Analysis of Hospitalizing Behaviors Based on Big Trajectory Data

Yongdong Wang, Dongwei Xu, Peng Peng, Qi Xuan, *Member, IEEE*, and Guijun Zhang

***Abstract*— With the improvement of living standards, people pay more attention to health, which is significant to analyze people's hospitalizing behaviors. The wide use of mobile devices generates a great deal of data, which contains a lot of travel information about residents. Many people would like to see a doctor through calling an online car hailing for its convenience. Thus, based on big trajectory data generated by the online car hailing, the hospitalizing behaviors of residents are analyzed in this paper. The hospitalizing behaviors are analyzed from two aspects. One is performed from the temporal aspect, in which the daily numbers of trips of hospitalizing behaviors under different modes are analyzed. The other one is performed from the spatial aspect, in which the hot hospitals, popularity, and gravity distribution of hospitals are analyzed. Based on the spatial analysis, the network constructed by the hot hospitals is also analyzed. The results show that the hospitalizing behavior analysis can reflect the hospitalizing behaviors in detail, which can make contributions to the decision-making of infrastructure configuration for institutions, such as urban planning departments and hospitals.**

***Index Terms*— Behavior analysis, hospitalizing behavior, road network, road traffic, trajectory data.**

I. Introduction

DATA are the basis for the study of the intelligent traffic system. Both urban traffic control and induction are limited by the scale of traffic data. With the rapid development of science and technology in China, many mobile devices have been used in our daily life. The Global Positioning System (GPS) devices equipped in the vehicles can record the dynamic performance of urban traffic and the crowd mobility.

Manuscript received February 4, 2019; revised April 23, 2019; accepted May 23, 2019. Date of publication June 26, 2019; date of current version August 8, 2019. This work was supported in part by the Zhejiang Provincial Natural Science Foundation under Grant LQ16E080011, in part by the China Postdoctoral Science Foundation under Grant 2018M632501, in part by the Major Project of Zhejiang Lab under Grant 2019DH0ZX01, and in part by the National Natural Science Foundation of China under Grant 61572439. *(Corresponding author: Dongwei Xu.)*

Y. Wang and P. Peng are with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China (e-mail: wydong_112@163.com; pp_2017_zjut_edu@163.com).

D. Xu is with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China, also with the Research Institute, Enjoyor Co., Ltd., Hangzhou 310030, China, also with the Institute of Cyberspace Security, Zhejiang University of Technology, Hangzhou 310023, China, and also with the Zhejiang Laboratory, Hangzhou 311121, China (e-mail: dongweixu@zjut.edu.cn).

Q. Xuan is with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China, also with the Institute of Cyberspace Security, Zhejiang University of Technology, Hangzhou 310023, China, and also with the Zhejiang Laboratory, Hangzhou 311121, China (e-mail: xuanqi@zjut.edu.cn).

G. Zhang is with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China (e-mail: zgj@zjut.edu.cn).

Digital Object Identifier 10.1109/TCSS.2019.2920696

Many people like to travel by online car hailing for its convenience, which generates great trajectory data in road traffic field and provides the data support for the study of urban traffic. The trajectory data contain a great deal of information, which can be used to perform the travel analysis of residents. Compared to the loop data, the trajectory data can be collected easily and they contain more information. Thus, the travel information of urban residents can be better mined based on the trajectory data.

The trajectory data after the preprocessing of computation and knowledge mining can be used to identify and detect the urban traffic states [1], [2]. The hot topics of traffic states research include the estimations of link travel time and average travel speed, and the analysis and prediction of traffic congestion. Traffic anomaly detection refers to the detection of traffic accident, bad traffic events, and large-scale traffic congestion prediction. For example, Liu *et al.* [3] presented a mobility-based clustering method to detect congestion, which used the instantaneous vehicle speed to detect the congestion of the surrounding condition. Kong *et al.* [4] realized the urban traffic congestion estimation and prediction based on the trajectory data of the floating car, which proved the merits of the method in accuracy, real-time performance, and stability. D'Andrea and Marcelloni [5] presented an expert system to detect traffic congestion and accidents based on the real-time GPS data collected from the intelligent phones of drivers. Wang *et al.* [6] designed a congestion detection system based on the trajectory data, which mapped the trajectory data into the map to automatically obtain the speed values and the traffic congestion conditions in the links. Besides, the condition of the current road network has also been provided in multiperspectives.

The basic indicators and information, such as spatial and temporal distributions and running modes, can also be obtained based on the GPS data. The information not only can be applied to the daily taxi managements but also can be used to provide decision-making for the taxi market. Besides, the valuable knowledge can also be mined to provide guidelines and decision-making for urban traffic planning and management and improve urban traffic. For example, Liao *et al.* [7] proposed a directed density clustering method, which can mine the distribution in both temporal and spatial perspectives. Gong *et al.* [8] presented a trip purpose inferring framework based on the trajectory data. The spatiotemporal model, combined Monte Carlo simulation method, is modeled based on the interesting points to study the trajectory data in Shanghai. The experiments results showed nine types of spatiotemporal

characteristics in daily activities, such as temporal regularity, spatial dynamics, travel distance, and travel directions. The semantic mobility patterns can also be mined based on the trajectory data [9]–[12]. Through mining trajectory data, the traffic running managements and decision-making can be realized. In addition, the trajectory data contribute to the analysis of social network [13], [14].

Route planning and prediction mainly refer to the selection and estimation of vehicle driving paths in terms with trajectory data [15]–[18]. Route prediction focuses on the prediction of driving paths and the estimation of destination based on the historical driving data [19]. For example, Lin and Li [20] proposed a night bus route planning method to select the optimal route. To solve the positioning draft and jitter problems of mobile phones, Zhang *et al.* [21] presented a trip identification method based on mobile phone positioning data, which has high identification accuracy and can provide relevant information for transportation planning. Alexander *et al.* [22] clustered the users' clustered location based on a large amount of trajectory data and then inferred departure time based on the arrival time, which constructed the users' daily travel circumstances. The experiments proved that this method can provide a way of route planning.

Besides the applications mentioned earlier, the trajectory data can also be applied to update urban road network, identify travel mode, obtain the construction information of road network, and so on [23]–[26]. The updated urban road network based on the trajectory data can reduce the cost of road mapping [27]. For example, Xiao *et al.* [28] modeled the Bayesian network of travel mode based on the trajectory data. Yang *et al.* [29] utilized the trajectory data of floating car to model the construction of lanes, which can obtain the number of lanes and the characteristics of lane construction. Kong *et al.* [30] inferred the behavior of entering the town based on the trajectory data, which combined the behavioral mode characteristics and real labeled data set to train a multiple classification model. Du *et al.* [31] judged whether users stopped in a certain location based on mobile phone trajectory data and geographical information. Lin *et al.* [17] presented a graph convolutional neural network with data-driven graph filter model to predict the station-level hourly demand in a large-scale bike-sharing network, which can learn the hidden heterogeneous pairwise correlation between the stations.

Due to the improvement of living standards, people pay more attention to health. The urban traffic mode and the travel regularity have great significance for residents and urban traffic managements. Yet, the study of hospitalizing behaviors mainly concentrates on the factors that influence the treatment and the social structure of medical care [32]. The hospitalizing behaviors do not have a clear description, which has a great significance for urban traffic resources, route planning, and so on. In this paper, the analysis of hospitalizing behaviors is achieved based on the big trajectory data, the contributions of which mainly include three aspects.

1) The hospitalizing behaviors are analyzed in the temporal aspect, which can both macroscopically and microscopically depict the hospitalizing behaviors in the transportation aspect. The results help urban traffic managements understand the regularity of hospitalizing behaviors on the whole.
2) The hospital rank can be acquired based on the spatial analysis of hospitalizing behaviors, which can induce both hospitals and traffic managements to allocate the infrastructures.
3) The network analysis with nodes composed of hospitals is conducted. In this paper, the aggregation and dissipation processes, the analysis of gravity, and the transfer analysis have been analyzed. The network analysis can assist both traffic departments and residents to make reasonable choices.

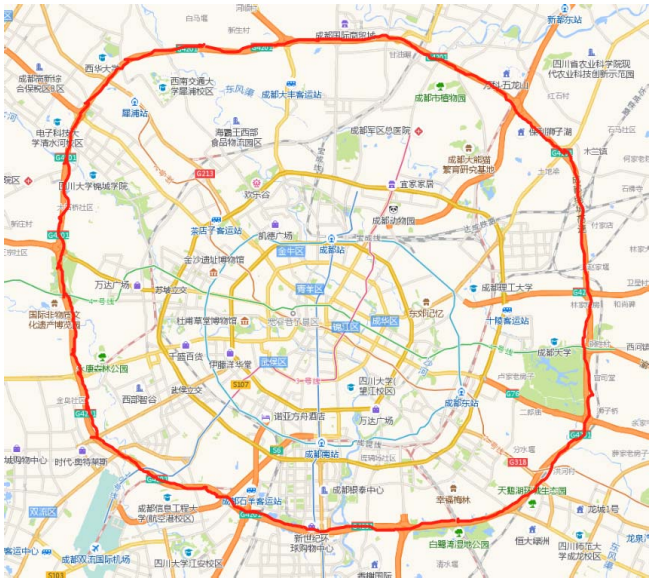


Fig. 1. Study area of Chengdu.

The rest of this paper is organized as follows. The data have been described in detail in Section II. The temporal and spatial analyses of hospitalizing behaviors are conducted in Section III and IV, respectively. The study is concluded in Section V.

II. Data

The big trajectory data used in our research are derived from DiDi Chuxing, on which residents call cars on the Internet by mobile phones. The data are in November, 2017 in Chengdu, Sichuan. The ranges of latitude and longitude of the area are from east 102°54′ to 102°53′ and from North 30°05′ to 31°26′, respectively. The size of data used in our study is approximately 200 000 after progressing. The specific area is visual in the red cycle in the QGIS platform, which can be seen in Fig. 1. The trajectory data located region contains the prosperous districts, which can reflect the travel behaviors of urban residents and are useful for the analysis of hospitalizing behaviors.

The data includes seven data types, i.e., order ID, the time getting on vehicle, the time getting off vehicle, the longitudes getting on and off vehicle, and the latitudes getting on and off vehicle. Since not all data can be used for the analysis

etAoBlnj	2016-11-24 19:27:00	2016-11-24 19:50:08	104.08274	30.68405	104.0481	30.67216
qvzmJkk	2016-11-24 08:14:21	2016-11-24 08:39:10	103.96845	30.67348	104.07345	30.68898
jonkRiw	2016-11-24 13:58:59	2016-11-24 14:07:10	104.07789	30.64156	104.08307	30.65785
ipxkTjr	2016-11-24 18:46:10	2016-11-24 18:56:37	104.044487	30.69768	104.07865	30.70056
fBvhBkil	2016-11-24 14:06:36	2016-11-24 14:22:17	104.0834	30.65981	104.05957	30.64816
rDruJmwl	2016-11-24 16:10:58	2016-11-24 16:42:23	104.073723	30.689259	104.15396	30.64291
sxnhAok	2016-11-24 20:52:08	2016-11-24 21:11:23	104.09412	30.66614	104.06338	30.62973
kuzsAft	2016-11-24 15:41:43	2016-11-24 15:52:45	104.06042	30.65939	104.060896	30.677415
przpEst	2016-11-24 11:14:28	2016-11-24 11:30:54	104.09657	30.65752	104.14896	30.61878
lBziFjx	2016-11-24 15:46:41	2016-11-24 15:56:44	104.06352	30.67014	104.08007	30.67814
nqgmCipJ	2016-11-24 15:59:32	2016-11-24 16:13:36	104.08213	30.67819	104.07403	30.6863
nDCqEmt	2016-11-24 13:58:46	2016-11-24 14:12:10	104.04447	30.69802	104.05513	30.72829
fzAvFyt	2016-11-24 18:11:03	2016-11-24 18:52:17	104.08349	30.67751	104.09664	30.63999
loygJoxR	2016-11-24 14:07:30	2016-11-24 14:33:51	104.037651	30.689217	104.08151	30.7949
jzAjxsu	2016-11-24 09:48:09	2016-11-24 10:02:58	104.0669	30.6564	104.07363	30.6899
oOwjBst	2016-11-24 22:36:55	2016-11-24 23:03:18	104.06193	30.64906	104.01949	30.7026
zBosDil	2016-11-24 17:12:42	2016-11-24 17:24:00	104.07187	30.65003	104.08339	30.66738
ntslzos	2016-11-24 11:34:21	2016-11-24 11:47:56	104.01867	30.66213	104.04955	30.67252

Fig. 2. Parts of the data for hospitalizing behaviors.

of hospitalizing behaviors, some preprocesses should be conducted. First, the information of the order ID is desensitized to protect the privacy of the passengers. Second, the records with the NA or empty value for longitude or latitude are deleted. Third, the records with longitudes or latitudes beyond the ranges of the area should be deleted. Fourth, the distance threshold ε is provided to select the records that regard hospitals as origin or destination, which can be illustrated in the following equations:

$$\begin{aligned} |D(p_i, q_i) - D(x_{\text{ON}}, y_{\text{ON}})| &\leq \varepsilon \;\; D(x_{\text{ON}}, y_{\text{ON}}) \in \text{hospital } i \\ |D(p_i, q_i) - D(x_{\text{ON}}, y_{\text{ON}})| &\geq \varepsilon \;\; D(x_{\text{ON}}, y_{\text{ON}}) \notin \text{hospital } i \end{aligned} \tag{1}$$

$$\begin{aligned} |D(p_i, q_i) - D(x_{\text{OFF}}, y_{\text{OFF}})| &\leq \varepsilon \;\; D(x_{\text{OFF}}, y_{\text{OFF}}) \in \text{hospital } i \\ |D(p_i, q_i) - D(x_{\text{OFF}}, y_{\text{OFF}})| &\geq \varepsilon \;\; D(x_{\text{OFF}}, y_{\text{OFF}}) \notin \text{hospital } i \end{aligned} \tag{2}$$

where $D(p_i, q_i)$ and p_i and q_i denote the location of hospital i and the longitude and latitude of hospital i, respectively. $D(x_{\text{ON}}, y_{\text{ON}})$ and x_{ON} and y_{ON} represent the location of getting on a vehicle and the longitude and latitude of the location, respectively. $D(x_{\text{OFF}}, y_{\text{OFF}})$ and x_{OFF} and y_{OFF} represent the location of getting off a vehicle and the longitude and latitude of the location, respectively.

Parts of the data of hospitalizing behaviors can be seen in Fig. 2.

III. Temporal Distribution Analysis of Hospitalizing Behaviors

The temporal analysis of hospitalizing behaviors can reflect the temporal traffic demand. In this section, three parts of hospitalizing behaviors, i.e., the daily average number of trips, the time distribution, and the travel time analysis, are conducted to depict the temporal distribution from different perspectives. The residents represent that the trips are relevant to the hospitals, that is, getting on or off vehicles at hospitals.

A. Daily Average Number of Trips

A trip begins when a resident gets on a vehicle and ends when the resident gets off the vehicle. The analysis of the daily average number of trips reflects the macroscopic traffic demands for hospitalizing behaviors. Due to different travel regularities, job categories, and so on, the traffic demands under workday and weekend modes are different. Taking the trip number of one day for analysis cannot reflect the whole

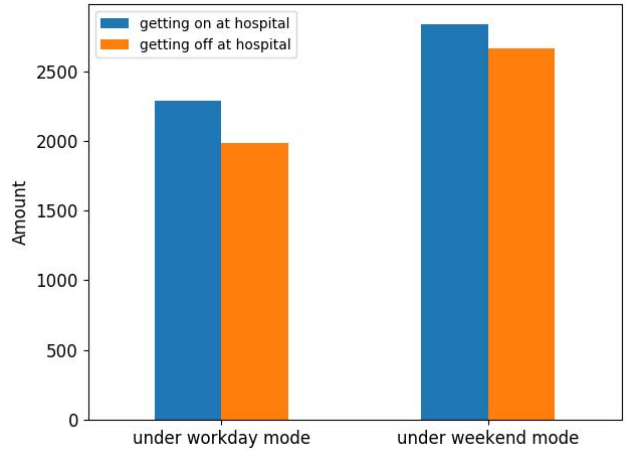


Fig. 3. Daily average number of trips of hospitalizing behaviors.

hospitalizing behaviors regularity. Besides, since the numbers of days of workdays and weekends for analysis are 22 and 8, respectively, the comparison of the total number of trips under the two modes is not reasonable. Thus, we analyze the daily average number of trips under the two modes.

Let $N(x, y, t_{\text{ON}}, t_{\text{OFF}})$ denote the number of trips that start at location x on time t_{ON} and end at location y on time t_{OFF}. The average numbers of trips of getting on and off vehicles at hospitals can be obtained through the following equations:

$$N_{\text{ON}}^{M_p} = \sum_{t_{\text{ON}} \in M_p, x \in DS} N(x, y, t_{\text{ON}}, t_{\text{OFF}}) \tag{3}$$

$$N_{\text{OFF}}^{M_p} = \sum_{t_{\text{OFF}} \in M_p, y \in DS} N(x, y, t_{\text{ON}}, t_{\text{OFF}}) \tag{4}$$

where $N_{\text{ON}}^{M_p}$ and $N_{\text{OFF}}^{M_p}$ denote the average numbers of trips of getting on and off vehicles at hospitals under M_p mode, respectively, DS is composed of D_{H_i}, D_{H_i} denotes the ε neighborhood of hospital i, $p \in \{1, 2\}$, and M_1 and M_2 denote workday and weekend modes, respectively.

Based on (3) and (4), the daily average number of trips of hospitalizing behaviors under workday and weekend modes can be obtained. The results are shown in Fig. 3.

From Fig. 3, we can see that the numbers of trips under workday and weekend modes are different. The number of trips of getting on vehicles at hospitals on weekends is larger than that on workdays. Consistently, the number of trips of getting off vehicles at hospitals on weekends is larger than that on workdays. Some reasons can be explained for these results. As some office workers have to work and make an appointment with the doctors on weekends, the daily average number of trips on weekends is larger than that on workdays. Besides, since some trips begin at the current day and end at the next day, the number of trips of getting on vehicles at hospitals is larger than that of getting off vehicles at hospital no matter on workday or weekend. We can also conclude that more residents prefer to see the doctor under the weekend's mode. These findings suggest that hospitals will receive more patients and there maybe exist resource shortage on weekends. Thus, for hospitals, there should allocate more medical workers and

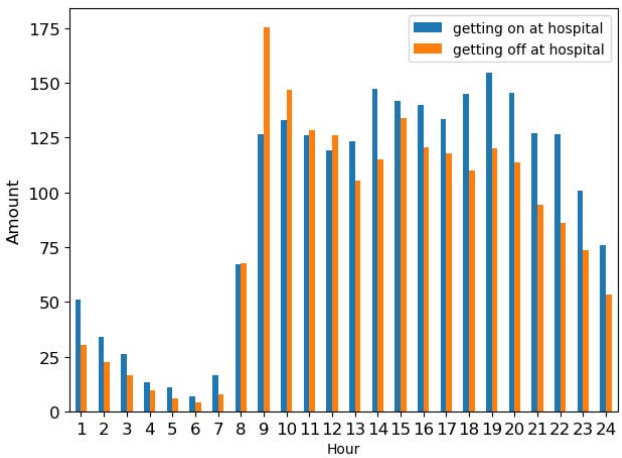


Fig. 4. Time distribution in time series on workday

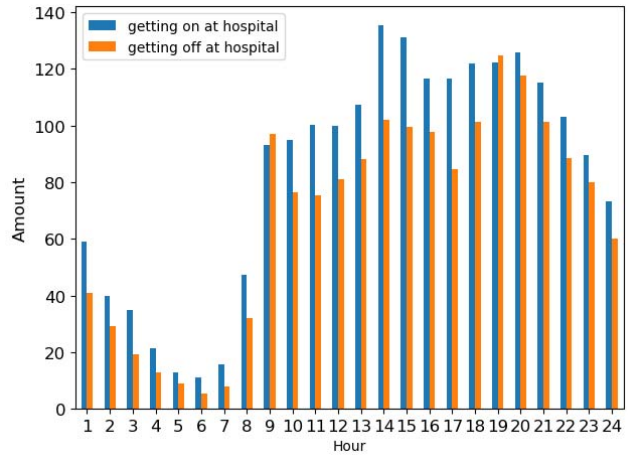


Fig. 5. Time distribution in time series on weekend

equipments. For patients, if possible, it would be better to see a doctor on the workday.

B. Time Distribution Analysis of Hospitalizing Behaviors

To further depict the hospitalizing behaviors, the time distribution analysis of hospitalizing behaviors is performed. The purpose is to mine the traffic demands of hospitalizing behaviors in time series.

The time distribution of hospitalizing behaviors of getting on and off vehicles at hospitals is shown in the following equations:

$$N_{\mathrm{ON}}^{M_p}(t_h) = \sum_{t_{\mathrm{ON}} \in (M_p, t_h), x \in DS} N(x, y, t_{\mathrm{ON}}, t_{\mathrm{OFF}}) \tag{5}$$

$$N_{\mathrm{OFF}}^{M_p}(t_h) = \sum_{t_{\mathrm{OFF}} \in (M_p, t_h), y \in DS} N(x, y, t_{\mathrm{ON}}, t_{\mathrm{OFF}}) \tag{6}$$

where $h \in \{0, 1, 2, \ldots, 23\}$ and t_h denotes the time period during h o'clock to $(h+1)$ o'clock. $N_{\mathrm{ON}}^{M_p}(t_h)$ and $N_{\mathrm{OFF}}^{M_p}(t_h)$ denote the average numbers of trips of getting on and off vehicles at hospitals during the t_h time period under M_p mode.

The results of time distribution analysis can be seen in Figs. 4 and 5.

From Figs. 4 and 5, we can see that the travel regularities of hospitalizing behaviors under workday and weekend modes are similar in time series. The number of travel trips reduces until 6:00 and then increases. In particular, the number of travel trips increases sharply at 8:00. The reason is that residents begin to travel for seeking medical treatment at work hours. The number of trips gradually reduces from 20:00. Besides, the number of trips of getting off vehicles at hospitals is larger than that of getting on vehicles during 8:00–9:00 under the workday mode. This phenomenon can be explained by the residents' arrival at hospitals for seeking medical treatments. The situation is just opposite during 14:00–20:00, because many residents have finished the medical treatments and begin to leave. The similar situation can also be seen during 8:00–9:00 under the weekend mode, while the departure trend concentrates during 14:00–24:00. For taxi drivers, it is a better choice to wait for the passengers during 14:00–20:00 at hospitals, and for patients, they can see a doctor during 1:00–7:00, since there are a small number of patients at the time.

C. Travel Time Analysis of Hospitalizing Behaviors

The travel time analysis is mainly focused on the duration of trips. The meaning of the analysis is to investigate the residents' preference of trip duration. Since the detailed division of travel time does not have significance, we divide the travel time into three levels, i.e., shorter than 30 min, between 30 and 60 min, and larger than 60 min. The number of trips under different levels based on getting on and off vehicles at hospital can be illustrated in the following eqautions:

$$N_{\mathrm{ON}}^{M_p}(m_i) = \sum_{t_{\mathrm{ON}} \in M_p, x \in DS, |t_{\mathrm{OFF}} - t_{\mathrm{ON}}| \in m_i} N(x, y, t_{\mathrm{ON}}, t_{\mathrm{OFF}}) \tag{7}$$

$$N_{\mathrm{OFF}}^{M_p}(m_i) = \sum_{t_{\mathrm{OFF}} \in M_p, y \in DS, |t_{\mathrm{OFF}} - t_{\mathrm{ON}}| \in m_i} N(x, y, t_{\mathrm{ON}}, t_{\mathrm{OFF}}) \tag{8}$$

where $i \in \{1, 2, 3\}$ and m_i denotes the ith travel time level, m_1–m_3 denote shorter than 30 min, between 30 and 60 min, and larger than 60 min, respectively, and $N_{\mathrm{ON}}^{M_p}(m_i)$ and $N_{\mathrm{OFF}}^{M_p}(t_h)(m_i)$ denote the number of trips of getting on and off vehicles at hospital corresponding to m_i travel time level under the M_p mode, respectively.

The results under workday and weekend modes are shown in Figs. 6 and 7, respectively.

From Figs. 6 and 7, we can see that the main travel time is shorter than 60 min under both workday and weekend modes. This suggests that residents prefer to spend shorter time on the way to go to hospitals for medical treatment. As there is a delay between departure time and arrival time, the arrival time is later than the departure time. The number of trips of getting on vehicles at hospitals is larger than that of getting off vehicle at hospitals. Combined with the analysis results in Figs. 4 and 5, we can find that the delay is no more than 60 min, that is, most of the residents who get on vehicles during 8:00–9:00 will arrive their destinations during 9:00–10:00.

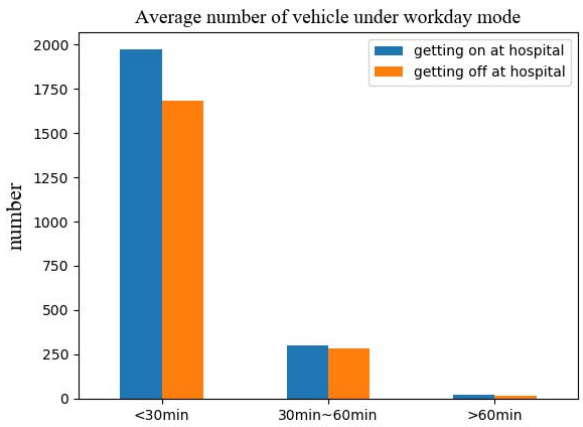


Fig. 6. Distribution of trip duration on workday.

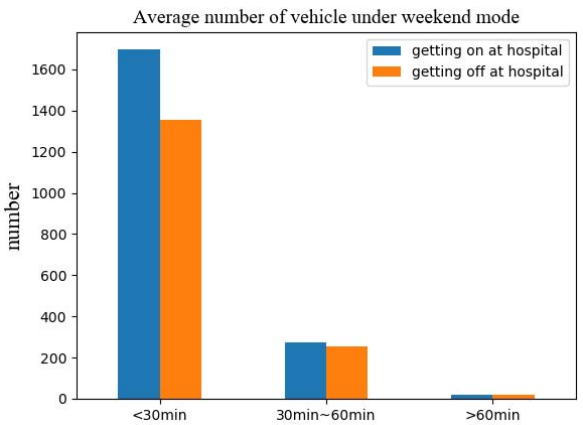


Fig. 7. Distribution of trip duration on weekend.

Generally, the trip duration is mainly focused on shorter than 30 min, which means that most residents would like to select a hospital close to the starting point.

IV. Spatial Analysis of Hospitalizing Behaviors

The spatial analysis is composed of three parts, i.e., the rank analysis of hospitals, the aggregation and dissipation processes, and the gravity and transfer analysis for hospitalizing behaviors. Compared to the temporal analysis in Section III, the spatial analysis of hospitalizing behaviors has more important significations. For example, the rank analysis of hospitals can help both patients and hospitals realize the rank of hospitals from the view of traffic. The analysis of aggregation and dissipation processes reveals the aggregation and dissipation processes of every hospital, which further assist public traffic management to reasonably arrange the traffic resources and provide early warning for hospitals to deal with the crowd gathering. In short, the network analysis attempts to mine the cause and effect from the big trajectory data.

A. Rank Analysis of Hospitals

The aim of the rank analysis is to conduct a degree of popularity based on the trips of hospitalizing behaviors.

TABLE I

Part of the Hospitals

Hospital ID	The name of hospital
H1	The fifth hospital of Sichuan province
H2	Sichuan province forestry center hospital
H3	Sichuan provincial fourth peoples hospital
H4	Chengdu university of TCM
H5	Chengdu eight peoples hospital
H6	Chengdu Tumor hospital
H7	The Fourth peoples hospital of Chengdu
H8	Sichuan Electric Power Hospital
H9	The third peoples hospital of Chengdu
H10	Sichuan Provincial People's Hospital
H11	West China Hospital, Sichuan University
H12	Chengdu western hospital
H13	Chengdu Tai Kun Tong Chinese medicine clinic
H14	Chengdu Ping An hospital
H15	maternal and child health care hospital of Chenghua District

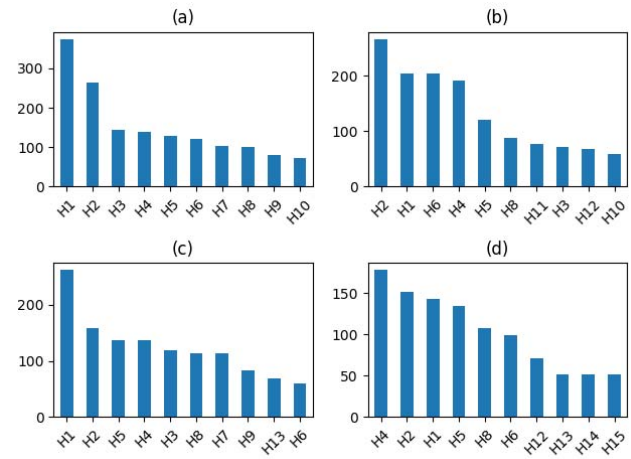


Fig. 8. Rank analysis of hospitalizing behaviors. (a) and (b) Rank of getting on and off vehicles under the workday mode, respectively. (c) and (d) Rank of getting on and off vehicles under the weekend mode, respectively. The x- and y-axes represent the hospitals and the counts of trips, respectively.

Though every hospital can advocate the popularity and medical level of their hospitals, the data will tell the truth. In this section, we will rank the hospitals of getting on and off vehicles at hospitals under different modes, which can be seen in the following equations:

$$N_{\mathrm{ON}}^{M_p}(H_i) = \sum_{t_{\mathrm{ON}} \in M_p, x \in D_{H_i}} N(x, y, t_{\mathrm{ON}}, t_{\mathrm{OFF}}) \tag{9}$$

$$N_{\mathrm{OFF}}^{M_p}(H_i) = \sum_{t_{\mathrm{OFF}} \in M_p, y \in D_{H_i}} N(x, y, t_{\mathrm{ON}}, t_{\mathrm{OFF}}) \tag{10}$$

where $N_{\mathrm{ON}}^{M_p}(H_i)$ and $N_{\mathrm{OFF}}^{M_p}(H_i)$ denote the number of trips of getting on and off vehicles at hospital H_i under the M_p mode, respectively, H_i denotes the hospital H_i, and D_{H_i} denotes the neighborhood of hospital H_i.

The rank of only top ten hospitals is listed. Parts of the hospitals can be seen in Table I.

The results of the top ten popular hospitals are shown in Fig. 8.

As shown in Fig. 8, we can see that the trend of getting on and off vehicles at hospitals under the workday mode

is similar in the top ten hospitals. However, the differences between Fig. 8(a) and (b)–(d) should not be ignored. The number of top two hospitals in Fig. 8(a) is larger than the latter eight hospitals, while this phenomenon is not clearly in Fig. 8(b)–(d). The reason can be explained that some healthy residents near the hospital should not be regarded as the trips of hospitalizing behaviors.

The top ten hospitals of getting on and off vehicles at hospitals are different under the weekend mode, which can be seen in Fig. 8(c) and (d). The reason for this phenomenon may be that many residents prefer to selecting the nearby hospitals for seeking medical treatments on the workday. When it is weekend, residents have more leisure time to choose the hospitals and more hospitals will be taken into consideration.

In addition, the rank analysis can also be used to validate the truth of popularity. From Fig. 8, we can find that the top two hospitals are H1 and H2, while the top two hospitals are H11 and H5 on the Internet. Besides, both H1 and H2 are Grade II Level hospitals compared with other hospitals, which means that other hospitals have a more comprehensive power. The rank of H1, H3, and H5 in the province is 141, 7, and 388, respectively. This discovery demonstrates that these suggest that the official rankings are only one thing, and the data tell the truth. People can take the public praise into consideration of the hospitals when seeing a doctor, for example, H1 and H2.

B. Aggregation and Dissipation Analysis of Hospitalizing Behaviors

The aggregation and dissipation analysis of hospitalizing behaviors can be seen in time series, which can be used for the allocation of infrastructure and resources in hospitals. Besides, the aggregation and dissipation processes also reflect the traffic demands of hospitals, which can help traffic control departments optimize the traffic plans.

In this section, the location of getting on and off vehicles represents the origin and destination nodes, respectively, and the number of trips represents the weight. Then, the network $\mathbf{Net_A}$ can be constructed through (11). There are two types of nodes in $\mathbf{Net_A}$, i.e., hospital nodes and non-hospital nodes

$$\mathbf{Net_A} = (\boldsymbol{N}, \boldsymbol{E}, \boldsymbol{ES}) \tag{11}$$

where $N = H1, H2, \ldots, Hn, \ldots$ is the set of nodes in the records that contain hospital nodes and non-hospital nodes, n is the number of hospitals, $\boldsymbol{E} = \{e_{xy} \| x, y \in \boldsymbol{D}\}$ is the set of edges, which means a trip leave from location x and arrive at location y, $e_{\mathrm{xy}} \neq e_{\mathrm{yx}}$, $x \in hospital$ or $y \in hospital$, and $\boldsymbol{ES} : \boldsymbol{E} \rightarrow \boldsymbol{S}$ is the mapping function from an edge to the trip number.

The distribution of trips in geography and social network can be seen in Figs. 9 and 10, respectively. From Fig. 9, we can have an intuitive understanding of the trips' distribution in geographic space. We can find that the trips close to the urban center have more trips. Besides, an intuitive understanding of the distribution of entering and exiting hospitals can be seen in Fig. 10, from which the in-degree and out-degree of hospitals can be seen.

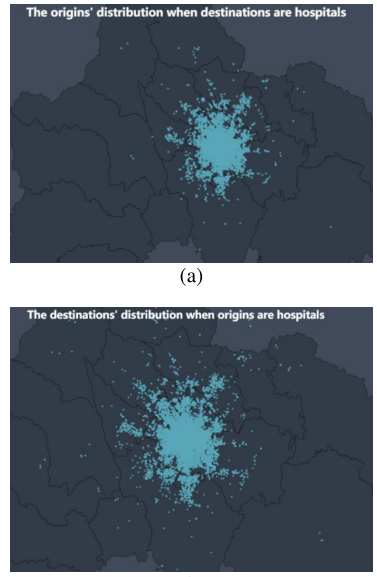


Fig. 9. Distribution of trips. (a) Origins' distribution when origins are hospitals. (b) Destinations' distribution when origins are hospitals.

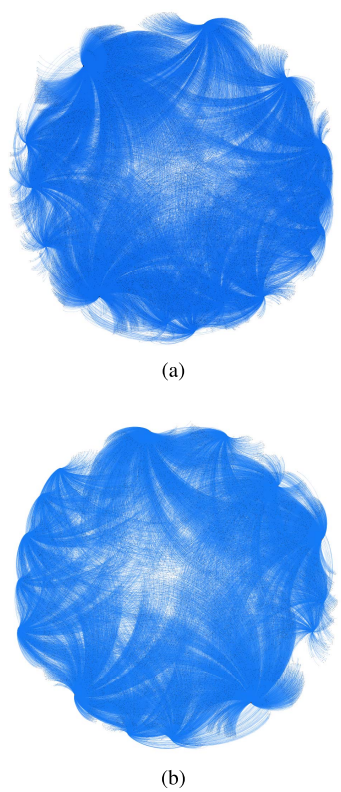


Fig. 10. Distribution of the trips. (a) Getting on vehicles at hospitals. (b) Getting off vehicles at hospitals.

The in-degree and out-degree of the hospital nodes represent the dynamic traffic demands of hospitalizing behaviors, which can be obtained through the following equations:

$$VN(t_h, H_i)_{\mathrm{in}}^{M_p} = \sum_{t_{\mathrm{OFF}} \in (t_h, M_p), y \in D_{H_i}} N(x, y, t_{\mathrm{ON}}, t_{\mathrm{OFF}}) \tag{12}$$

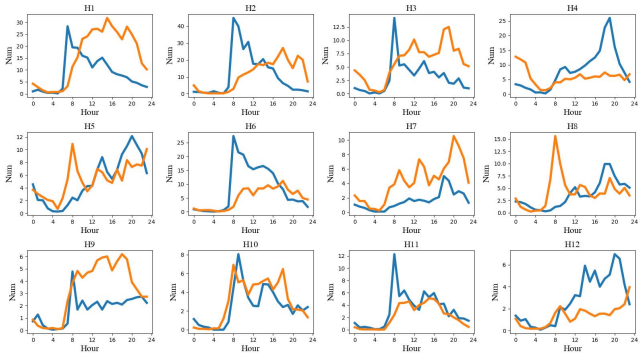


Fig. 11. In-degree and out-degree in time series under the workday mode. The blue and orange lines represent the in-degree and out-degree, respectively.

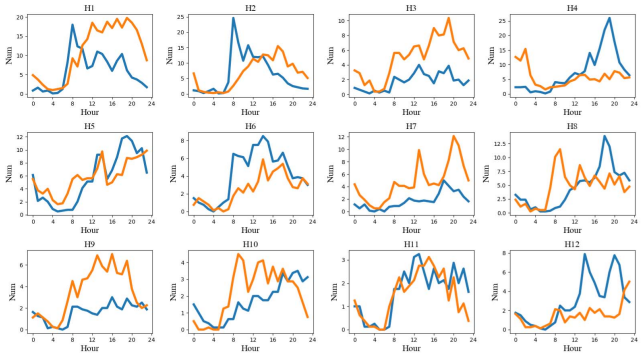


Fig. 12. In-degree and out-degree in time series under the weekend mode. The blue and orange lines represent the in-degree and out-degree, respectively.

$$VN(t_h, H_i)_{\text{out}}^{M_p} = \sum_{t_{\text{ON}} \in (t_h, M_p), x \in D_{H_i}} N(x, y, t_{\text{ON}}, t_{\text{OFF}}) \quad (13)$$

where $VN(t_h, H_i)_{\text{in}}^{M_p}$ and $VN(t_h, H_i)_{\text{out}}^{M_p}$ denote the in-degree and out-degree of hospital H_i during the t_h time period under the M_p mode, respectively.

The in-degree and out-degree of top 12 hospitals in Table I are investigated here, which can be seen in Figs. 11 and 12, respectively.

From Figs. 11 and 12, we can see the in-degree and out-degree of hospitalizing behaviors distributed in each hour, while the change rules are different under the workday and weekend modes. For the workday mode, we can see that there exist three patterns. The first pattern is that the in-degree sharply increases around 8:00 and then gradually reduced until 24:00. Meanwhile, the out-degree increases around 8:00 until 12:00 and then fluctuates in time series. This pattern can be seen in some hospitals, for example, H1–H3, H6, H9, and H11. In the second pattern, the peaks of in-degree and out-degree concentrated around 8:00 and 20:00, respectively. This pattern can be seen in some hospitals, for example, H5, H8, and H12. The last pattern is that both the trends of in-degree and out-degree are approximately synchronous in time series. The trend can be seen in some hospitals, for example, H4, H7, and H10.

Based on the division of patterns under the workday mode, there are also three patterns under the weekend mode. We can see that comparing Figs. 11 and 12, the patterns of in-degree and out-degree under the workday and weekend modes are basically the same for some hospitals, for example, H1, H2, H4, H7–H9, and H12, while the patterns for other hospitals are different under the workday and weekend modes.

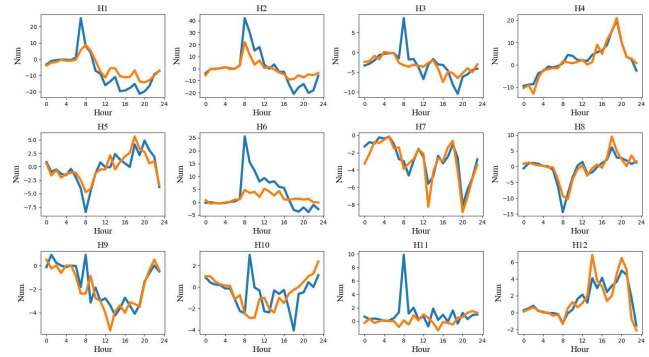


Fig. 13. Processes of aggregation and dissipation. The blue and orange lines represent the processes aggregation and dissipation under the workday and weekend modes, respectively.

The analysis of in-degree and out-degree reflects the residents' arrival and departure of hospitals for seeking medical treatments. The cumulative effect is more valuable compared to the analysis of in-degree and out-degree, because the cumulative effect shows the dynamically cumulative traffic demand information in time series. The information can be used for patients to choose alternative hospitals and hospitals to arrange resources.

The analysis of aggregation and dissipation is taken the historical effects into consideration. The aggregation and dissipation of hospitalizing behaviors can be described in the following equation:

$$\Delta VN(t_h, H_i)^{M_p} = \sum_{t=0:00}^{t=t_h} VN(t, H_i)_{\text{in}}^{M_p} - \sum_{t=0:00}^{t=t_h} VN(t, H_i)_{\text{out}}^{M_p}. \quad (14)$$

The results of the aggregation and dissipation of hospitalizing behaviors can be seen in Fig. 13.

According to the results shown in Fig. 13, three patterns can be concluded under the workday mode. The first pattern is that the aggregation bursts and then quickly disappears, which can be seen in some hospitals, for example, H1–H3, H6, H10, and H11. The second pattern is that the dissipation gradually disappears and, then, the aggregation grows up in time series. This phenomenon can be found in some hospitals, for example, H4 and H12. The third pattern is that only a slight aggregation appears in a small time period, while other time is in dissipation, which can be found in some hospitals, for example, H5, H7, and H8.

Compared with the aggregation and dissipation processes under the workday mode, we can find that there is no explicit aggregation in H1–H3 under the weekend mode. Another pattern under the weekend mode is that just a little aggregation appears. We can also find that for some hospitals, the patterns

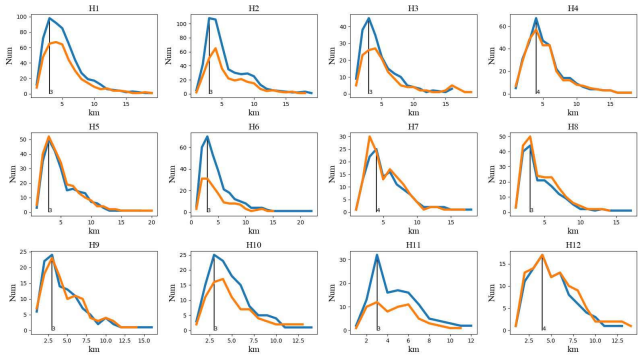


Fig. 14. Gravity analysis of H1–H12 under the workday and weekend modes. The blue and orange lines represent the gravity under the workday and weekend modes, respectively.

under the workday and weekend modes are similar, which can be found in H5, H8, and H9.

In short, the analysis of in-degree, out-degree, aggregation, and dissipation shows the dynamics of hospitalizing behaviors in hospitals. Based on the analysis, hospitals can be categorized into several groups based on hospitalizing behaviors. To some extent, the classification reflects the pattern of hospitalizing behaviors. In addition, the classification also exhibits the property of the hospitals. For example, H1, H6, and H10 have the top comprehensive influence overall. H1 and H5 are famous for aged diseases' treatment.

C. Gravity Analysis of Hospitalizing Behaviors

The gravity analysis exhibits the travel psychology of seeking medical treatments in terms with the travel distance, that is, there is a certain relationship between the choice of hospitals and the travel distance to hospitals. To obtain the relationship, we take the logarithm operation for both distance and frequency meanwhile. The gravity analysis under the workday and weekend modes is shown in Fig. 14.

According to the rank analysis in Section IV-A, the most popular hospitals for medical treatments are H1 and H2, which can also be found in Fig. 14. Specifically, the biggest number of trips is close to 100 in H1 and H2. For other hospitals, the number is not exceeding 70. Besides, a peak, for all the hospitals shown in Fig. 14, is appeared within a travel distance (about 4 km). Before the peaks occur, the number of trips increases sharply with the increase in travel distance, the curve of which is approximately linear. After the peak, the trend is gradually reduced with the increase in travel distance. Thus, we can conclude that residents take the distance to the hospitals into consideration in terms of selecting hospitals for seeking medical treatments. Most people prefer 3 or 4 km of travel distances. When beyond the distance, the number of trips begins to reduce. According to the above-mentioned analysis, many residents choose a hospital that no more than 4 km from the location getting on vehicle. Combined with the aggregation and dissipation patterns of the hospitals, residents can choose a suitable hospital and a better time to see a doctor based on their location and time.

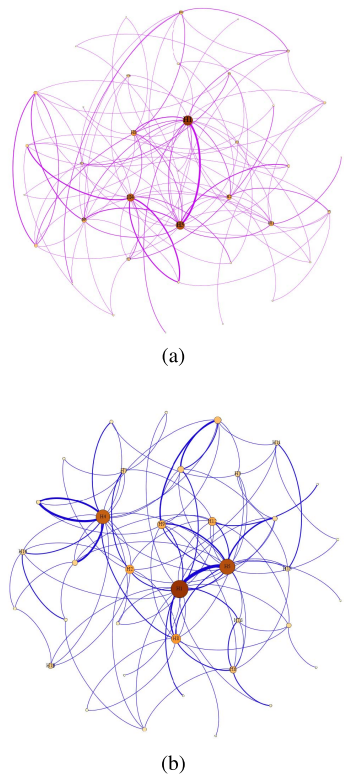


Fig. 15. Transfer network of hospitalizing behaviors. (a) Workday. (b) Weekend.

D. Transfer Analysis of Hospitalizing Behaviors

The transfer analysis is based on the trips, the origin and destination of which are both hospitals. The hospitals are regarded as the nodes, and the number of trips between them is the weight. The aim of transfer analysis is to study the trip number of transfer hospitals. The transfer network can be constructed, which can be described in the following equation:

$$\mathbf{Net_B} = (\boldsymbol{N_{hos}}, \boldsymbol{E}, \boldsymbol{ES}) \tag{15}$$

where $\boldsymbol{N}_{\text{hos}} = \{H1, H2, \ldots, H\text{n}\}$ is the set of nodes in the records, which is composed by hospitals, n is the number of nodes, $\boldsymbol{E} = \{e_{\text{ij}} \| i, j \in \boldsymbol{N_{Hos}}\}$ is the set of edges, which means a trip leave from location hospital i to hospital j, $e_{\text{ij}} \neq e_{ji}$, and $\boldsymbol{ES} : \boldsymbol{E} \rightarrow \boldsymbol{S}$ is the mapping function from an edge to the trip number.

The transfer network of hospitalizing behaviors can be seen in Fig. 15.

From Fig. 15, we can see that the transfer number from H1 to H5 is larger than other transfers. Besides, H4 also has a larger transfer between other hospitals. To explore the internal relationship, the distribution of common doctors between two hospitals is constructed, which can be seen in Fig. 16.

The labeled nodes are the top 15 hospitals listed in Table I. The size of nodes represents the number of total common doctors among the hospitals. Intuitively, H4, H9, H10, and H12 have common doctors with many hospitals. H1 and H2 are shown in the lower right in Fig. 16 and both of them are small, but they all have common doctors with other hospitals. To analyze the relationship between the number of common

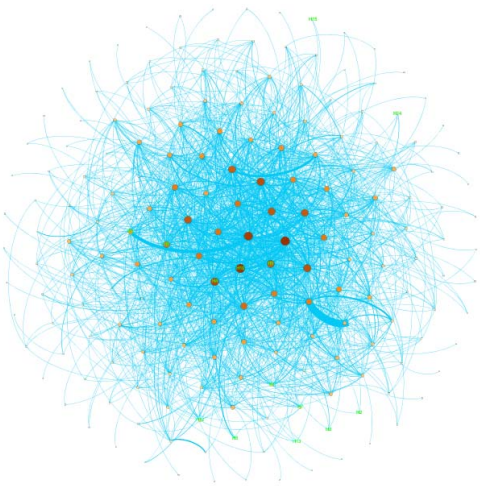


Fig. 16. Distribution of common doctors among hospitals.

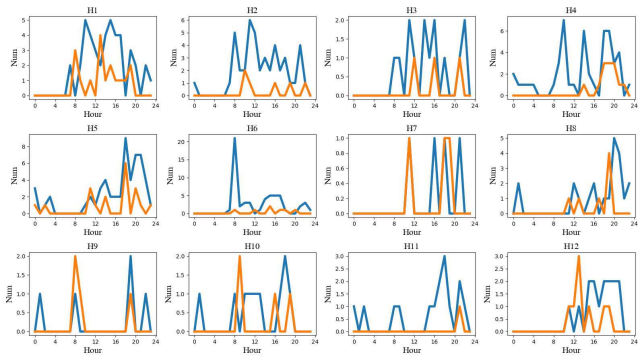


Fig. 17. Transfer of hospitalizing behaviors in time series. The blue and orange lines represent the transfer trips under the workday and weekend modes, respectively.

doctors and the transfer number for hospitalizing behaviors, the transfer analysis is essential.

The transfer number of trips of hospitalizing behaviors in time series under the M_p mode mode can be shown in the following equation:

$$N(t_h, H_i)_{\text{tran}}^{M_p} = \sum_{x \in DS, y \in H_i, t_{\text{OFF}} \in (t_h, M_P)} N(x, y, t_{\text{ON}}, t_{\text{OFF}}) \quad (16)$$

where $N(t_h, H_i)_{\text{tran}}^{M_p}$ denotes the transfer number of trips for hospitalizing behaviors in network Net_B in time series under the M_p mode.

The transfer results in time series can be seen in Fig. 17.

From Fig. 17, we can see that the transfer number of trips for hospitalizing behaviors is not very much. The transfer number of H1 and H2 is larger than that of other hospitals, which is mainly caused by their praise degree. For H1 and H2, the transfer trips occur in various time under the workday and weekend modes, and the number under the workday mode is significantly larger than that under the weekend mode. Besides, the transfer analysis generally happens in commuting time and at night, especially concentrated around 8:00 and 20:00. According to the results shown in Fig. 17, we should pay more attention to some hospitals, for example, H1, H2, and H6, because there are more transfer trips to these hospitals. On the whole, the transfer analysis is not explicit, which is mainly caused by the small data size for transfer analysis of hospitalizing behaviors. In spite of this, the transfer number of H1 and H2 is larger, which can further validate the popularity of two hospitals, while the transfer number of H4, H9, and H10 is not explicit. Combining with Fig. 16, we can see that the number of common doctors is not proportional to the transfer numbers, but it is relevant to common doctors.

However, there are some shortcomings that are worth improving in the following.

1) There are many ways for residents to travel, such as private cars, public transit, bicycles, subway, and so on. There are some limitations in analyzing the hospitalizing behavior distribution only by the data obtained from online car-hailing service. It is more comprehensive to combine other kinds of data to conduct the hospitalizing behaviors' analysis.
2) The traditional investigation method of travel has many shortcomings, for example, low efficiency. However, the collected information is more comprehensive, including the way travel, occupation, age, and class. Therefore, we can consider the combination of traditional survey methods to conduct a more comprehensive analysis in the next study.

V. Conclusion

The analysis of hospitalizing behaviors based on the big trajectory data in temporal and spatial lays the foundation for urban traffic management departments. In addition, the analysis results can provide references for hospitals to reasonably allocate infrastructures, and residents and taxi drivers to make full use of their time. In this paper, some research and analysis of the temporal and spatial distribution of hospitalizing behaviors are conducted and the traffic demands of hospitalizing behaviors are analyzed. Based on the rank analysis, the network analysis is also conducted to study the travel regularity of hospitalizing behaviors.

Considering the reasonable analysis of the hospitalizing behavior distribution based on the big trajectory data in temporal and spatial, we will explore the hospitalizing behavior distribution based on multi-source data fusion method in our next study.

VI. Acknowledgment

The trajectory data are retrieved from the Didi Chuxing of China.

References

[1] L. Lin, Q. Wang, and A. Sadek, "Data mining and complex network algorithms for traffic accident analysis," *Transp. Res. Rec.*, vol. 2460, no. 1, pp. 128–136, Jan. 2014.
[2] F. Xia, A. Rahim, X. Kong, M. Wang, Y. Cai, and J. Wang, "Modeling and analysis of large-scale urban mobility for green transportation," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1469–1481, Apr. 2018.
[3] S. Liu, Y. Liu, L. Ni, M. Li, and J. Fan, "Detecting crowdedness spot in city transportation," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1527–1539, May 2013.

[4] X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," *Future Generat. Comput. Syst.*, vol. 61, pp. 97–107, Aug. 2016.

[5] E. D'Andrea and F. Marcelloni, "Detection of traffic congestion and incidents from GPS trace analysis," *Expert Syst. Appl.*, vol. 73, pp. 43–56, May 2017.

[6] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. van de Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2159–2168, Dec. 2013.

[7] L. Liao, X. Jiang, F. Zou, L. I. Luming, and H. Lai, "A fast method of fcd trajectory data clustering based on the directed density," *J. Geo Inf. Sci.*, vol. 17, no. 10, pp. 1152–1161, 2015.

[8] L. Gong, X. Liu, L. Wu, and Y. Liu, "Inferring trip purposes and uncovering travel patterns from taxi trajectory data," *Cartogr. Geogr. Inf. Sci.*, vol. 43, no. 2, pp. 103–114, Mar. 2016.

[9] C. Wan, Y. Zhu, J. Yu, and Y. Shen, "SMOPAT: Mining semantic mobility patterns from trajectories of private vehicles," *Inf. Sci.*, vol. 429, pp. 12–25, Mar. 2018.

[10] M. Chen, Y. U. Xiaohui, and Y. Liu, "Mining mobility patterns based on deep representation model," *J. Comput. Appl.*, vol. 36, no. 1, pp. 33–38, 2016.

[11] K. Robert, T. Dennis, and E. Thomas, "Semantic enrichment of movement behavior with foursquare—A visual analytics approach," *IEEE Trans. Visualizat. Comput. Graph.*, vol. 21, no. 8, pp. 903–915, Aug. 2015.

[12] L. Sun and K. W. Axhausen, "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transp. Res. B, Methodol.*, vol. 91, pp. 511–524, Sep. 2016.

[13] Q. Xuan, H. Fang, C. Fu, and V. Filkov, "Temporal motifs reveal collaboration patterns in online task-oriented networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 91, no. 5, p. 052813, 2015.

[14] C. Fu *et al.*, "Link weight prediction using supervised learning methods and its application to yelp layered network," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 8, pp. 1507–1518, Aug. 2018.

[15] H. Huang, D. Bucher, J. Kissling, R. Weibel, and M. Raubal, "Multimodal route planning with public transport and carpooling," *IEEE Trans. Intell. Transp. Syst.*, to be published.

[16] W. Sheng, Z. Bao, J. S. Culpepper, T. Sellis, and C. Gao, "Reverse *k* nearest neighbor search over trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 4, pp. 757–771, Apr. 2017.

[17] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 258–276, Dec. 2018.

[18] X. J. Kong *et al.*, "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3874–3886, May 2018.

[19] R. A. Stegmann, I. V. Z. E. liobait, T. Tolvanen, J. Hollmén, and J. Read, "A survey of evaluation methods for personal route and destination prediction from mobility traces," *Wiley Interdiscipl. Rev. Data Mining Knowl. Discovery*, vol. 8, no. 2, 2018, Art. no. e1237.

[20] N. Lin and J. Li, "Floating car data mining and its application in path planning," *Comput. Eng. Design*, vol. 37, no. 7, pp. 1952–1957, 2016.

[21] J. Zhang, P. Qiu, X. U. Zhijie, and D. U. Mingyi, "A method to identify trip based on the mobile phone positioning," *J. Wuhan Univ. Technol.*, vol. 37, no. 5, pp. 934–938, 2013.

[22] L. Alexander, J. Shan, M. Murga, and M. C. González, "Origin-destination trips by purpose and time of day inferred from mobile phone data," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 240–250, Sep. 2015.

[23] Y. Zheng, "Trajectory data mining: An overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, p. 29, May 2015.

[24] R. Narayana, K. Pallav, J. Aayush, S. S. Arkatkar, and J. Gaurang, "Application of trajectory data for investigating vehicle behavior in mixed traffic environment," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2672, no. 43, pp. 122–133, 2018.

[25] X. Kong *et al.*, "Big trajectory data: A survey of applications and services," *IEEE Access*, vol. 6, pp. 58295–58306, 2018.

[26] F. Xia, J. Wang, X. Kong, Z. Wang, J. Li, and C. Liu, "Exploring human mobility patterns in urban scenarios: A trajectory data perspective," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 142–149, Mar. 2018.

[27] Z. Shan, W. Hao, W. Sun, and B. Zheng, "COBWEB: A robust map update system using GPS trajectories," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2015, pp. 927–937.

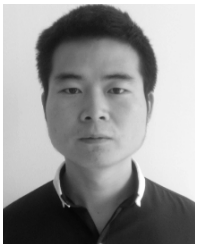
[28] G. Xiao, Z. Juan, and C. Zhang, "Travel mode detection based on GPS track data and Bayesian networks," *Comput., Environ. Urban Syst.*, vol. 54, pp. 14–22, Nov. 2015.

[29] L. Yang and Z. J. Zou, "Study of urban intersection lane structure based on floating car tracking data," *Sci. Surv. Mapp*, vol. 39, pp. 76–80, Aug. 2016.

[30] Y. Kong, C. Jin, and X. Wang, "Population flow analysis based on cellphone trajectory data," *J. Comput. Appl.*, vol. 36, no. 1, pp. 44–51, 2016.

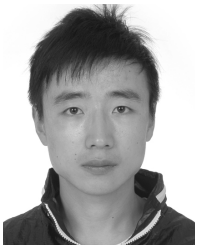
[31] R. Du, J. Huang, N. Zhong, and Z. Huang, "Stay point identification of mobile phone trajectory," *J. Frontiers Comput. Sci. Technol.*, vol. 8, no. 2, pp. 200–206, 2014.

[32] R. A. Manners, *Professional Dominance: The Social Structure of Medical Care*. Evanston, IL, USA: Routledge, 2017.



Yongdong Wang received the B.E. degree from the Nanyang Institute of Technology, Nanyang, Henan, China, in 2013. He is currently pursuing the Ph.D. degrees with the College of Information Engineering, Institute of Information Processing and Automation, Zhejiang University of Technology, Hangzhou, China.

His current research interests include intelligent traffic and control engineering and theory.



Dongwei Xu received the B.E. and Ph.D. degrees from the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China, in 2008 and 2014, respectively.

He is currently a Lecturer with the College of Information Engineering, Institute of Information Processing and Automation, Zhejiang University of Technology, Hangzhou, China. His current research interests include intelligent transportation control, management, and traffic safety engineering.

Peng Peng received the B.E. degree from the Suzhou University of Science and Technology, Suzhou, China, in 2017. He is currently pursuing the M.E. degree with the College of Information Engineering, Institute of Information processing and Automation, Zhejiang University of Technology, Hangzhou, China.

His current research interests include intelligent traffic, and control engineering and theory.

Qi Xuan (M'18) received the B.S. and Ph.D. degrees in control theory and engineering from Zhejiang University, Hangzhou, China, in 2003 and 2008, respectively.

He was a Post-Doctoral Researcher with the Department of Information Science and Electronic Engineering, Zhejiang University, from 2008 to 2010, and a Research Assistant with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, in 2010 and 2017. From 2012 to 2014, he was a Post-Doctoral Researcher with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He is currently a Professor with the College of Information Engineering, Zhejiang University of Technology, Hangzhou. His current research interests include network-based algorithm design, social network data mining, social synchronization and consensus, reaction–diffusion network dynamics, machine learning, and computer vision.

Guijun Zhang is currently a Professor with the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. His current research interests include intelligent information processing, optimization theory, and algorithm design.