

# Mining Urban Congestion Evolution Characteristics Based on Taxi GPS Trajectories

Weiyan Xu<sup>\*</sup>, Yumei Huang

School of Mathematics and Statistics, Lanzhou University, Lanzhou, China

**Email address:**

xuwy2017@lzu.edu.cn (Weiyan Xu), huangym@lzu.edu.cn (Yumei Huang)

<sup>\*</sup>Corresponding author

**To cite this article:**

Weiyan Xu, Yumei Huang. Mining Urban Congestion Evolution Characteristics Based on Taxi GPS Trajectories. *American Journal of Traffic and Transportation Engineering*. Vol. 5, No. 1, 2020, pp. 1-7. doi: 10.11648/j.ajtte.20200501.11

**Received:** February 7, 2020; **Accepted:** February 24, 2020; **Published:** February 28, 2020

---

**Abstract:** The taxi GPS trajectories involve sufficient temporal and spatial characteristics and make it easy for us to obtain potential knowledge for understanding human mobility pattern and urban traffic network dynamics. Sensing urban traffic conditions not only enables traffic management authority to improve urban traffic management. It can also provide decision-making for residents and taxi drivers. A spectral clustering method is proposed for sensing traffic congestion using taxi GPS trajectories. First, taxi GPS trajectories are pre-processed and matched with the urban road network established based on the primal graph representation. Second, the average speed of the road segments is obtained according to the taxi GPS trajectories and a dynamic weighted graph of urban road network is constructed to capture complicated urban traffic network. Then, a spectral clustering method is developed to detect the urban traffic congestion. Finally, the congestion evolution characteristics in Lanzhou, China are visualized and analyzed during different periods in the weekdays and weekends. Experimental results show that the proposed method can effectively detect traffic congestion, and the results are consistent with the usual actual experience. Compared with other traffic congestion methods, the proposed method can detect urban traffic congestion with wider coverage and lower cost. Therefore, the proposed method can be integrated into the classic intelligent traffic system, assisting urban traffic prediction, personal travel route plan, route planning and navigation application.

**Keywords:** Urban Traffic, Traffic Congestion Evolution, Spectral Clustering, Dynamic Weighted Graph, Visualization, Taxi GPS Trajectories

---

## 1. Introduction

Due to the increase in urban population, the load-bearing pressure on cities is increasing, especially in the severe condition of urban traffic load, which has caused various social and environmental problems. Traffic congestion is an important urban issue affecting urban development and the people's daily lives [1]. On the one hand, the traffic congestion on urban roads greatly reduces the traffic efficiency of the road network and brings great inconvenience to residents' travel. On the other hand, long-term congestion also increases vehicle emissions and energy consumption. The effective prediction of traffic congestion can divert and regulate the upcoming urban congestion. Accurate traffic prediction is of great research significance and application value for urban traffic

management [2].

Because of rapid growth of network communication technology, it is easy to obtain spatiotemporal big data with great mining value. The global positioning system (GPS) data is a major component of urban big data. It involves sufficient temporal and spatial characteristics and makes it easy for us to obtain potential knowledge for understanding urban functional structure, human mobility pattern, and traffic network dynamics. The research results based on GPS big data have been widely used in smart cities and related applications, such as transportation network [3], public safety [4], and urban planning and management [5], etc.

Taxis are the most convenient means of transportation in the urban areas. Liu et al. [6] built a spatial embedded network to study the interactions between urban areas using taxi GPS in Shanghai. The community detection method was

used to divide the urban area into a two-level hierarchical structure, and the mobility of each level was further investigated. Cui et al. [7] established an urban travel model describing characteristics such as travel demand, speed, and direction of travel routes, and used it to estimate the capacity of urban road network. Zhang et al. [8] employed fuzzy c-means clustering and spatial autoregressive moving average models to study the relationship between traffic congestion and the built environment. They applied Shanghai taxi GPS data to verify the validity of their model. Based on GPS trajectories of Beijing taxis and Didi shared cars, Dong et al. [9] studied the travel characteristics of taxi and shared service modes. Zheng et al. [10] excavated hotspot routes at different periods, and studied hotspots areas and residents' hotspot routes in Chongqing based on 10,000 taxi GPS trajectories. Chen et al. [11] provided a recommendation algorithm for carpooling and conventional taxi services based on GPS trajectories. They evaluated the recommendation algorithm using 14747 taxi GPS trajectories. The experimental results showed that the total mileage of all passengers was greatly reduced. Liu et al. [12] proposed a spatial and temporal analysis method based on an income per unit time and the time for seeking passengers. The results showed that the designed method could better reflect the distribution of high profitable passengers.

Effective traffic congestion prediction can greatly improve the quality of public transportation road management. At present, there are many researches on the prediction of urban traffic congestion, and some research results have appeared. Kong et al. [13] provided a fuzzy evaluation algorithm to forecast the urban traffic congestion based on vehicle GPS trajectories. Andrea et al. [14] proposed an expert system for predicting road congestion and accidents using the GPS trajectories from Pisa. Wang et al. [15] raised a three-phase framework to study traffic congestion correlation between road segments based on multi-source data, which included the road network, POIs, and the taxi GPS trajectories of Beijing. Wang et al. [16] analyzed the causes of traffic congestion and several metrics of congestion evaluation. They also gave the architecture of intelligent transportation system based on big Data. Four deep learning models, which include convolutional neural network, recurrent neural network, long short-term memory, and gated recurrent unit, were used for predicting traffic congestion based on GPS trajectories [17]. The results showed that deep learning models obtained higher prediction accuracy compared with conventional machine learning models. Kan et al. [18] provided a method for sensing traffic congestion at the turns using taxi GPS trajectories. Compared with other methods, their method allowed a more granular analysis of traffic congestion.

Lanzhou, the capital of Gansu province, is one of the important central cities in the western region and an important node city in the Silk Road Economic Belt. Lanzhou is a typical valley city, and the Yellow River runs through the city from west to east. The city of Lanzhou is narrow from north to south and long from east to west.

Nowadays, traffic congestion is becoming a more and more serious problem in this valley city. Lanzhou is one of the top ten congested cities in China in 2019 [19]. How to improve the efficiency of congestion prediction and assist individual and government decision-making is particularly important. The primary purpose of this paper is to develop a spectral clustering method to predict the urban traffic congestion using taxi GPS trajectories. We focus on congestion evolution characteristics during different periods in the weekday and weekend.

The remainder of this paper is structured as follows. Section 2 describes the GPS trajectories, preprocessing and the study areas. A spectral clustering method is developed to predict the urban traffic congestion in Section 2. In Section 3 the experimental results are discussed. Finally, the conclusion and our future work are concluded in Section 4.

## 2. Materials and Methods

### 2.1. GPS Trajectories and Preprocessing

In general, taxi GPS trajectories contain taxi identification numbers, timestamp, location coordinates (i.e., latitude and longitude), instantaneous speed, direction, vehicle status (occupied or unoccupied), accumulated mileage et al. The raw GPS trajectories used in this paper contain trajectories of 3000 taxis in Lanzhou (29.6% of the fleet), China from March 6 to 12, 2017.

Raw GPS trajectories usually contain anomalous data caused by multipath and other uncertainties. In order to make the research results more accurate, the raw GPS data need to be preprocessed. Firstly, we will eliminate anomalous GPS records of taxis that exceed 100 km/h due to speed limits in urban areas. Secondly, the road network of Lanzhou is obtained using Minnesota Traffic Generator (MNTG) [20], which is an extensible web-based road network traffic generator. Then, we implement the map-matched GPS trajectories on the urban road network and eliminate the singularities that are out of range. Finally, for some points with time granularity greater than 30s, these missing points are supplemented based on linear interpolation of their neighbor coordinates.

### 2.2. Study Areas

The main urban areas of Lanzhou include Chengguan District, Qilihe District, Anning District and Xigu District. The urban road network structure of Lanzhou is complicated, including some mountain roads. There are few taxi GPS trajectories on some roads. Therefore, the study areas include only urban areas covered by taxi GPS trajectories. The part trajectories of GPS matching with urban road network are visualized in Figure 1. The taxi GPS trajectories are plotted by the red point. The road marked in green will be discussed in Section 3.

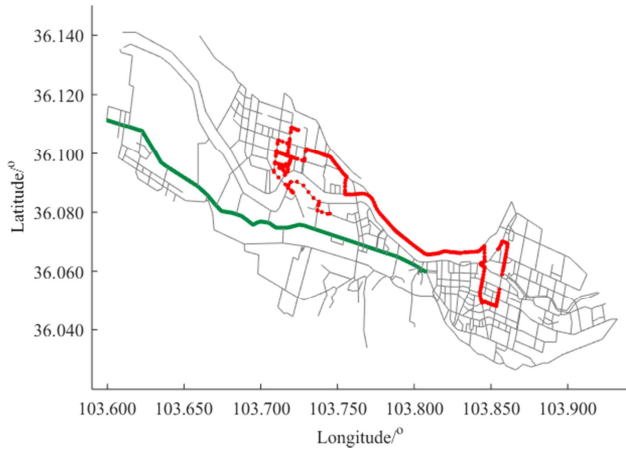


Figure 1. Part trajectories of GPS matching with urban road network.

### 2.3. Calculating the Average Speed of Road Segments

There are currently no unified metrics to evaluate traffic conditions. Conventional metrics employed to estimate traffic congestion include level-of-service (LOS), travel time index (TTI), vehicle speed, average commute time, saturation flow rate, queuing length, average travel time, etc. [21]. In fact, various traffic congestion evaluation metrics are established according to the traffic conditions of different countries. Vehicle speed is a considerable metric of urban traffic conditions and directly reflects the traffic condition of the urban area. Many studies explored urban traffic conditions using the vehicle speed obtained through the loop detector [22-23]. Therefore, vehicle speed is used to study urban congestion in this paper.

There are 227 roads in the specific study area. These roads contain 1022 road segments. The average speed of the  $j$ -th road segment during time period  $t$  can be calculated by the formula as follows:

$$\bar{x}_j(t) = \frac{\sum_i \sum_k x_{jik}(t)}{n}$$

where  $x_{jik}(t)$  is the  $k$ -th instantaneous speed of the  $i$ -th taxi on the  $j$ -th road segment during period  $t$ .  $n$  is the number of GPS point which located in the  $j$ -th road segment during time period  $t$ .

### 2.4. Traffic Congestion Prediction Based on Spectral Clustering

Spectral clustering is a clustering method based on spectrum theory and graph theory [24]. Compared with traditional clustering methods, its advantage is that it can cluster on a sample space of any shape and converge to a global optimal solution. It clusters the feature vectors of the Laplacian matrix of the sample data to achieve the purpose of clustering the sample data. The spectral clustering algorithm translates each sample data to the vertex of the graph, and converts the similarity between the vertices into the weight of the corresponding vertex connecting edge, thus the graph is an undirected weighted graph  $G=(V,E,W)$ , where

$V=\{v_1, v_2, \dots, v_N\}$  is a vertex set,  $E=\{e_{ij}, i, j=1, 2, \dots, N\}$  denotes the edge set,  $W=\{w_{ij}, i, j=1, 2, \dots, N\}$  denotes the weighted affinity matrix and  $N$  is the number of nodes.

The clustering process is to find the partitions of the graph to maximize the similarity within the sub-graphs and minimize the similarity between the sub-graphs. First, we consider the road segments of the road network as nodes and construct the dynamic undirected weighted graph  $G_t=(V, E, W_t)$ . It should be noted the number of nodes and edges is relatively constant because the urban roads of Lanzhou have been completed and will not change in the short term. The weights change over time. Therefore, the graph  $G_t$  is a temporal graph. According to the temporal graph, we predict the congestion segments based on spectral clustering. The detailed steps are given in Table 1.

Table 1. Traffic congestion prediction based on spectral clustering.

Input: Dataset $X_t = \{\bar{x}_1(t), \bar{x}_2(t), \dots, \bar{x}_N(t)\}$ , number $k$ of clusters to construct
Output: congestion road segments
Step 1: Construct graph and build its weighted affinity matrix $W_t$ (i.e., the similarity matrix $S_t$ ) using the Gaussian similarity function:
$w_{ij}^t = s_{ij}^t = \exp\left(\frac{-\ \bar{x}_i(t) - \bar{x}_j(t)\ ^2}{2\sigma^2}\right)$
where $\sigma$ is a hyper-parameter, $w_{ii}^t = 0$ .
Step 2: Calculate the degree matrix $D_t = \{d_i^t, i=1, 2, \dots, N\}$ , $d_i^t$ is the sum of weights of edges connecting to the node:
$d_i^t = \sum_{j=1}^N w_{ij}^t$
Step 3: Obtain the Laplacian matrix $L_t = D_t - W_t$
Step 4: Construct the normalized Laplacian matrix $L_{sym}^t = D_t^{-1/2} L_t D_t^{1/2}$
Step 5: Find the eigenvalues and eigenvectors of Laplacian: compute the first $k$ eigenvectors $u_1^t, u_2^t, \dots, u_k^t$ of $L_{sym}^t$ . Let $U_t = \{u_1^t, u_2^t, \dots, u_k^t\}$ , $U_t \in R^{N \times k}$ , $u_i^t$ as the columns for $U_t$ .
Step 6: Obtain the normalization matrix $M_t \in R^{N \times k}$ from $U_t$ by normalizing the rows to norm 1:
$m_{ij}^t = \frac{u_{ij}^t}{\sqrt{\sum_k (u_{ik}^t)^2}}$
Step 7: Let $y_i \in R^k$ be the vector corresponding to the $i$ th row of $M_t$ .
Step 8: Cluster the points $y_i \in R^k, i=1, 2, \dots, N$ into clusters $C_1^t, C_2^t, \dots, C_k^t$ by the k-means algorithm.
Step 9: Calculate the average value of sample data in each cluster and determine congestion road segments.
Step 10: Return congestion road segments

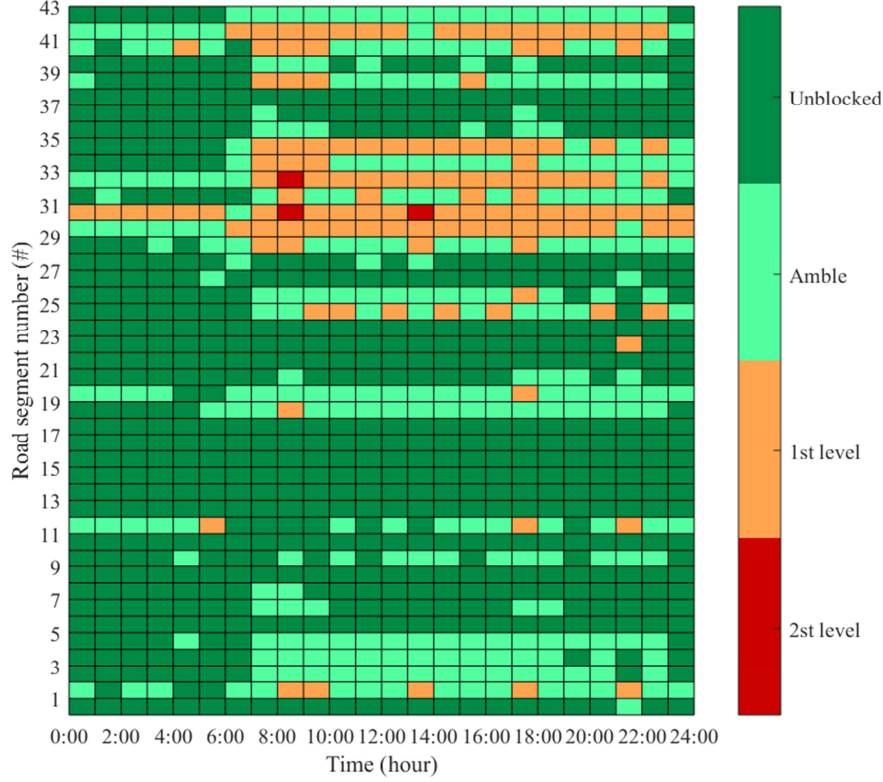
## 3. Results and Discussions

### 3.1. The Congestion Evolution of One Road During 24 Hours

We first study the traffic conditions of one road for 24

hours. The selected road is one of the main roads in Lanzhou and is marked with green color in Figure 1. The road spans three districts (i.e. Xigu District, Qilihe District, and Chengguan District) and is approximately 22 kilometers, accounting for two-thirds of the total length of Lanzhou City.

It starts at Chemical Road and ends at Jiefangmen, contains Xigu West Road (segment 1 to segment 5), Xigu Middle Road (segment 6 to segment 8), Xigu East Road (segment 9 to segment 10), Xijin West Road (segment 11 to segment 30), and Xijin East Road (segment 31 to segment 43).



**Figure 2.** The congestion evolution of one road during 24 hours.

The congestion evolution of this road during 24 hours is shown in Figure 2. From Figure 2, there are slightly congestion (1st level) on segment 2 during 8:00–10:00, 13:00–14:00, 17:00–18:00 and 21:00–22:00. The other segments of Xigu West Road, Xigu Middle Road, and Xigu East Road are not congested during 24 hours. This result shows the traffic condition in Xigu District is smooth except for commuting hours. The segment 30 remaining slightly congestion during 6:00–24:00 is part of Xijin West Road, which is the necessary road to Xigu District. Therefore, the traffic volume on this segment is relatively large. Next, we study the traffic condition of Xijin East Road. Except for the segment of Xihu Park, we can see that the most segments of Xijin East Road have been mild during 7:00–22:00. On one hand, the Xijin East Road is not only a transportation hub in Qilihe District, but also the way to connect the commercial centers including Lanzhou Center, Wanhui Plaza, Xitaihua, Xiaoxihu, etc. On the other hand, there are four subway stations were being built on this road in 2017. In particular, construction has the greatest impact on segment 31. The segment 31 connects West Station Cross which is one of the intersections with the heavy traffic. The subway station of West Station Cross is the largest subway station, and multiple lanes are blocked during period of construction, leaving only one lane for vehicles. In addition, poor road conditions can

cause vehicles to travel slowly. These reasons lead to the segment 31 is almost always congested during 24 hours. In particular, traffic was severely congested (2st level) during 8:00–9:00 and 13:00–14:00.

### 3.2. The Congestion Evolution of Urban Road Networks

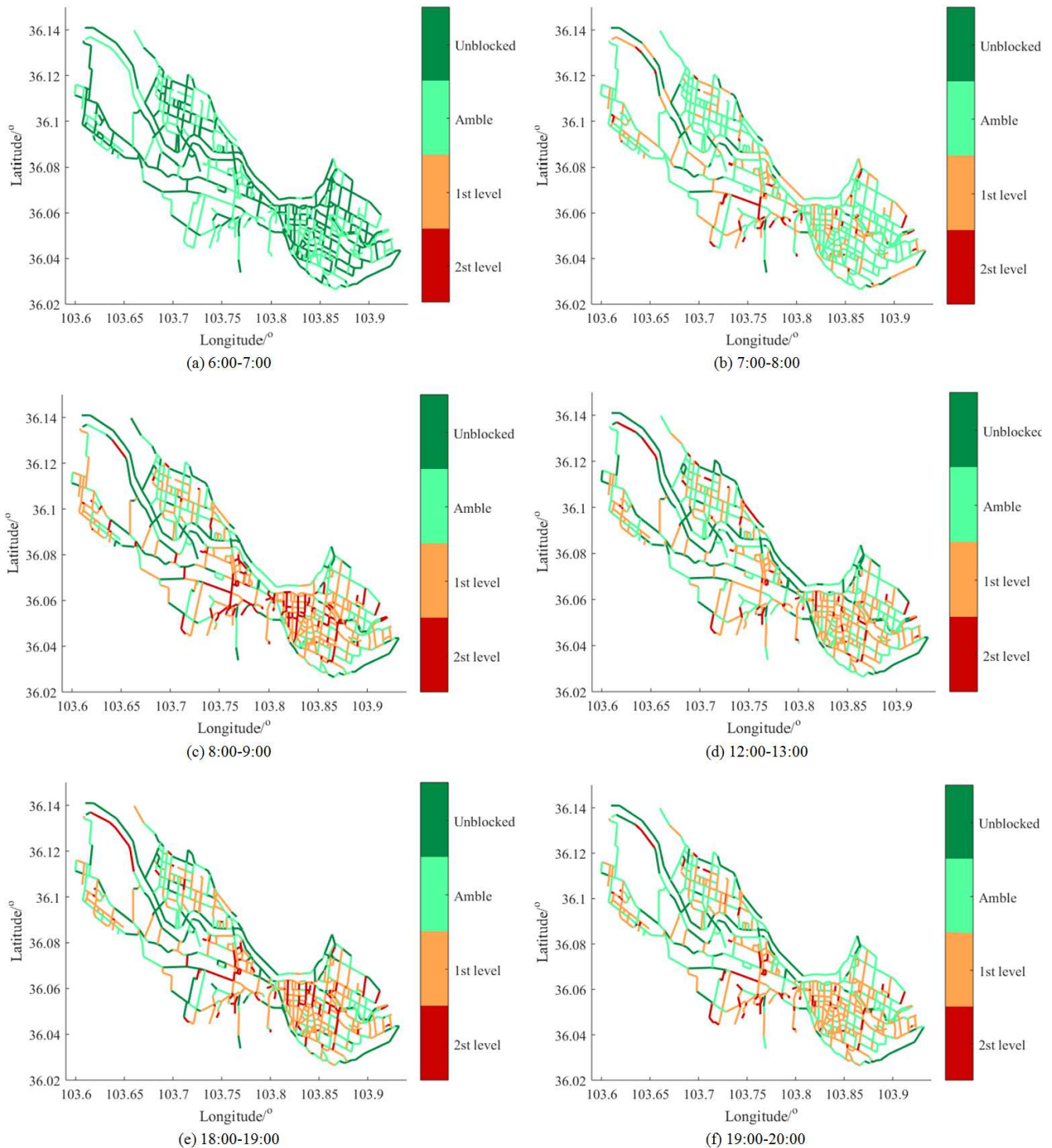
Further, we study the traffic conditions of the Lanzhou city. The evolutions of road segment congestion during the six time periods 6:00–7:00, 7:00–8:00, 8:00–9:00, 12:00–13:00, 18:00–19:00, and 19:00–20:00 in the weekdays are present in Figure 3, where the road traffic conditions during each 60-minutes period remain stabilizing. Figure 3 (a) shows that all roads are unblocked from 6:00 to 7:00. In the morning rush hour (7:00–8:00, 8:00–9:00, Figure 3 (b) and (c)), some road segments become severe congested, including Zhongshan Road, Xijin East Road, Tongwei Road, Jingchang Bei Road, jingning Bei Road, et al. During the midday (12:00–13:00, Figure 3 (d)), the severe congestion is relieved, except for a few segments. During the evening rush hour 18:00–19:00 and 19:00–20:00 (Figure 3 (e) and (f)), congestion is become more serious than that of midday. Most congested segments are in Chengguan District. It can be seen that severe traffic congestion segments appear in urban commercial centers, and slightly congestion (1st level) is



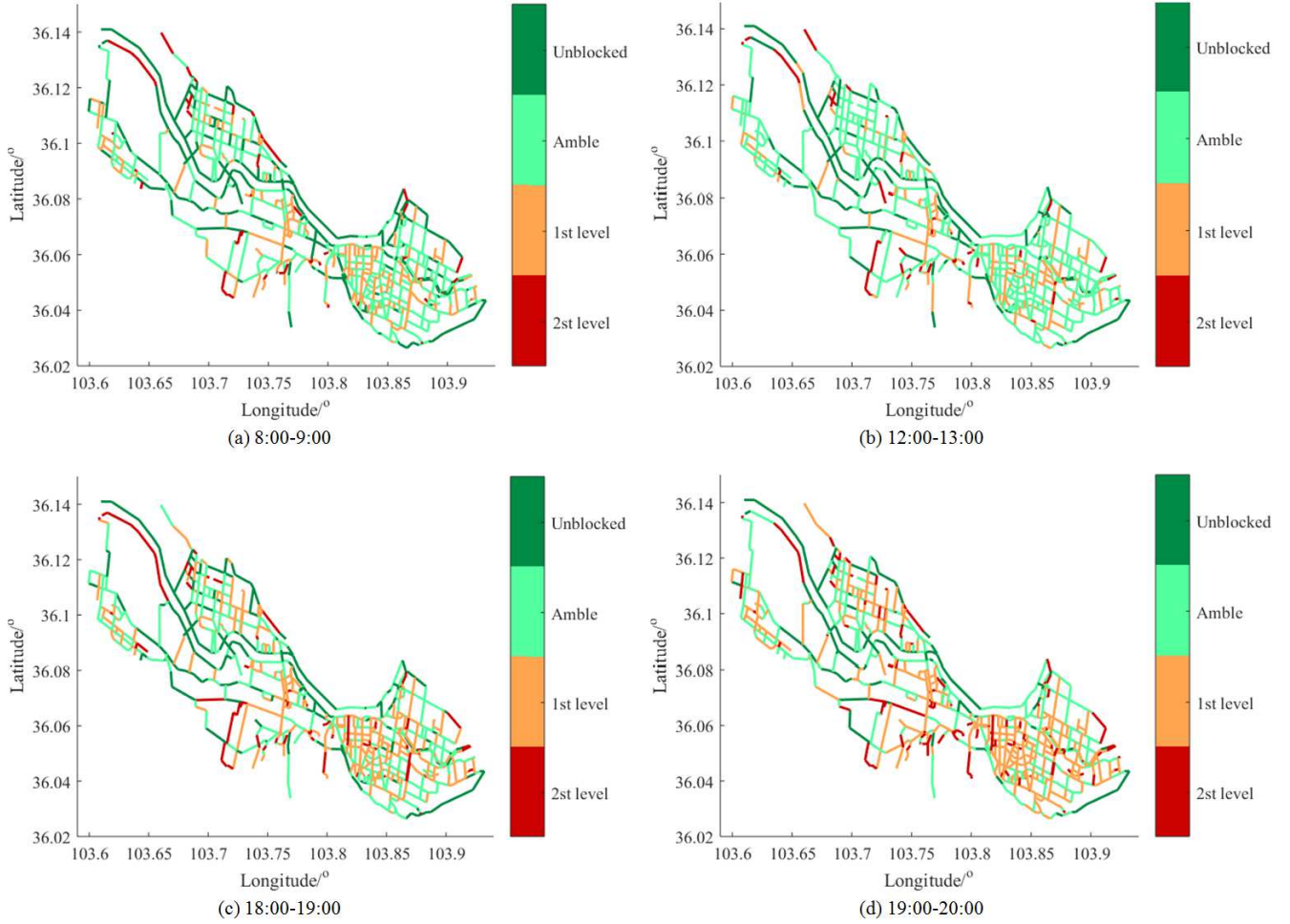
usually centered on severe traffic congestion segments. This hierarchical congestion pattern is fully compatible with gradual changes in traffic conditions. We can also find that the severe congestion changes in most roads during the three peak periods. There are many reasons for this result, but the most important one is that in the morning, people usually have a single destination: getting to school or work. In the evening, people leave work at different times and drive to different places to watch, to eat, and to shop, etc.

Figure 4 depicts the evolutions of road segment congestion

during four time periods 8:00–9:00, 12:00–13:00, 18:00–19:00, and 19:00–20:00 in the weekends. According to Figure 2 and Figure 3, on the whole, we can conclude that the congestion on weekends is not as severe as on weekdays. Compared with the traffic condition on weekdays, there are no significant morning rush hours on the weekends. Many people go out for lunch or shopping after 12:00. Correspondingly, the traffic congestion increases after lunch. The period between 18:00 and 20:00 can be considered as the rush hour on the weekends.



**Figure 3.** The congestion evolution of road segment in the weekdays.



**Figure 4.** The congestion evolution of road segment in the weekends.

## 4. Conclusions

The taxi GPS trajectories contain enough information about residents' travel characteristics and urban traffic conditions. The purpose of the study is to propose an effective method for predicting urban congestion in an urban area. We developed a spectral clustering method to detect the urban traffic congestion using taxi GPS trajectories. We investigated the congestion evolution characteristics during different periods in the weekday and weekend. Experimental results showed that the method proposed in this paper could effectively detect traffic congestion, and the results are consistent with the usual actual experience. Moreover, the proposed method is built based on actual taxi GPS trajectories and urban road network rather than simulation data. Our research results have reference significance to solve urban traffic congestion problem. Therefore, the proposed method can be integrated into the classic intelligent traffic system (ITS), formulating a reasonable traffic management and providing decision-making for residents and taxi drivers.

There are various traffic congestion evaluation indicators, such as average traffic speed, travel time index, commute duration, level of service (LOS), etc., are established according

to the traffic conditions of different countries. As future work, we plan to improve traffic congestion algorithms considering multiple congestion evaluation indicators rather than single indicator. In addition, our next step will take the lane-extraction method based on multiple trajectories big data to provide more accurate and finer-grained traffic congestion prediction via machine learning.

## Acknowledgements

We would like to thank to Prof. Huifang Feng for providing the GPS data used in this work. This work is partially supported by the National Natural Science Foundation of China under Grant (11571156, 71761031).

## References

- [1] Zheng, Y., Capra, L., Wolfson, O., et al. (2014). Urban computing: concepts, methodologies, and applications, ACM Trans. Intell. Syst. Technol. 5 (3), Article 38, 1-55.
- [2] Yang, Q. Y., Wang, J. Z., Song, X. M. et al. (2015). Urban traffic congestion prediction using floating car trajectories data, Algorithms and Architectures for Parallel Processing, Springer International Publishing, pp. 18-30.

- [3] Feng, H. F., Bai, F. S., Xu, Y. J. (2019). Identification of critical roads in urban transportation network based on GPS trajectory data. *Physica A: Statistical Mechanics and its Applications*, Article ID 122337, 535: 1-13.
- [4] Catlett, C., Cesario, E., Talia, D., Vinci, A. (2019). Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments. *Pervasive and Mobile Computing*, 53, 62-74.
- [5] Hao J. W., Zhu J., Zhong R. (2015). The rise of big data on urban studies and planning practices in China: Review and open research issues. *Journal of Urban Management*, 4, 92-124.
- [6] Liu, X., Gong, L., Gong, Y., Liu, Y. (2015). Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*, 43, 78-90.
- [7] Cui, J., Liu, F., Hu, J., et al. (2016). Identifying mismatch between urban travel demand and transport network services using GPS data: A case study in the fast growing Chinese city of Harbin. *Neurocomputing*, 181, 4-18.
- [8] Zhang, K. S., Sun, J., Shen, S. W., et al. (2017). Analyzing spatiotemporal congestion pattern on urban roads based on taxi GPS data. *Journal of Transport and Land Use*, 10 (1), 675-694..
- [9] Dong, Y., Wang, S., Li, L., Zhang, Z. (2018). An empirical study on travel patterns of internet based ride-sharing. *Transportation Research Part C*, 86, 1-22.
- [10] Zheng, L., Xia, D., Zhao, X., et al. (2018). Spatial-temporal travel pattern mining using massive taxi trajectories data. *Physica A*, 501, 24-41.
- [11] Chen P. P., Lv H. G., Gao S. W., et al. (2017). A Real-time taxicab recommendation system using big trajectories data. *Wireless Communications and Mobile Computing*, Article ID 5414930, 18 pages.
- [12] Liu, L., Zhang, F., Du, Z. H., et al. (2018). The analysis of high profitable strategy for seeking passengers based on taxi GPS trajectories data of Shenzhen city. *Journal of Zhejiang University (Science Edition)*, 45 (1), 82-91
- [13] Kong, X. J., Xu, Z. Z., Shen, G. J., et al. (2016). Urban traffic congestion estimation and prediction based on floating car trajectories data. *Future Generation Computer Systems*. 61, 97-107.
- [14] Andrea, E. D., Marcelloni, F. (2017). Detection of traffic congestion and incidents from GPS trace analysis. *Expert Systems with Applications*, 73 (1), 43-56.
- [15] Wang, Y. Q., Cao, J. N., Li, W. G., et al. (2017). Exploring traffic congestion correlation from multiple data sources. *Pervasive and Mobile Computing*, 41, 470-483.
- [16] Wang, Z. H., Yu, Y. S., Ju, D. C. (2018). Analysis and Prediction of Urban Traffic Congestion Based on Big Data. *International Journal on Data Science and Technology*, 4 (3), 100-105.
- [17] Sun, S. N., Chen J., Sun, J. (2019). Traffic congestion prediction based on GPS trajectories data. *International Journal of Distributed Sensor Networks*, 15 (5), 1-18.
- [18] Kan Z. H., Tang, L. L Kwan M. P., et al. (2019). Traffic congestion analysis at the turn level using Taxis' GPS trajectories data. *Computers, Environment and Urban Systems* 74, 229-243.
- [19] Top 10 most congested cities in China. Available online: <https://baijiahao.baidu.com/s?id=1639840213234339231&wfr=spider&for=pc>, (accessed on 16 January 2019).
- [20] Mokbel, M. F., Alarabi, L., Bao, J., et al. (2013). MNTG: An extensi-ble web-based traffic generator, *Advances in Spatial and Temporal Databases. SSTD 2013. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg.
- [21] Litman T. (2011). Smart congestion reductions II: Reevaluating the role of public transit for improving urban transportation. 2011/1/1, Available online: [https://www.researchgate.net/publication/254608957\\_Smart\\_congestion\\_reductions\\_II\\_Reevaluating\\_the\\_role\\_of\\_public\\_transit\\_for\\_improving\\_urban\\_transportation](https://www.researchgate.net/publication/254608957_Smart_congestion_reductions_II_Reevaluating_the_role_of_public_transit_for_improving_urban_transportation)
- [22] Jia, S. Q., Pend, H. Q., Liu, S., (2011). Urban traffic state estimation considering resident travel characteristics and road network capacity. *Journal of Transportation Systems Engineering and Information Technology*, 11 (5), 81-85.
- [23] He F. F., Yan, X. D., Liu, Y., Ma, L. (2016). A traffic congestion assessment method for urban road networks based on speed performance index. *Procedia Engineering*, 137, 425- 433.
- [24] Xia T., Cao, J., Zhang, Y. D, et al. (2009). On defining affinity graph for spectral clustering through ranking on manifolds. *Neurocomputing* 72 (13), 3203-3211.