weekdays, whereas subway line 1 is relatively busy on weekends.

The Spatio-Temporal Mobility Patterns by Taxi

As shown in Fig. 3, we use a heat map with colored spots and lines to show hotspots and relatively busy routes for taxis. On weekends, we discover that HongQiao Railway Station and Pudong International Airport are busy from 7:00 to 19:00, and the passenger volumes in the former reach the peak value in the rush hours rather than during lunch breaks. There is mainly a long trip between the downtown regions and Pudong Airport, as shown in Figs. 3a–3c. Nonetheless, as the second airport in Shanghai, Hongqiao Airport only has taxi trips at noon, which is just related to its more convenient location. In addition, we notice that there are three hotspots, that is, Jiangsu Road, South Shanxi Road, and West Nanjing Road, in the evening rush hours in Fig. 3c. The regions are located in flourishing districts that have restaurants, coffee shops, karaoke clubs, ballrooms, bars, and a modern business center. People love to go to these places for relaxation on weekend nights.

As shown in Figs. 3d–3f, we can infer that the trips by taxi in morning rush hours are obviously more than in other time slots, while HongQiao Railway Station and Pudong Airport are still hotspots. In the evening rush hours, Fudan University becomes an extremely busy region, which may be affected by commuting time after school. More specifically, there are relatively more trips from Vanke Qibao International Garden to Hongqiao Railway Station and from Hongqiao Airport to South Zhuxiang in the morning rush hours, and from Hongqiao Railway Station to Shenkun Road in the evening rush hours. The regularities of human mobility provide valuable information for taxi operation. Moreover, we analyze human movement patterns by taxi between weekends and weekdays comparatively. We find more individuals travel by taxi on weekends than weekdays. All taxi trips have a high level of aggregation in the city center. Besides, by studying Figs. 2 and 3, we can notice that the distribution by subway is more regular than that by taxi, due to the fixed routes by subway.

Detailed Analysis of Human Mobility Patterns

We leverage three metrics to analyze the patterns of human mobility in the two datasets: trip displacement, trip duration, and trip interval. Through statistical analysis and data fitness, we obtain several important human mobility patterns.

Trip Displacement

Definition 2: A trip displacement is the spherical distance with a pair (v_i, v_{i+1}) in an HMN. The spherical distance is the inferior arc length by using great-circle between v_i and v_{i+1} on the surface of a sphere.

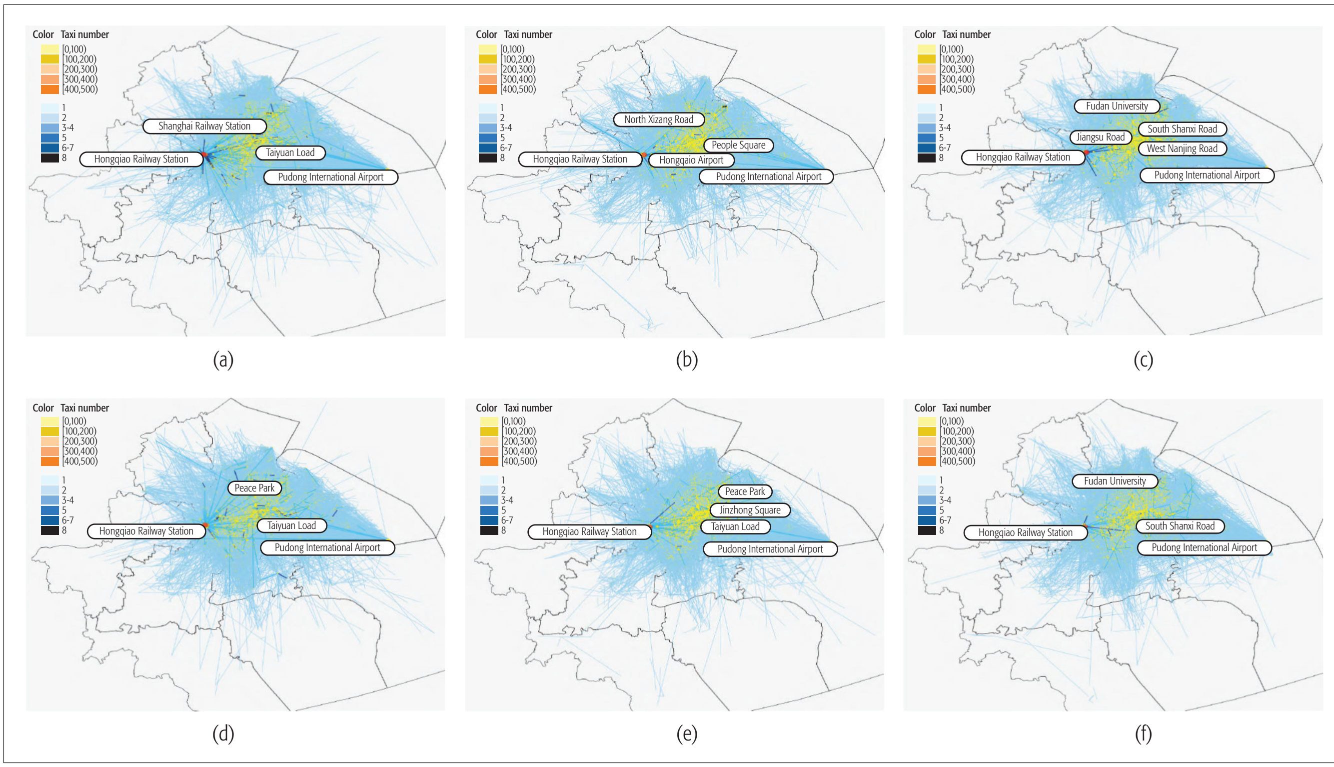

Figure 3. The spatio-temporal evolution of passenger flow by taxi: a) 7:00–9:00 on weekends; b) 11:00–13:00 on weekends; c) 17:00–19:00 on weekends; d) 7:00–9:00 on weekdays; e) 11:00–13:00 on weekdays; f) 17:00–19:00 on weekdays.

Trip displacement is an extremely important measurement indicator. Although the metric does not represent the actual length, it is determined by an origin and a destination rather than a travel route, and reflects urban human mobility patterns.

As shown in Fig. 4, we show three distributions to fit the trip displacement by subway and taxi, which are power law distribution, exponential distribution, and log-normal distribution. As the figure shows, the distributions of trip displacement have largely similar evolution law between weekends and weekdays by subway and taxi, while exponential distribution and log-normal distribution are better fitting curves than power law distribution. Furthermore, we discover that the best fitting distribution is log-normal law by using BIC in data fitness. The parameters of log-normal law are deduced by leveraging MLE. By comparing the parameters, we can infer that D_{t1} and D_{t2} are a little better fitted than D_{s1} and D_{s2}, and the latter has only 2 parameters less than the former, which means lower complexity.

To show a slice of interesting phenomena hidden in two datasets more clearly, we have a quantitative analysis between trip displacement by subway and taxi on weekends and weekdays. The displacement of the longest trip is 6 km with 6 percent and 7 km with 7 percent. We discover that 90 percent of occupied trips with displacement are less than or equal to 24 km and 21 km on weekends and weekdays, respectively. Therefore, it is obvious that most folks travel around 20 km for their social activities by subway, and the highest percentage of occupied trips is about 7 km. It is closely associated with advanced urban public transportation and the urban functional structure in Shanghai. In addition, individuals usually travel much further on weekends than weekdays, which implies that people prefer to go for an outing or other social activities rather than work on weekends.

We note that 90 percent of occupied trips are correspondingly within 11 km and 12 km on weekends and weekdays, and the most displacement is 2 km with 19 percent on weekends and 18 percent on weekdays. Furthermore, the difference in displacement between weekends and weekdays is just 1 km by taxi. However, the distance by taxi is far less than by subway, which contributes to the relatively high cost of the trip. Thus, most folks are more willing to ride the subway for a longer journey.

Trip Duration

Definition 3: A trip duration (t_d) is the elapsed time between an origin v_i and a destination v_j in an HMN.

Trip duration is another important evaluation metric and closely tied to trip displacement. According to Fig. 5, we discover that the distribution of trip duration by subway is obviously different from that by taxi. We leverage the four distributions to fit human mobility data and calculate the corresponding parameters. Based on empirical analysis, we conclude that the law of trip duration is fitted to log-normal distribution for taking a taxi and Weibull distribution for taking a subway. Nevertheless, there are similar laws between weekends and weekdays for the two transportation modes. Furthermore, we find that individuals usually spend 26 and 28 min taking the subway on weekends and weekdays, respectively. However, most people spend much less time (8 min) traveling by taxi.

We also consider the relationship between the trip displacement and the average trip duration,

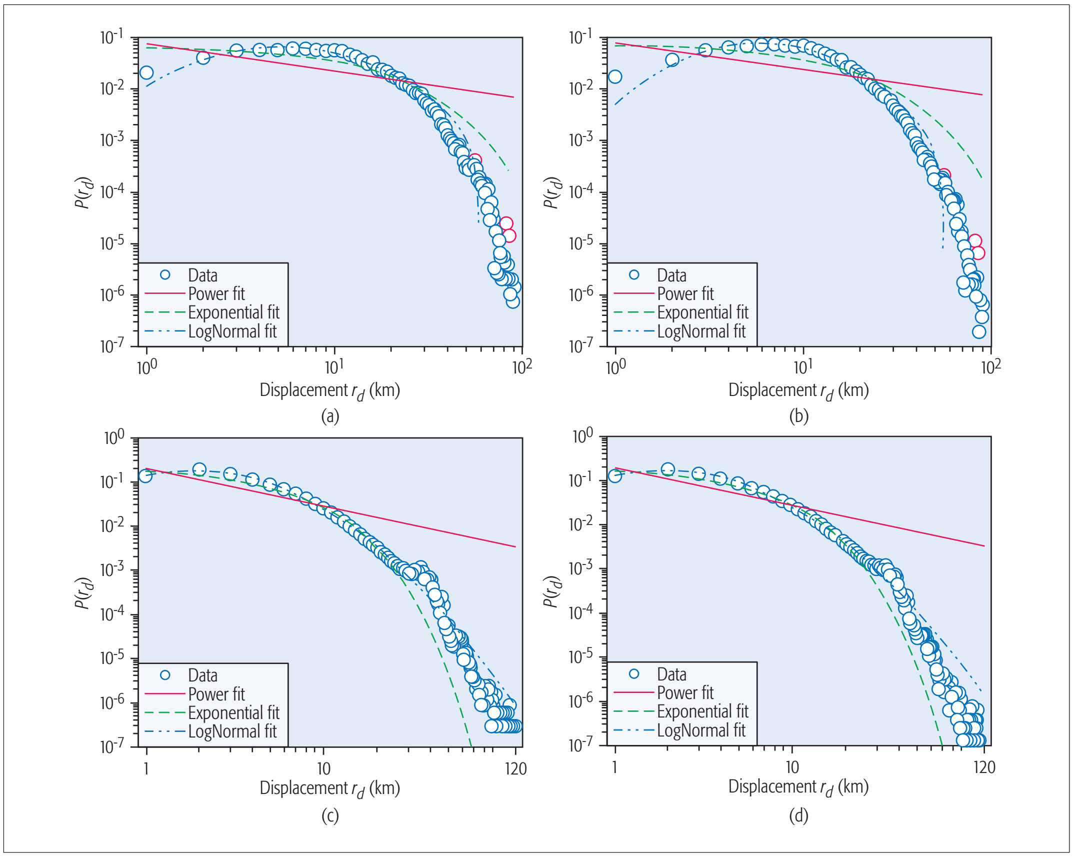

Figure 4. Patterns of trip displacement: a) subway on weekends; b) subway on weekdays; c) taxi on weekends; d) taxi on weekdays.

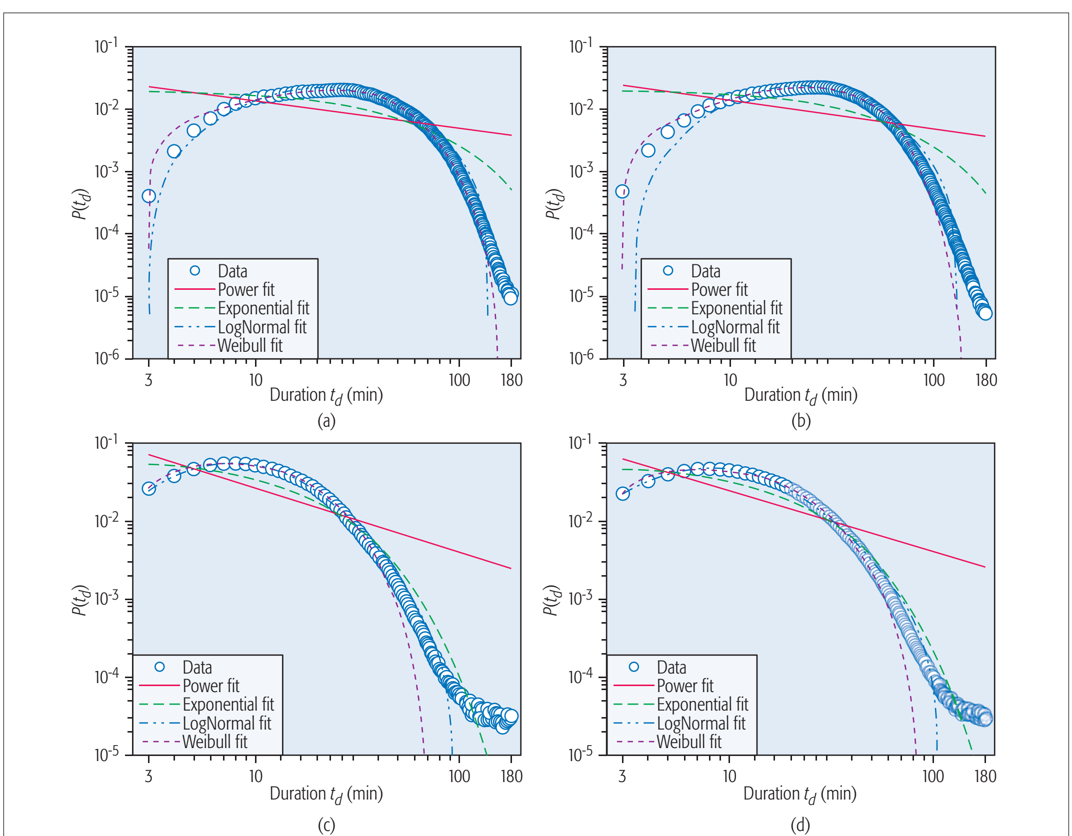

Figure 5. Patterns of trip duration: a) subway on weekends; b) subway on weekdays; c) taxi on weekends; d) taxi on weekdays.

Trip displacement is an extremely important measurement indicator. Although the metric does not represent the actual length, it is determined by an origin and a destination rather than a travel route, and reflects urban human mobility patterns.

Based on the empirical analysis, we conclude that the law of trip duration is fitted to log-normal distribution for taking a taxi and Weibull distribution for taking a subway. Nevertheless, there are similar laws between weekends and weekdays for the two transportation modes.

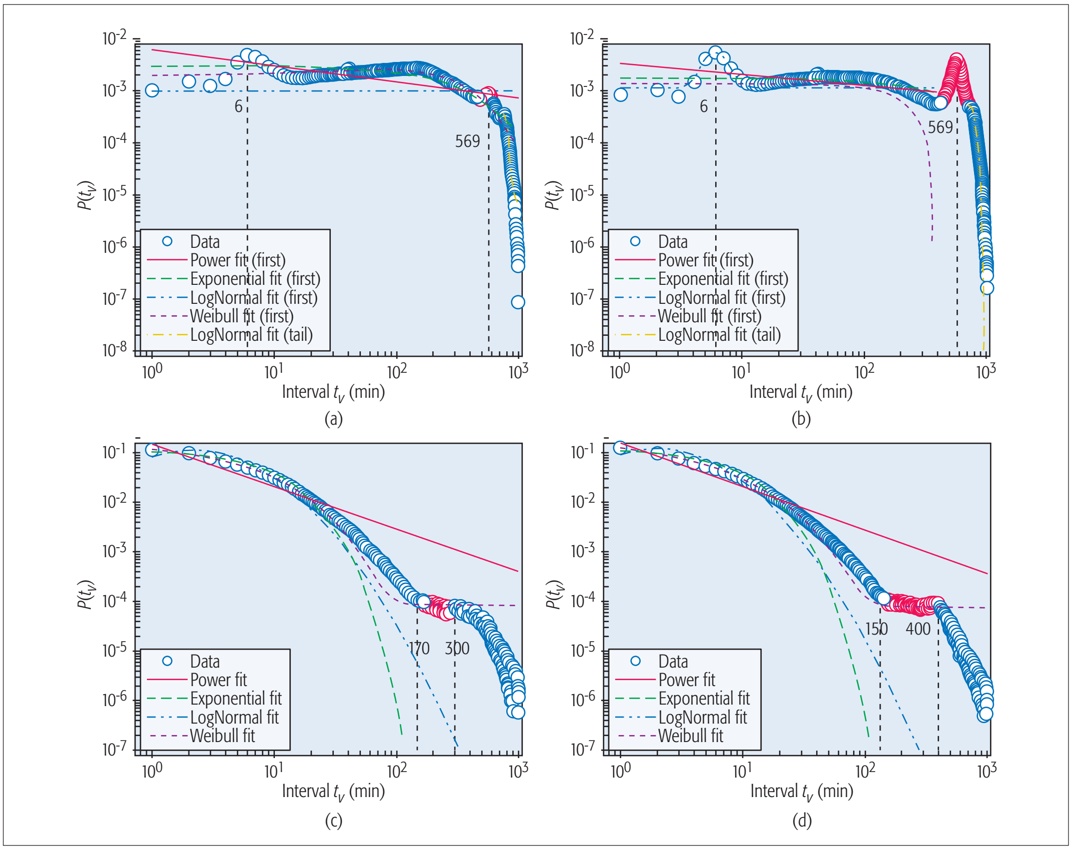

Figure 6. Patterns of trip interval: a) subway on weekends; b) subway on weekdays; c) taxi on weekends; d) taxi on weekdays.

which implies the level of urban traffic conditions more deeply. Through analyzing the regularities of displacement and duration, we note that there are quite a few interesting laws in Shanghai transportation. The average duration constantly increases with the rise of the displacement by subway, and approximately follows a linear function with the average speed changing from 18 km/h to 35 km/h. Although the subway operation is not impacted by traffic jams, the curve shows a slight fluctuation when the trip displacement is more than 46 km. It is probably related to subway transfers and the subway network, which makes travel speed go down.

In contrast, taxi operation is more complicated than subway operation. We notice that the average trip duration reaches the peak value (388 min) with the trip displacement of 93 km by taxi on weekends, whereas it is correspondingly 97 km on weekdays. This suggests that people prefer to travel on weekends, but unfortunately encounter bad traffic circumstances. In particular, we discover that the average speed by taxi is faster than by subway in general, while it fluctuates up and down more obviously than the latter. This may be attributed to a slice of factors such as the road network, holidays, traffic incidents, and weather conditions.

Based on a thorough analysis, we discover that more time needs to be spent at a transfer station for a subway trip over 120 min and results in an lower average speed. For taxis, it is related to the urban road networks and the shift of taxi drivers, which leads to more time spent on a short range of straight-line distance.

Trip Interval

Definition 4: A trip interval (t_v) is the elapsed time between two consecutive trips by the same person taking the subway or the same taxi.

To a certain extent, the evaluation metric also reflects urban human mobility patterns. More specifically, it refers to the interval for the same person by subway. For a taxi, it refers to the interval of two consecutive trips. As shown in Fig. 6, the scaling law in the trip interval by subway has its characteristics compared to those by taxi. By processing traffic data and analyzing inherent laws, we determine that Weibull distribution fits better than the others. Nonetheless, as seen from Figs. 6a and 6b, two peaks exist in the distribution of interval that contribute to only fitting the first part with the Weibull law. The tail part of the distribution follows the log-normal law exceedingly well.

For dataset D_s, the probability function increases quickly to a high value at 6 min and falls down slowly as trip interval varies uniformly. The second peak of trip interval is reached with the interval value of 569 min. Specifically, the percentage of the second peak on weekdays is much more than that on weekends, which indicates that most individuals spend the interval working on weekdays. With respect to the distance from home or company to subway station, the interval exactly coincides with eight work hours.

For dataset D_t, the probability function continues to drop from the starting point with the interval 1 min as trip interval increases. However, as shown in Figs. 6c and 6d, the probability of trip interval remains unchanged from 170 to 300 min on weekends and 160 to 400 min on weekdays, which shows that the taxi operation on weekends is more efficient than that on weekdays.

In addition, we discover that 90 percent of trip intervals by subway are within 580 min on weekends and 643 min on weekdays. In other words, folks go to another consecutive trip about 1 h earlier on weekends than on weekdays, which implies that most individuals travel for social activity rather than working, and the traffic is relatively smooth on weekends. Conversely, 90 percent of the trip intervals by taxi are not much different between weekends and weekdays. In particular, the interval by taxi is much less than that by subway because the evaluation object is different.

Conclusion

In this work, we propose an integrated analysis method to find the characteristics of human mobility in Shanghai, China. The approach utilizes MLE and BIC to fit the patterns of human movement based on subway transit data and taxi GPS data. Except for our mentioned datasets, our proposed method can also be used for other traffic trajectory data such as mobile phone calls, social media check-ins, and private car data. Through qualitative and quantitative analysis, we notice that trip displacement is better fitted to log-normal distribution rather than an exponential model. In particular, the level of data fitness for taxis is better than for subways. Nevertheless, no matter the travel method, the displacement on weekends is different from that on weekdays.

For the trip duration, we verify that Weibull distribution fits the elapsed time by subway, whereas log-normal distribution approximates the trip time by taxi. We also notice that most folks spend much less time on traveling by taxi (8 min) than by subway (28 min). In addition, the correlation between trip displacement and trip duration is discussed. As regards the trip interval, the patterns by taxi follow Weibull distribution, whereas the laws by subway show a two-regime scheme, which is the first part following Weibull distribution and the tail part obeying log-normal distribution. Specifically, the interval by subway is intensely close to 8 h more than that on weekdays, which manifests that most folks spend the interval on working days. We also discover that taxi operation is more efficent on weekends than on weekdays.

In future work, we plan to study the patterns of human mobility with respect to factors such as weather conditions and public holidays. To go a step further to mine patterns of human travel, we desire to integrate other transportation mode data and leverage other evaluation indicators to analyze traffic trajectory data more thoroughly.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China under Grant No. 61572106, the Natural Science Foundation of Liaoning Province, China, under Grant No. 201602154, and the Dalian Science and Technology Planning Project under Grants No. 2015A11GX015 and 2015R054.

References

[1] Q. Yang and H. Wang, "Toward Trustworthy Vehicular Social Networks," *IEEE Commun. Mag.*, vol. 53, no. 8, Aug. 2015, pp. 42–47.

[2] G. Xiong *et al.*, "Cyber-Physical-Social System in Intelligent Transportation," *IEEE/CAA J. Automatica Sinica*, vol. 2, no. 3, 2015, pp. 320–33.

[3] X. Kong *et al.*, "Urban Traffic Congestion Estimation and Prediction Based on Floating Car Trajectory Data," *Future Generation Computer Systems*, vol. 61, 2016, pp. 97–107.

[4] X. Kong *et al.*, "Time-Location-Relationship Combined Service Recommendation Based on Taxi Trajectory Data," *IEEE Trans. Industrial Informatics*, vol. 13, no. 3, 2017, pp. 1202–12.

[5] I. Rhee *et al.*, "On the Levy-Walk Nature of Human Mobility," *IEEE/ACM Trans. Networking*, vol. 19, no. 3, 2011, pp. 630–43.

[6] D. Brockmann, L. Hufnagel, and T. Geisel, "The Scaling Laws of Human Travel," *Nature*, vol. 439, no. 7075, 2006, pp. 462–65.

[7] K. Zhao *et al.*, "Explaining the Power-Law Distribution of Human Mobility Through Transportation Modality Decomposition," *Scientific Reports*, vol. 5, 2015, p. 9136.

[8] V. Tanuja and P. Govindarajulu, "Application of Trajectory Data Clustering in CRM: A Case Study," *Int'l. J. Computer Science Network Security*, vol. 17, no. 1, 2017, p. 137.

[9] Z. Xiao *et al.*, "Identifying Different Transportation Modes from Trajectory Data Using Tree-Based Ensemble Classifiers," *ISPRS Int'l. J. Geo-Info.*, vol. 6, no. 2, 2017, p. 57.

[10] J. Tkacík and P. Kordík, "Neural Turing Machine for Sequential Learning of Human Mobility Patterns," *Proc. 2016 IEEE Int'l. Joint Conf. Neural Networks*, 2016, pp. 2790–97.

[11] J. Tang *et al.*, "Uncovering Urban Human Mobility from Large Scale Taxi GPS Data," *Physica A: Statistical Mechanics and Its Applications*, vol. 438, 2015, pp. 140–53.

[12] C. Song *et al.*, "Limits of Predictability in Human Mobility," *Science*, vol. 327, no. 5968, 2010, pp. 1018–21.

[13] M. Lenormand *et al.*, "Influence of Sociodemographics on Human Mobility," *Scientific Reports*, vol. 5, 2015, p. 10,075.

[14] S. Jiang, J. Ferreira, and M. C. Gonzales, "Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore," *IEEE Trans. Big Data*, vol. 3, no. 1, 2016, pp. 208–19.

[15] W. Wang *et al.*, "A Comparative Analysis of Intra-City Human Mobility by Taxi," *Physica A: Statistical Mechanics and Its Applications*, vol. 420, 2015, pp. 134–47.

Biography

Feng Xia [M'07, SM'12] received his B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He is currently a full professor in the School of Software, Dalian University of Technology, China. His research interests include computational social science, network science, data science, and mobile social networks.

Jinzhong Wang received his B.S. degree in computer education from Anshan Normal University, China, in 2002 and his M.S. degree in computer application technology from Liaoning University, Shenyang, China, in 2005. He is currently working toward a Ph.D. degree with the School of Software, Dalian University of Technology. Since 2005, he has been with Shenyang Sport University, China. His research interests include traffic data analytics and mobile social networks.

Xiangjie Kong [M'13, SM'17] received his B.Sc. and Ph.D. degrees from Zhejiang University. He is currently an associate professor in the School of Software, Dalian University of Technology. His research interests include computational social science, urban big data, and mobile computing.

Zhibo Wang [M] received his B.E. degree in automation from Zhejiang University in 2007 and his Ph.D degree in electrical engineering and Computer Science from University of Tennessee, Knoxville, in 2014. He is currently an associate professor with the School of Computers, Wuhan University, China. His current research interests include mobile crowdsensing systems, cyber-physical systems, recommender systems, and privacy protection. He is a member of ACM.

Jianxin Li received his Ph.D. degree in computer science from Swinburne University of Technology, Australia, in 2009. He is currently a senior lecturer in the School of Computer Science and Software Engineering, UWA. His research interests include database query processing, social network analytics, and traffic data analytics.

Chengfei Liu [M] received his B.S., M.S., and Ph.D. degrees in computer science from Nanjing University, China, in 1983, 1985, and 1988, respectively. Currently he is a professor in the Department of Computer Science and Software Engineering, Swinburne University of Technology. His current research interests include keyword search on structured data, graph databases, social networks, query processing and refinement for advanced database applications, and data-centric workflows. He is a member of ACM.

In future work, we plan to study the patterns of human mobility with respect to factors such as weather conditions and public holidays. To go a step further to mine patterns of human travel, we desire to integrate other transportation mode data and leverage other evaluation indicators to analyze traffic trajectory data more thoroughly.