IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 23, 2018, accepted February 19, 2018, date of publication February 27, 2018, date of current version March 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2810140

The Impact of the Transportation Network Companies on the Taxi Industry: Evidence from Beijing's GPS Taxi Trajectory Data

WEIWEI JIANG[1], (Student Member, IEEE), AND LIN ZHANG[2], (Senior Member, IEEE)
[1]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[2]Shenzhen Engineering Laboratory for Data Science and Information Technology, Tsinghua-Berkeley Shenzhen Institute, Shenzhen 518055, China

Corresponding author: Lin Zhang (linzhang@tsinghua.edu.cn)

This work was supported by the Shenzhen Municipal Development and Reform Commission, Shenzhen Engineering Laboratory for Data Science and Information Technology, under Grant SDRC [2015]1872.

ABSTRACT To gain insight into how transportation network companies, such as Uber and Didi, impact the taxi industry, we conduct a multi-period analysis of taxi drivers' behaviors, based on GPS trajectory data collected from three time periods in Beijing, *i.e.*, November 2012, November 2014, and November 2015. We extract both passenger-delivery and passenger-searching trip information from GPS trajectories and compare the spatial, temporal, densification, and poolability properties of taxi trips in different time periods. Our results reveal that the taxi industry was adversely influenced by the competition between transportation network companies; as compared with that of 2012, the average passenger-delivery trip number per day per taxi dropped by 18.08% and the average daily profit per taxi dropped by 19.29% in the year 2015, respectively. We also compare passenger-searching strategies, passenger-delivery strategies, and service area preferences between taxi drivers with top and bottom efficiency in different time periods. We find that compared with drivers with lower efficiency, drivers with high efficiency tend to search locally, have a higher delivery speed, and serve more often within the inner part of Beijing.

INDEX TERMS Transportation, human factors, performance analysis.

I. INTRODUCTION

In the past few years, the largest disruption to the taxi industry is the rise of transportation network companies. A transportation network company (TNC), such as Uber, Lyft, and Didi Chuxing, provides prearranged transportation services for compensation of traditional taxi market using an online-enabled application (app) or platform to connect passengers with drivers using their personal vehicles [1]. Adopting more efficient driver-passenger matching technology and surge pricing rules [2], these TNCs have become strong competitors for traditional taxis.

Previous studies have shown that Uber has a higher efficiency than traditional taxis, by comparing the capacity utilization rate of UberX drivers with that of traditional taxi drivers in five cities [3], which leads to a significant decrease in traffic congestion and carbon dioxide emissions in the urban areas of the United States [4]. There is also evidence showing that Uber makes it easier to get a ride in rainy hours, by increasing supply with surge pricing [5].

The growth of TNCs is accompanied with the depression of taxi demand. Previous study has shown that after Uber entered the New York market in May 2011, the number of NYC taxi rides per hour decreased by approximately 8% in 2014 and 2015 [5]. Besides, there have been protests against TNCs (*e.g.*, Uber) from taxi drivers all over the world since 2014 [6].

In China's market, Didi Chuxing (formerly Didi Kuaidi) has become the largest transportation network company after its merging with other TNCs, including Kuaidi Dache and Uber China in 2015 and 2016, respectively. Nowadays, Didi offers a diverse range of transportation services, including private car hailing, taxi hailing, Hitch (a social commuting program), and Chauffeur (a designated driver service), for about 300 million users across over 400 cities in China [7].

Didi started with its taxi hailing service in 2012. Unlike Uber's automatic dispatch system [8], Didi's taxi hailing service gives taxi drivers the privilege of choosing passengers and competing with each other. In early 2014, Didi and Kuaidi

Dache fought with each other for taxi hailing market share in China. Backed by China's Internet giants Tencent and Alipay, both companies gave promotion fees to taxi drivers for each deal made and greatly changed the taxi service pattern [9]. Since then, Didi launched its Chauffeur service, private car hailing service, and Hitch service in August 2014, May 2015, and June 2015, respectively. In November 2015, Didi added ridesharing service to its private car hailing service, bringing more competition to taxi drivers.

From the perspective of taxi drivers, although Didi's taxi hailing service improves their income by increasing the number of rides and reducing idle time, the competition from Didi's Chauffeur, private car hailing, and Hitch services cannot be ignored. In this paper, we conduct a comparative analysis and answer the following questions: (1) how does the taxi industry in Beijing change before and after the emergence of TNCs and whether they have caused the taxi industry to fall into a recession; (2) how do taxi drivers' behaviors change, in terms of the passenger-searching strategy, the passenger-delivery strategy, and the service-area preference.

To answer these questions, we use GPS trajectories collected from thousands of Beijing taxis in November for three different years:

- November 2012: when DiDi was in the early stage and competed with other small startups, and Uber had not entered into China's market;
- November 2014: when the two largest taxi-hailing companies at that time, Didi and Kuaidi Dache, were fighting with each other for taxi-hailing market share, and Didi just launched its Chauffeur service;
- November 2015: when Didi and Uber China were competing with each other for private car-hailing market share, using subsidies and lower prices.

During these time periods, the number of resident population in Beijing is 20.693, 21.516, and 21.705 million in years 2012, 2014, and 2015, respectively [10]. The population growth rate is less than 4.9% from 2012 to 2015. The number of private cars in Beijing is 4.075, 4.372, and 4.403 million in years 2012, 2014, and 2015, respectively, while the number of licensed taxis in Beijing has been controlled around 67,000 in the past few years. In the meantime, the public transit passenger volume (including bus and subway) is 7,615.78, 8,158.49, and 7,383.84 million in years 2012, 2014, and 2015 [11]. The decrease of public transit volume in 2015 might be another result of the competition between TNCs, since a lower price would attract passengers from both taxis and public transit.

Based on GPS trajectories of taxi drivers, we extract trip records and estimate taxi fares based on Beijing's taxi pricing rule. We quantify the taxi drivers' performance using the average efficiency, and take both the revenue from delivering passengers and the gasoline cost of delivering and searching passengers into consideration.

Our contributions in this study are threefold:

1. We conduct a comparative analysis of taxi trips and reveal the change of the taxi industry in three time periods, which makes us one of the first papers to present a comprehensive analysis on transitions of taxi drivers' behaviors during different phases of TNCs' development.

2. Our quantitative research approach based on GPS trajectory analysis reveals that the taxi industry is adversely influenced by transportation network companies, in terms of the number of daily orders and profits.

3. We compare taxi service strategies between taxi drivers with top and bottom efficiency and observe some significant differences. Our findings give an insight for taxi drivers on how to make improvements.

Compared to our preliminary work [12], we add new results of spatial analysis, densification and poolability properties, new definitions of service area preferences, and many other minor revisions in this paper.

The rest of this paper is organized as follows. Section II gives a literature review of related work on analysis of GPS trajectory and taxi drivers' performance. Section III presents the data description, trip extraction process and fare estimation method used in this paper. Section IV delivers our analyses and results. We conclude this paper in Section V.

II. RELATED WORK

With the development of mobile sensing devices such as smart phones and GPS navigators and location-based services, we now have the ability to collect and analyze digital footprints of human mobility, such as GPS trajectories [13], mobile phone records [14], and check-ins [15]. With more and more datasets becoming available, previous studies have addressed a variety of research issues, such as activity recognition [16], road map making [17], [18], urban human mobility understanding [19]–[22], anomalous trajectory detection [23], city region function identification [24]–[26], and location-based social networks [27]. Some concerns about privacy and location cheating are also discussed in previous studies [14], [28].

In this study, we would focus on the analysis of GPS trajectories collected from taxi drivers. These trajectories could help us gain a better understanding of taxi drivers' behaviors, compare taxi drivers' performance, and monitor the change of the taxi industry. In this section, we would give a literature review from these three aspects.

A. STUDY OF TAXI DRIVERS' BEHAVIORS

In this part, we classify the previous studies of taxi drivers' behaviors into four aspects.

The first aspect is about taxi drivers' driving style. Previous studies have shown that by analyzing GPS trajectories, driving style could be assessed and even recognized [29], [30]. In [31], drivers are divided into 4 types as aggressive, relatively aggressive, relatively cautious and cautious, and the driving style could be further used in a lane change warning system. Abnormal or dangerous driving behaviors could also be identified using machine learning techniques [32].

The second aspect is about taxi drivers' passenger-delivery behaviors. By utilizing GPS trajectories collected from taxis, previous studies have used the extracted knowledge for trip time estimation [33] and route planning [34].

The third aspect is about taxi drivers' passenger-searching behaviors. Previous studies have discussed whether taxi drivers should actively hunting or passively waiting for passengers [35]. To improve the passenger-searching process, previous studies have proposed recommendation systems that are designed for both taxi drivers and passengers [36], taxi dispatching services that send a nearby taxi to a passenger [37], and taxi ridesharing services that pair more than one passenger to a taxi [38].

The last aspect is about taxi drivers' service-area preference. Clustering algorithms have been used to identify pickup/dropoff hotspots [39] and discover the spatial distribution of taxis [40].

In this study, we would leave the driving style for our future research and focus on the latter three aspects.

B. COMPARISON AMONG TAXI DRIVERS

Instead of taking taxi drivers as a whole, some studies split taxi drivers into different groups, and compare the behaviors among taxi drivers from different groups. To evaluate and split taxi drivers, different metrics of performances have been used already, such as daily income [41], occupied distance [35], proportion of occupied time [42], and average passenger-finding time [43]. Based on the performances, there are two common practices of grouping: top and ordinary drivers [44], [45], or top and bottom drivers [46]. For example, in [45], the authors find that compared with ordinary drivers, top drivers seek passengers more actively and usually navigate through the whole city.

In this study, we propose to use the average efficiency as the performance metrics, which is calculated with not only the revenue from delivering passengers, but also the gasoline cost from delivering and searching passengers. We would compare top and bottom drivers, not only because it is less studied before, but also because we want to know why the top drivers are successful, as well as why the bottom drivers are less successful.

C. CHANGE OF TAXI INDUSTRY

With the development of transportation network companies, there have been studies about their impacts on the taxi industry, as we discussed in Section I. In [9], the authors use a comparative approach and analyze the taxi drivers' behaviors within the battle between Didi and Kuaidi Dache. Their results show that the money promotion used by these taxi-hailing companies increases the number of taxi trips per day per taxi and shortens the total idle time, while it also brings inconvenience to the passengers who travel for longer distance or go to unpopular locations. In [47], the author uses a large taxi GPS trajectory dataset to examine the impact of ridesourcing on the taxi industry, and finds out that taxis compete effectively in peak/off-peak periods and in areas with high density.

III. DATA

A. DATA DESCRIPTION

We use the taxi GPS trajectory data collected in Beijing in three time periods: November 2012, November 2014, and November 2015. The taxi GPS trajectory is sampled at a rate of one minute. Each sample contains the anonymous taxi identity, timestamp, latitude, longitude, azimuth, spot speed, and operation status (occupied, vacant or stopped). Since human traveling patterns are notably different in weekdays and weekends/holidays, we only consider weekdays in this study and leave the situations on weekends for future research. The data is summarized in Table 1 (some weekdays are omitted in this study due to lack of data) and more details are discussed in Section IV.

As we could see from Table 1, the average passenger-delivery trip number per day per taxi dropped by **18.08%** and the average profit per day per taxi dropped by **19.29%** in year 2015, respectively, compared with that of year 2012. These results reveal that the taxi industry was adversely influenced by the competition between Uber and Didi in 2015.

TABLE 1. Data summary.

Year	2012	2014	2015
Number of weekdays used in this study	22	18	18
Number of taxis	8,879	17,749	20,067
Average passenger-delivery trip number per day per taxi	17.64	17.31	14.45
Average working hour per day per taxi	12.07	11.93	11.39
Average profit per day per taxi (RMB)	515.89	519.53	416.38
Average efficiency per day per taxi (RMB/hour)	42.25	42.96	36.25
Average efficiency per day per taxi of top drivers (RMB/hour)	56.55	59.11	51.53
Average efficiency per day per taxi of bottom drivers (RMB/hour)	26.04	25.96	20.19

B. TRIP EXTRACTION

Based on the raw GPS data, our first objective is to extract the trip information from the data. We define a trip record as the tuple (origin longitude&latitude, destination longitude&latitude, begin time, end time, trip distance, anonymous taxi identity). For a *passenger-delivery trip*, the begin time is when a taxi becomes occupied from a previous vacant status, *i.e.*, pick-up, and the end time is when a taxi becomes vacant from a previous occupied status, *i.e.*, drop-off. The origin/destination longitude and latitude correspond to the location of the taxi at start/end time. A *passenger-searching trip* is the opposite process. We estimate the trip distance by the length of trace, which is represented as a polyline that connects a set of GPS points. As the location errors caused by inaccurate GPS signals (less than 10 meters) are much smaller than the traveling distances of trips, these errors are omitted in this paper.

A trip's *duration* is defined as the interval between begin time and end time. However, the dataset is imperfect, as the information about whether the taxi is occupied or vacant may be inaccurate. To alleviate this problem, we define that a trip is *valid* only if the trip lasts for at least 1 minute.

This specification filters out the cases when a vacant taxi is labeled as occupied mistakenly, and we only consider valid trips in this study. A taxi driver's *working hour* is defined as the summation of duration of both passenger-delivery and passenger-searching trips.

C. FARE ESTIMATION

Even though the pricing rule may be slightly different in different time periods, we use the latest pricing rule (Table 2) for trips in this study to make a fair comparison.

TABLE 2. Beijing taxi pricing rule since 2016.

Basic fee	13 RMB for the first 3km. 2.3 RMB/km (mileage fare) after the first 3km.
Congestion fee	4.6 RMB for each 5 minutes in peak hours or 2.3 RMB for each 5 minutes in regular hours after the first 3km at speed lower than 12km/h. Peak hours are defined as time periods from 7:00 to 9:00 and from 17:00 to 19:00.
Extra fee for long-distance trips	Extra 50% mileage fare for the part of trip over 15km. This extra fare is waived if the distance between the pick-up and drop-off locations is less than 2km, *e.g.*, a round trip.
Extra fee for late-night trips	Extra 20% mileage fare for the part of trip at night from 23:00 to 5:00.

We consider the gasoline consumption as the main cost of taxi drivers. The gasoline consumption rate is estimated as 6.9 liters per 100 kilometers, based on China's Phase IV fuel consumption standard for passenger vehicles, and we use the gasoline price in December 2016, which is 5.92 RMB per liter. The overall gasoline consumption takes both passenger-delivery and passenger-searching trips into consideration. A taxi driver's *daily profit* is then calculated as his/her daily income minus the gasoline cost.

To quantify taxi drivers' performance, we introduce the definition of *efficiency* as the profit a taxi driver earns per working hour. Note that in some cities of China, a taxi may be shared by two drivers, and these drivers take shift handover twice every day (i.e., once in the morning and once in the afternoon) [44]. We would not consider this situation for two reasons. Firstly, this practice is not common in Beijing, as compared with the total number of taxis. Secondly and more importantly, we are focusing on the efficiency of taxi drivers, rather than the working hour or the daily profit. While a taxi shared by two drivers may get longer working hour and higher daily profit, it may not be necessary to be more efficient. Since the daily profit is divided by the working hour, we are actually evaluating an averaged performance of the two drivers.

IV. ANALYSIS

In this section, we present our analyses and results about both the taxi industry and the taxi drivers' behaviors.

A. PASSENGER-DELIVERY TRIP INFORMATION

We start with passenger-delivery trips in this part, as passenger-delivery trips are the most important indicators of both human mobility as well as the taxi drivers' income. We would firstly give some statistics, and then study the spatial and temporal properties. We would also use a model named as space-time graph and validate the densification property. We further introduce a concept of trajectory-based poolability and study the potential of sharing these trips.

1) TRIP STATISTICS

In Figure 1, we show the histogram that plots the daily passenger-delivery trip numbers served by each taxi. In 2014, there is a larger proportion of drivers with bigger number of trips, compared with 2012 and 2015. This phenomenon may be explained by the money promotion policy in the battle of Didi and Kuaidi Dache. During the promotion period, more trips served would bring more rewards, regardless of the trip distances. Interested readers may refer to [9] for a similar result.

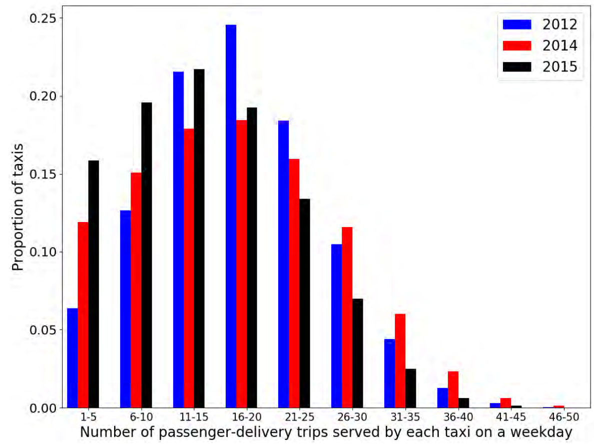


FIGURE 1. The histogram plot of the daily passenger-delivery trip numbers served by each taxi on a weekday.

In Figure 2, we show the empirical distribution of traveling distances. The distribution exhibits a shape of the long-tailed distribution, as the trips are more concentrated on the short-distance part. Besides, the proportion of shorter trips (*e.g.*, 0-3 km) increases noticeably in 2015.

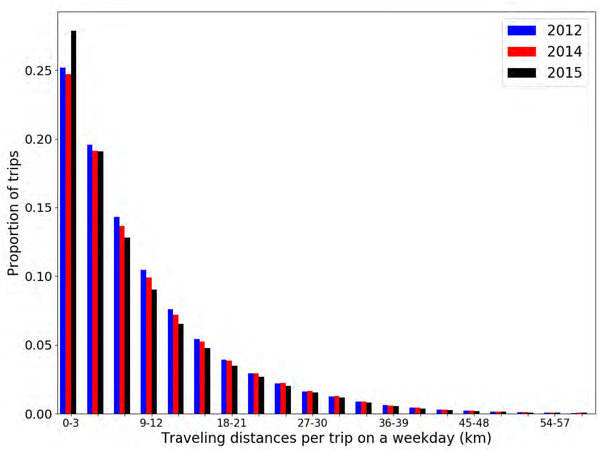


FIGURE 2. The empirical distribution of traveling distances on a weekday.

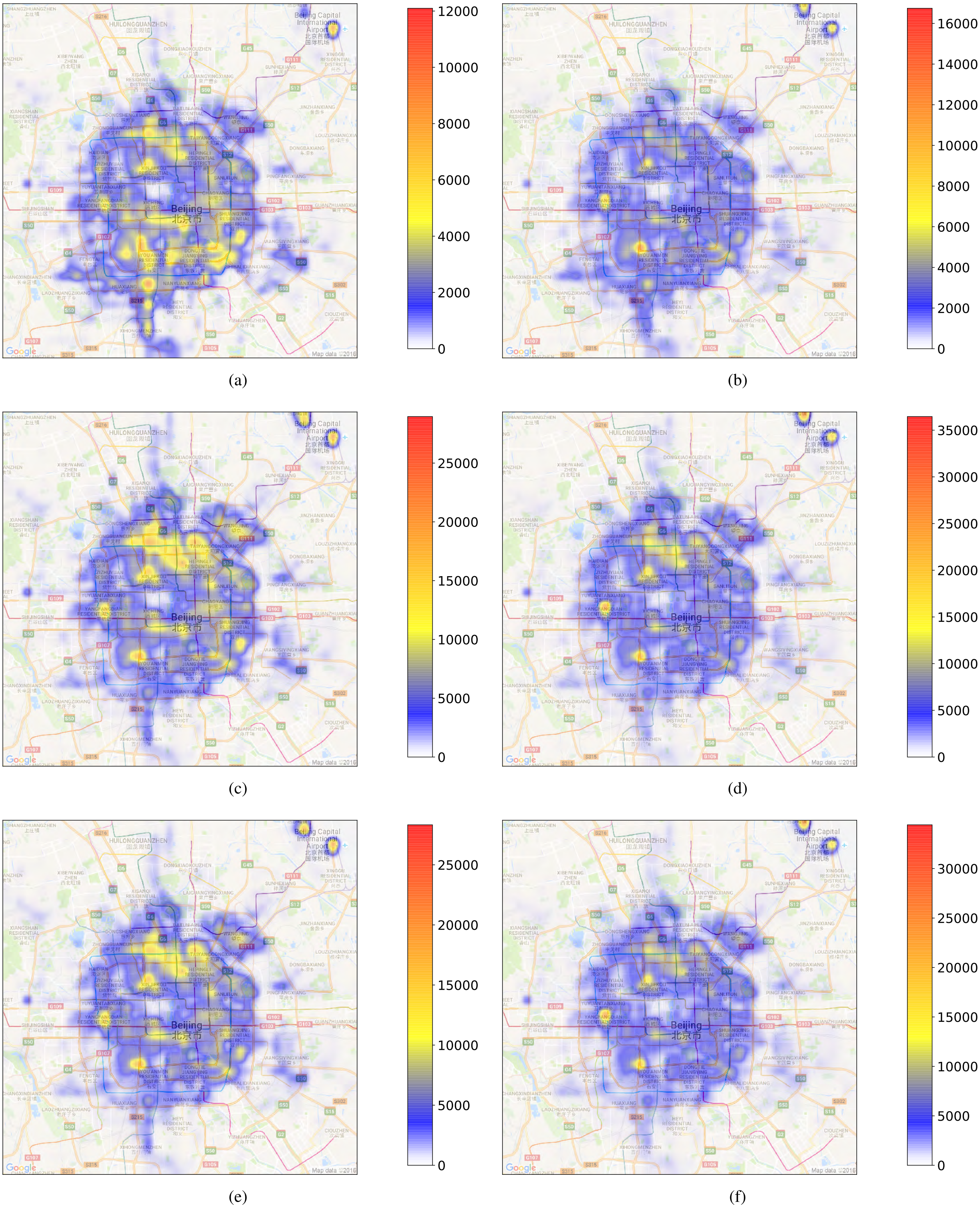


FIGURE 3. **(a) Heatmap of pick-up points in 2012; (b) Heatmap of drop-off points in 2012; (c) Heatmap of pick-up points in 2014; (d) Heatmap of drop-off points in 2014; (e) Heatmap of pick-up points in 2015; (f) Heatmap of drop-off points in 2015.**

2) SPATIAL PROPERTY

To show the spatial properties of passenger-delivery trips in different time periods, we plot the pick-up and drop-off heatmaps of the whole month in three years from Figure 3(a) to Figure 3(f). The heatmaps are noticeably different in 2012, and higher proportions of both pick-up and drop-off points are located in the northern part of Beijing (the upper side on the map) in 2014 and 2015.

3) TEMPORAL PROPERTY

To recall, the begin time of a passenger-delivery trip is defined as the pick-up time of the trip. In Figure 4, we show the average numbers of passenger-delivery trips per day per taxi served in each hour during the three time periods. In 2012 and 2014, the numbers are basically on a similar level, while there is a significant drop in 2015 during daytime. This drop of average trip numbers in 2015 reflects the lower demand for taxis, when the competition from transportation network companies is an important factor.

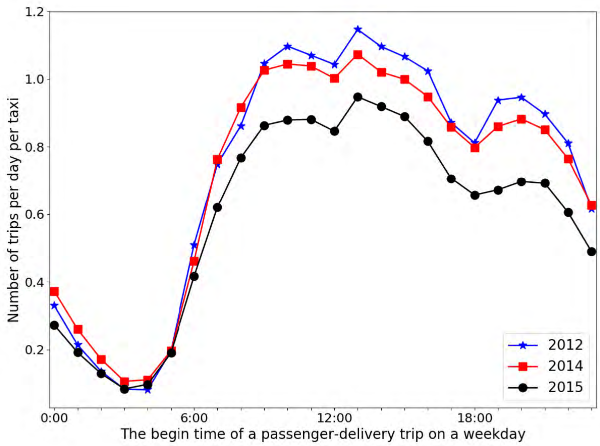


FIGURE 4. The average numbers of passenger-delivery trips per day per taxi served in each hour.

To better understand the supply-demand relationship in the temporal domain, we define the occupancy rate as the ratio of occupied taxis over the total number of taxis which are occupied or vacant. Notice that the definition excludes the taxis which are stopped. We show the occupancy rate sampled every half hour in Figure 5. During daytime, the occupancy rate is the lowest in 2015, while during nighttime, the occupancy rate is the lowest in 2012. From this result, we can find that although it becomes harder for the taxi drivers to become occupied with a passenger during daytime, less drivers would stay vacant during nighttime in 2015. This could be explained from the phenomenon that taxi drivers prefer to work late into the night or early in the morning, when there are prearranged trips, *e.g.*, a trip to the airport for catching up an early morning flight. With smartphone-based taxi-hailing services, it would be much easier and more convenient for the passengers to book a taxi trip.

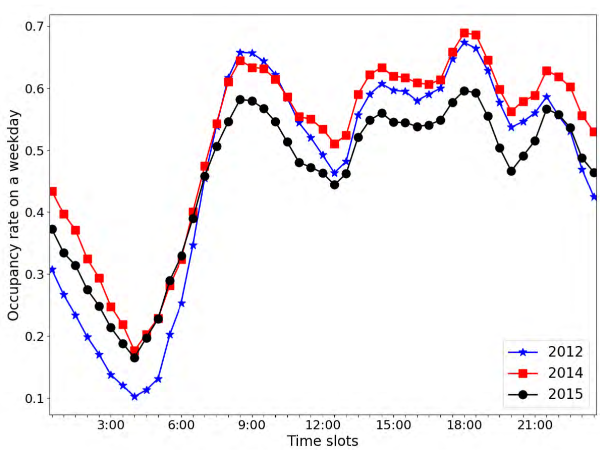


FIGURE 5. The occupancy rate every half hour on a weekday.

4) SPACE-TIME GRAPH AND DENSIFICATION PROPERTY

To capture the spatial and temporal variability of passenger-delivery trips, we use a graph model introduced in [48] and discover that the taxi trips exhibit the well known property called "densification power law", which are found in graphs modelling human behaviors, such as social network graphs and publication citation graphs [49].

To build a space-time graph, we firstly split a day into time intervals of 5 minutes and the geographical map into grid cells of 100 meters $\times$100 meters. For each trip, we would map the trip's origin and destination into the corresponding grid cell ids. For example, in Figure 6, the trip with origin O_1 and destination D_1 is mapped to N_1 and N_2, and the trip with origin O_2 and destination D_2 is mapped to N_1 and N_3. In each time interval t, we would build a space-time graph, and a cell is considered as a node in the graph only if the origin or destination of a passenger-delivery trip with a begin time in this specific time interval falls within that cell. As in the example in Figure 6, two trips are mapped to a space-time graph consisting of three nodes, *i.e.*, N_1, N_2, and N_3, and two edges, *i.e.*, $N_1 \rightarrow N_2$ and $N_1 \rightarrow N_3$.

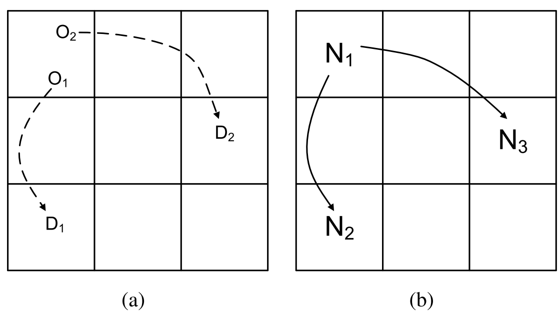
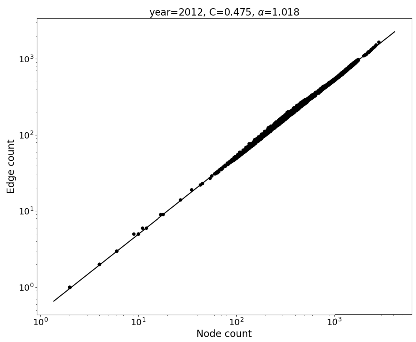


FIGURE 6. (a) Two trips, (*O*1; *D*1) and (*O*2; *D*2), within the same time interval; (b) The corresponding space-time graph, which consists three nodes and two edges.

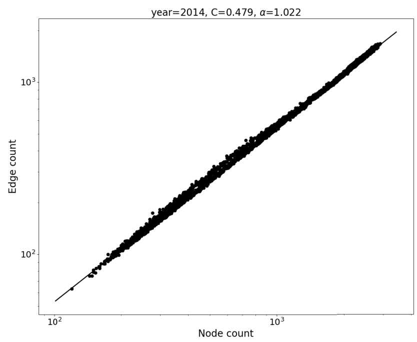
We discover that the passenger-delivery trips exhibit the densification power law (DPL), which means the number of edges grow as a power of the number of nodes. Specifically, for each time interval t, we have:

$$e(t) \propto n(t)^{\alpha} \text{ or } e(t) = Cn(t)^{\alpha}$$

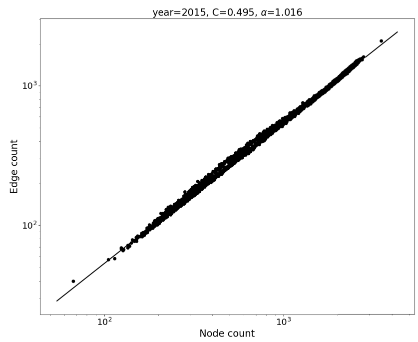
where $e(t)$ and $n(t)$ represent the number of edges and nodes in the space-time graph, respectively. C and α are constants. We show the specific values of C and α learned from data in Figure 7(a) to Figure 7(c). A smaller C or a larger α indicates that the trips are "denser" in the spatial space,



(a)



(b)



(c)

FIGURE 7. Densification power law (DPL) plots of the passenger-delivery trips. (a) DPL plot of the space-time graph in 2012; (b) DPL plot of the space-time graph in 2014; (c) DPL plot of the space-time graph in 2015.

which gives a higher probability of carpooling, as we would evaluate next.

5) POOLABILITY PROPERTY

While the densification power law reveals how the space-time graphs evolve, a more practical descriptor would be how likely these taxi trips could be pooled, as previous studies have shown that carpooling, or ride-sharing, could generate both significant environmental and economic benefits [50]. This is particularly meaningful for a city like Beijing, which is bothered by severe air pollution in the past few years and the emissions of automotive exhaust have been proved to be a non-negligible factor that produces air pollution.

In the previous study, the concept of poolability is built on road network [51] or origin-destination pair [48]. In this study, we propose a trajectory-based definition of poolability. Considering Trip 1 with origin O_1, destination D_1, start time s_1, and end time e_1, we denote its trajectory as $\{(x_1, y_1, t_1), (x_2, y_2, t_2), \ldots, (x_N, y_N, t_N)\}$, where x_i and y_i represents longitude and latitude, and t_i represents the timestamp of GPS sample i, N is the total number of GPS samples for this trajectory, $i = 1, 2, \ldots, N$. Equivalently, we have $O_1 = (x_1, y_1)$, $D_1 = (x_N, y_N)$, $s_1 = t_1$, and $e_1 = t_N$. For Trip 2 with origin O_2, destination D_2, start time s_2, and end time e_2, we say that Trip 1 is poolable with Trip 2 as long as the following constraints are satisfied:

(1) $s_1 \leq s_2$

(2) $e_1 \geq e_2$

(3) $\text{distance}((x_j, y_j), O_2) < \Delta O$, where $j = \arg\min_{i\in\{1,2,\ldots,N\}} |t_i - s_2|$

(4) $\text{distance}((x_k, y_k), D_2) < \Delta D$, where $k = \arg\min_{i\in\{1,2,\ldots,N\}} |t_i - e_2|$

where the first constraint requires that Trip 1 starts no later than Trip 2, the second constraint requires that Trip 1 ends no later than Trip 2, the third and fourth constraints require that the origin and destination of Trip 2 are close to the trajectory of Trip 1, as shown in Figure 8. We use Euclidean distance as the distance function shown above, and ΔO and ΔD are predefined parameters. In this study, we choose $\Delta O =$ 200 meters and $\Delta D = 200$ meters.

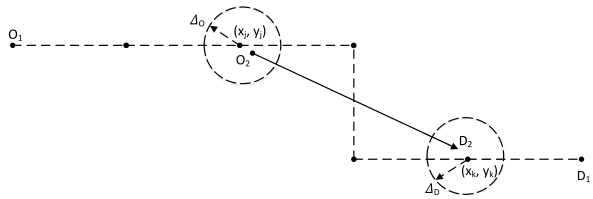


FIGURE 8. An example of a pair of poolable trips.

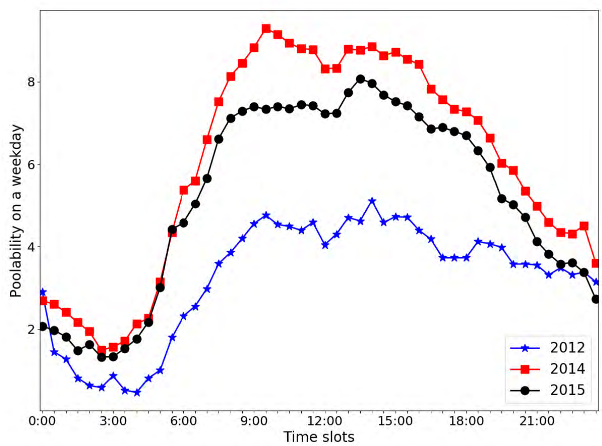


FIGURE 9. The average poolability every half hour.

For a trip, we say it can be pooled, as long as there is another trip that is poolable with it. We define *poolability* as the percentage of trips that can be pooled and evaluate poolability every half hour. The result is shown in Figure 9. As we can tell from the result, poolability is higher during daytime, when there are more trips and thus higher opportunity of being pooled. However, since we have different subsets of taxis used in this study for different time periods, the poolability in 2012 may not be necessarily lower than 2014 and 2015, if we could evaluate the trips from all taxi drivers in Beijing.

B. PASSENGER-SEARCHING TRIP INFORMATION

Now we discuss the passenger-searching trips. For simplicity, we refer the duration of a passenger-searching trip as its *idle time* and the travelling distance as its *idle distance*. The empirical distributions of the idle time and distance are shown in Figure 10 and Figure 11, respectively. Even though the idle time lengths become larger in 2014 and 2015, the idle distances have only a small change. It is possible that the drivers do not want to go far away to search for a passenger, or they just want to save the gasoline cost.

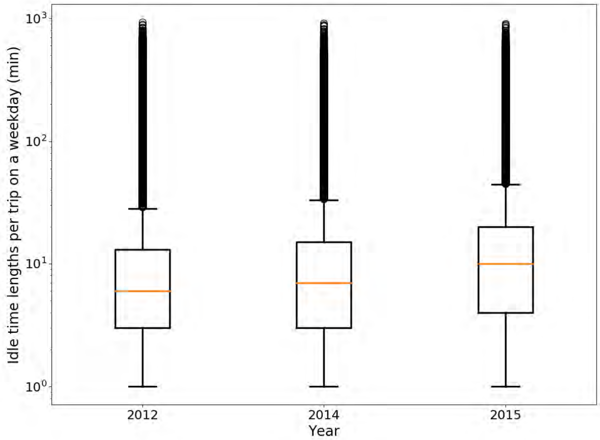


FIGURE 10. The empirical distribution of the idle time lengths on a weekday.

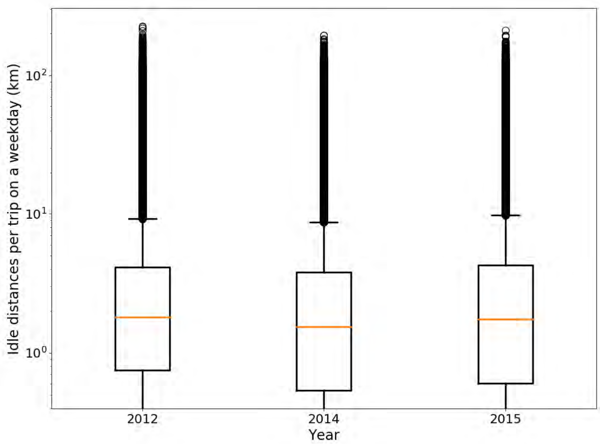


FIGURE 11. The empirical distribution of the idle distances on a weekday.

C. WORKING HOUR, PROFIT & EFFICIENCY

Now we pay our attention to the working hour, profit and efficiency of taxi drivers.

1) WORKING HOUR

As we defined in Section III-B, the working hour is defined as the accumulated duration when the taxi driver is occupied or vacant. As shown in Table 1, the average working hours per day per taxi are 12.07, 11.93, and 11.39, in 2012, 2014, and 2015, respectively. The distribution of the working hour exhibits a bell shape of the normal distribution, as shown in Figure 12. The distribution in 2012 has a higher mean and a lower variance, indicating that the working hour difference among taxi drivers becomes larger in 2014 and 2015.

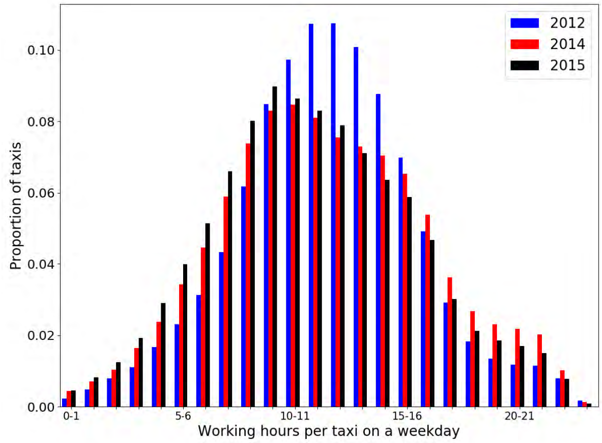


FIGURE 12. The empirical distribution of the working hour on a weekday.

2) PROFIT

As discussed in Section III-C, a taxi driver's daily profit is calculated as the income minus the gasoline cost. The average profits per day per taxi (in RMB) are 515.89, 519.53, and 416.38, in 2012, 2014, and 2015, respectively. The empirical distribution of the daily profit is shown in Figure 13. The decrease in 2015 can be interpreted as a result of the competition between Didi and Uber China. Lower prices and different choices of service provided by these TNCs attract passengers from traditional taxi services.

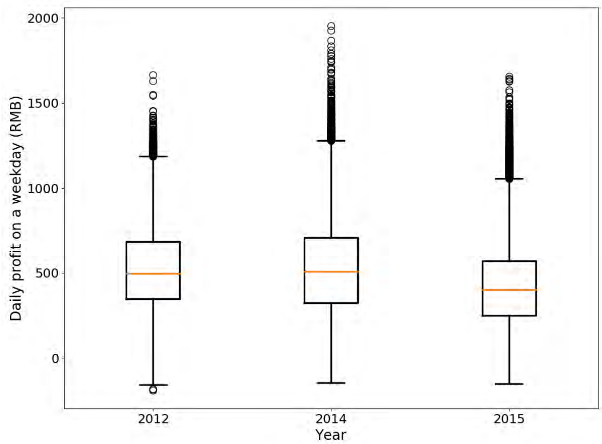


FIGURE 13. The empirical distribution of the daily profit on a weekday.

3) EFFICIENCY

Recall that the efficiency is defined as the profit a taxi driver earns per working hour. The empirical distribution of the efficiency is shown in Figure 14. The average efficiency per day per taxi (in RMB/hour) is 42.25, 42.96, and 36.25, in 2012, 2014, and 2015, respectively. In 2015, fewer

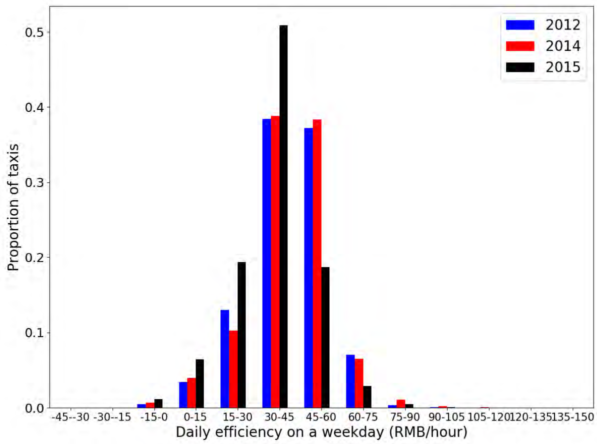


FIGURE 14. The empirical distribution of the efficiency on a weekday.

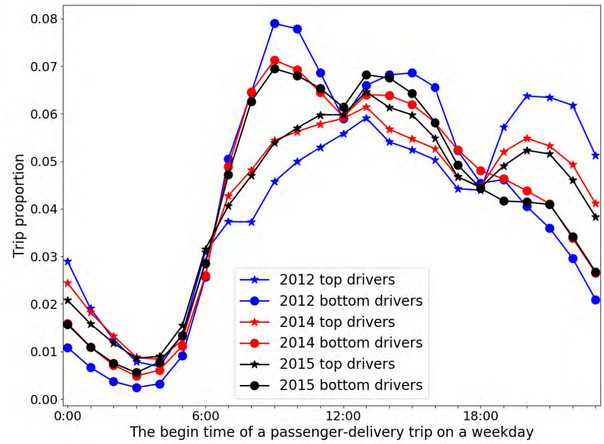


FIGURE 15. The temporal distribution of both top and bottom drivers' trips in three years.

taxi drivers can achieve a high efficiency (*e.g.*, greater than 60 RMB/hour). This result again supports the observation that taxi drivers experienced a hard time in 2015, compared with 2012 or 2014.

D. COMPARISON BETWEEN TOP AND BOTTOM DRIVERS

To compare the taxi drivers' performance, we choose the taxi drivers with top 20% efficiency as our *top drivers* and those with bottom 20% efficiency as our *bottom drivers*. The average efficiency of top and bottom drivers in different years is listed in Table 1. We regard efficiency as a better measurement of a taxi driver's intelligence rather than the daily profit, as each taxi driver's working hour may be different. In this part, we would firstly show some spatial and temporal properties of the trips from both top and bottom drivers, and then compare their passenger-searching strategies, passenger-delivery strategies, and service area preferences.

1) SPATIAL AND TEMPORAL PROPERTIES

The temporal distribution of both top and bottom drivers' passenger-delivery trips in three years is shown in Figure 15. There is a remarkable pattern that shows the top drivers have a higher proportion of trips during nighttime, while the bottom drivers have more trips during daytime. This phenomenon may have different interceptions, *e.g.*, the top drivers work harder during nighttime, or the competition among taxi drivers during daytime is more intense. Further research is needed to explain this pattern.

To describe the spatial property, we show the pick-up heatmaps of both top and bottom drivers' trips in the whole month in Figure 16(a) to Figure 16(f). In 2012, the bottom drivers pick up more passengers in the southern part of Beijing (lower side on the map), and in 2014 and 2015, the bottom drivers pick up more passengers in the southeastern part of Beijing (lower right side on the map). Compared with the result from Section IV-A.2, there is a higher demand for taxis in the northern part of Beijing and top drivers have a closer spatial connection with this result.

2) PASSENGER-SEARCHING STRATEGY

Passenger-searching strategies refer to the cruising behaviors of taxi drivers during vacant time. We classify the passenger-searching strategies into three types, *i.e.*, *hunting locally*, *waiting locally*, and *going distant*. The definitions are stated as follows:

$$d_{\text{idle}} \begin{cases} > \tau_d, & \text{going distant} \\ \leq \tau_d & \begin{cases} t_{\text{idle}} > \omega_d, & \text{waiting locally} \\ t_{\text{idle}} \leq \omega_d, & \text{hunting locally} \end{cases} \end{cases} \tag{1}$$

where d_{idle} and t_{idle} denote the idle distance and time, respectively; and τ_d and ω_d are threshold parameters. In this study, we empirically set $t_d = 1.5$ kilometers and $\omega_d = 5$ minutes.

We count the number of going-distant trips s^y_{gd}, hunting-locally trips s^y_{hl}, and waiting-locally trips s^y_{wl} in different years, where $y \in \{2012, 2014, 2015\}$. Then we could construct the passenger-searching strategy vector for a particular taxi driver (with normalization) as $[s^y_{\text{gd}}, s^y_{\text{hl}}, s^y_{\text{wl}}]/(s^y_{\text{gd}} + s^y_{\text{hl}} + s^y_{\text{wl}})$. The pure strategy can be represented as $[1, 0, 0]$ for going distant, $[0, 1, 0]$ for hunting locally, and $[0, 0, 1]$ for waiting locally.

We show the average result of normalized searching strategies for both top and bottom drivers in Figure 17. As we can tell from Figure 17, top drivers usually have a lower probability of going distant, because this searching strategy might waste more gasoline.

3) PASSENGER-DELIVERY STRATEGY

Passenger-delivery strategies refer to the behaviors of taxi drivers when delivering passengers. It is a complex process which involves factors such as route planning and driving style. For simplicity, we use the average passenger-delivery speed as an indicator of the passenger-delivery strategy, as shown in Figure 18. We can find that top drivers tend to have a higher passenger-delivery speed, which may indicate better route-planning abilities and driving skills.

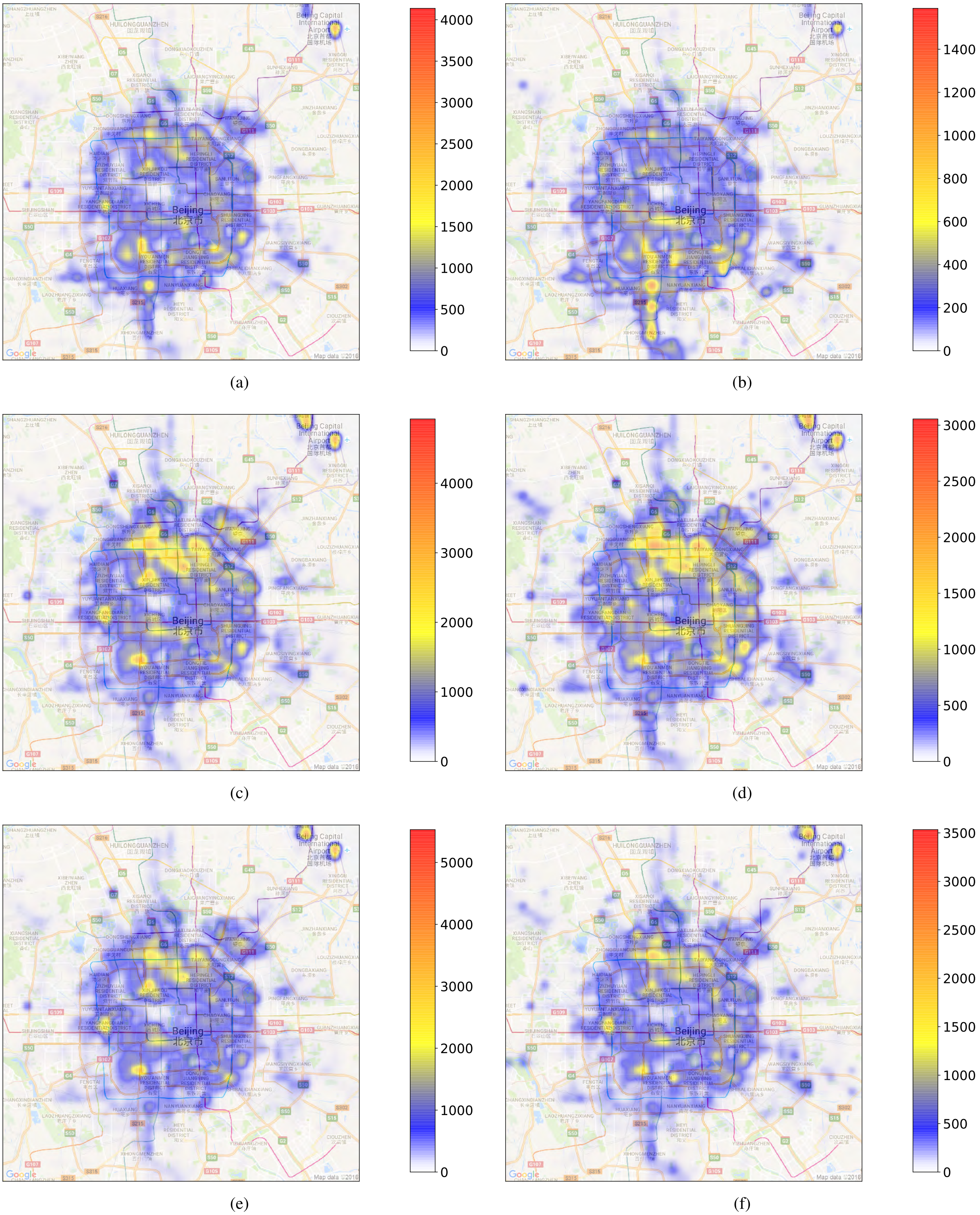


FIGURE 16. (a) Heatmap of top drivers' pick-up points in 2012; (b) Heatmap of bottom drivers' pick-up points in 2012; (c) Heatmap of top drivers' pick-up points in 2014; (d) Heatmap of bottom drivers' pick-up points in 2014; (e) Heatmap of top drivers' pick-up points in 2015; (f) Heatmap of bottom drivers' pick-up points in 2015.

4) SERVICE-AREA PREFERENCE

Beijing has a regular chessboard pattern and possesses multiple ring roads, as shown in Figure 19. We would consider the four major ring roads (from 2nd to 5th Ring Road) and divide the whole city into five areas:

- Area 1: within the 2nd Ring Road;
- Area 2: within the 3rd Ring Road and outside the 2nd Ring Road;
- Area 3: within the 4th Ring Road and outside the 3rd Ring Road;

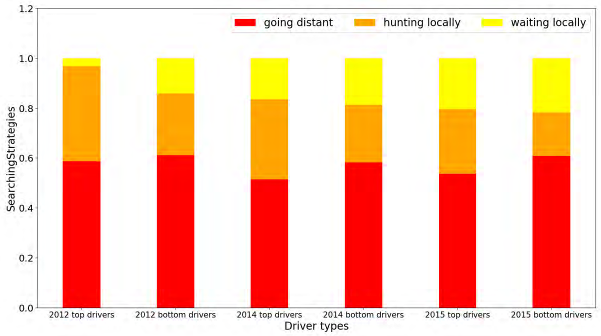


FIGURE 17. **The passenger searching strategies.**

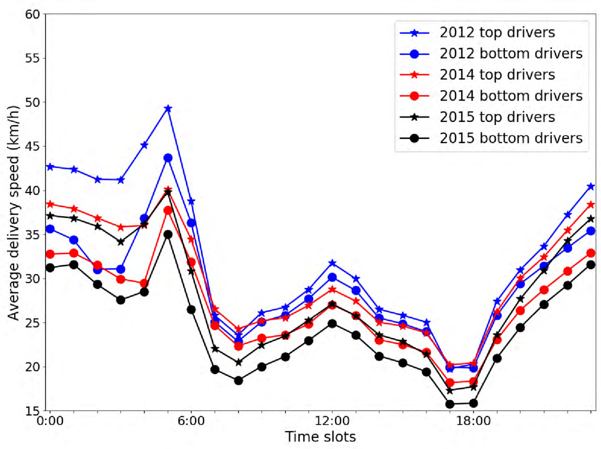


FIGURE 18. **The average passenger-delivery speed.**

FIGURE 19. **Map of four ring roads in Beijing.**

- Area 4: within the 5th Ring Road and outside the 4th Ring Road;
- Area 5: outside the 5th Ring Road.

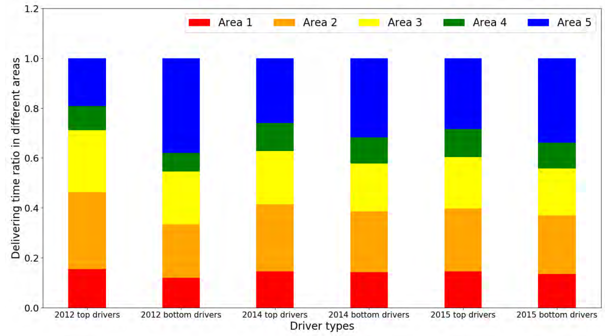


FIGURE 20. **Delivering time ratio in different areas.**

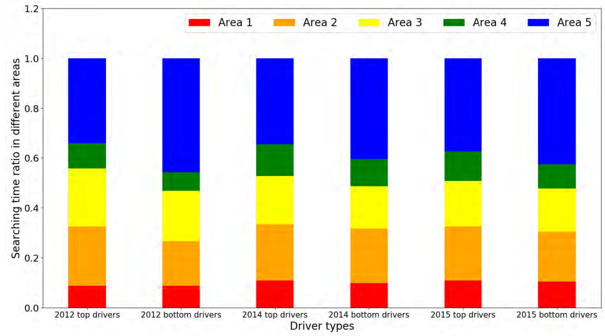


FIGURE 21. **Searching time ratio in different areas.**

In Figure 20 and Figure 21, we show the time ratios of both top and bottom drivers spending in these areas, when they are delivering passengers or searching passengers, respectively. As we can tell from Figure 20 and Figure 21, top drivers usually spend more time delivering and searching passengers in the inner part of Beijing.

V. CONCLUSION

Taxis play an irreplaceable role in city road transportation. Recent transportation network companies such as Uber and Didi have brought a remarkable impact on the taxi industry. To better understand the influences, we conduct a multi-period analysis of the taxi drivers' behaviors, based on GPS trajectories collected from three time periods, November 2012, November 2014, and November 2015 in Beijing.

Our results support the opinion that the growth of TNCs is accompanied with the depression of taxi demand. Especially, the lower price caused by the competition between these TNCs in November 2015 results in a significant reduction of taxi trips. A lower usage of taxis also decreases the poolability of taxi trips and damages the taxi drivers' enthusiasm, as indicated by the shorter working hours per day.

We also compare the strategies between taxi drivers with top and bottom efficiency in different time periods. We find that drivers with top efficiency tend to search locally, have a higher delivery speed, and serve more often within the inner part of Beijing, compared to drivers with lower efficiency. These findings may be used to explain the efficiency difference and help the taxi drivers to make improvements.

REFERENCES

[1] California Public Utilities Commission. (2013). *Decision Adopting Rules and Regulations to Protect Public Safety While Allowing New Entrants to the Transportation Industry*. Accessed: Jan. 22, 2018. [Online]. Available: http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M077/K112/77112285.PDF

[2] M. K. Chen and M. Sheldon, "Dynamic pricing in a labor market: Surge pricing and flexible work on the Uber platform," in *Proc. ACM Conf. Econ. Comput. (EC)*, New York, NY, USA, 2016, p. 455. [Online]. Available: http://doi.acm.org/10.1145/2940716.2940798

[3] J. Cramer and A. B. Krueger, "Disruptive change in the taxi business: The case of Uber," *Amer. Econ. Rev.*, vol. 106, no. 5, pp. 177–182, 2016.

[4] Z. Li, Y. Hong, and Z. Zhang, "An empirical analysis of on-demand ride sharing and traffic congestion," in *Proc. Int. Conf. Inf. Syst.*, 2016, pp. 4–13.

[5] A. Brodeur and K. Nield, "Has Uber made it easier to get a ride in the rain?" Dept. Econ., IZA Inst. Lab. Econ., Bonn, Germany, IZA Discussion Paper 9986, 2016. [Online]. Available: https://ssrn.com/abstract=2797557

[6] The Washington Post. (2014). *How the Anti-Uber Backlash is Spreading Around the World*. Accessed: Jan. 22, 2018. [Online]. Available: https://www.washingtonpost.com/news/worldviews/wp/2014/12/09/how-the-anti-uber-backlash-is-spreading-around-the-world

[7] Didi Chuxing. *About Didi*. Accessed: Jan. 22, 2018. [Online]. Available: http://www.xiaojukeji.com/website/about.html

[8] B. Voytek. (2016). *Optimizing a Dispatch System Using an AI Simulation Framework*. Accessed: Jan. 22, 2018. [Online]. Available: https://newsroom.uber.com/semi-automated-science-using-an-ai-simulation-framework/

[9] B. Leng, H. Du, J. Wang, Z. Xiong, and L. Li, "Analysis of taxi drivers' behaviors within a battle between two taxi apps," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 296–300, Jan. 2015.

[10] Beijing Statistical Information Net. (2016). *Beijing Resident Population*. Accessed: Jan. 22, 2018. [Online]. Available: http://www.bjstats.gov.cn/tjsj/

[11] Beijing Municipal Bureau of Statistics. (2016) *Usage Statistics of Beijing's Public Facilities*. Accessed: Jan. 22, 2018. [Online]. Available: http://www.bjstats.gov.cn/zt/tjkfr2016/szbj/201609/t20160918_358920.html

[12] W. Jiang, J. Lian, M. Shen, and L. Zhang, "A multi-period analysis of taxi drivers' behaviors based on GPS trajectories," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1115–1120.

[13] Y. Zheng and X. Zhou, *Computing with Spatial Trajectories*, 1st ed. New York, NY, USA: Springer, 2011.

[14] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Sci. Rep.*, vol. 3, no. 6, 2013, Art. no. 1376.

[15] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," in *Proc. Int. Conf. Weblogs Soc. Media*, Jul. 2011, pp. 81–88.

[16] L. Huang, Q. Li, and Y. Yue, "Activity identification from GPS trajectories using spatial temporal POIs' attractiveness," in *Proc. 2nd ACM SIGSPATIAL Int. Workshop Location Based Soc. Netw. (LBSN)*, New York, NY, USA, 2010, pp. 27–30. [Online]. Available: http://doi.acm.org/10.1145/1867699.1867704

[17] L. Cao and J. Krumm, "From GPS traces to a routable road map," in *Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. (GIS)*, New York, NY, USA, 2009, pp. 3–12. [Online]. Available:http://doi.acm.org/10.1145/1653771.1653776

[18] Y. Chen and J. Krumm, "Probabilistic modeling of traffic lanes from GPS traces," in *Proc. ACM 18th SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. (GIS)*, New York, NY, USA, 2010, pp. 81–88. [Online]. Available:http://doi.acm.org/10.1145/1869790.1869805

[19] M. Veloso, S. Phithakkitnukoon, and C. Bento, "Urban mobility study using taxi traces," in *Proc. Int. Workshop Trajectory Data Mining Anal.*, 2011, pp. 23–30.

[20] X. Li *et al.*, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers Comput. Sci.*, vol. 6, no. 1, pp. 111–121, 2012.

[21] B. Jiang, J. Yin, and S. Zhao, "Characterizing the human mobility pattern in a large street network," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 1, p. 021136, 2009.

[22] G. Chen, X. Jin, and J. Yang, "Study on spatial and temporal mobility pattern of urban taxi services," in *Proc. Int. Conf. Intell. Syst. Knowl. Eng. (ISKE)*, 2010, pp. 422–425.

[23] C. Chen, D. Zhang, P. S. Castro, N. Li, L. Sun, and S. Li, "Real-time detection of anomalous taxi trajectories from GPS traces," in *Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Berlin, Germany: Springer, 2011.

[24] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li, "Land-use classification using taxi GPS traces," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 113–123, Mar. 2013.

[25] G. Qi, X. Li, S. Li, Z. Wang, D. Zhang, and G. Pan, "Measuring social functions of city regions from large-scale taxi behaviors," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2011, pp. 384–388.

[26] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 186–194.

[27] N. Li and G. Chen, "Analysis of a location-based social network," in *Proc. IEEE Int. Conf. Comput. Sci. Eng. (CSE)*, Vancouver, BC, Canada, Aug. 2009, pp. 263–270.

[28] W. He, X. Liu, and M. Ren, "Location cheating: A security challenge to location-based social network services," in *Proc. Int. Conf. Distrib. Comput. Syst.*, 2011, pp. 740–749.

[29] C. Renso, M. Baglioni, J. A. F. de Macêdo, R. Trasarti, and M. Wachowicz, "How you move reveals who you are: Understanding human behavior by analyzing trajectory data," *Knowl. Inf. Syst.*, vol. 37, no. 2, pp. 331–362, Nov. 2013. [Online]. Available: http://dx.doi.org/10.1007/s10115-012-0511-z

[30] O. Derbel and R. J. Landry, "Driving style assessment based on the GPS data and fuzzy inference systems," in *Proc. Int. Multi-Conf. Syst., Signals Devices*, 2015, pp. 1–8.

[31] C. Wang, R. Fu, J. S. Peng, and J. Mao, "Driving style classification method for lane change warning system," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 14, no. 3, pp. 186–187, 2014.

[32] Z. Chen, J. Yu, Y. Zhu, Y. Chen, and M. Li, "D^3: Abnormal driving behaviors detection and identification using smartphone sensors," in *Proc. 12th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2015, pp. 524–532.

[33] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2014, pp. 25–34. [Online]. Available: http://doi.acm.org/10.1145/2623330.2623656

[34] J. Yuan *et al.*, "T-drive: Driving directions based on taxi trajectories," in *Proc. 18th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. (GIS)*, New York, NY, USA, 2010, pp. 99–108. [Online]. Available: http://doi.acm.org/10.1145/1869790.1869807

[35] B. Li *et al.*, "Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2011, pp. 63–68.

[36] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-Finder: A recommender system for finding passengers and vacant taxis," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 10, pp. 2390–2403, Oct. 2013.

[37] M. Maciejewski, J. Bischoff, and K. Nagel, "An assignment-based approach to efficient real-time city-scale taxi dispatching," *IEEE Intell. Syst.*, vol. 31, no. 1, pp. 68–77, Jan./Feb. 2016.

[38] S. Ma, Y. Zheng, and O. Wolfson, "T-share: A large-scale dynamic taxi ridesharing service," in *Proc. IEEE 29th Int. Conf. Data Eng. (ICDE)*, Apr. 2013, pp. 410–421.

[39] X. Wan, J. Wang, Y. Du, and Y. Zhong, "DBH-CLUS: A hierarchal clustering method to identify pick-up/drop-off hotspots," in *Proc. 15th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput.*, May 2015, pp. 890–897.

[40] X. Wan, M. Gao, J. Kang, and J. Zhao, "Taxi origin-destination areas of interest discovering based on functional region division," in *Proc. 3rd Int. Conf. Innov. Comput. Technol. (INTECH)*, Aug. 2013, pp. 365–370.

[41] L. Liu, C. Andris, A. Biderman, and C. Ratti, "Revealing taxi driver's mobility intelligence through his trace," in *Movement-Aware Applications for Sustainable Mobility: Technologies and Approaches*. Hershey, PA, USA: IGI Global, 2010, pp. 105–120.

[42] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani, "An energy-efficient mobile recommender system," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 899–908.

[43] P. S. Castro, D. Zhang, C. Chen, S. Li, and G. Pan, "From taxi GPS traces to social and community dynamics: A survey," *ACM Comput. Surv.*, vol. 46, no. 2, p. 17, 2013.

[44] D. Zhang *et al.*, "Understanding taxi service strategies from taxi GPS traces," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 123–135, Feb. 2014.

[45] L. Tang, F. Sun, Z. Kan, C. Ren, and L. Cheng, ''Uncovering distribution patterns of high performance taxis from big trace data,'' *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 5, p. 134, 2017.
[46] H. Rong, X. Zhou, C. Yang, Z. Shafiq, and A. Liu, ''The rich and the poor: A Markov decision process approach to optimizing taxi driver revenue efficiency,'' in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, 2016, pp. 2329–2334.
[47] Y. M. Nie, ''How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China,'' *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 242–256, Jun. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X17301018
[48] A. Jauhri *et al.* (Jan. 2017). ''Space-time graph modeling of ride requests based on real-world data.'' [Online]. Available: https://arxiv.org/abs/1701.06635
[49] J. Leskovec, J. Kleinberg, and C. Faloutsos, ''Graph evolution: Densification and shrinking diameters,'' *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, p. 2, 2007.
[50] B. Caulfield, ''Estimating the environmental benefits of ride-sharing: A case study of Dublin,'' *Transp. Res. D, Transp. Environ.*, vol. 14, no. 7, pp. 527–531, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361920909000893
[51] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti, ''Quantifying the benefits of vehicle pooling with shareability networks,'' *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 37, pp. 13290–13294, 2014.

WEIWEI JIANG (S'16) was born in Nantong, China, in 1991. He received the B.Sc. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2013, where he is currently pursuing the Ph.D. degree. He has several technical papers in the area of intelligent transportation system and GPS trajectory analysis.

LIN ZHANG (S'05–M'06–SM'13) received the B.Sc., M.Sc., and Ph.D. degrees from Tsinghua University, Beijing, in 1998, 2001, and 2006, respectively. He was a Visiting Professor with the University of California at Berkeley, Berkeley, CA, USA, from 2011 to 2013. He is currently a Professor and Co-Director with the Tsinghua-Berkeley Shenzhen Institute, Tsinghua University. He has co-authored over 60 peer-reviewed technical papers and five U.S. or Chinese patents applications. His current research interests include efficient protocols for sensor networks, statistical learning and data mining algorithms for sensory data processing, and information theory. He is a Committee Member of the Chinese Information Theory Society. He was a recipient of the IEEE/ACM SenSys 2010 Best Demo Award, the IEEE/ACM IPSN 2014 Best Demo Award, and the IEEE CASE 2013 Best Paper Award.

In 2006, he led a 2008 Beijing Olympic Stadium (the ''Bird's Nest'') structural security surveillance project, which deployed over 400 wireless temperature and tension sensors across the stadium's steel support structure and dome. The system adopted a flexible spectrum sensing and adaptive multihop routing algorithm to overcome strong radio interference and long-distance transmission channel fading, and played a critical role in the construction of the stadium. Since then, he has implemented wireless sensor networks in a wide range of application scenarios, including underground mine security, precision agriculture, and industrial monitoring. Since 2008, he has been in close association with CISCO to develop a metropolitan area sensing and operating network (MASON). MASON utilizes moving vehicles as carriers of the sensor platform so as to gain good sensing coverage and low cost at the same time. Shenzhen has adopted the technology and is deploying the MASON system in the urban area to monitor the microclimate environment, as an important part of the information service to its citizens. At Tsinghua University, he has been teaching the courses Selected Topics in Communication Networks and Information Theory to senior undergraduate and graduate students. He received the Excellent Teacher Award from Tsinghua University in 2004 and 2010, respectively.