Physica A 604 (2022) 127871

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Uncovering the spatiotemporal patterns of traffic congestion from large-scale trajectory data: A complex network approach

Jie Zeng [a], Yong Xiong [b,*], Feiyang Liu [a], Junqing Ye [a,*], Jinjun Tang [a]

[a] Smart Transportation Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha, 410075, China
[b] Hunan Lianzhi Technology Co., Ltd, Changsha, 410200, China

ARTICLE INFO

Article history:
Received 24 May 2022
Received in revised form 25 June 2022
Available online 28 June 2022

Keywords:
Traffic congestion modeling
Spatiotemporal pattern mining
Complex networks
Community detection
Trajectory data

ABSTRACT

Understanding the spatiotemporal characteristics of traffic congestion is the cornerstone of generating traffic management and control strategies. Based on the large-scale taxi trajectory data in Shenzhen, China, this study designs an effective framework to explore the spatiotemporal patterns of traffic congestions. To bridge trajectory data with urban road networks, we develop a two-stage map-matching method from the aspects of distance and angle. Then, the free-flow speed of each road segment is extracted and employed to identify traffic congestion. In this way, a novel complex network method, named chronological network (chronnet), is utilized for traffic congestion modeling, and we employ an overlapping community detection algorithm to identify region-level bottlenecks. According to the network properties, we explore the influence scope of traffic congestions and uncover the role of each road segment in the propagation process. Meanwhile, community detection results indicate that there are typical local clustering structures in traffic congestions, and each community also has its unique traffic characteristics. Overall, these findings reveal that the complex network can effectively mine the consecutive patterns of traffic congestion.

1. Introduction

How to alleviate traffic congestion is a consistent challenge that has puzzled traffic scholars and engineers for several decades. During these years, the intelligent transportation systems (ITS) have provided a new approach to address this problem. Based on emerging traffic collection devices, traffic managers can conduct timely analyses about traffic states and design corresponding control strategies. Nowadays, numerous novel traffic management technologies are applied to real-world traffic systems, such as information provision [1,2], signal optimization [3,4], congestion pricing [5,6], etc. These approaches have been demonstrated as effective countermeasures to relieve traffic congestion [7]. However, the urban road network consists of massive road segments, so it is impractical to implement traffic control on all road segments in such a comprehensive environment. Therefore, it is essential to understand its spatiotemporal patterns and conduct targeted countermeasures.

Existing studies on traffic congestion modeling can be classified into two categories, i.e., simulation-based models and data-driven methods. Here, the former focuses on understanding the microscopic characteristics of traffic congestion from simulation models, such as traffic flow theory, epidemic diseases, etc. Bando et al. [8] learned the dynamics of

* Corresponding authors.
E-mail addresses: xiongyong36@qq.com (Y. Xiong), yjq29@csu.edu.cn (J. Ye).

https://doi.org/10.1016/j.physa.2022.127871
0378-4371/

traffic flow as a collective motion problem, and then a simulation model was established to represent the motion of each vehicle for traffic congestion modeling. Olmos et al. [9] utilized the cellular automaton model to represent the vehicle dynamics and explored the collapse of urban traffic. Saberi et al. [10] employed the susceptible–infected–recovered (SIR) model to explain the dynamics of traffic congestion propagation and dissipation. With the development of connected automated vehicles (CAVs), researchers also turn to congestion modeling under the environment of CAVs and mixed traffic flow [11]. However, simulation-based models always suffer limitations in computational costs and parameter calibration [10]. Therefore, researchers and engineers gradually paid attention to data-driven methods for congestion analysis.

In traffic congestion analysis, the foundation of the data-driven method is the accurate traffic states. During these years, various traffic state collection devices are equipped in the urban road network, providing opportunities to solve this problem from a data-driven perspective. Overall, two major traffic detectors are installed in urban road networks, i.e., fixed sensors and mobile devices. The former is mainly installed at fixed locations (e.g., intersections), and the latter can be equipped on individual vehicles (i.e., taxis, buses, etc.). Based on the fixed detectors, several efforts have been made in traffic congestion analysis [12,13]. However, limited by installation and maintenance costs, fixed detectors are always sparsely installed in urban road networks. Thus, only limited areas are covered in these methods [14]. According to the advantages of these two detectors, trajectory data from the mobile devices is more competent for traffic state analysis and travel patterns mining in large-scale road networks [10,15–21]. During these years, this type of data has won the favor of researchers in traffic congestion analysis and propagation inference [22–24].

Due to the superiorities of trajectory data, several efforts are applied to traffic congestion analysis in urban road networks [25–27]. In traffic congestion identification, Wang et al. [28] designed a trajectory-based traffic congestion visualization system, and it can help extract traffic states in a large-scale road network intuitively. Kan et al. [14] proposed a turn-level traffic congestion detection method based on taxi trajectory data. In this method, trajectories were classified into three intensities and employed to identify congestion events. Meanwhile, He et al. [29] also proposed a trajectory-based method to identify turn-level congestion without the digital map. And in traffic congestion analysis, Li et al. [30] applied the percolation theory to reveal the transition phenomenon in the urban road network. An et al. [31] aggregate trajectory data into grid-based areas and calculated traffic states inside each grid to mine the recurrent evolution patterns of urban congestions. Çolak et al. [32] mined travel patterns from mobile phone traces to analyze congested travel in urban areas. They found that traffic congestion can be significantly alleviated by the centralized routing scheme with awareness of social good.

However, although critical efforts have been put into this field, challenges and limitations still exist, such as:

(1) Essentially, traffic congestions can be regarded as spatiotemporal events with comprehensive characteristics, driven by the operation law of the traffic system. In addition to the well-known local evolution characteristics, traffic congestions also exhibit similar variation patterns in distant road segments, e.g., parallel roads, areas with similar functions, etc. Revealing these spatiotemporal characteristics and global synchronizations are vital for network-scale traffic prediction and management, but these patterns are rarely involved in the current studies.

(2) Bottleneck identifications for urban road networks are always limited to segment-level. However, traffic congestion always produces at a certain point and spreads to nearby road segments. So, it is easy to cause regional congestion in the urban road network. In recent years, traffic researchers and managers gradually focused on region-scale or network-scale control instead of the single point [33,34]. Therefore, besides identifying critical road segments, it is also an essential task to identify urban areas prone to traffic congestion, i.e., region-level bottlenecks.

(3) Evolution patterns of traffic congestion still need further exploration. For instance, the roles of road segments in the propagation process vary from road to road, such as alleviating or exacerbating congestion. Therefore, it is essential to distinguish the role of each road segment. Meanwhile, learning the influence scope of traffic congestion is also necessary for congestion management and control. Here, this scope can be understood from two aspects, including the influence of distance and intensity. Further analysis of these characteristics can help traffic managers to formulate refined and effective measures to relieve traffic pressure.

To solve these challenges, we propose a data-driven method based on complex networks to understand traffic congestions from large-scale trajectory data. Overall, the main contributions of this study can be summarized as three-fold. (1) We propose a framework to learn the spatiotemporal patterns and consecutive characteristics of traffic congestions by the complex network theory, i.e., the chronological network (chronnet). (2) An overlapping community detection algorithm is employed to extract the propagation characteristics of traffic congestion and identify region-level bottlenecks. (3) Based on the structural properties analysis of the constructed complex network, we explore the spatiotemporal patterns and analyze the influence scope of traffic congestion in the large-scale road network.

The rest part of this study is organized as follows. Section 2 introduces the traffic congestion modeling framework in detail. The trajectory data involved in this study and congestion count analysis are described in Section 3. Section 4 explores the structural properties and discusses the influence scope of traffic congestions from the constructed complex network. Meanwhile, region-level bottlenecks are also identified in this section. Finally, we conclude this study and future works in Section 5.

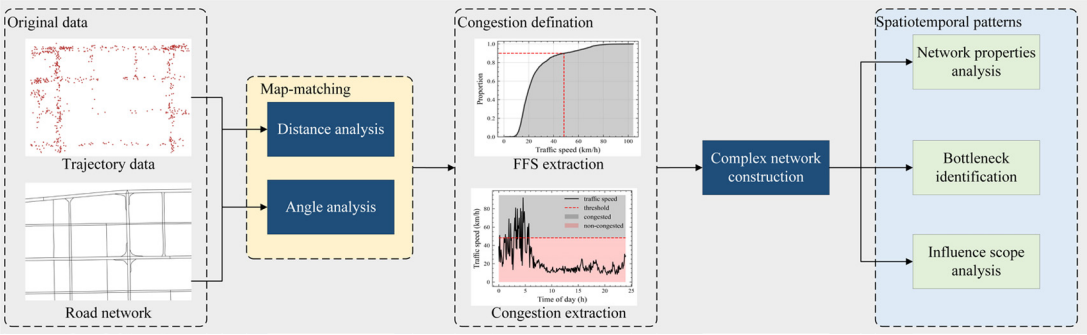


Fig. 1. Framework of spatiotemporal pattern analysis of traffic congestions.

2. Methodology

2.1. Framework

The framework of this study is illustrated in Fig. 1. From the overview of this figure, this work comprises four critical parts, i.e., map-matching, traffic congestion identifying, complex network construction, and spatiotemporal pattern analysis. First, we develop a two-stage map-matching method to bridge trajectory data with each road segment. After that, the free-flow speed is defined and employed to identify segment-level traffic congestions. Based on the extracted congestion matrix, a complex network construction method is adopted to transform consecutive traffic congestions into a directed weighted graph. Finally, we employ an overlapping community detection method to identify the critical areas in the complex network and conduct further analysis of influence scope based on network properties.

2.2. Map-matching

Map-matching aims to map GPS trajectories to road segments. Since this study focuses on identifying traffic congestions at the segment level, map-matching becomes an essential foundation in this work. The map-matching method developed by this study is described below.

Motivated by [35,36], the main idea of this method is to conduct map-matching from two aspects, i.e., distance analysis and angle analysis. To accelerate the map-matching process in the large-scale road network, this study proposes an angle analysis to replace the original transmission probability in the previous study [36]. For the distance analysis, supposing p_i denotes a GPS point, we first select the candidate road segment set $c(p_i)$ from the surrounding roads. In $c(p_i)$, all the perpendicular distances to p_i should be no more than 50 m. For each candidate road segment $r \in c(p_i)$, this method regards the vertical foot between p_i and r as the candidate projected point cp_i^r. Then, as shown in Eq. (1), a normal distribution with $\mu = 0, \sigma = 20$ is utilized to measure the likelihood between GPS point p_i and the corresponding cp_i^r. Here, d_i^r denotes the distance between p_i and cp_i^r.

$$N(cp_i^r) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(d_i^r - \mu)^2}{2\sigma^2}) \tag{1}$$

Afterward, we define another likelihood $V(cp_i^r \to cp_{i+1}^s)$, named angle analysis, to represent whether the angle between two consecutive candidate points cp_i^r and cp_{i+1}^s follows the corresponding angle from p_i to p_{i+1}. To achieve this goal, Eq. (2) is employed to calculate the angle between GPS point p_i and p_{i+1}.

$$A(p_i \to p_{i+1}) = \arctan(\frac{x_{p_{i+1}} - x_{p_i}}{y_{p_{i+1}} - y_{p_i}} * \cos y_{p_{i+1}}) \tag{2}$$

Here, $A(p_i \to p_{i+1})$ denote the angle from GPS point p_i to p_{i+1} based on the great circle theory. Meanwhile, x_{p_i} and y_{p_i} represent the longitude and latitude of p_i, respectively. The angle between the consecutive candidate points cp_i^r and cp_{i+1}^s can also be calculated by this equation.

After that, the angle similarity $V\left(cp_i^r \to cp_{i+1}^s\right)$ between the truly path (i.e., from p_i to p_{i+1}) and the candidate path (i.e., from cp_i^r to cp_{i+1}^s) is mathematically expressed by Eq. (3).

$$V(cp_i^r \to cp_{i+1}^s) = \min(\frac{A(p_i \to p_{i+1})}{A(cp_i^r \to cp_{i+1}^s)}, \frac{A(cp_i^r \to cp_{i+1}^s)}{A(p_i \to p_{i+1})}) \tag{3}$$

According to this equation, values of all the angle similarities are in [0, 1]. For each two consecutive GPS points, the smaller the deviation with the angle of road segments, the larger the angle probability. In this way, the final likelihood

of the candidate projected points can be written in Eq. (4).

$$F_s(cp_i^r \rightarrow cp_{i+1}^s) = N(cp_i^r) \cdot V(cp_i^r \rightarrow cp_{i+1}^s), i \in [1, n-1] \tag{4}$$

2.3. Traffic congestion identification

Identifying traffic congestion is the foundation for congestion-based studies. According to the industry standard GA/T 115-2020 in China, traffic congestion can be determined by actual speed and limited maximal velocity. Based on the results of map-matching, we aggregate traffic speed of each road segment into 5-min period and employ it to identify traffic congestion. However, collecting the limited velocities of all the road segments in a large-scale road network is an extremely expensive task. Hence, the cumulative density distribution of traffic speed at each road segment can be employed to estimate its limited velocity [30]. Specifically, we utilize the 90th percentile of traffic speed at each segment to represent the corresponding free-flow speed. In this way, as defined in Eq. (5), this study introduces a traffic congestion index (TCI) to measure traffic congestion.

$$\mathbf{TCI}[i,j] = \frac{\boldsymbol{V}[i,j]}{\mathbf{FFS}[j]} \tag{5}$$

Hence, $\boldsymbol{V}[i,j]$ denotes the average traffic speed of road segment j at period i. Meanwhile, **FFS** and **TCI** stand the free-flow speed vector and traffic congestion index matrix, respectively. This equation indicates that traffic congestion is more severe if **TCI** is closer to 0. Finally, as shown in Eq. (6), a threshold is applied to determine traffic congestion based on **TCI**. It means that if **TCI** of road segment j at time i is lower than the congestion threshold r_c, then it will be regarded as congestion at period i. And according to the industry standard GA/T 115-2020 in China, we assign the congestion threshold as $\mathrm{r}_c = 0.6$.

$$\boldsymbol{C}[i,j] = \begin{cases} 1, & \mathbf{TCI}[i,j] < \mathrm{r}_c \\ 0, & \mathbf{TCI}[i,j] \geq \mathrm{r}_c \end{cases} \tag{6}$$

It is noted that the coverage of floating vehicles is relatively low, making the average speed matrix $\boldsymbol{V}$ significantly sparse. According to Eqs. (5) and (6), if $\boldsymbol{V}[i,j]$ is 0, it will be regarded as congestion. Therefore, we design a correction strategy to deal with this problem. For each period, road segments without vehicle passing are considered unobstructed (i.e., $\mathbf{TCI}[i,j] = 1$ and $\boldsymbol{C}[i,j] = 0$).

2.4. Complex network construction

As described in Eq. (6), traffic congestions can be discretized according to the spatial and temporal dimensions. Thus, we can regard them as special events with spatiotemporal characteristics. During these years, the complex network has been treated as a powerful tool for spatiotemporal data mining [37,38]. The chronological network (chronnet) [39] is a novel complex network for spatiotemporal event modeling. It has the advantages of convenient understanding and efficient calculation. Therefore, we employ it to model the network-scale evolution patterns of traffic congestions. The original chronnet needs to divide the study area into grid-based data, and each grid is regarded as a node in the complex network. However, the topology information is overlooked in this method. Therefore, this study considers each road segment as a node in chronnet instead of the grid-based area.

The principle of this complex network method is shown in Fig. 2. Here, we take traffic congestions at 10 road segments in 4 continues periods as an example. To apply this method in the road network, we regard each road segment as a node in the complex network. Hence, road segment with traffic congestion is marked as yellow in Fig. 2a, and nodes in Fig. 2b denote the corresponding road segments with Fig. 2a. In Fig. 2b, the numbers on the edges denote the corresponding edge weights w_{ij}. Meanwhile, since several road segments never suffer congestions in this example, they are not illustrated in Fig. 2b (e.g., node 4 and 5). Overall, the detail definition of this method is summarized as follows. Supposing the road segment x_i is congested at period t, $\mathbf{x}_{t+1}$ represents the congestion road segments list at period $t+1$. Then, this method will add a directed edge from x_i to each element of $\mathbf{x}_{t+1}$ if the distance between these two nodes is less than a given threshold $d_{\max}$. Therefore, the edge weight w_{ij} can denote the number of congestion propagation from node i to node j.

In the original chronnet [39], all the edges are undirected. However, considering the actual traffic significance, we use the direct edges to represent the congestion propagation process among road segments. Therefore, a directed edge will be added from x_i to all the elements of $\mathbf{x}_{t+1}$. It is noted that the distance threshold $d_{\max}$ will limit the distance between node pairs with consecutive congestions. To explore the consecutive patterns of traffic congestions from a global view, following the setting in the original chronnet [39], we set the distance threshold $d_{\max}$ as $+\infty$. Results in Section 4.4 can demonstrate the reasonableness and validity of this setting. Meanwhile, since traffic congestion is time-varying, congestion propagations will occur between the same pair of road segments at different periods. Also, under severe congestion, a road segment may continue to congest during consecutive periods. So, multiplex edges and self-loops are allowed in this network.

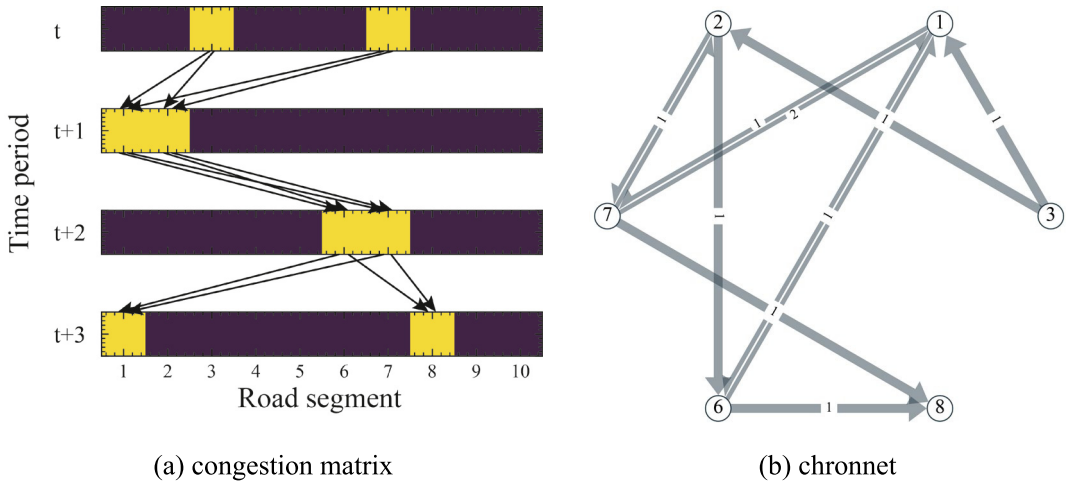


Fig. 2. Example of the chronnet construction method.

2.5. Bottleneck identification

Community in the complex network denotes local node groups with strong connections. Based on the constructed chronnet, we employ community detection to identify region-level bottlenecks in the large-scale urban road network. Supposing w_{ij} and w_{ji} denote edge weights between node i and node j, each weight can only represent the propagation relationship in a specific direction. Therefore, to measure the local connection relationship effectively and reasonably, the chronnet is transformed into an undirected graph to identify region-level bottlenecks. Specifically, all the directed edges are replaced by undirected edges, and each weight on these undirected edges is calculated by the summation of the corresponding directed weights (i.e., $w_{ij} + w_{ji}$).

Epasto et al. [40] proposed a scalable and flexible framework, named the ego-splitting method, to address the overlapping community detection problem. They have demonstrated the effectiveness and accuracies of this method from theoretical proof and experiment results. Considering the advantages of this method, we employ it to identify region-level bottlenecks in this study. The detail calculation process is summarized in Fig. 3. Here, ego-net and persona (shown in Fig. 3b and Fig. 3c) are the cores of this method, and they are employed as guidance for overlapping community detection. In this way, this method transforms the overlapping communities into a non-overlapping problem.

The first step of this method is selecting the ego-net of each node, where the ego-net denotes the subgraph generated by the neighborhood of each node. For example, the ego-net of node 2 consists of $[0, 3, 4, 6, 7, 9]$. Then, as shown in Fig. 3b, a local community detection method is applied to the ego-net of each node to partite it into sub-ego-nets. After that, according to Fig. 3c, each node is incorporated with its sub-ego-net to construct a persona. Supposing the number of sub-ego-nets for node i is k_i ($k_i \geq 1$), we will add $k_i - 1$ virtual nodes and construct k_i personas, e.g., node 2^* in Fig. 3c. Furthermore, the personas of all the nodes are integrated to generate a larger graph, and a global community detection method is employed to conduct community detection on the final persona graph (i.e., the topological network shown in Fig. 3d and Fig. 3e). Finally, all the virtual nodes are merged with the corresponding nodes, which means these nodes are overlapped into multiple communities. It is worth noting that although the calculation process described in Fig. 3 is about unweighted graphs, this method can be also applied on weighted graphs [40]. A more detailed description and theoretical guarantees of this method can be found in [40,41].

In this study, we employ the connected components algorithm for the local community detections. But for the global community detections, since the constructed network is a weighted graph, the connected components algorithm utilized in the original study is unsuitable for this task. So, we replace the connected components algorithm with the Louvain algorithm [42] to improve detection capabilities by involving the edge weights. In the Louvain algorithm, the modularity Q [43] described in Eq. (7) is the core to determine the closeness of communities. Here, m is the summation of the edge weights in the graph. Meanwhile, s_p and c_p denote the strength and community of node p.

$$Q = \frac{1}{2m} \sum_{p,q} [w_{pq} - \frac{s_p s_q}{2m}] \delta(c_p, c_q) \tag{7}$$

$$\delta\left(c_p, c_q\right) = \begin{cases} 0, & c_p \neq c_q \\ 1, & c_p = c_q \end{cases} \tag{8}$$

The calculation process of this algorithm is displayed in Fig. 4. Here, we utilize an unweighted network as an example. Overall, there are two phases in the calculation process of the Louvain algorithm. The first phase aims to find local

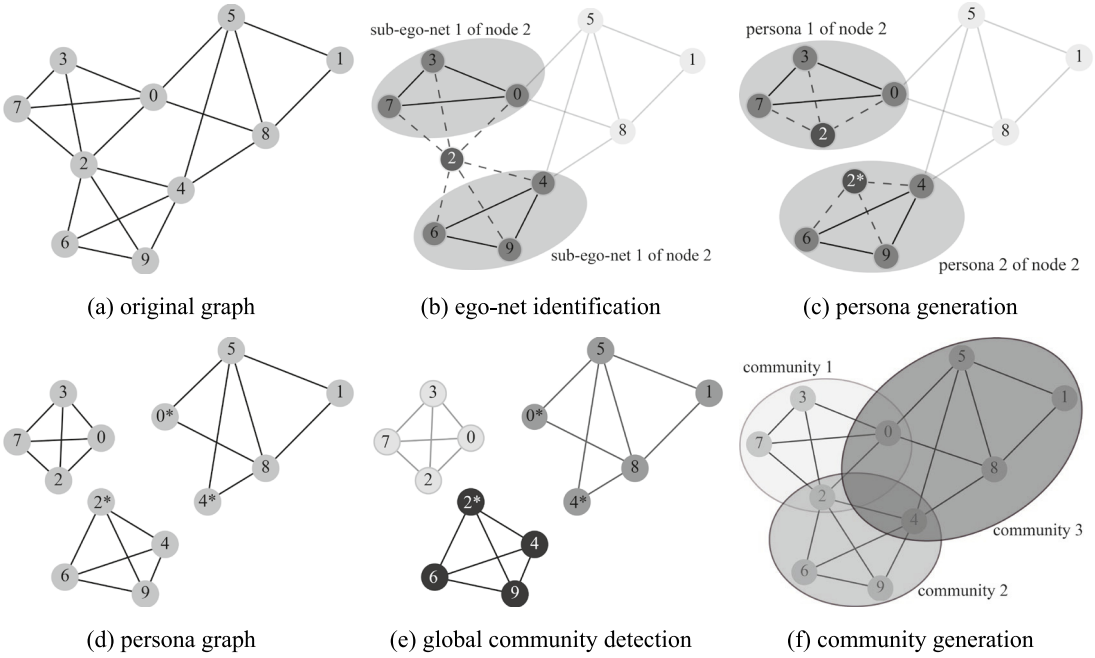


Fig. 3. The calculation process of the ego-splitting method.

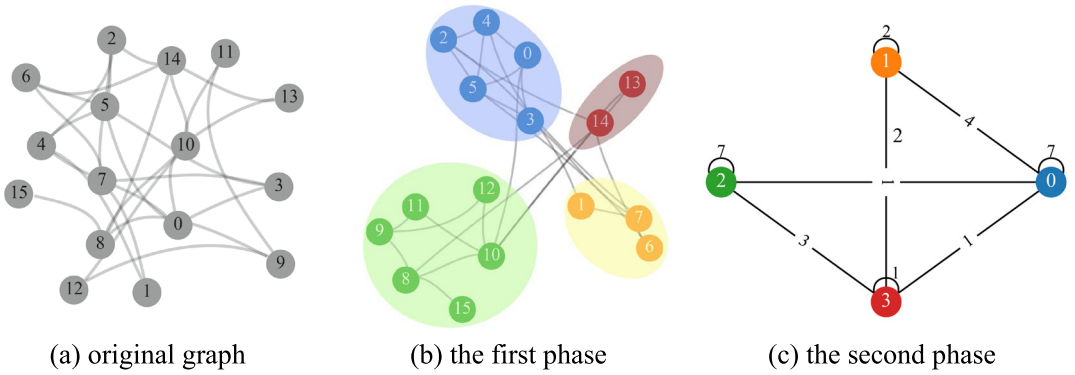


Fig. 4. The calculation process of the Louvain algorithm.

communities based on the increment of modularity ΔQ, and the second phase will aggregate each community as a super-node and reconstruct the network. Then, each super-node in the second phase will be regarded as a node in the next iteration, and the modularity-based principle will be applied to it again to generate a new first phase. This algorithm will iteratively conduct these two phases until the termination condition, e.g., the modularity no longer changes, reaches the maximum iterations, etc.

3. Data description and analysis

3.1. Data description

This study employs trajectory data from more than 16,000 taxis in Shenzhen, China, for traffic congestion modeling in a large-scale road network. This dataset ranges from February 1st to 28th in 2019, including over 620 million records. Columns of this dataset include license plate information, speed, longitude, latitude, angle, and passenger state, so that various traffic information can be extracted, such as average speed, travel demand, travel time, etc.

Since Nanshan, Luohu, and Futian District are the most prosperous in Shenzhen city [44], this study selects these three districts for further analysis. Fig. 5 shows the selected area of this study. It locates in the southwest of Shenzhen and reaches 334.32 km^2. Therefore, only trajectories in this area are extracted, and others are discarded. In this way,

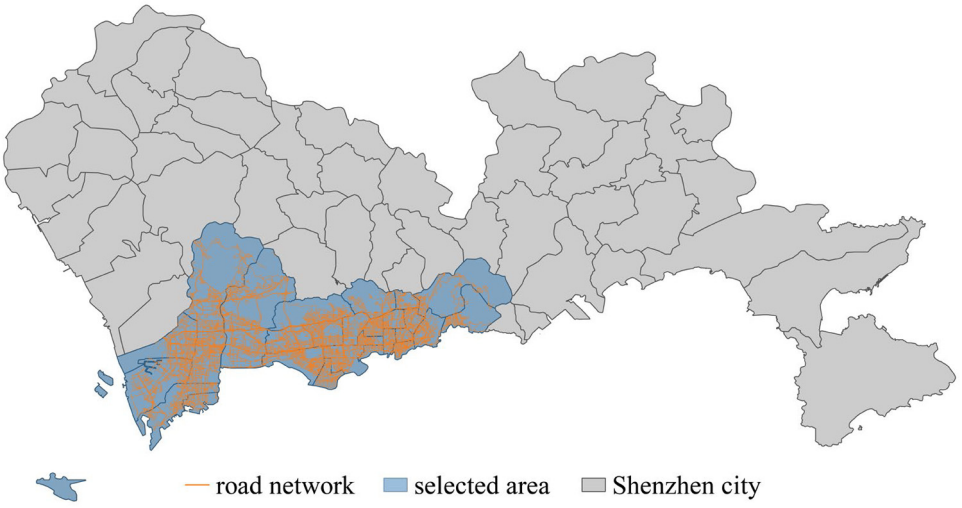


Fig. 5. The selected road network in Shenzhen city.

350,452,943 trajectory records and 17,007 road segments are obtained from this area. Then, the map-matching method described in Section 2.2 is developed to bridge trajectory data with road segments. Afterward, we aggregate taxi speed on each road segment into 5-min and employ it for further analysis.

Due to the low coverage of floating vehicles, missing data is an inevitable challenge for trajectory-based researches and applications [45]. Therefore, in this study, the historical average (HA) method is employed to counter this problem. For example, if traffic speed at 8:00–8:05 on a Monday is missing, we will employ the average speed at the same period of remaining Mondays (where speed data is not missing) to recover it. All the experiments of this study are implemented by python 3.8.12 on a Linux workstation (CPU: AMD 3700X @ 3.6 GHz, RAM: 64 GB memory).

3.2. Congestion count distribution

According to the traffic congestion identification method in Section 2.3, we can obtain the number of traffic congestions at each road segment, named congestion count. The spatial distribution of congestion count is illustrated in Fig. 6. From this figure, we can find that traffic congestion is a widespread phenomenon that significantly impacts almost all road segments. Meanwhile, the spatial distribution of traffic congestion seems to be random, so it is quite difficult to understand spatiotemporal patterns and identify bottlenecks for traffic management. Therefore, it is essential to apply an effective data mining method to solve this problem.

Afterward, this figure also indicates that although most road segments have good traffic conditions, a few segments still suffer extreme congestion. Fig. 7a displays the time-varying characteristics of average speeds at the most congested road segment, where 5,888 congestions occurred. Here, the black line and gray shading represent the average speed and standard deviations at this road segment of one month. It is evident that traffic speeds at this road segment during the daytime are chronically low, but it allows vehicles to drive fast at night. Meanwhile, we further illustrate its cumulative density distribution in Fig. 7b and display the estimated free-flow speed in Fig. 7a (i.e., the red line). From these figures, we can find that the estimated free-flow speed of this segment is about 50 km/h, and it complies with general regulations for urban roads. Meanwhile, the congestion threshold $r_c \cdot$ **FFS** on this road segment is 29.99 km/h. However, during the daytime, traffic speeds at almost all the periods are much lower than this value. Hence, we can regard that it is always stuck in traffic congestion.

4. Numerical experiments and discussion

4.1. Analysis of network properties

According to Section 2.4, a directed weighted chronnet is constructed to model traffic congestions in the urban road network. Overall, there are 17,007 nodes and 1,022,158 edges in this complex network. In this subsection, we introduce several widely-used metrics to further understand the properties of this complex network, e.g., degree, clustering coefficient, and centralities. The definitions of these metrics are described as follows.

(1) Degree and strength

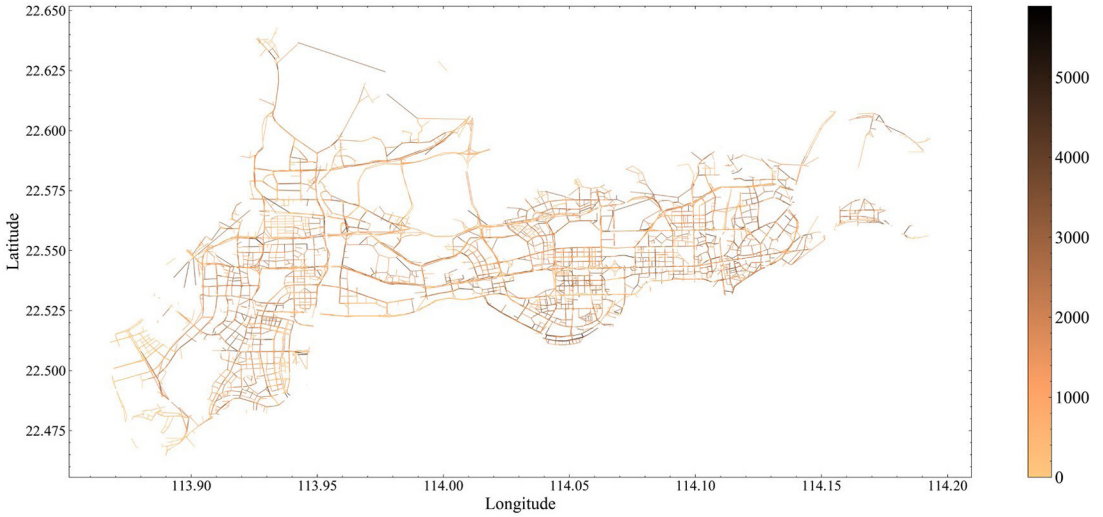


Fig. 6. Spatial distribution of congestion count in the study area.

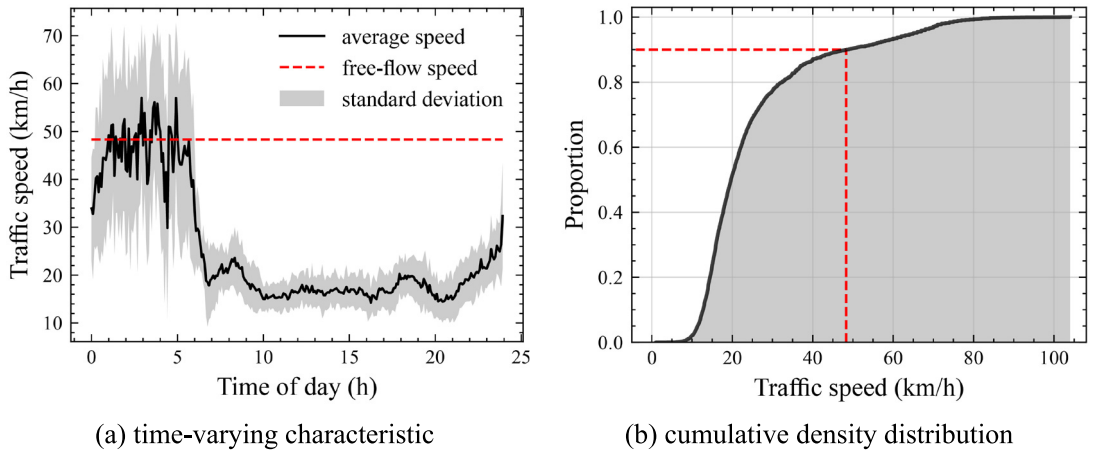


Fig. 7. Speed distribution at the most congested road segment in the study area.

Since the constructed chronnet is a directed weighted graph, it can be understood in two aspects, i.e., node degree and strength. The definition of these two metrics can be mathematically expressed by the following equations.

$$k_i^{out} = \sum_{j=1}^{N} a_{ij},\ k_i^{in} = \sum_{j=1}^{N} a_{ji} \tag{9}$$

$$s_i^{out} = \sum_{j=1}^{N} w_{ij},\ s_i^{in} = \sum_{j=1}^{N} w_{ji} \tag{10}$$

Here, $\boldsymbol{A} = (a_{ij})_{N\times N}$ and $\boldsymbol{W} = (w_{ij})_{N\times N}$ denote the adjacent matrix and weight matrix of the chronnet, respectively. Meanwhile, k_i^{out} stands the out-degree of node i, and s_i^{out} represents the out-strength. According to their definitions, the former expresses how many road segments interact with the target node, and the latter explains the interaction intensity. Furthermore, from the definition of chronnet [39], road segments with long periods of congestion are likely to treat as high degree nodes. Meanwhile, road segments with numerous traffic congestions are always assigned high strengths.

(2) Clustering coefficient

For directed graphs, the clustering coefficient is defined as Eq. (11) [46]. Here, the clustering coefficient denotes the fraction of all possible directed triangles in the neighbors of each node. In this equation, $T(u)$ denotes the number of possible directed triangles. Meanwhile, $k(u)$ is the total degree of node u, and $k^{\leftrightarrow}(u)$ denotes the reciprocal degree. According to its definition, the value of each c_u belongs to [0, 1], and it can represent the local connection density among

Table 1
Network properties of the constructed chronnet.

	Mean	Min	Max	S.D.
Degrees (1×10^2)	1.202	0.000	8.160	1.561
Strengths (1×10^4)	2.281	0.000	49.487	4.424
Clustering coefficient	0.332	0.000	1.000	0.181
Degree centrality	0.007	0.000	0.048	0.009
Betweenness centrality	0.000	0.000	0.000	0.000
Closeness centrality	0.004	0.000	0.036	0.006

neighborhoods.

$$c_u = \begin{cases} \dfrac{2T(u)}{k(u)\left(k(u)-1\right) - 2k^{\leftrightarrow}(u)}, & k(u) \geq 2 \\ 0, & else \end{cases} \tag{11}$$

(3) Degree centrality

For node u, supposing the number of its neighbors is n_u, the degree centrality (i.e., DC_u) is defined as the following equation, where N is the number of total nodes.

$$DC_u = \frac{n_u}{N} \tag{12}$$

(4) Betweenness centrality

Betweenness centrality aims to represent how many shortest paths pass through a specific node. Supposing $\sigma(s,t)$ denotes the number of shortest paths from node s to node t, the betweenness centrality of node u can be mathematically expressed by Eq. (13). In this equation, V denote the node set, and $\sigma(s,t|v)$ represent the number of shortest paths passing through node u.

$$BC_u = \sum_{(s,t)\in V} \frac{\sigma(s,t|u)}{\sigma(s,t)} \tag{13}$$

(5) Closeness centrality

Closeness centrality is defined as the reciprocal of the average distance from node u to the remaining nodes in the network. Supposing node u has $n-1$ reachable nodes, the calculation principle of this metric can be summarized below [47].

$$CC_u = \frac{n-1}{N-1} \cdot \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)} \tag{14}$$

Table 1 summarizes the properties of the constructed chronnet. From the results, two conclusions can be intuitively extracted. First, according to the centrality measures, almost all the centralities are close to zero. It indicates that rare road segments will severely impact the whole road network. This is because the influence of traffic congestion is easily diminished with distance. Meanwhile, we can obtain another finding from the clustering coefficient, i.e., there are apparent clustering structures in this complex network. This metric in the chronnet is relatively high, indicating that there are strong interactions among the neighbors of each node. In other words, it means that if two nodes have the same neighborhood, congestion propagation is likely to occur between these two road segments. Therefore, these results indicate that traffic congestion has typical local propagation patterns.

4.2. Degree and strength analysis

4.2.1. Degree distribution

The degree distribution is a valuable metric that can distinguish different network structures and understand network properties. Since we transform the consecutive patterns of traffic congestions into a directed weighted graph, it is vital to explore the distribution characteristics of in-degree and out-degree, respectively. From the traffic significance, in-degree and out-degree in the chronnet can further reflect the propagation direction of traffic congestions. For instance, if a node has a high in-degree, that means traffic congestions on other road segments are likely to spread to this node. Figs. 8 and 9 display the statistical and spatial distributions of node degrees, respectively. From these figures, several interesting conclusions can be summarized.

(1) The statistical distribution of traffic congestions (i.e., Fig. 8a) is similar to degree distribution (i.e., Fig. 8b). Thus, it indicates that the degree distribution of the chronnet can effectively capture that of traffic congestion. Meanwhile, the probability distributions indicate that both in-degree and out-degree approximately follow the power-law distribution in the early stage. This phenomenon reveals that the constructed chronnet has a typical feature of scale-free networks. Thus, only a few nodes have high degrees, while there are numerous nodes with low degrees. In other words, several

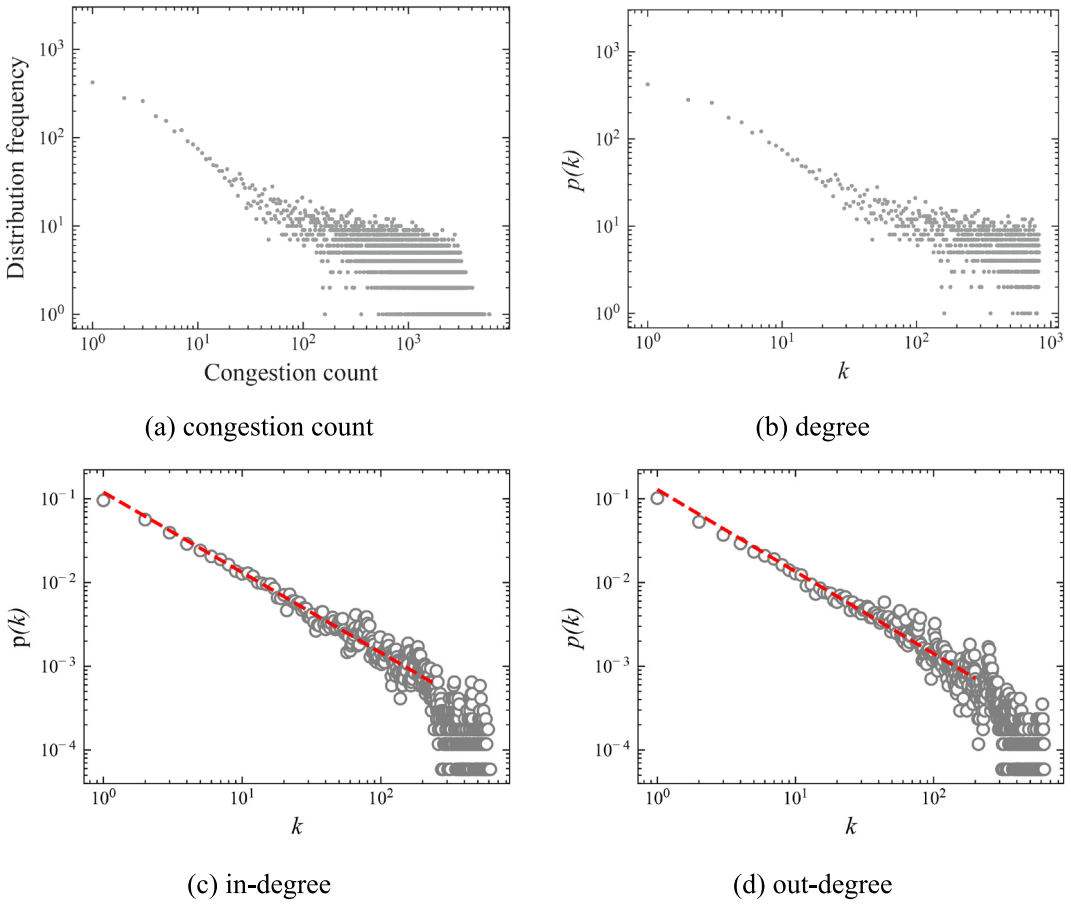


Fig. 8. Statistical distributions of congestion counts and node degrees.

road segments play a role as the hubs in the chronnet, and it always propagates congestion to other nodes. Therefore, traffic conditions will significantly improve if we can identify these hubs and take the corresponding measures to cut the propagation.

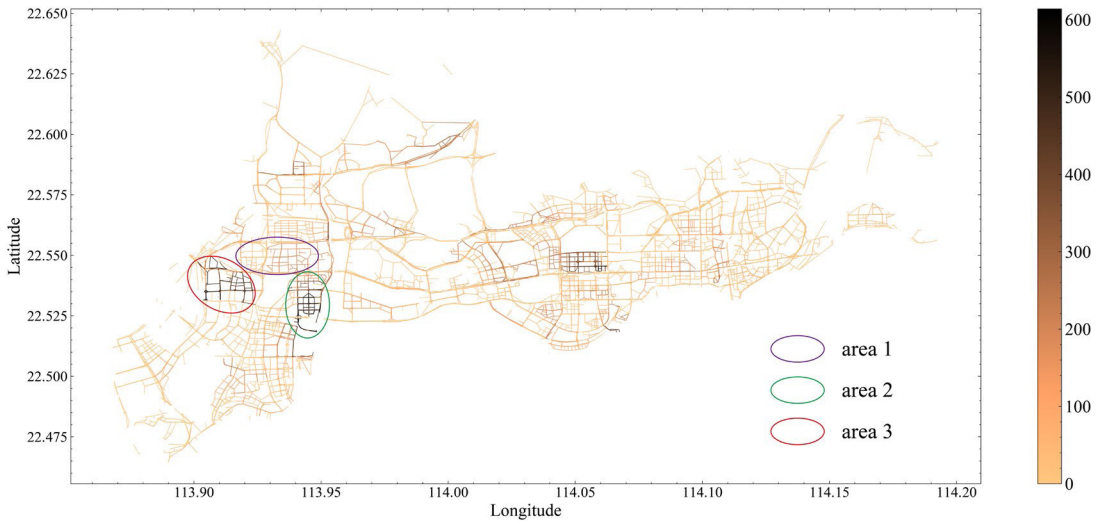
(2) Node degree is a widely-used factor for critical node identification [48]. From the spatial distribution of node degrees, we can visually identify the bottlenecks in such a large-scale road network. Compared with the congestion count distribution shown in Fig. 6, the chronnet significantly highlights the importance of critical road segments. Therefore, it reveals that the chronnet can effectively capture the hidden spatiotemporal patterns of traffic congestion.

(3) As mentioned above, in-degree and out-degree can represent the propagation directions of traffic congestion. Following this opinion, Fig. 9 further expresses the propagation directions from the macro level. In Fig. 9a, road segments in area 1 have low in-degrees, and those in area 2 and 3 are higher. However, as shown in Fig. 9b, the opposite situation happens in the out-degree. This phenomenon reveals that numerous traffic congestion originates from area 1. Since its neighbor areas (i.e., area 2 and area 3) have high in-degrees, it means that they always receive congestion. So, the propagation direction of traffic congestion is likely to be from area 1 to area 2 and area 3. Although all of these areas always suffer the impacts of traffic congestion, the reasons for this influence are different. Therefore, it may be more helpful to relieve traffic pressure by conducting effective measures in the origin of traffic congestions, i.e., area 1.

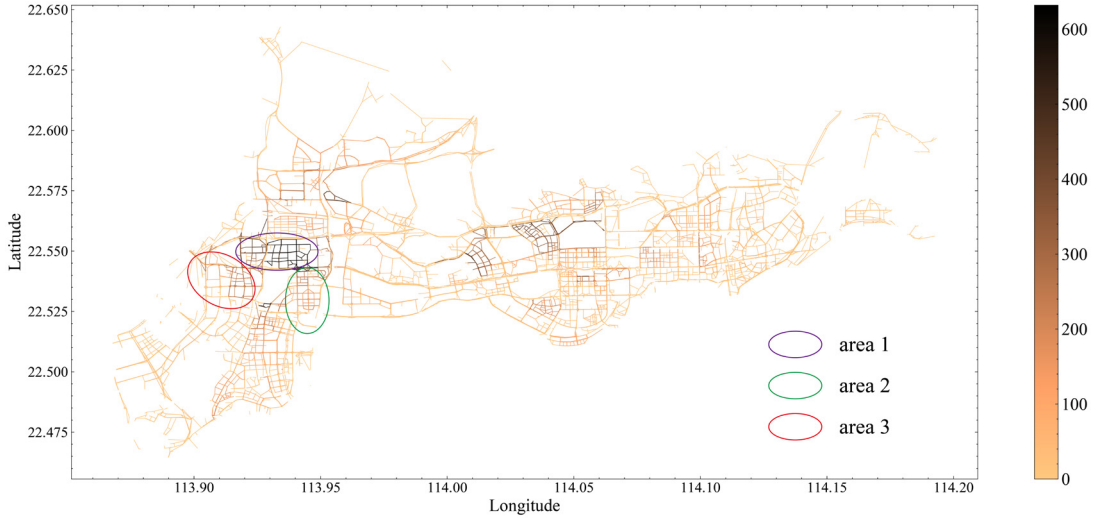
4.2.2. Strength distribution

In the chronnet, degree and strength can denote the interaction with other road segments from different aspects. Therefore, after the exploration of degree distributions, Figs. 10a and 10b illustrate the spatial distribution of in-strength and out-strength in the study area, respectively. Here, the inner plots show the statistical distributions of node strengths. As mentioned above, the strength of each node denotes its interaction intensity with other nodes. Hence, we can also regard the in-strength and out-strength as the frequencies of incoming and outgoing traffic congestion.

Overall, this figure also expresses similar opinions to the degree distributions mentioned above. From the inner plots, we can find that the node strengths in the constructed chronnet also have typical heavy-tailed characteristics. It reveals that there exist several traffic congestion hubs in the urban road network. Although the proportion of these hubs is



(a) in-degree



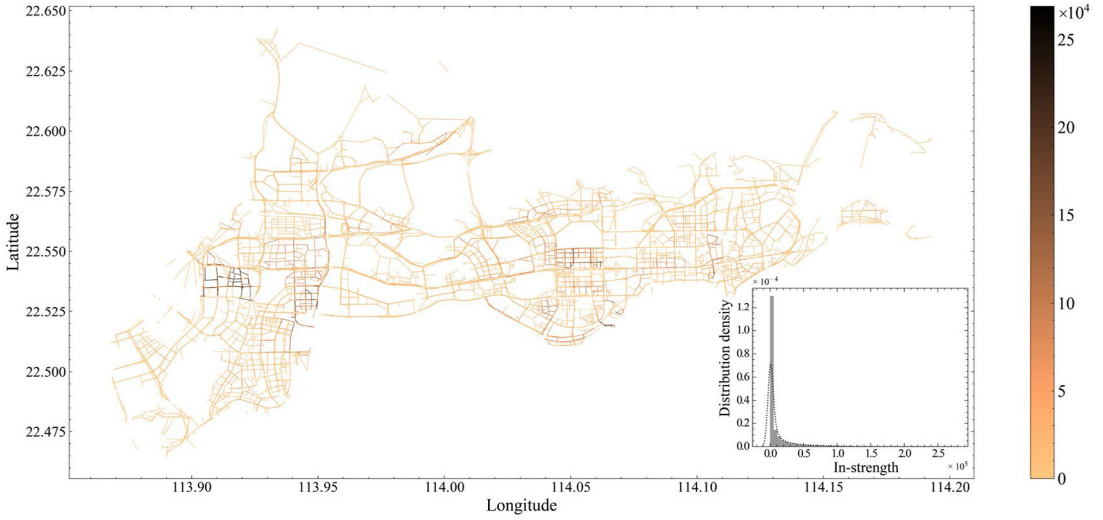
(b) out-degree

Fig. 9. Spatial distributions of node degrees.

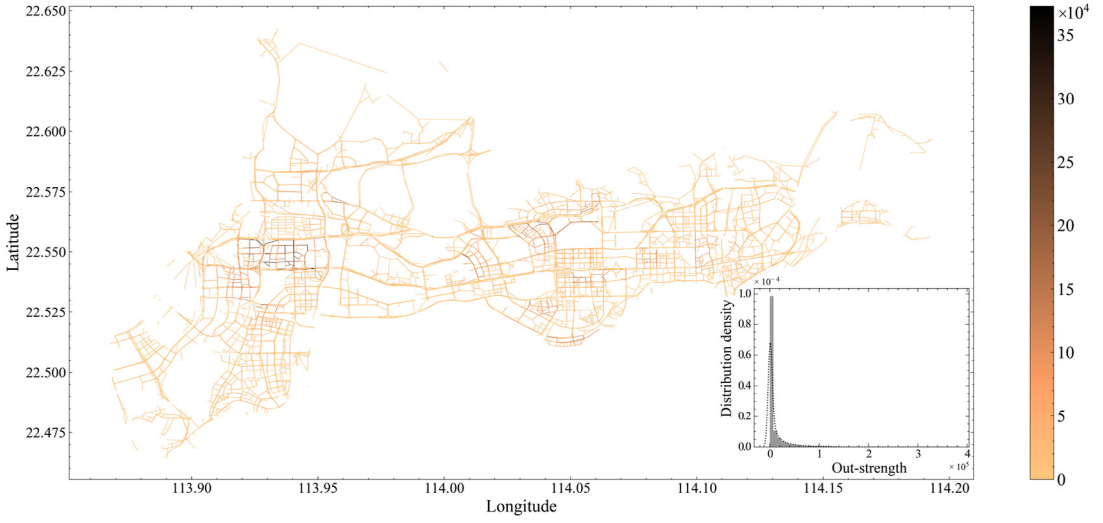
deficient, they always play a critical role in congestion propagation. Meanwhile, according to the spatial distributions of node strengths, bottleneck regions and propagation directions of traffic congestion also express similar trends with degree distributions.

Meanwhile, we can further identify the role of each road segment in the congestion propagation process. Specifically, for a road segment with high in-strength and low out-strength, we can treat it as a critical node in alleviating congestion. Otherwise, we can regard it may exacerbate traffic congestion. Following this opinion, we calculate $\Delta s_i = s_i^{out} - s_i^{in}$ on all the road segments and employ the Gaussian Mixture Model (GMM) to find cluster structures based on this measure. Fig. 11 shows the spatial distribution of the clustering results, where the inner plot displays the scatter between in-strength and out-strength.

It is worth noting that the original results of GMM are symmetrical, i.e., cluster 1 and cluster 4 are divided into the same cluster. Considering the traffic significance in congestion propagation, we distinguish them into different clusters and merge several clusters with similar distribution characteristics. Finally, as shown in this figure, the roles of road segments in congestion propagation can be classified into four clusters. For instance, cluster 1 (i.e., black nodes) and 4 (i.e., green nodes) denote road segments that can significantly alleviate and exacerbate congestion, respectively. Therefore, transportation administration should focus on relieving pressures on road segments of cluster 1. Meanwhile, scatters in cluster 2 and 3 are closely distributed around the baseline (i.e., red lines), indicating that the out-strengths of these nodes



(a) in-strength



(b) out-strength

Fig. 10. Distribution of strengths in the study area.

are almost equal to the in-strengths. Hence, we can regard these nodes as transfer nodes in the propagation process. Furthermore, from the spatial distribution of clusters in Fig. 11, we can find that these transfer nodes are mainly located on arterial roads. In contrast, road segments in urban blocks are more likely to affect the propagation process of traffic congestion.

4.3. Influence scope analysis

In this subsection, we explore the spatial influence scope of traffic congestions based on the chronnet. Fig. 12 displays the relationship between distance and edge features in this complex network. Specifically, Fig. 12a expresses the statistical distribution of the real-world distance between two connected nodes. And in Fig. 12b, the x-axis denotes distance among road segments, and the y-axis represents the average weight on the corresponding edges. Here, the distance matrix is calculated by the longitude and latitude of the midpoint of each road segment. In these two subfigures, breakpoints for piecewise functions are determined by the open-source python package named Piecewise Linear Functions (*pwlf*) [49].

According to Fig. 12a, four different distribution characteristics are identified, and the slopes of these four lines express a continuously decreased trend with distance. Specifically, the boundaries of these four lines are 675.00 m, 1962.34 m, and 5949.38 m. Since edge in the chronnet means congestion propagation between these two road segments,

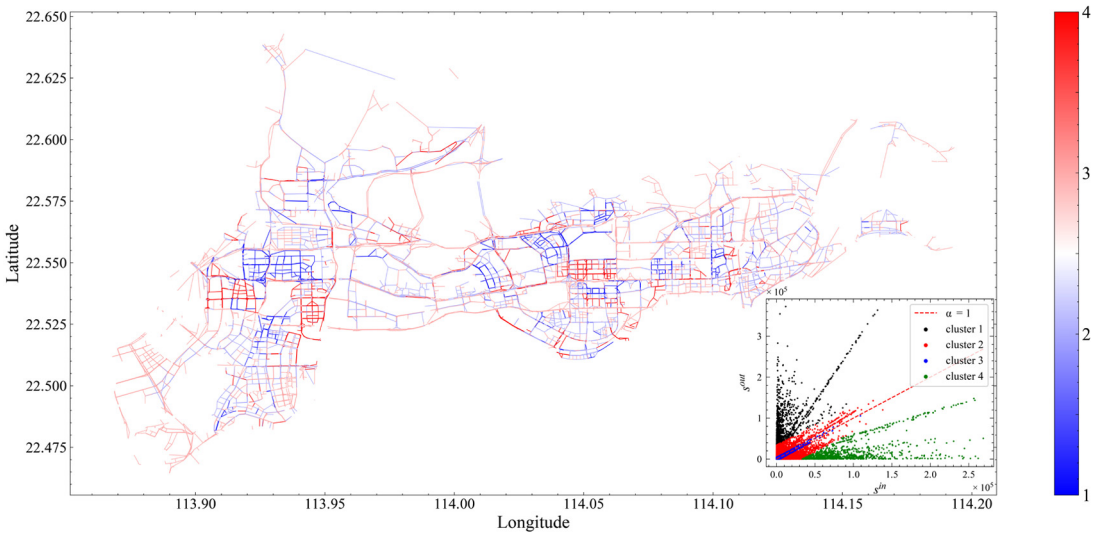
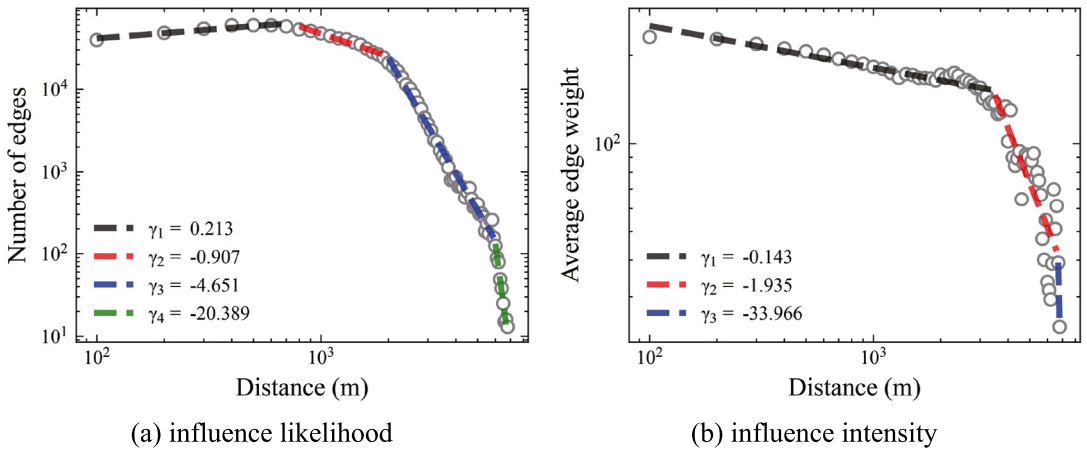


Fig. 11. Distributions of the clustering results based on Δs.



(a) influence likelihood (b) influence intensity

Fig. 12. Influence scope of traffic congestion in the chronnet.

this relationship between distance and edge count is named as influence likelihood. There is an interesting finding that the scope of the fitted line is positive in the first interval, while all the remaining intervals have negative scopes. This phenomenon indicates that the influence likelihood of congestion propagation peaks at 675.00 m. Within this threshold, traffic congestion can continuously spread to other road segments, so that this interval can be regarded as the expansion stage of traffic congestion. After this value, since the surrounding road segments can help relieve traffic pressure, the impact between two road segments with far distance will weaken. With the increment of distance, it becomes more and more difficult to spread congestion, so the slopes of fitted lines sharply drop.

Moreover, we can draw a similar conclusion from the influence intensity variation shown in Fig. 12b. In this subfigure, the scatters can be fitted by three piecewise lines, and the slopes of these lines can be obtained from the legend. These results are consistent with the propagation patterns of traffic congestion, i.e., it is likely to spread to the surrounding road segments during a short time. Similar to Fig. 12a, the average weights on the edges and line slopes show an overall downward trend. Since surrounding roads act as a bridge in congestion propagation between distant road segments, the influence intensity on the nearby roads is stronger than others. Therefore, the interaction intensity among road segments significantly decreases with distance increment, leading to sharply descending edge weights.

Furthermore, a difference also exists in these two subfigures. In Fig. 12b, all the fitted slopes are negative, demonstrating that the average edge weight negatively correlates with distance. Oppositely, the scope of the first interval in Fig. 12a is positive. Therefore, we can conclude that although traffic congestion is more likely to propagate to road segments within 675 m, influence intensities also decrease with distance.

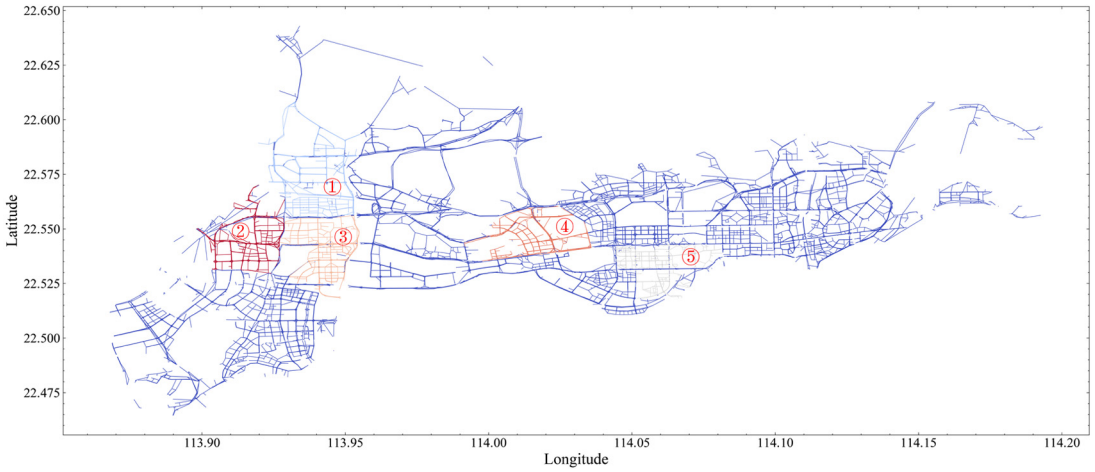


Fig. 13. Spatial distribution of the top 5 communities.

4.4. Results of bottleneck identification

According to the spatial distributions of degrees and strengths, nodes with extremely high degrees (or strengths) are closely located, e.g., area 1 in Fig. 9b. Here, we further explore the local clustering structures of traffic congestion by the overlapping community detection algorithm.

Fig. 13 displays the spatial distributions of the top five communities. It is worth noting that following the settings in [39], we set the values of the distance threshold $d_{\max}$ as $+\infty$, so almost all the consecutive congestions are possible to be connected. Interestingly, even though only the node strengths are employed in the community detection algorithm, the results in Fig. 13 still reveal that road segments of each community are still located in the same local area. It indicates that the constructed chronnet can effectively reflect the actual evolution patterns of traffic congestion. The main reason for this phenomenon is that traffic congestion is always spread locally in both the spatial and temporal dimensions. Hence, consecutive congestions are more likely to occur on the surrounding road segments with the congested road segments, making each local area has its unique evolution pattern. Overall, although the chronnet method seems simple, its principle is consistent with traffic congestion. Therefore, it is correct and reasonable to apply it for congestion modeling.

Meanwhile, we can extract another interesting finding from this figure. The spatial distributions of top communities are similar to node degrees and strengths. For instance, area 2 and 3 in Fig. 9 are almost extracted as community 2 and 3 in Fig. 13, respectively. It reveals that there are strong local connections within congestion areas. Specifically, since there are significant congestion interactions between area 1 and area 2, area 2 and the right part of area 1 are clustered into the same community. Similarly, this phenomenon can also be found in area 3 and the left part of area 1. Therefore, this finding further demonstrates the evolution process described in Section 4.2.1.

According to the construction process of the chronnet, communities denote node groups where spatiotemporal events often occur in the same period [39]. In this study, it means that traffic congestion in each community may have its unique time-varying characteristics. Therefore, we illustrate the temporal evolution patterns of traffic congestions in the top five communities in Fig. 14a. In this figure, the locations of these five communities match the spatial distribution in Fig. 13, and each line denotes the kernel density estimate (KDE) distribution of congestion time in each community. In this figure, congestion time is defined as the period (i.e., 5-min) when the road segment is congested. This figure indicates that traffic congestions in these communities have different temporal evolution patterns. For example, community 3 is the only community prone to congestion during morning and evening rush hours. This is because numerous companies locate in this area, especially high-tech companies. So, commuters always bring critical traffic pressure to the road network, resulting in significant congestion during rush hours. Besides, we can also find that during the early morning, traffic congestions are more likely to occur in community 2. This phenomenon is because many residential areas distribute in this area. So, vehicles leaving these residential areas in the early morning cause traffic congestion.

Furthermore, we also compare the KDE distributions of daily congestion count, degree, and strength in Fig. 14. These figures show several differences among different communities. Based on Fig. 14b, we can find that the daily congestion count of community 3 is similar to community 4, and both of them are always at a low level compared to other communities. However, the degrees of road segments in community 3 are much higher than in community 4 and other communities, and strength distribution also holds the same view. Therefore, it means traffic congestions in community 3 have more powerful propagation capabilities. Overall, the community analysis results demonstrate that the chronnet can effectively identify consecutive congestions and have a typical community structure. Meanwhile, the applied community detection method can also distinguish the different characteristics of different areas.

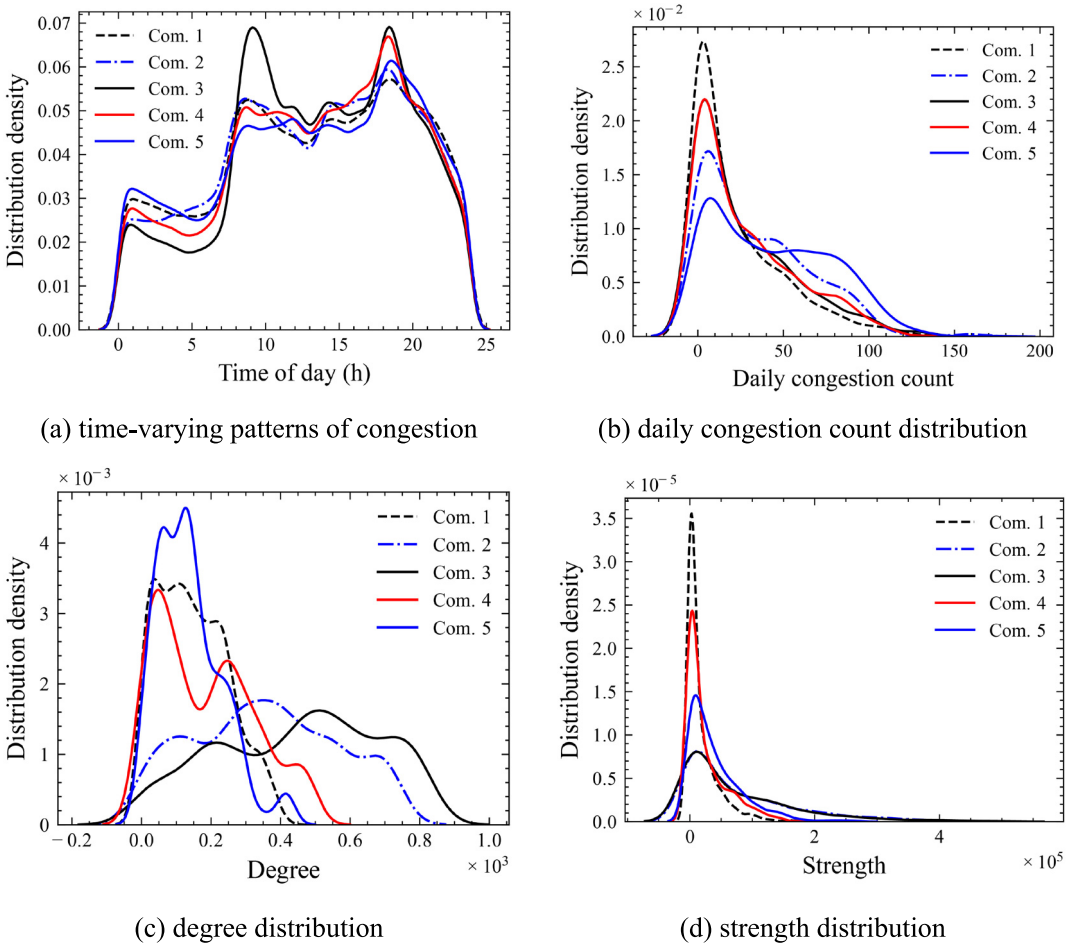


Fig. 14. KDE distributions of network properties in the top 5 communities.

4.5. Comparisons between weekdays and weekends

Due to commute flow, travel patterns on weekdays and weekends always express different characteristics. In this subsection, we further explore their different congestion patterns. It is worth noting that the Chinese New Year holiday is included in this dataset, so residents worked on February 2nd and 3rd (i.e., Saturday and Sunday) and then took holiday in the following week. Hence, we regard traffic congestion on this holiday as weekends. In this way, we utilize traffic speed on weekdays and weekends to construct two chronnets, respectively.

Fig. 15a shows the daily congestion count distribution during weekdays and weekends. According to this figure, the average daily congestion counts on weekday and weekend are similarly distributed. Meanwhile, we can also see that 60 is a demarcation point. Distribution density of traffic congestion on weekends always higher than weekdays before 60, while the opposite occurs after 60. This phenomenon indicates that it is more likely to occur serious congestions on weekdays. However, according to the $\Delta s'$ distribution shown in Fig. 15b, it reveals that both in-strength and out-strength on weekdays are much higher than that on weekends. Here, $\Delta s'_i$ stands the difference between node strength on weekdays and weekends. It means that although traffic congestion count between weekdays and weekends is not significantly different, there exist stronger interactions between congested road segments on weekdays.

Furthermore, the spatial distribution of $\Delta s'$ is displayed in Fig. 16. In order to reduce the visual impact of extreme values, we set a threshold (i.e., $\Delta s'[\Delta s' > 5000] = 5000$) to draw this figure. From this figure we can obtain that there exist several extreme congested areas on weekdays. Meanwhile, it is interesting to find that the most congested area is also similar with the community 2 and community 3 in Fig. 13. As mentioned in Section 4.4, numerous high-tech companies, including the Shenzhen High-tech Industrial Park, locate in this area, making the road network extremely busy on weekdays. Hence, this area always suffers the significant pressure of logistics and commuter flow. But in the weekends, this phenomenon is significantly alleviated, so the big difference appears between weekdays and weekends.

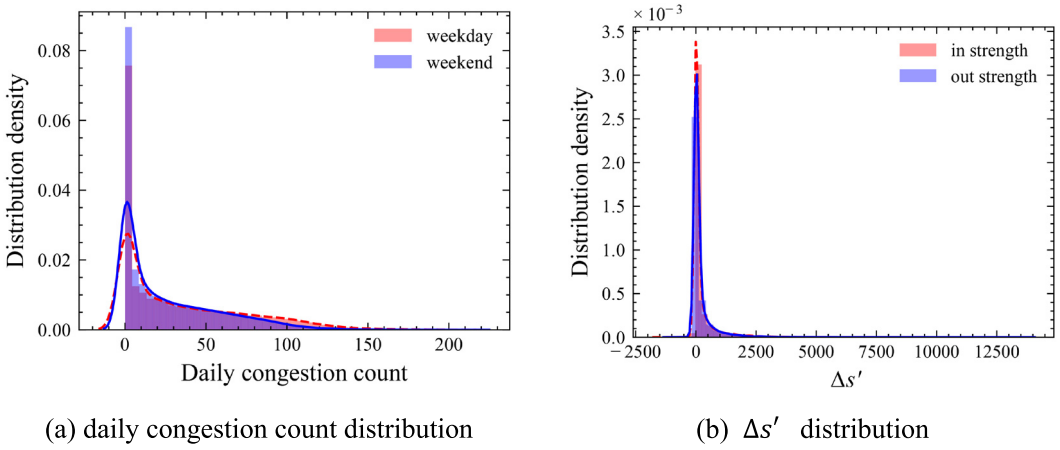


Fig. 15. Congestion characteristics comparisons between weekdays and weekends.

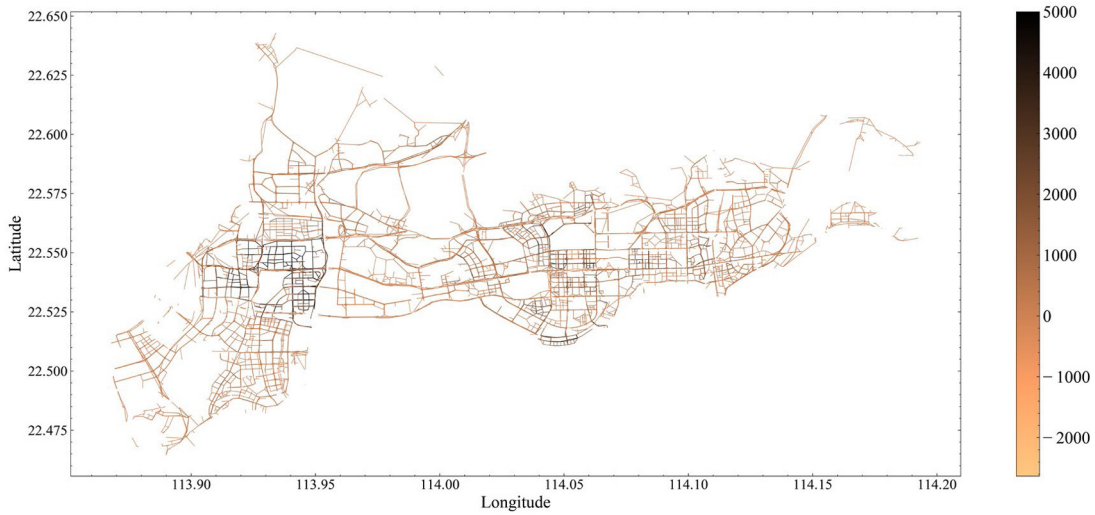


Fig. 16. Spatial distribution of $\Delta s'$ in the selected road network.

5. Conclusions

This study regard traffic congestions as spatiotemporal events and employ a complex network method to uncover their propagation patterns. A two-stage map-matching method is developed to bridge taxi trajectory with the road network. Based on the extracted traffic speed on each road segment, we define a free-flow speed and employ it to determine traffic congestion. In this way, the chronnet method is utilized to explore the consecutive patterns in network-scale traffic congestions. The main conclusion of this study is summarized below.

(1) This study provides new insight into traffic congestion modeling in the large-scale road network. It demonstrates that it is suitable to regard traffic congestions as spatiotemporal events and explore their spatiotemporal patterns from the aspect of complex networks.

(2) Both the degree and strengths follow the long-tailed distribution. It means that there are several hubs for congestion propagation in the urban road networks. Meanwhile, these two metrics can help identify the role of road segments in congestion propagation. According to the influence scope analysis, traffic congestions are more likely to spread to the road segments within 675.00 m, but the influence intensity always continues to decrease with distance.

(3) The community detection results indicate that traffic congestions likely express apparent local clustering structures. The KDE distribution of traffic features demonstrates congestion in each community has its unique consecutive characteristics.

(4) It is interesting to find that there are rarely significant differences in traffic congestion counts between weekends and weekdays. However, traffic congestions on the weekdays have a more substantial influence on the surrounding road segments.

From the overview of this study, there still exist several limitations. For instance, since taxis do not account for a high proportion of motor vehicles, their trajectory data is always sparsely distributed in the urban road network. Due to critical barriers in data collection, only taxi trajectories are involved in this study. Future studies may consider introducing multi-source traffic data and emerging technologies [36] in traffic congestion analysis [50]. Meanwhile, since traffic congestion has different levels, classifying it into different levels and integrating it into the chronnet construction method may also be an interesting direction for future works. Furthermore, complex networks have been rapidly developed in recent years. Future works can pay attention to other advanced complex network methods [38,51] for congestion modeling.

CRediT authorship contribution statement

Jie Zeng: Conceptualization, Methodology, Software, Writing – original draft. **Yong Xiong:** Data curation, Writing – review & editing, Methodology. **Feiyang Liu:** Conceptualization, Data curation, Software. **Junqing Ye:** Writing – review & editing, Visualization. **Jinjun Tang:** Methodology, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This research is funded by the National Natural Science Foundation of China (No. 52172310), Humanities and Social Sciences Foundation of the Ministry of Education, China (No. 21YJCZH147), Innovation-Driven Project of Central South University, China (No. 2020CX041).

References

[1] S. Spana, L. Du, Optimal information perturbation for traffic congestion mitigation: Gaussian process regression and optimization, Transp. Res. Part C Emerg. Technol. 138 (2021) 103647, http://dx.doi.org/10.1016/j.trc.2022.103647, 2022.
[2] Y. Liu, Z. Yang, Information provision and congestion pricing in a risky two-route network with heterogeneous travelers, Transp. Res. Part C Emerg. Technol. 128 (March) (2021) 103083, http://dx.doi.org/10.1016/j.trc.2021.103083.
[3] J.J. Sánchez-Medina, M.J. Galán-Moreno, E. Rubio-Royo, Traffic signal optimization in la almozara district in saragossa under congestion conditions, using genetic algorithms, traffic microsimulation, and cluster computing, IEEE Trans. Intell. Transp. Syst. 11 (1) (2010) 132–141, http://dx.doi.org/10.1109/TITS.2009.2034383.
[4] Z. Li, M. Shahidehpour, S. Bahramirad, A. Khodaei, Optimizing traffic signal settings in smart cities, IEEE Trans. Smart Grid 8 (5) (2017) 2382–2393, http://dx.doi.org/10.1109/TSG.2016.2526032.
[5] S. Vosough, A. de Palma, R. Lindsey, Pricing vehicle emissions and congestion externalities using a dynamic traffic network simulator, Transp. Res. Part A Policy Pract. 161 (2) (2022) 1–24, http://dx.doi.org/10.1016/j.tra.2022.04.017.
[6] A. de Palma, R. Lindsey, Traffic congestion pricing methodologies and technologies, Transp. Res. Part C Emerg. Technol. 19 (6) (2011) 1377–1399, http://dx.doi.org/10.1016/j.trc.2011.02.010.
[7] L. Lehe, Downtown congestion pricing in practice, Transp. Res. Part C Emerg. Technol. 100 (2018) (2019) 200–223, http://dx.doi.org/10.1016/j.trc.2019.01.020.
[8] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, Y. Sugiyama, Dynamical model of traffic congestion and numerical simulation, Phys. Rev. E 51 (2) (1995) 1035–1042, http://dx.doi.org/10.1103/PhysRevE.51.1035.
[9] L.E. Olmos, S. Çolak, S. Shafiei, M. Saberi, M.C. González, Macroscopic dynamics and the collapse of urban traffic, Proc. Natl. Acad. Sci. USA 115 (50) (2018) 12654–12661, http://dx.doi.org/10.1073/pnas.1800474115.
[10] M. Saberi, et al., A simple contagion process describes spreading of traffic jams in urban networks, Nature Commun. 11 (1) (2020) 1–9, http://dx.doi.org/10.1038/s41467-020-15353-2.
[11] Z. Liu, Y. Liu, J. Wang, W. Deng, Modeling and simulating traffic congestion propagation in connected vehicles driven by temporal and spatial preference, Wirel. Netw. 22 (4) (2016) 1121–1131, http://dx.doi.org/10.1007/s11276-015-1021-1.
[12] T. Tsekeris, N. Geroliminis, City size, network structure and traffic congestion, J. Urban Econ. 76 (1) (2013) 1–14, http://dx.doi.org/10.1016/j.jue.2013.01.002.
[13] R. Arnott, E. Inci, The stability of downtown parking and traffic congestion, J. Urban Econ. 68 (3) (2010) 260–276, http://dx.doi.org/10.1016/j.jue.2010.05.001.
[14] Z. Kan, L. Tang, M.P. Kwan, C. Ren, D. Liu, Q. Li, Traffic congestion analysis at the turn level using Taxis' GPS trajectory data, Comput. Environ. Urban Syst. 74 (2018) (2019) 229–243, http://dx.doi.org/10.1016/j.compenvurbsys.2018.11.007.
[15] X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, B. Zhang, Urban traffic congestion estimation and prediction based on floating car trajectory data, Futur. Gener. Comput. Syst. 61 (2016) 97–107, http://dx.doi.org/10.1016/j.future.2015.11.013.
[16] F. Zong, T. Wu, H. Jia, Taxi drivers' cruising patterns-insights from Taxi GPS traces, IEEE Trans. Intell. Transp. Syst. 20 (2) (2019) 571–582, http://dx.doi.org/10.1109/TITS.2018.2816938.
[17] Z. Yang, M.L. Franz, S. Zhu, J. Mahmoudi, A. Nasri, L. Zhang, Analysis of Washington, DC taxi demand using GPS and land-use data, J. Transp. Geogr. 66 (January) (2018) 35–44, http://dx.doi.org/10.1016/j.jtrangeo.2017.10.021.
[18] T. Seo, T. Kusakabe, Probe vehicle-based traffic state estimation method with spacing information and conservation law, Transp. Res. Part C Emerg. Technol. 59 (2015) 391–403, http://dx.doi.org/10.1016/j.trc.2015.05.019.

[19] J. Tang, F. Liu, Y. Wang, H. Wang, Uncovering urban human mobility from large scale taxi GPS data, Phys. A Stat. Mech. Appl. 438 (2015) 140–153, http://dx.doi.org/10.1016/j.physa.2015.06.032.
[20] A.S. Dokuz, Weighted spatio-temporal taxi trajectory big data mining for regional traffic estimation, Phys. A Stat. Mech. Appl. 589 (2022) 126645, http://dx.doi.org/10.1016/j.physa.2021.126645.
[21] A.S. Dokuz, Fast and efficient discovery of key bike stations in bike sharing systems big datasets, Expert Syst. Appl. 172 (2020) 114659, http://dx.doi.org/10.1016/j.eswa.2021.114659, 2021.
[22] S. Luan, R. Ke, Z. Huang, X. Ma, Traffic congestion propagation inference using dynamic Bayesian graph convolution network, Transp. Res. Part C Emerg. Technol. 135 (2021) 103526, http://dx.doi.org/10.1016/j.trc.2021.103526, 2022.
[23] T. Erdelić, T. Carić, M. Erdelić, L. Tišljarić, A. Turković, N. Jelušić, Estimating congestion zones and travel time indexes based on the floating car data, Comput. Environ. Urban Syst. 87 (March) (2021) http://dx.doi.org/10.1016/j.compenvurbsys.2021.101604.
[24] Y. Huang, D. (Jian) Sun, L.H. Zhang, Effects of congestion on drivers' speed choice: Assessing the mediating role of state aggressiveness based on taxi floating car data, Accid. Anal. Prev. 117 (800) (2018) 318–327, http://dx.doi.org/10.1016/j.aap.2018.04.030.
[25] X. Qian, T. Lei, J. Xue, Z. Lei, S.V. Ukkusuri, Impact of transportation network companies on urban congestion: Evidence from large-scale trajectory data, Sustain. Cities Soc. 55 (January) (2020) 102053, http://dx.doi.org/10.1016/j.scs.2020.102053.
[26] Z. Kan, M.P. Kwan, D. Liu, L. Tang, Y. Chen, M. Fang, Assessing individual activity-related exposures to traffic congestion using GPS trajectory data, J. Transp. Geogr. 98 (2021) 103240, http://dx.doi.org/10.1016/j.jtrangeo.2021.103240, 2022.
[27] L. Yan, H. Shen, K. Chen, MobiT: Distributed and congestion-resilient trajectory-based routing for vehicular delay tolerant networks, IEEE/ACM Trans. Netw. 26 (3) (2018) 1078–1091, http://dx.doi.org/10.1109/TNET.2018.2812169.
[28] Z. Wang, M. Lu, X. Yuan, J. Zhang, H. Van De Wetering, Visual traffic jam analysis based on trajectory data, IEEE Trans. Vis. Comput. Graphics 19 (12) (2013) 2159–2168, http://dx.doi.org/10.1109/TVCG.2013.228.
[29] Z. He, G. Qi, L. Lu, Y. Chen, Network-wide identification of turn-level intersection congestion using only low-frequency probe vehicle data, Transp. Res. Part C Emerg. Technol. 108 (October) (2019) 320–339, http://dx.doi.org/10.1016/j.trc.2019.10.001.
[30] D. Li, et al., Percolation transition in dynamical traffic network with evolving critical bottlenecks, Proc. Natl. Acad. Sci. USA 112 (3) (2015) 669–672, http://dx.doi.org/10.1073/pnas.1419185112.
[31] S. An, H. Yang, J. Wang, N. Cui, J. Cui, Mining urban recurrent congestion evolution patterns from GPS-equipped vehicle mobility data, Inf. Sci. (Ny). 373 (2016) 515–526, http://dx.doi.org/10.1016/j.ins.2016.06.033.
[32] S. Çolak, A. Lima, M.C. González, Understanding congested travel in urban areas, Nature Commun. 7 (2016) http://dx.doi.org/10.1038/ncomms10793.
[33] L. Zhao, et al., T-GCN: A temporal graph convolutional network for traffic prediction, IEEE Trans. Intell. Transp. Syst. (2019) 1–11.
[34] Z.-C. Li, X.-Y. Ge, Traffic signal timing problems with environmental and equity considerations, J. Adv. Transp. 48 (8) (2014) 1066–1086, http://dx.doi.org/10.1002/atr.1246.
[35] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, Y. Huang, Map-matching for low-sampling-rate GPS trajectories, in: GIS Proc. ACM Int. Symp. Adv. Geogr. Inf. Syst., 2009, pp. 352–361, http://dx.doi.org/10.1145/1653771.1653820.
[36] P. Wang, J. Lai, Z. Huang, Q. Tan, T. Lin, Estimating traffic flow in large road networks based on multi-source traffic data, IEEE Trans. Intell. Transp. Syst. (2020) 1–12, http://dx.doi.org/10.1109/TITS.2020.2988801.
[37] F. Falasca, A. Bracco, A. Nenes, I. Fountalis, Dimensionality reduction and network inference for climate data using δ-MAPS: Application to the CESM large ensemble sea surface temperature, J. Adv. Model. Earth Syst. 11 (6) (2019) 1479–1515, http://dx.doi.org/10.1029/2019MS001654.
[38] N. Boers, B. Goswami, A. Rheinwalt, B. Bookhagen, B. Hoskins, J. Kurths, Complex networks reveal global pattern of extreme-rainfall teleconnections, Nature 566 (7744) (2019) 373–377, http://dx.doi.org/10.1038/s41586-018-0872-x.
[39] L.N. Ferreira, et al., Spatiotemporal data a nalysis with chronological networks, Nature Commun. 11 (1) (2020) 4036, http://dx.doi.org/10.1038/s41467-020-17634-2.
[40] A. Epasto, S. Lattanzi, R. Paes Leme, Ego-splitting framework: from non-overlapping to overlapping clusters, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug, 2017, pp. 145–154, http://dx.doi.org/10.1145/3097983.3098054.
[41] B. Rozemberczki, O. Kiss, R. Sarkar, Karate club, 2020, pp. 3125–3132, http://dx.doi.org/10.1145/3340531.3412757.
[42] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) 2008, http://dx.doi.org/10.1088/1742-5468/2008/10/P10008.
[43] M.E.J. Newman, Modularity and community structure in networks, Proc. Natl. Acad. Sci. USA 103 (23) (2006) 8577–8582, http://dx.doi.org/10.1073/pnas.0601602103.
[44] Y. Guo, S.Y. He, Built environment effects on the integration of dockless bike-sharing and the metro, Transp. Res. Part D Transp. Environ. 83 (2020) http://dx.doi.org/10.1016/j.trd.2020.102335.
[45] J. Liu, G.P. Ong, X. Chen, GraphSAGE-Based traffic speed forecasting for segment network with sparse data, IEEE Trans. Intell. Transp. Syst. April (2020) 1–12, http://dx.doi.org/10.1109/TITS.2020.3026025.
[46] G. Fagiolo, Clustering in complex directed networks, Phys. Rev. E - Stat. Nonlinear, Soft Matter. Phys. 76 (2) (2007) 1–8, http://dx.doi.org/10.1103/PhysRevE.76.026107.
[47] S. Wasserman, K. Faust, Social Network Analysis, Cambridge University Press, 1994.
[48] Z. Du, J. Tang, Y. Qi, Y. Wang, C. Han, Y. Yang, Identifying critical nodes in metro network considering topological potential: A case study in shenzhen city—China, Phys. A Stat. Mech. Appl. 539 (2020) 122926, http://dx.doi.org/10.1016/j.physa.2019.122926.
[49] C.F. Jekel, G. Venter, Piecewise linear fitting : A python library for fitting 1D continuous piecewise linear functions, no. March, 2019, pp. 1–15, [Online]. Available: https://github.com/cjekel/piecewise_linear_fit_py.
[50] C. Ma, J. Zhou, X.D. Xu, J. Xu, Evolution regularity mining and gating control method of urban recurrent traffic congestion: A literature review, J. Adv. Transp. 2020 (2020) 1–13, http://dx.doi.org/10.1155/2020/5261580.
[51] N. Boers, B. Bookhagen, N. Marwan, J. Kurths, J. Marengo, Complex networks identify spatial patterns of extreme rainfall events of the South American Monsoon System, Geophys. Res. Lett. 40 (16) (2013) 4386–4392, http://dx.doi.org/10.1002/grl.50681.