Wang et al. 2018. *Int. J. Vehicle Structures & Systems*, 10(2), 150-159
ISSN: 0975-3060 (Print), 0975-3540 (Online)
doi: 10.4273/ijvss.10.2.15

International Journal of Vehicle Structures & Systems
Available online at www.maftree.org/eja

Exploring Intra-Urban Travel Mobility using Large-Scale Taxi Global Positioning System Trajectories

Haixiao Wang[a], Fang Liu[a] and Jinjun Tang[b]

[a]*School of Energy and Transportation Engg., Inner Mongolia Agricultural University, Hohhot, China*
[b]*School of Traffic & Transportation Engg., Central South University, Changsha, China*
Corresponding Author, Email: jinjuntang@csu.edu.cn

ABSTRACT:

Using taxi GPS trajectories data is of very importance to explore Spatio-temporal features of human mobility in transportation designing and planning. The data were collected from taxi GPS devices in Harbin city during a week. The taxi trips are extracted from GPS data, and travel distance and time in occupied and vacant states are firstly used to investigate the human mobility. Then, the urban area is divided into 400 grids. Furthermore, travelling network corresponding to taxi trips are designed to further examine the dynamics of mobility, in which the grid are considered as nodes and edge weights are defined as total number of trips among nodes. We observe some basic statistical features of network: degree, edge weights, clustering coefficients and network structure entropy. We also use the correlation between strength and degree to analyze the significance of nodes. Based on network analysis, we select two grids, a central business district and a residential district with high degree and strength, to study the spatial and temporal properties of trips that start from and end at these two grids. Finally, the correlation between trip volume and operation efficiency is explored and we find that hourly trip volume express negative correlation with operation efficiency.

KEYWORDS:

Urban mobility; Taxi GPS trajectories; Travel time; Travel network; Spatial-temporal property

CITATION:

H. Wang, F. Liu and J. Tang. 2018. Exploring Intra-Urban Travel Mobility using Large-Scale Taxi Global Positioning System Trajectories, *Int. J. Vehicle Structures & Systems*, 10(2), 150-159. doi:10.4273/ijvss.10.2.15.

1. Introduction

Urban travel behavior reflects general moving regularity of citizens, and it is an important factor to evaluate the rationality of city structure and planning. Thus, deep exploring traveling characteristics is significant to improve urban planning and enhance people living quality. From the aspect of transportation, analyzing the patterns of travel behavior can help researchers and agency to understand OD (Origin and Destination) distribution in urban city and it is also beneficial to road network planning, public transit planning and traffic management. Researchers from different research fields implement applications to understand human mobility, such as spread of diseases [1], city planning [2], traffic engineering [3], financial market forecasting [4], and economic well-being [5]. In all these applications, various data sources are used to analyze human mobility including e-mail [6-7], network traces [8], GPS data from floating cars [9-10], mobile phone data [11-12], banking notes [13], social media check-in data [14] and smart subway fare card [15].

Residents' travel rely on a certain transportation tool. In the city, taxis undertake large part of travel because of its flexibility. For its convenient, comfortable and fast, the taxi is frequently treated as an ideal traffic tool to complete long-distance travel in urban area. Meanwhile, in order to effectively supervise real-time operation status of taxi vehicles, GPS devices are equipped to collect taxi temporal and spatial information. This information provides us rich data resources to discover spatial-temporal patterns of travel behavior, hot spots analysis and even OD distribution. Thus, comparing with traditional data collection approaches, such as questionnaire or artificial counting, taxi GPS and status data can accurately reflect traveling features of passengers. Recently, abundant researches are conducted to study urban human mobility based on taxi GPS trajectories [16-25]. Liu et al [16] reported the distance distribution follow the truncated power-law function using mobile phone data and taxi GPS data, respectively.

Liang [17] explored an exponential function to fit the displacements of taxi trajectories. Qian et al [18] studied urban dynamics using the taxi GPS data in New York City, and they contributed two findings as follows: applied two-step clustering method to recover trip pattern; used an exponential distribution to fit trips patterns. Qi et al [19] explored the relationship between the taxi passengers travelling characteristics and social function. Liu et al [20] analyze city spatial structure features of Shanghai city using taxi data. Castro et al [21] explored human mobility from taxi location information. Through integrating mobile phone and taxi data together, Kang et al [22] explored human movements in Singapore. They also compared the difference between human mobility patterns from two data source. Sagarra [23] proposed a super sampling method to estimate mobility based on an entropy

maximization procedure using taxi data in the New York City. They also used a network-based metrics to predict vehicle flows.

Liu et al [24] analyzed urban spatial travelling behavior using GPS traces from 3000 taxis. Based on taxi GPS traces in the Lisbon City, Veloso et al [25] analyzed spatio-temporal variation of taxi services and behaviours. The purpose of this study is to analyze human mobility using taxi GPS data collected in Harbin city. The travel distance and time of trips are firstly used to explore the human mobility. Then, we build travel networks corresponding to taxi trips on different days, in which the girds are treated as nodes, and trips between grids are regarded as edges. Furthermore, some statistical quantities are calculated to uncover the dynamics of human mobility and hot spots in urban city. Finally, we further discuss spatial distribution of trips in several specific hot spots.

2. Data description

2.1. Data source

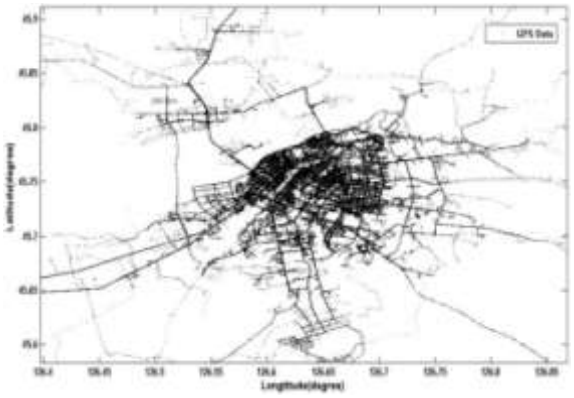
The total number of taxis was about thirteen thousands of Harbin fleet in 2012, and there were around one hundred taxi companies in Harbin city. The daily average travel per taxi reaches to approximately 330 kilometres. The taxi GPS data were collected in a week from 1st to 7th August in 2012. The data are recorded at a rate of 30 seconds, and total samples are 2880 in one day. Each data sample includes 7 attributes: Taxi ID, time, latitude, longitude, speed, orientation and status, which are shown in Table 1.

Table 1: Data structures, Date 2012/8/1

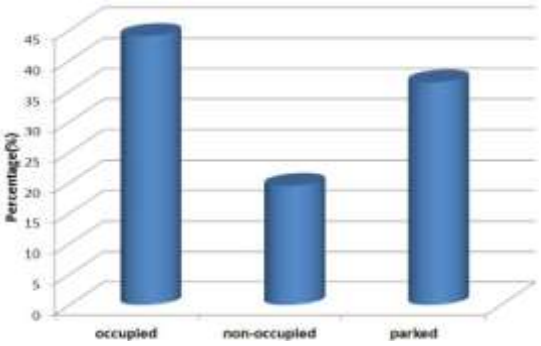
Taxi ID	Time	Latitude	Longitude
100300002	6:59:00	45.740090	126.619156
100300002	6:59:30	45.738384	126.616920
100300002	7:00:00	45.736588	126.614845
100300002	7:00:30	45.736557	126.614820
100300010	11:08:00	45.756557	126.604000
100300010	11:08:30	45.757168	126.604280
100300010	11:09:00	45.759000	126.605290
100300010	11:09:30	45.759712	126.605644
Taxi ID	Speed	Orientation	Status
100300002	38	112	0
100300002	35	109	0
100300002	29	110	0
100300002	30	70	0
100300010	45	10	1
100300010	40	8	1
100300010	33	12	1
100300010	37	13	1

2.2. Data analysis for travelling status and profit

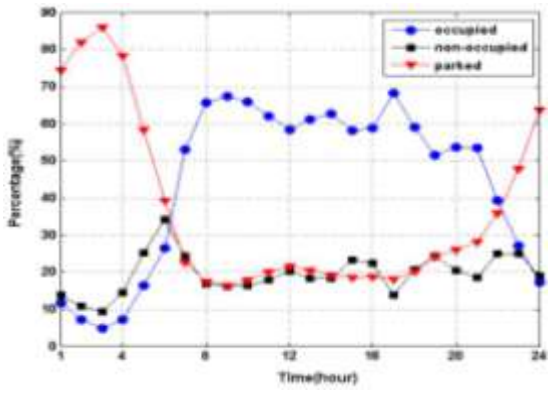
Fig. 1 shows the location data of thousand drivers on Aug. 1st, the trajectories almost cover the road network in Harbin. Aim to detect the interesting points, an important data should be utilized: status records, which record the taxi is vacant or occupied. That is the changes of status mean the drivers send a passenger to goal successfully or they just take a new passenger. This information is helpful to recognize hot spots. Fig. 1(a) shows the percentage of three statuses of all the drivers on Aug. 1st. The percentage of occupied by passengers reaches to about 45%, there are nearly 19% of time the drivers are roaming on the road to find passengers, and about 36% of time the vehicle is parked without driving data. Fig. 1(b) displays the variation of average percentage for three statuses in 24 hours. From 23:00 pm to 6:00 am, the percentage of parked is much higher than the other two statuses. In 3:00 am, its value reaches to the peak point, and percentages of non-occupied and parked drop to their lowest point. From 7:00 am to 22:00 pm, the percentage of occupied tends to higher than the other two statues.



(a): Trajectory distribution for around one thousand taxi vehicles in a day



(b): Percentage of three statuses

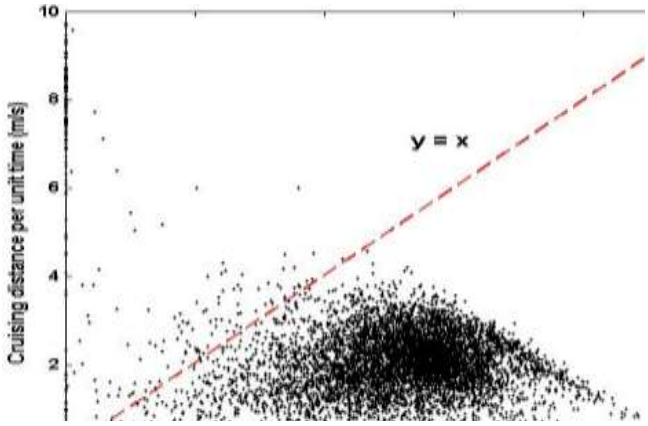


(c): Average percentage of three statuses in hours

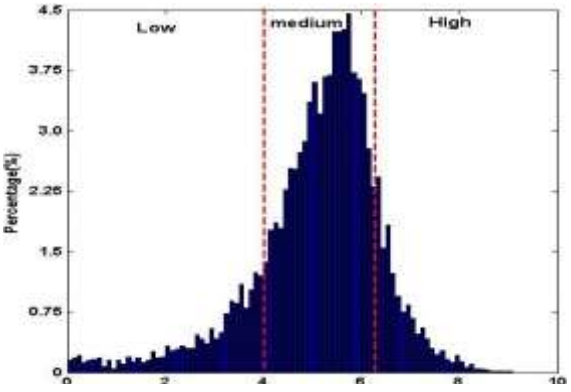
Fig. 1: Trajectory distribution and statistical comparison

Fig. 2(a) shows the density distribution between the cruising distance and fare distance per unit time on Aug. 1st. As we can see, most of drivers are profitable, see the hot kernel on the bottom. The fare distance per unit time are higher than that of the cruising distance, this means these drivers only cruise few to pick up next passenger.

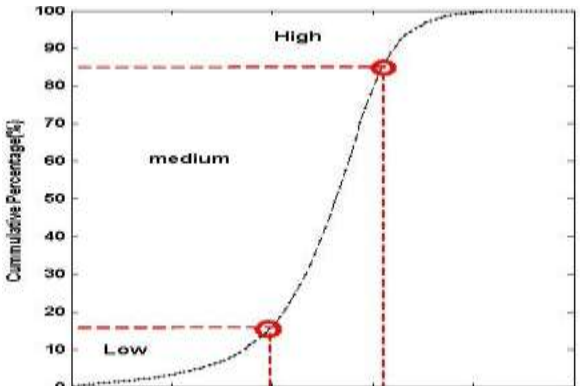
However, some driver's profit is low, they need cruise more than the majority. The last part of drivers who cruise less can earn more than the majority. These drivers are experienced. When they drop off a passenger, they directly head for the places with higher probability to pick up a passenger based on their experience rather than randomly cruise around drop-off locations. Fig. 2(b) presents the proportion distribution of drivers with different profit level. The cumulative percentage to classify all the drivers into low, medium and high profit is shown in Fig. 2(c). The top 15% are considered as high-profit-level drivers, the bottom 15% are considered as the low-profit-level drivers and the rests drivers are the medium-profit-level part. The fare distance per unit time of low part is lower than 4m/s, the profit range of the medium part starts from 4m/s to 6.2m/s, the profit of high parts is higher than 6.2m/s.



(a): Density scatter of profit between cruising and fared trips



(b): Distribution of profit



(c): Cumulative percentage of profit

Fig. 2: Density scatter and statistical distribution of profit

3. Human mobility based on taxi trajectories

Taxi trips are extracted from GPS dataset. As we aforementioned, there are two states in operation process of taxi drivers: load up passengers or occupied and vacant or non-occupied. For trips in different states, we calculate travel distance and time respectively. For the dataset in occupied state, $R^o = (k, l^o, \tau^o)$, it includes the location information l^o at time period τ for taxi k, and $l^o = (x^o, y^o)$ indicates longitude and latitude information. Similarly, $R^n = (k, l^n, \tau^n)$ and $l^n = (x^n, y^n)$ express the dataset in occupied state. The travel distance can be calculated as,

$$d = \sum_{i=1}^{N-1} |l_{i+1} - l_i| \tag{1}$$

Where N means the total no. of data samples in a unique trip on different states, "| |" indicates the Euclidean distance between two adjacent locations. The trip travel time can be defined as,

$$t = \tau_N - \tau_1 \tag{2}$$

Where N indicates the total no. of data samples in a unique trip on different states, τ_1 and τ_N represent the starting and ending time. In order to compare different distribution of travel time and distance in weekday and weekend, we eventually collect 31823 and 31828 trips on weekday (Aug. 1st) for occupied and vacant states. On weekend (Aug. 4th), we collect 33228 and 33232 trips for occupied and vacant states respectively.

3.1. Travel distance

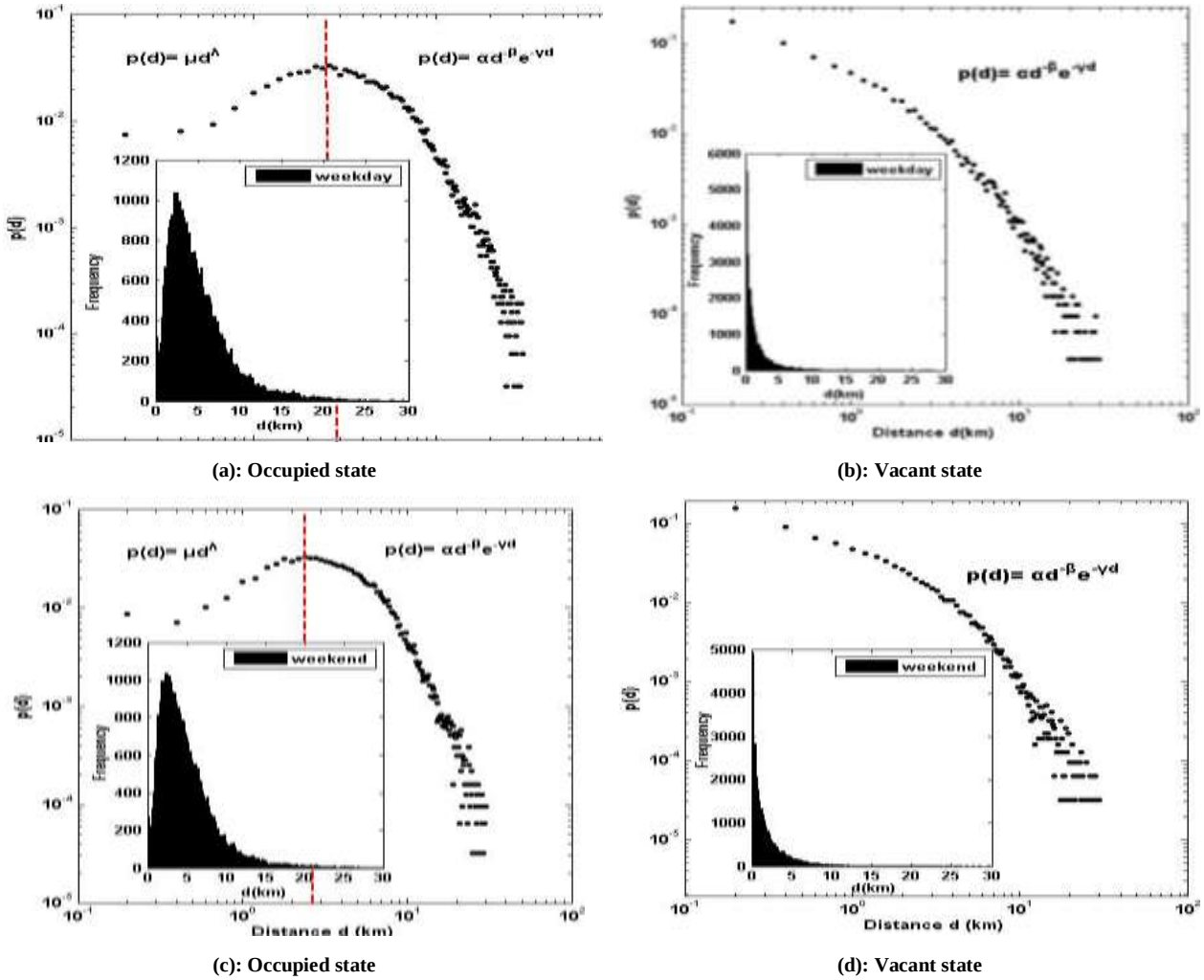
In this section, we calculate the frequency distribution of travel distance and fit its probability p(d) distribution under double logarithmic scale in Fig. 3. From the observation of the Fig., the trips collected from different days, weekday and weekend, express similar travel distance distribution. However, distance distribution in two states, occupied and non-occupied, exhibits obviously different patterns In Fig. 3a, the p(d) of occupied trips firstly increases gradually to the peak, and then it descends in the travel distance range from 3km to 30km. The distribution of p(d) can be fitted by a power-law function for the first part:

$$p(d) = \mu d^{\lambda} \tag{3}$$

Also, the second part can be fitted by truncated power-law function:

$$p(d) = \alpha d^{-\beta} e^{-\gamma d} \tag{4}$$

All the parameters and indicator to evaluate the fitting performance (coefficient of determination: R^2) are shown in Table 2. The p(d) of vacant trips decreases gradually in the travel distance range from 0 to 30km. We can observe the frequency value of short distance is relatively higher than that of long distance. It indicates the roaming area for the vacant taxi is limited in a small range to decrease travel distance or time for finding next passenger to maximize their profit. The distribution is fitted well by a truncated power-law function. The parameters are shown in Table 1. The similar results can be concluded in Figs. 3(c) & 3(d) and Table 2.



(a): Occupied state **(b): Vacant state**

(c): Occupied state **(d): Vacant state**

Fig. 3: Trip distance distribution and fitting results of trips

Table 2: Fitting parameters for travel distance and travel time distribution

Travel distance			
Status	Day of week	Fitting parameters	
		Power law	Truncated power law
Occupied	Weekday	μ =0.017;λ=0.797;R2=0.988;	α =0.128;β=0.553;γ=0.197;R2=0.966
	Weekend	μ=0.018;λ=0.693;R2=0.943;	α=0.175;β=0.751;γ=0.195;R2=0.962
		Truncated power law	
Non-occupied	Weekday	α=0.055; β=1.204; γ=0.134; R2=0.965	
	Weekend	α=0.062; β=1.155; γ=0.148; R2=0.953	
Travel time			
Status	Day of week	Fitting parameters	
		Power law	Truncated power law
Occupied	Weekday	μ=0.013;λ=0.746;R2=0.955;	α=2.701;β=1.536;γ=0.022;R2=0.985
	Weekend	μ=0.012;λ=0.812;R2=0.958;	α=6.108;β=1.848;γ=0.021;R2=0.978
		Truncated power law	
Non-Occupied	Weekday	α=0.312; β=1.251; γ=0.021; R2=0.973	
	Weekend	α=0.341; β=1.243; γ=0.022; R2=0.968	

3.2. Travel time

Compare with travel distance, travel time reflects actual traffic statues in road network. The frequency and probability distribution are computed with different states on weekday and weekend in Fig. 4. The p(t) of occupied trips firstly increases gradually to the peak, and then decreases in the travel time range from 9 to 100 min. For the 1st part of the trips, The distribution of p(d) can be fitted by a power-law function for the 1st part,

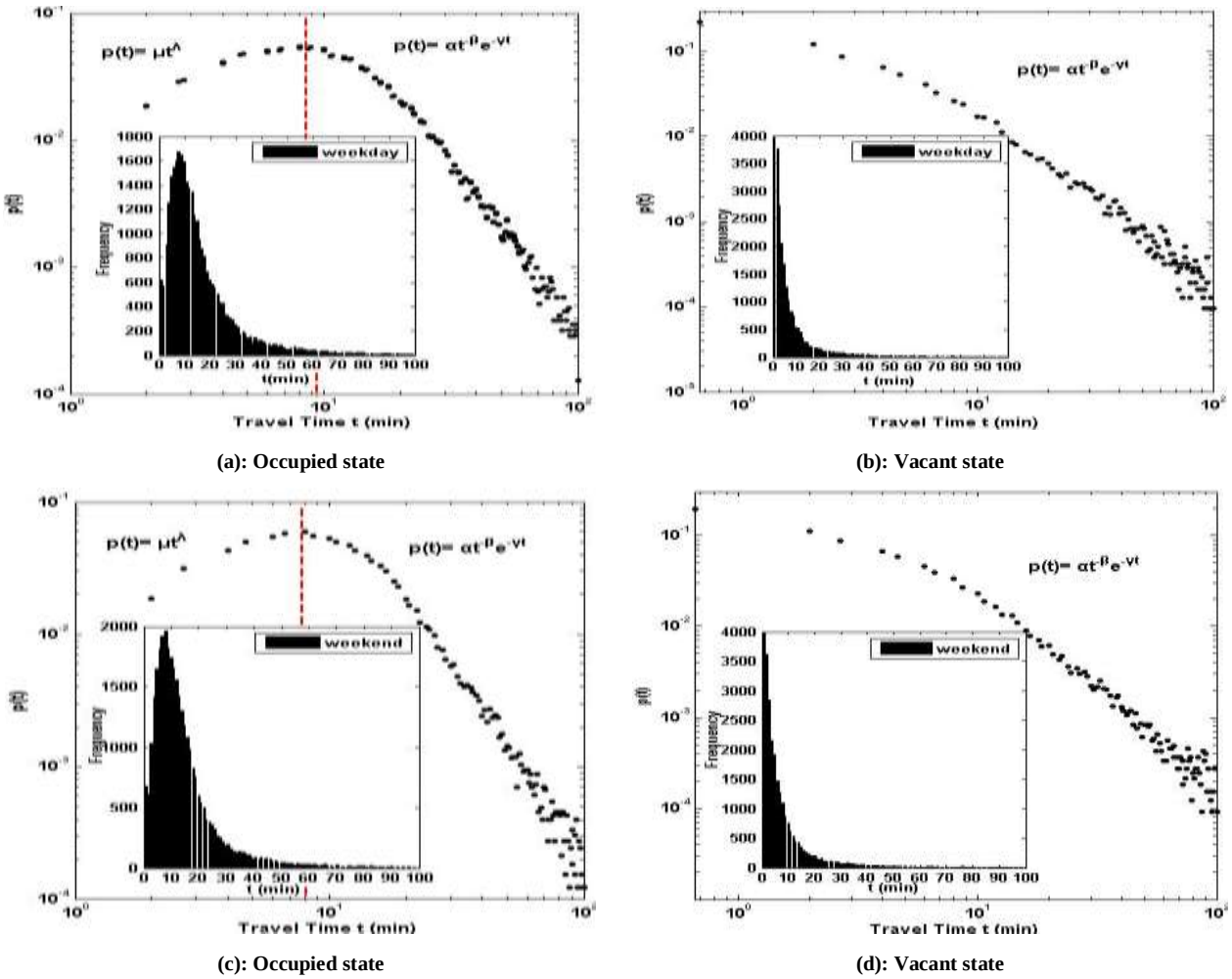
$$p(t) = \mu t^{\lambda} \tag{5}$$

The 2nd part can be fitted by truncated power-law function,

$$p(t) = \alpha t^{-\beta} e^{-\gamma t} \tag{6}$$

All fitting coefficients are displayed in Table 2. The p(t) of vacant trips decreases in the travel time range from 0

to 100min. The distribution is fitted by a power-law with exponential cut off. For the trips collected on weekend, the distribution and coefficients are shown in the Figs. 4c & 4d and Table 2. Different with the results in [13, 26, 27] reported that the travel distance can be fitted well by power-law function from cell phone communication data, our study shows travel distance and travel time of occupied trips express a hybrid patterns. The reason includes following two aspects: (1) The travelling space of taxi vehicle is limited by the size of city and the structure of road. (2) The economic factor is another important condition for the passengers. In summary, as a significant tools for traveling, taxi vehicle expresses its own advantages and disadvantages. . The trips extracted from taxi GPS traces behave unique characteristics to explore human mobility and improve transportation planning.



(a): Occupied state **(b): Vacant state**

(c): Occupied state **(d): Vacant state**

Fig. 4: Travel time distribution and fitting results of trips

4. Dynamics in travelling network

4.1. Network based on traveling trips

In this section, we design two travel network from the data recorded on weekday (Aug. 1st) and weekend (Aug. 4th). The main land of Harbin city is divided into 400 grids. Each grid covers the land area of 0.075° (longitude) × 0.005° (latitude). The grid is treated as node, and the trips connecting two nodes are treated as edges. Accordingly, we construct weekday travel network (WDTN) and weekend travel network (WETN) with both 400 nodes and 25006 and 25920 edges respectively. Some statistical properties are shown in Table 3. The degree is the no. of edges for a node connects with other nodes. From the observation of quantities in Table 3, we can see that WDTN and WETN express similar structure, and in the networks, a large amount of nodes only share small number of all the edges. Furthermore, the entropy is applied to evaluate the structural feature. The entropy is a central role in the information theory and be used to measure the uncertainty [28]. In term of the definition of NSE, when the value of E_N is small, the network has good connectivity. When the value of E_N is high, it means the network is divided into several local network and no. of nodes are not connected with the others. For the weekday, E_N = 0.831, and for the weekend E_N = 0.818. Furthermore, the average shortest path length of a network is defined as [29],

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij} \tag{7}$$

Where N means the total no. of nodes and d_{ij} represents the shortest path length between nodes i and j. The average shortest path length L is 2.389 for the WDTN.

For WETN, L is 2.527. A network with "small world" feature generally has a low average shortest path length and high clustering coefficients. However, two networks analyzed in this study don't show "small word" property. The reason is that a lot of nodes cannot be directly connected, for the low clustering coefficients, this results in the increase of values of average distance within the network. Moreover, the values of E_N for two corresponding networks are relatively high. This indicates the network is divided into different random local networks, and no. of nodes do not connect with the others. Thus, the high values of E_N for two networks also indicate that two networks do not express "small word" features and the shortest distance among nodes become long, even sometimes be unconnected.

Table 3: Statistical results of travel network for two days

Day of week	Statistical indicators	Degree	Edge weight	Clustering coefficients
Weekday	Mean	125.03	0.3408	0.3919
	Standard deviation	101.99	1.5465	0.1663
Weekend	Mean	129.60	0.3617	0.3990
	Standard deviation	104.79	1.5677	0.1788

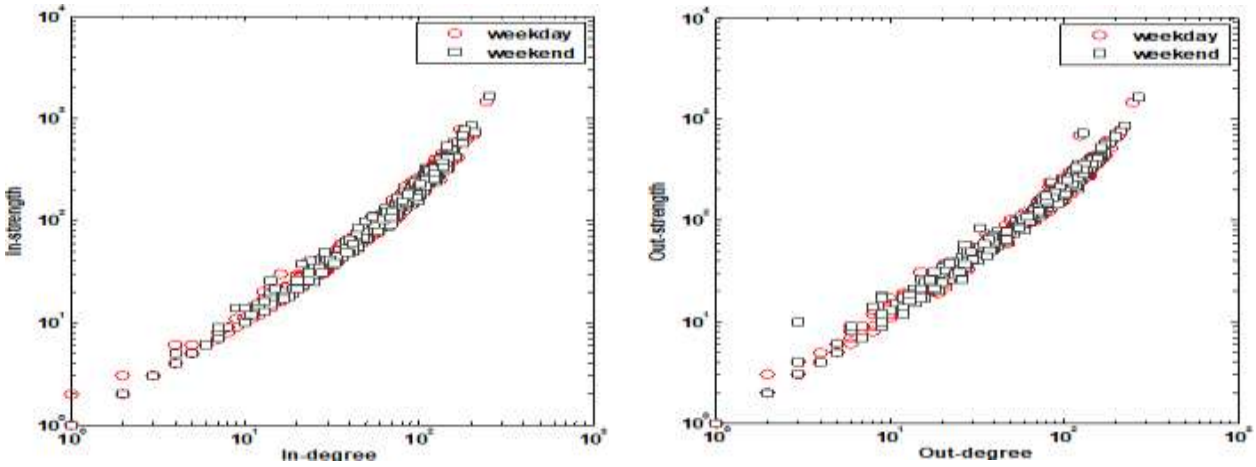
4.2. Correlation between the strength and degree

The strength of node i is defined as the sum for weights of edges which connect with it. As the networks designed in this study are directed and weighted, the strength can be further categorized into in-strength and out-strength. In-strength is the sum for weight of edges which link the node as destination, and out-strength is the sum for weight of edges which link the node as origin. They are denoted as:

$$s_i^{in} = \sum_{j=1}^{n} a_{ji} \cdot w_{ji} \tag{8}$$

$$s_i^{out} = \sum_{j=1}^{n} a_{ij} \cdot w_{ij} \tag{9}$$

Where n means the no. of the nodes, w_{ij} expresses the weight of edge starts from node i to j, a is the adjacent matrix. Similarly, the degree can be also classified into the in-degree and out-degree in a network. In the complex network, the strength and degree are used to evaluate the significance of node from intensity and extent. The Fig. 5 shows the degree-strength correlation of two networks. The correlation between in-degree and in-strength, and the correlation between out-degree and out-strength both can be fitted by power law distribution. This results indicate the grids in urban city with wide communications express an important role in the networks.



(a): Correlation between in-degree and in-strength

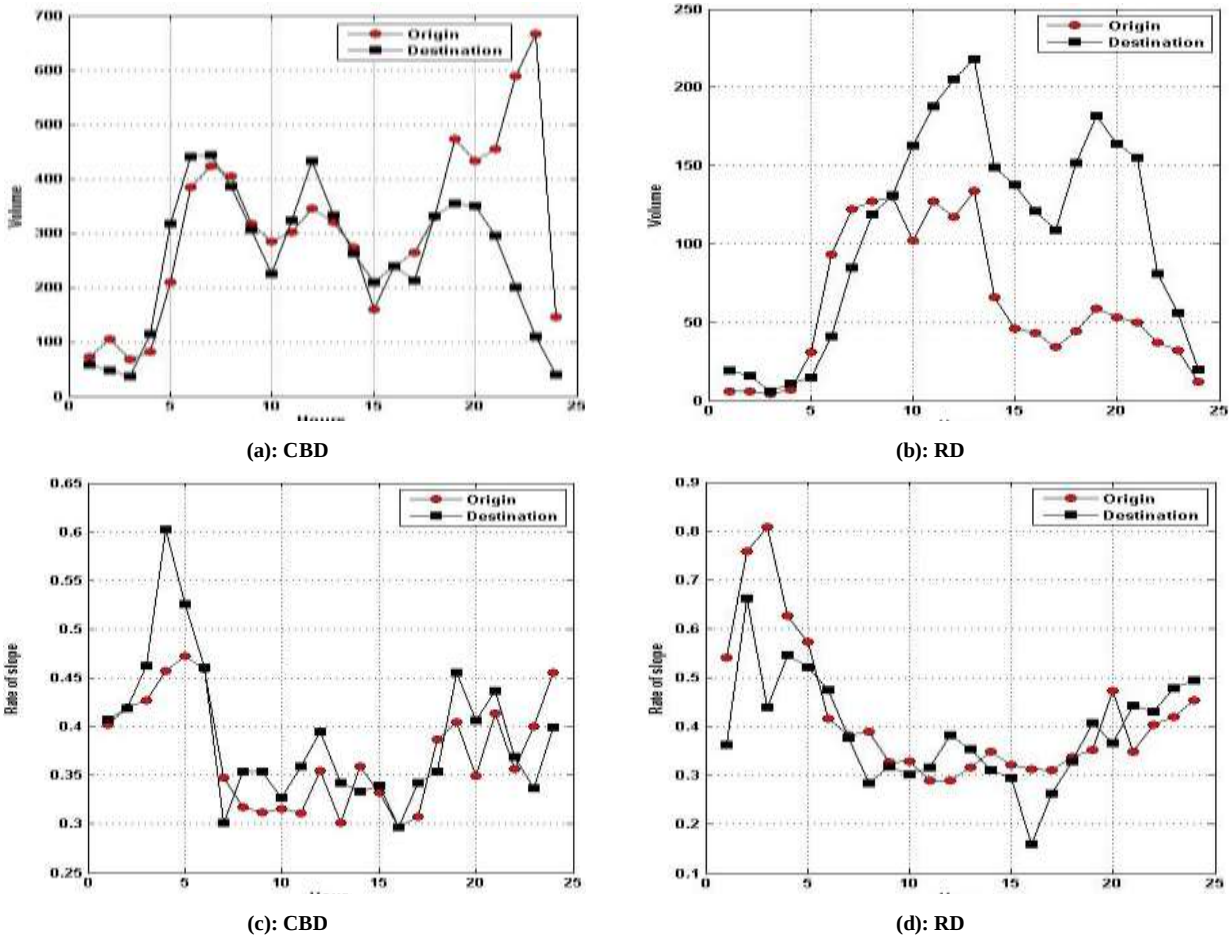
(b): Correlation between out-degree and out-strength

Fig. 5: Power-law distributions for the correlation between degree and strength

5. Trips distribution for hot spots

From the network-based analysis in above sections, we find that some nodes or grids play key role in transportation system. In this section, we focus on spatial and temporal property of trips that generated from and attracted to these grids. We choose two grids or nodes with high degree and strength as cases study. The first one is a Central Business District (CBD), and the second one is a Residential District (RD) in Harbin city (longitude from 126.63° to 126.6375° and latitude from 45.7775° to 45.78°). Figs. 6a & 6b shows the variation of trip volume in 24 hours during a week from Aug. 1 to Aug. 7 in 2012. The origin and destination in Fig. mean trips generated from and attracted to the selected grids. In business district, the volumes of origin and destination trips have similar distribution patterns, but after 20 pm, the volume of origin trips are significantly higher than that of destination trips, which indicates people leave the shopping center before it is closed. For the residential area, the no. of destination trips is obviously higher than that of origin trips during 10 to 12 am and 17 to 20 pm, as people generally come back home rather than travel outside in this time period. Figs. 6c & 6d shows hourly variation of operational efficiency in two grids. We define the rate of slope (r = distance/time) to characterize the operational efficiency of taxi vehicles, namely the traveling distance per unit time. Thus, the high value of rate r means high operational efficiency. From 1 am to 6

am, due to the traffic system be in a good condition, the efficiency of the taxi is significantly higher, then, with the increasing of traffic flow, the efficiency declines sharply and turns to be stable around r = 0.3. After 20 pm, the efficiency increases again to a high level with relief of traffic congestion.



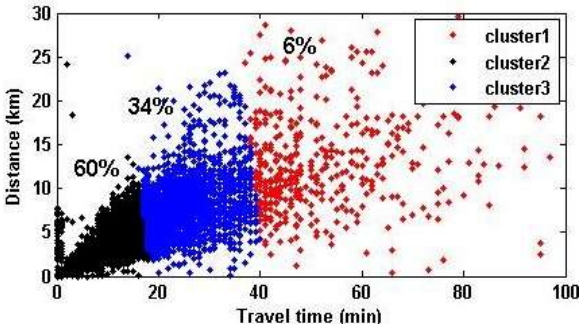
(a): CBD **(b): RD**

(c): CBD **(d): RD**

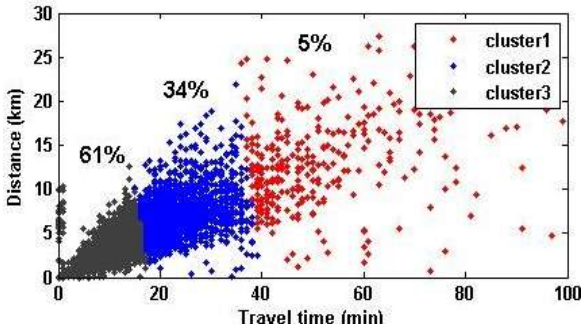
Fig. 6: Trips volume and operational efficiency distribution of two grids in 24 hours

Fig. 7 shows the correlation of the travel time and distance in two grids. We can see that the distance and travel time exhibit positive correlation, that is, as distance of trips increases, the trips travel time accordingly lengthens. Moreover, we adopt K-means algorithm to divide the samples into three clusters, and the nos. in the Fig. is the proportion of different categories. As we can see, more than 60% trips are short distance travel cluster, there are about 1/3 of all trips belong to the medium distance travel cluster, and less than 10% trips belong to long distance travel part. Fig. 8 provides the spatial distribution of three clusters calculated from Fig. 7 in urban city area, and it also represents that the taxi, as an important transportation tool for urban travel, its traveling range is mainly limited in urban area. In the Fig., grid (9, 8) represents the CBD and grid (5, 9) means the RD in the study. In order to further evaluate the efficiency of taxi operation in time of day, we display the correlation between distance and travel time of grid (5, 9) or RD at different hours (here, we just show 8 hours in a day) in Fig. 9. We also define the rate of slope (r = distance/time) to characterize the operational efficiency of taxi vehicles, namely the traveling distance per unit time. Thus, the high value of rate r means high operational efficiency.

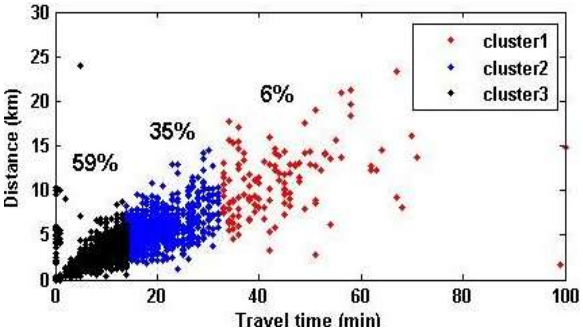
In order to explore the correlation between trip volume and operation efficiency, we scatter 96 samples in the Fig. 10. (For each grid, we can obtain 24 hourly samples from origin or destination trips) due to the different no. of volume (shown in Fig. 6) in two grids, we normalize the trip volume into the range of [0, 1]. As we can see, these two variables display negative relationship. Most of trips efficiency distribute in the range of [0.3, 0.45]. Trips with low volume have high efficiency, as the volume increases, the efficiency is gradually reduced. From Fig. 6, we can see the distribution of trip volume and operation efficiency both express hourly variation patterns in 24 hours. Therefore, during peak hours, high traveling demand of citizens causes increase of taxi trip volume. Meanwhile, as traffic system is under the condition of high pressure during peak hours, the operation efficiency of taxi will be affected by variation of traffic states and tends to be in a low level. During the non-peak hours, low traffic pressure and low traveling demand will result in the decrease of taxi trip volume and enhancement of operation efficiency.



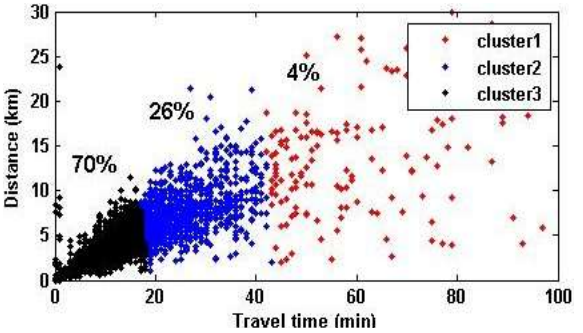
(a): Trips generated from CBD



(b): Trips attracted to CBD

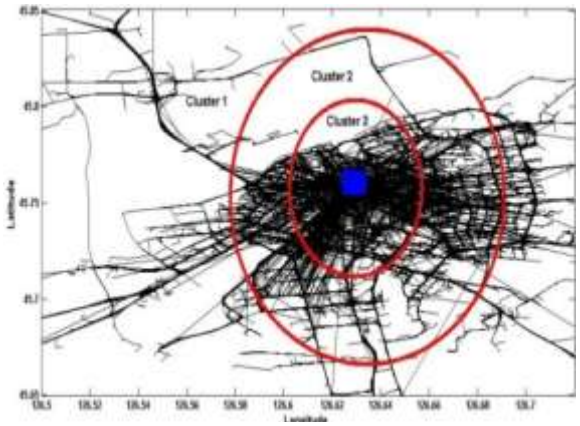


(c): Trips generated from RD

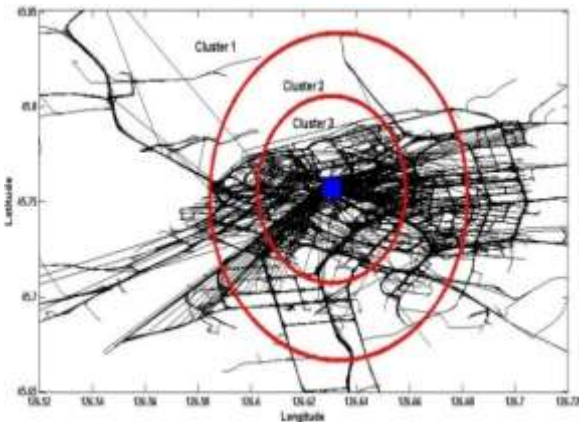


(d): Trips attracted to RD

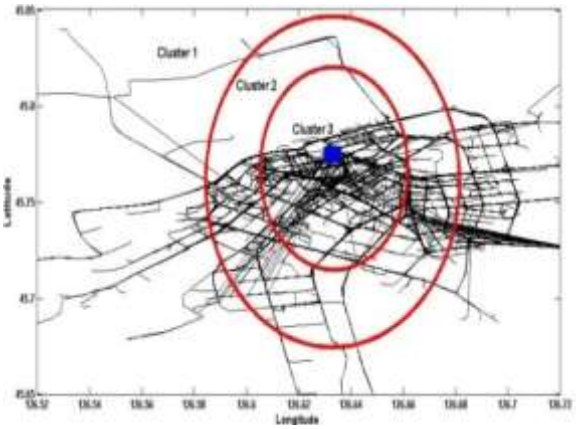
Fig. 7: Correlation between travel time and distance of trips in two districts



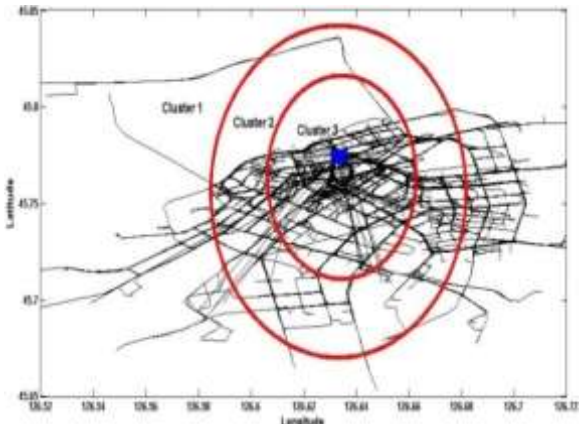
(a): Trips generated from grid (9, 8) or CBD



(b): Trips attracted to grid (9, 8) or CBD



(c): Trips generated from grid (5, 9) or RD



(d): Trips attracted to grid (5, 9) or RD

Fig. 8: Spatial distributions of trips in two transportation zones

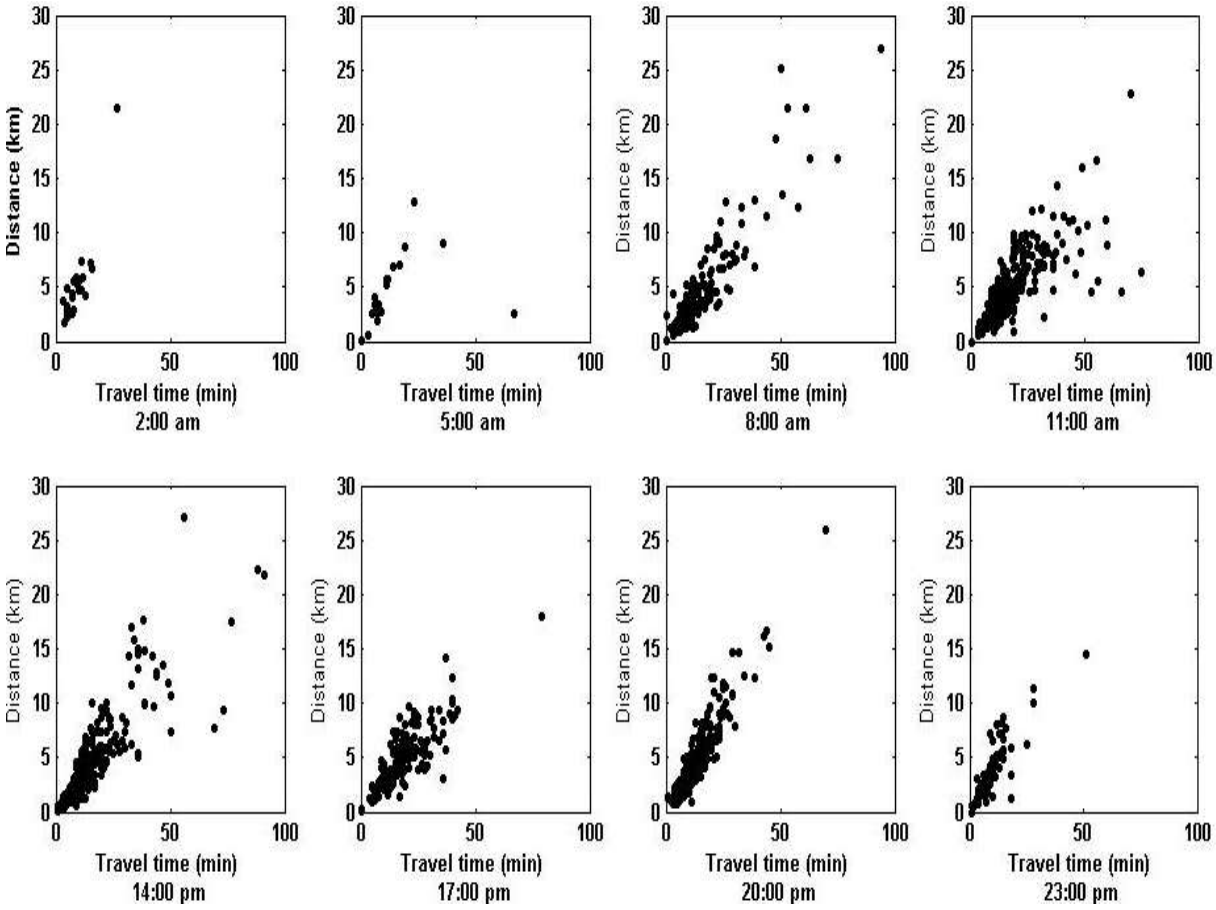


Fig. 9: Correlation between travel time and distance of trips in grid (5, 9) or RD for eight hours

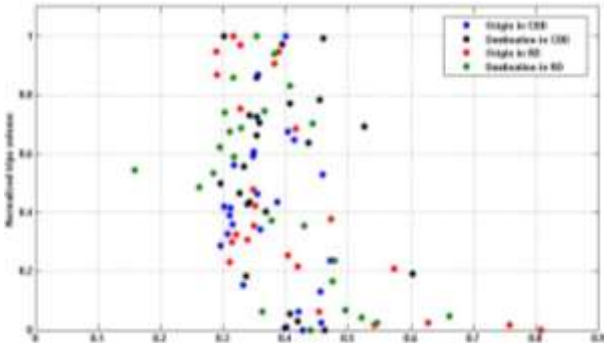


Fig. 10: Correlation between normalized trips volume and operational efficiency for two districts

6. Conclusions

In this paper, we analyze the human mobility using taxi GPS data. The distribution of taxi trips under occupied status express two patterns: increasing part and decreasing part. The curve of increasing part can be well fitted by power law function, and distribution of in decreasing part is fitted by truncated power law function. Furthermore, the city area is divided into 400 grids. Accordingly, we design two travelling network based on taxi traces recorded on weekday and weekend: weekday travel network (WDTN) and weekend travel network (WETN). By observing some basic statistical quantities, we find that two networks express similar structure. In order to analyze the importance of nodes in networks, the correlation between strength and degree is examined, and we find that the in-strength and out-strength of nodes have similar increasing pattern with the in-degree and out-degree. The structure information entropy for two networks indicates that the travel network have no small word property and the shortest distance between nodes become long, even sometimes be unconnected. Finally, taking two grids with high degree and strength as example, Central Business District (CBD) and Residential District (RD), we analyze the spatial and temporal property of trips that generates from and attracts to these grids. We find that trip volume express negative correlation with operation efficiency.

ACKNOWLEDGEMENTS:

This research was financially supported by the Foundation of Central South University (Grant No. 502045002) and National Natural Science Foundation of China (Grant No. 71403068 and 61573009).

REFERENCES:

[1] V. Colizza, A. Barrat, M. Barthélémy, A.J. Valleron and A. Vespignani. 2007. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions, *PLoS Med.*, 4, 95-110. https://doi.org/10.1371/journal.pmed.0040013.

[2] H.D. Rozenfeld, D. Rybski, A.J. Jr, M. Batty, H.E. Stanley and H.A. Makse. 2008. Laws of population growth, *Proc. Natl Acad. Sci.*, USA, 105, 18702-18707. https://doi.org/10.1073/pnas.0807435105.

[3] P. Wang, T. Hunter, A.M. Bayen and K. Schechtner. 2012. Understanding road usage patterns in urban areas, *Scientific Reports*, 2(12), 1001. https://doi.org/10.1038/srep01001.

[4] X. Gabaix, P. Gopikrishnan, V. Plerou and H.E. Stanley. 2003. A theory of power-law distributions in financial market fluctuations, *Nature*, 423, 267-270. https://doi.org/10.1038/nature01624.

[5] L. Pappalardo, Z. Smoreda, D. Pedreschi and F.L. Giannotti. 2015. Using big data to study the link between human mobility and socio-economic development, *IEEE Int. conf. Big Data*, Santa Clara, CA, USA, 871-878.

[6] G. Kossinets and D.J. Watts. 2006. Empirical analysis of an evolving social network, *Science*, 311, 88-90. https://doi.org/10.1126/science.1116869.

[7] H. Ebel, L.I. Mielsch and S. Bornholdt. 2002. Scale-free topology of e-mail networks, *Phys. Rev. E*, 66, 035103. https://doi.org/10.1103/PhysRevE.66.035103.

[8] M. Kim, D. Kotz and S. Kim. 2006. Extracting a mobility model from real user traces, *25th IEEE Int. Conf. Computer Communications*, 1-13.

[9] Y. Zheng, Q. Li, Y. Chen, X. Xie and W.Y. Ma. 2008. Understanding mobility based on GPS data, *10th Int. Conf. Ubiquitous Computing*, Seoul, Korea, 312-321.

[10] A. Bazzani, B. Giorgini, S. Rambaldi, R. Gallotti and L. Giovannini. 2010. Statistical laws in urban mobility from microscopic GPS data in the area of Florence, *J. Stat. Mech.*, P05001. https://doi.org/10.1088/1742-5468/2010/05/P05001.

[11] F. Calabrese, M. Diao, G.D. Lorenzo, J. Ferreira and C. Ratti. 2013. Understanding individual mobility patterns from urban sensing data: a mobile phone trace example, *Transport. Res. C*, 26(1), 301-313. https://doi.org/10.1016/j.trc.2012.09.009.

[12] C. Song, Z. Qu, N. Blumm and A.L. Barabási. 2010. Limits of predictability in human mobility, *Science*, 327, 1018-1021. https://doi.org/10.1126/science.1177170.

[13] D. Brockmann, L. Hufnagel and T. Geisel. 2006. The scaling laws of human travel, *Nature*, 439, 462-465. https://doi.org/10.1038/nature04292.

[14] L. Wu, Y. Zhi and Z. Sui. 2014. Intra-urban human mobility and activity transition: evidence from social media check-in data, *PLoS One*, 9(5), e97010. https://doi.org/10.1371/journal.pone.0097010.

[15] S. Hasan, C.M. Schneider and S.V. Ukkusuri. 2013. Spatiotemporal patterns of urban human mobility, *J. Stat. Phys.*, 151(2), 304-318. https://doi.org/10.1007/s10955-012-0645-0.

[16] Y. Liu, C. Kang, S. Gao, Y. Xiao and Y. Tian. 2012. Understanding characteristics of intra-urban trips using taxi trajectory data, *J. Geogr. Syst.*, 14(4), 463-483. https://doi.org/10.1007/s10109-012-0166-z.

[17] X. Liang, X. Zheng, W. Lu, T. Zhu and K. Xu. 2012. The scaling of human mobility by taxis is exponential, *Physics A*, 391, 2135-2144. https://doi.org/10.1016/j.physa.2011.11.035.

[18] X. Qian, X. Zhan and S.V. Ukkusuri. 2013. Characterizing urban dynamics using large scale taxicab data, *Transport. Res. Rec.*, 35(36), 17-32.

[19] G. Qi, X. Li, S. Li, G. Pan and Z. Wang. 2011. Measuring social functions of city regions from large-scale taxi behaviours, *IEEE Int. Conf. Pervasive Computing and Communications Workshops*, Seattle, *USA*, 384-388.

[20] X. Liu, L. Gong, Y. Gong and Y. Liu. 2015. Revealing travel patterns and city structure with taxi trip data, *J. Transport Geogr.*, 43, 78-90. https://doi.org/10.1016/j.jtrangeo.2015.01.016.

[21] P.S. Castro, D. Zhang, C. Chen, S. Li and G. Pan. 2013. From taxi GPS traces to social and community dynamics: A survey, *ACM Comput. Surv.*, 46(2), 17. https://doi.org/10.1145/2543581.2543584.

[22] C. Kang, S. Sobolevsky, Y. Liu and C. Ratti. 2013. Exploring human movements in Singapore: A comparative analysis based on mobile phone and taxicab usages, *2nd ACM SIGKDD Int. Workshop on Urban Computing*, 1-8.

[23] O. Sagarra, M. Szell, P. Santi, A.D. Guilera and C. Ratti. 2015. Super sampling and network reconstruction of urban mobility, *PLoS One*, 10(8), e0134508. https://doi.org/10.1371/journal.pone.0134508.

[24] L. Liu, C. Andris and C. Ratti. 2010. Uncovering cabdriver's behavior patterns from their digital traces, *Computers, Environ. Urban Syst.*, 34(6), 541-548. https://doi.org/10.1016/j.compenvurbsys.2010.07.004.

[25] M. Veloso, S. Phithakkitnukoon and C. Bento. 2011. Urban mobility study using taxi traces, *Int. Workshop on Trajectory Data Mining and Analysis, ACM*, 23-30.

[26] B.C. Csáji, A. Browetc, V.A. Traag, J.C. Delvenn, E. Huensc, P.V. Doorenc, Z. Smoredae and V.D. Blondel. 2013. Exploring the mobility of mobile phone users, P*hysics A*, 392, 1459-1473.

[27] M.C. González, C. Hidalgo and A.L. Barabási. 2008. Understanding individual human mobility patterns, *Nature*, 453(7196), 79-82. https://doi.org/10.1038/nature06958.

[28] C. Shannon. 1948. A mathematical theory of communication, *Bell Syst. Tech. J.*, 27, 623-656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.

[29] D.J. Watts and S.H. Strogatz. 1998. Collective dynamics of 'small-world' networks, *Nature*, 393(6684), 440-442. https://doi.org/10.1038/30918.