ORIGINAL PAPER

Dynamic partitioning of urban traffic network sub-regions with spatiotemporal evolution of traffic flow

Fei Yan · Man Zhang · Zhongke Shi

Abstract The spatial partitioning problem of urban traffic network sub-regions has been mainly researched in static mode by considering traffic conditions at a certain time. Nevertheless, traffic flow has strong dynamic characteristics running in urban network; it is necessary to divide the sub-regions according to the correlation of traffic flow in spatiotemporal dimension. Dynamic partitioning is of great significance for capturing the evolution of traffic sub-region and designing the peripheral control strategies in real time. In this paper, a large heterogeneous traffic network is divided into several uniform sub-regions by analyzing the temporal similarity and spatial heterogeneity of the link characteristics to reveal the evolution process of traffic flow. Firstly, a similarity model is defined by the analysis of link autocorrelation based on Moran's I index, and the slow coherency theory combining with α-Cut theory (called α-SC) is implemented to establish the framework of static partitioning. Secondly, a fine-tuning procedure is designed to maintain the homogeneity and compactness of the sub-regions. Finally, the links with high heterogeneity in the sub-regions are identified by heuristic algorithm and then iteratively adjusted, merged and cut to realize the dynamic updating of sub-regions. A case study using the real data set of Farmers Branch city is conducted to verify the effectiveness of the proposed method.

Keywords Traffic network · Spatial partitioning · Congestion evolution · Moran's I index · Slow coherency theory · Dynamic updating

1 Introduction

In urban networks, due to the complexity and unpredictability of people's travel routes, traffic congestion increasingly grows in most cities. The congestion needs to be better managed through effective traffic control strategies or other management measures. To reduce the congestion from a network-wide point of view, network-wide traffic modeling are needed to better coordinate the traffic lights of intersections within the urban networks. Thus, the aggregation level network modeling of urban traffic has recently gained great attention. Geroliminis and Daganzo [1] found that the relationship between the traffic flow and occupancy of a single detector data is highly discrete through real traffic statistic data from Yokohama, Japan. However, when all the detector data of the entire network are aggregated, the relationship between the traffic flow and occupancy is fitted into a curve with a small dispersion, so the concept of macroscopic fundamental diagram (MFD) reflecting the state of network traffic flow is proposed. The impacts of sudden traffic demand,

F. Yan (✉) · M. Zhang
Department of Automation, Taiyuan University of Technology, Taiyuan 030024, China
e-mail: yanfei140222@163.com

Z. Shi
Department of Control Theory and Control Engineering, Northwestern Polytechnical University, Xi'an 710072, China

signal control methods and resource allocation problem on MFD have been investigated in [2–4].

Recent findings have shown that the spatial distribution of traffic flow is one of the most significant factors affecting the shape of MFD [5–7]. Buisson et al. [8] gave a conclusion that heterogeneity of traffic networks has a strong impact on the scatter and the shape of an MFD. This indicates that a well-defined MFD only exists for a traffic network when the traffic flows are homogeneously distributed in the network. This paper focuses on dividing the heterogeneous urban traffic network into multiple sub-regions with uniform density distribution based on the similarity of traffic characteristics between road segments, which provides the premise for MFD-based peripheral control strategies.

There have been several researches on static partitioning of urban traffic network sub-regions recently [9–15]. For example, Ji and Geroliminis [13] proposed a three-module continuous algorithm for traffic network partitioning based on the normalized cut algorithm (Ncut) [14]. Ji et al. [15] used the concept of Maximum Connected Component (MCC) to describe the heterogeneous network as several homogeneous subregions, which visualized the evolution of congestion. Normally, the traffic network is transformed into a road graph for static partitioning [16]. Graph partitioning is a well-studied subject in vast fields such as image segmentation [14], community detection [17] and regional growth [18]. Based on k-means algorithm, Anwar et al. [19] treated the road network as a weighted graph and proposed a k-way α-Cut space partitioning method by applying spectral clustering theory which has good performance in identifying homogeneous components. However, the following issue of k-means is remaining obvious that the selection of the initial clustering center has a great influence of the cluster quality. It also can be known that the solution based on spectral clustering produces good results but reveals high computing complexity [20]. The static partitioning model in this paper can obtain uniform sub-regions with different saturation levels by incorporating slow coherency theory into spectral clustering. Slow coherency theory is a clustering algorithm that theoretically solves the problem of identifying the weakest correlation between links in complex networks [21], avoiding the influence of initial values on the clustering quality.

Due to different traffic volume, number of lanes, signal timing at intersections, etc., the urban network shows different heterogeneity at different time-stamps,

which brings challenges to the correlation measurement between links or signal intersections. Research shows that the traffic flow fluctuation time series derived from adjacent links has strong power-law crosscorrelation [22], so the challenge stimulates the use of the concept of local Moran's I index [23] and the spatial autocorrelation. The local Moran's I index quantifies the linear correlation strength between a link and its neighbors in the same space. This definition is applied to the similarity model to indirectly implement the connection constraints in this paper.

Although there are many literatures on the partition of control sub-regions of urban road network, the partition of control sub-regions is mainly studied in a static mode. The common static partition is only for the problem of road network partition at a certain moment. The static partition at the previous moment is not necessarily valid at the latter moment as the strong time-varying characteristics of traffic flow, and the complete static partitioning of each moment is considered as a computationally intensive task. Since the traffic flow has strong dynamic characteristics in urban road network, it is necessary to divide the sub-regions according to the correlation of the traffic flow in space and time dimensions. The traffic network static partitioning research was extended to the dynamic domain in [24–27]. Saeedmanesh et al. [24] developed two heuristic partitioning algorithms utilizing the defined mixed integer linear optimization model to find the traffic sub-regions with uniform traffic level.

In recent years, some scholars have attempted to use the switched piecewise affine systems modeling approach to solve the partitioning problem of urban road networks, which can approximate nonlinear systems with arbitrary accuracy and is based on partitioning the state space into finite polyhedral regions. In [28], a new class of approximation models for switching nonlinearities was proposed to regulate the switching problems between nonlinear subsystems and within each subsystem. In [29], the optimal hybrid control problem for a regional MFD network was formulated as a mixed integer nonlinear optimization problem, which is reformulated as a mixed integer linear programming problem using a piecewise affine approximation technique. Reference [30] investigated the stability and calming problems of discrete switching systems under an improved S-any switching strategy, completing the design of a state feedback controller for the switching system.

This paper presents a comprehensive framework to capture the evolution of traffic flow by updating the sub-region utilizing static partitioning and dynamic updating procedures. Among them, we first propose two steps in the static partitioning, initial partitioning and fine tuning of the sub-region boundary. In the dynamic partitioning model of dynamic updating procedures, the target links are only a few high-heterogeneity links, which can improve the partition efficiency of the road network. In the model, we firstly utilize a top-down approach to achieve the static partitioning of sub-regions. Secondly, the links with large variance in the sub-region at the previous moment are taken as the decision objects at the next moment by the heuristic method, and they are rearranged through a dynamic updating procedure to minimize the heterogeneity of the sub-regions. During the spatiotemporal evolution of the sub-regions, spatial proximity constraints are explicitly integrated into the dynamic updating model, which can ensure great connectivity between the links within the sub-regions.

The main contributions of this study include as follows:

- A Moran's I index technique is utilized to measure the spatial autocorrelation statistic of links and calculate the link similarity from a neighborhood topology view.
- A Moran's I index-based sub-region initial partitioning algorithm (α-SC) is proposed to extract homogeneous sub-regions from the traffic network.
- Based on the α-SC algorithm, a α-SC-FT algorithm for enforcing the connectivity and uniformity of sub-regions is proposed.
- A dynamic updating procedure is established by using heuristics to effectively identify the sub-regions with different levels of saturation degree.

The rest of the article is organized as follows: Sect. 2 introduces the concept of local Moran's I index. In Sect. 3, the Moran's I index-based traffic sub-region static partitioning algorithm (α-SC) is developed, and the α-SC algorithm is extended to optimize the connectivity of sub-regions by the fine tuning process (α-SC-FT). The static partitioning framework is extended to dynamic partitioning that tracks the evolution of traffic flow in Sect. 4. Section 5 presents the case study and experimental results. Finally, a conclusion and the further research directions are given in Sect. 6.

2 Problem definition and local Moran's I index

2.1 Mathematical representation of traffic network

To introduce the topological structure of the traffic network, we utilize a mathematical representation of the weighted undirected graph. Since the traffic condition of link in the network can reflect the traffic state of the network and could be easily detected, the focus of partitioning of traffic sub-region is on the link, not on the signalized intersection. The actual network could be converted to its dual form as illustrated in Fig. 1. The link is modeled as a vertex while the edge represents the intersection between two adjacent vertexes [31].

Given a traffic network consisting of n signalized intersections connected by m undirected links, a weighted undirected graph $G = (V, \varepsilon, W)$ is built in which $V = \{v_1, \ldots, v_m\}$ is set of the vertexes (links). Each vertex v_i associates the degree of saturation $v_i.s$ with itself. Any pair of vertexes (v_i, v_j) are spatially adjacent if there exists at least one a common intersection, and the edge ε_{ij} is existed between v_i and v_j. $w_{ij} \in W$ represents the weight value of ε_{ij}, which is defined as the correlation of traffic characteristics between v_i and v_j. In this paper, the research object
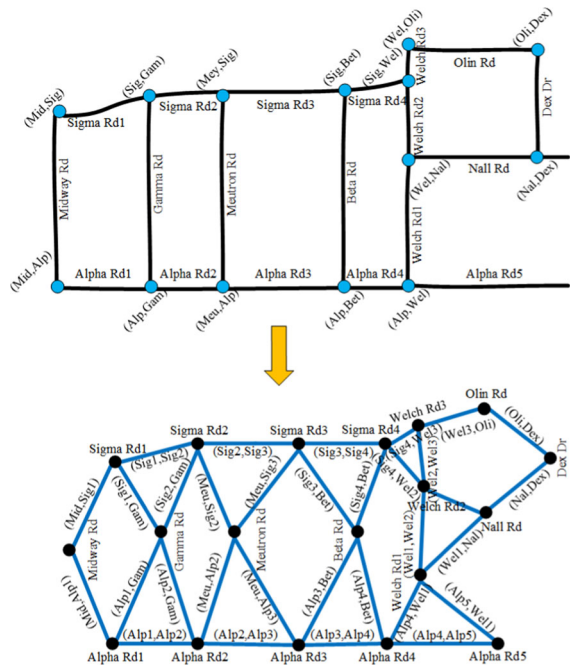

Fig. 1 Actual link map and dual graph

aims at the traffic state of the entire link, so the directionality is ignored.

According to the graph theory, the mathematical model of traffic networks proposed in this paper is based on a weighted undirected graph which derives from the topological relationships and weights of actual road network. In fact, any actual road network can be formulated as such a mathematical model. So, it is general in nature and applicable to a same class of urban traffic network.

2.2 Local Moran's I Index

The Moran's I index is an important measure of spatial autocorrelation [32], and local univariate Moran statistic quantifies the linear correlation strength between a sample vertex and sample vertexes in the neighborhood (spatial lag) [23]. Given a traffic network, the time dimension can be integrated into univariate spatial autocorrelation according to the time-varying characteristics of traffic flow. Then, the spatiotemporal sample vertex (link) with time t and space number i can be denoted as $v_{(t,i)}$. The univariate spatial autocorrelation statistics $T_{(t,i)}$ and $Wt_{(t,i)}$ at $v_{(t,i)}$ is given by:

$$Wt_{(t,i)} = \frac{\sum_{j=1}^{m} a_{(i,j)} \cdot T_{(t,j)}}{\sum_{j=1}^{m} a_{(i,j)}} \tag{1}$$

$$T_{(t,i)} = \frac{\left(s_{(t,i)} - \bar{s}_{(t)}\right)}{\sqrt{\frac{\sum_{i=1}^{m} \left(s_{(t,i)} - \bar{s}_{(t)}\right)^2}{m-1}}} \tag{2}$$

$$\bar{s}_{(t)} = \frac{1}{m} \sum_{i=1}^{m} s_{(t,i)} \tag{3}$$

$$a_{(i,j)} = \begin{cases} 1, & r\left(v_i, v_j\right) = 1 \\ 0, & r\left(v_i, v_j\right) > 1 \end{cases} \tag{4}$$

where $T_{(t,i)}$, $Wt_{(t,i)}$ denote the weighted value of the normalized attributes of the vertex $v_{(t,i)}$ and the normalized attributes summation of all vertexes which are neighbors to $v_{(t,i)}$ at time t, respectively. $a_{(i,j)}$ represents the adjacency relationship between $v_{(t,i)}$ and $v_{(t,j)}$. $r\left(v_i, v_j\right)$ denotes the length of the shortest path between $v_{(t,i)}$ and $v_{(t,j)}$. If $r\left(v_i, v_j\right) = 1$, $a_{(i,j)}$ is equal to 1 or vice versa, as shown in Eq. (4). $s_{(t,i)}$, $\bar{s}_{(t)}$ represent the saturation of $v_{(t,i)}$ and the mean saturation of all sample vertexes in G at t. Here, the saturation is defined as the ratio of the actual traffic flow on a link to its saturation capacity.
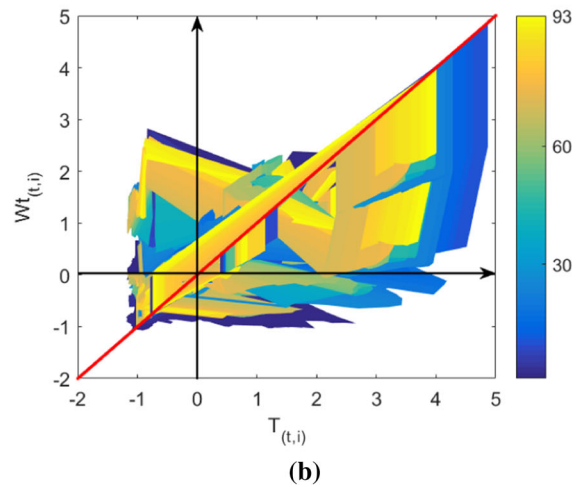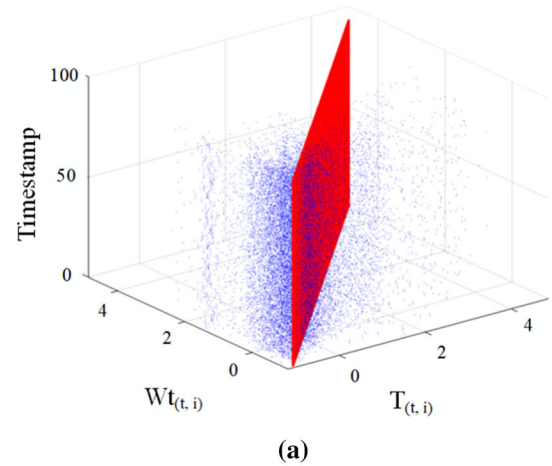

(a)

(b)

Fig. 2 **a** Moran's I scatter plot at 93 time-stamps; **b** contour map of Moran's I. (Color figure online)

A Moran's I scatter plot in 93 time-stamps and the corresponding contour map are depicted in Fig. 2. The slope of the regression line could visualize the spatial autocorrelation statistic with spatial lag (the average characteristics of the samples in the adjacent position) on the vertical axis and the standardized characteristics on the horizontal axis (Fig. 2b). The quadrant to which the scatter is mapped can identify the property of spatial autocorrelation: two types of positive spatial correlations (high–high and low–low aggregation) or two types of negative spatial correlations (high–low and low–high heterogeneous). Aggregation and heterogeneous states, respectively, reflect the same and different evolution law of links.

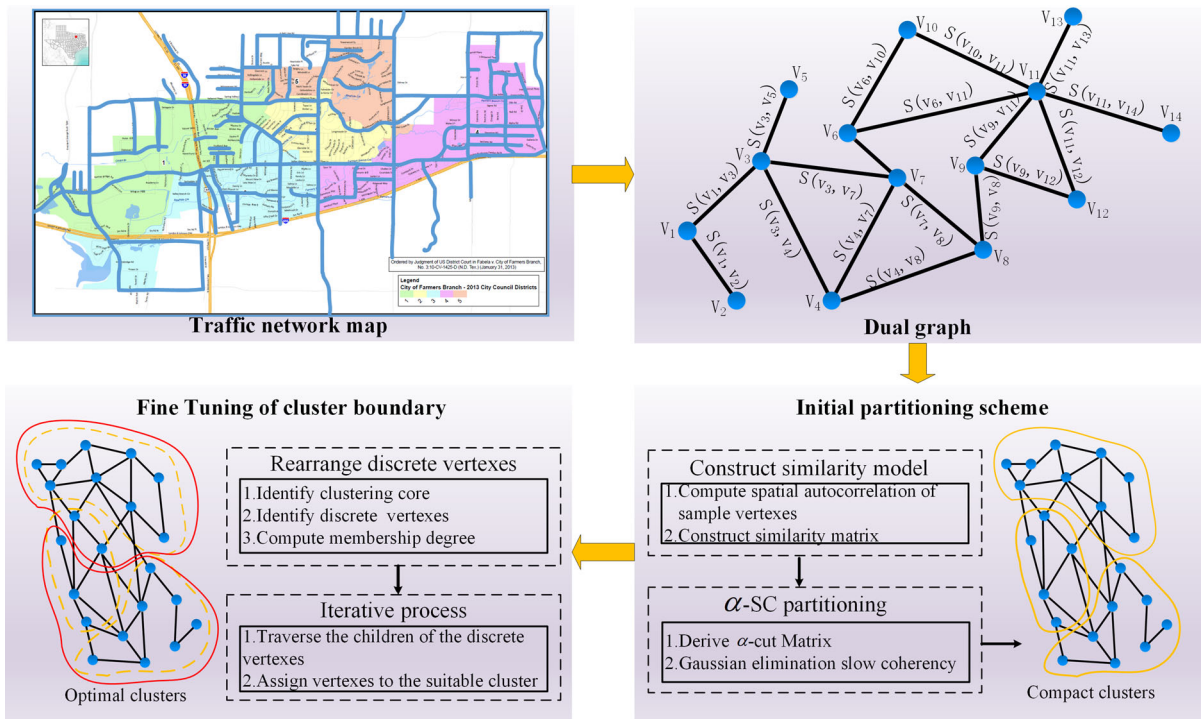In Fig. 2b, various colored regions denote the autocorrelation statistics for vertexes on different times.

Fig. 3 Architecture of the proposed method Framework. (Color figure online)

When the link has the same average characteristics as its neighborhood, the corresponding point will fall on the standard line of the function $y = x$ with a slope of 1 (red straight line). If there has a short vertical distance between the point and the standard line, it indicates that traffic properties of the link are similar to those in the neighborhood. During $n = 0$–60 time intervals, the dark blue and blue regions are relatively dispersed with respect to the red line, and it can be deduced that there is a strong heterogeneity in the traffic network. Relatively, the heterogeneity of the network is small during $n = 61$–93 time intervals.

3 Methodology for static partitioning

This section introduces the main steps of the static partitioning of traffic sub-region based on Moran's I index. A clustering method is applied to the traffic network partitioning, which can achieve the following goals:

- Identify the homogeneous sub-regions with high correlation of link saturation degree and minimize the cutting cost.

- The coverage of the sub-regions for the network confined to 100%, and the overlap rate between the sub-regions is 0%.
- Establish smooth boundaries for the sub-regions to facilitate the perimeter control strategies.

In fact, the first and third goals are conflicting constraints that need to be considered simultaneously. The approach of this multi-objective problem is to use homogeneity as the main object and incorporate other constraints. In the first step, a similarity model between each pair of links is defined based on the spatial autocorrelation property, with connectivity being considered as a constraint; Secondly, a mathematical partitioning framework is implemented by combining the slow coherency algorithm with α-Cut theory; Finally, a fine tuning procedure of the sub-region boundary is applied to rearrange the discrete links to the appropriate sub-regions layer by layer, and two evaluation indexes are introduced to judge the partition quality.

Figure 3 displays the proposed static partitioning framework containing "Initial partitioning scheme" and "Fine Tuning of cluster boundary". In Fig. 3, traffic network map shows the Farmers Branch urban road

network containing n intersections and m links. Dual graph represents the dual graph of the Farmers Branch urban road network, and the establishment process is as follows: firstly, all intersections are abstracted as a node set V', and all links are abstracted as an arc set S', and the attributes W' of arc are represented as distance, capacity, speed, vehicle type information, etc. Thus, the Farmers Branch road network can be abstracted as a weighted undirected graph $G' = (V', S', W')$. Secondly, a dual graph $G = (V, S, W)$ for the Farmers Branch road network is established, where $V = \{v_1, \ldots, v_m\}$ is the link set, S_{ij} is defined as the intersection between link v_i and link v_j; W_{ij} denotes the weight of S_{ij}, representing the correlation of traffic characteristics between the link v_i and link v_j.

3.1 Spatial autocorrelation-based similarly model

Here, the saturation degree of link is taken as the characteristics of the link, which is similar to the intensity of the image. The similarity model $S_{t(i,j)}$ was constructed by calculating the increment $D_{t(i,j)}$ of spatial autocorrelation of traffic properties before and after the vertex pair $\{v_i, v_j\}$ fusion at time t, as shown in Eqs. (5)– (12). The variable σ^2 is a constant with the range of (0, 1], which is taken as 0.5 here. The parameters used to describe the model are introduced in Table 1.

$$S_{t(i,j)} = \begin{cases} \exp\left(-\frac{D_{t(i,j)}}{2\sigma^2}\right), & a_{(i,j)} = 1 \\ 0 & \text{else} \end{cases} \tag{5}$$

Equation (5) calculates the similarity matrix, which is a quantitative representation of the similarity between two links that have an adjacency relationship.

$$D_{t(i,j)} = \left\| \left(T_{(t,G_u)}, Wt_{(t,G_u)}\right) - \left(T_{(t,i)}, Wt_{(t,i)}\right) \right\|_2^2 + \left\| \left(T_{(t,G_u)}, Wt_{(t,G_u)}\right) - \left(T_{(t,j)}, Wt_{(t,j)}\right) \right\|_2^2 \tag{6}$$

Equation (6) expresses the difference in local correlation before and after combining links v_i and v_j into sub-regions G_u, and $D_{t(i,j)}$ indicates the similarity of links v_i and v_j.

$$G_u = \{v_i, v_j\} \tag{7}$$

Equation (7) represents the merging of links v_i and v_j into sub-regions G_u.

$$Wt_{(t,G_u)} = \frac{\sum_{z=1}^{\hat{m}} a_{(G_u,z)} \cdot T_{(z)}}{\sum_{z=1}^{\hat{m}} a_{(G_u,z)}} \tag{8}$$

Table 1 Set of parameters for similarity model

Parameters	Defination
$Wt_{(t,Gu)}$	Weighted value of standardized traffic characteristics between G_u and its neighbors after fusion at time t
$T_{(t,Gu)}$	Standardized characteristic value between G_u and its neighbors after fusion at time t
$\hat{m}$	Number of links in network G after fusion
$\bar{s}_{Gu(t)}$	Mean value of the characteristics in G_u after fusion at time t
$s_{(t,Gu)}$	Characteristic value of G_u after fusion at time t
N_z	Number of elements in set z

Equation (8) denotes the weighted value of standardized traffic characteristics between G_u and its neighbors after fusion at time t.

$$T_{(t,G_u)} = \frac{s_{(t,G_u)} - \bar{s}_{G_u(t)}}{\sqrt{\frac{\sum_{z=1}^{\hat{m}} \left(s_{(t,z)} - \bar{s}_{G_u(t)}\right)^2}{\hat{m}-1}}} \tag{9}$$

Equation (9) calculates the standardized characteristics value between G_u and its neighbors after fusion at time t.

$$\hat{m} = m - 1 \tag{10}$$

Equation (10) represents the number of links within the control area after combining the links v_i and v_j into sub-regions G_u at time t.

$$\bar{s}_{G_u(t)} = \frac{\sum_{z=1, z \neq i,j}^{m} s_{(t,z)} + s_{(t,G_u)}}{\hat{m}} \tag{11}$$

Equation (11) calculates the mean value of the characteristics in G_u after fusion at time t.

$$s_{(t,G_u)} = \frac{1}{N_z} \sum_{z=i,j} s_{(t,z)} \tag{12}$$

Equation (12) represents the characteristic value of G_u after fusion at time t.

The similarity model considers both the correlation between the characteristics of the two links and the increments of their characteristics with neighbors after fusion. The small value of $D_{t(i,j)}$ indicates that there has a small increment in autocorrelation between before and after $\{v_i, v_j\}$ fusion, which further reflects the high similarity between them. The range for $S_{t(i,j)}$ is (0, 1]. $S_{t(i,j)} = 1$ represents that the characteristics of the two adjacent links are identical, and vice versa. The

formula defined above contains the connectivity constraints. At time t, the weight matrix $\mathbf{W} = \left\{w_{t(i,j)}\right\}$ of the network can be represented by the similarity matrix $\mathbf{S} = \left\{S_{t(i,j)}\right\}$.

Remark 1 In this section, the Moran's I index is firstly used to calculate the spatial autocorrelation statistics of the links and the link similarity from a neighborhood topology perspective.

3.2 Theory of α-Cut

The prerequisite for achieving good partition performance is to minimize the cutting cost of inter-partition while maximizing intra-partition correlation [14]. However, the two objects are naturally exclusive to one another and there is no guarantee that both objects can be optimized simultaneously. Here, the α-Cut theory [19] aims to balance the inter-partition cutting costs and intra-partition correlation. The objective function $\min_G \alpha - Cut(G)$ is optimized by Eq. (13).

$$\alpha - Cut(G) = \sum_{u=1}^{k} \left(\alpha_u \times \frac{W\left(G_u, \tilde{G}_u\right)}{|G_u|} - (1 - \alpha_u) \right.$$
$$\left. \times \frac{W(G_u, G_u)}{|G_u|} \right) \tag{13}$$

where k denotes the desired number of extracted sub-regions. The range of parameter α is (0, 1]. A small value of $\alpha - Cut(G)$ indicates the high degree of balance among the several sub-regions. Since α_u expressed by Eq. (14) is closely related to the characteristics of G_u, it is defined as the proportion of the summation of its weight to the summation of all vertex weights in G.

$$\alpha_u = \frac{W(G_u, V)}{W(V, V)} = \frac{W(G_u, G_u) + W\left(G_u, \tilde{G}_u\right)}{W(V, V)} \tag{14}$$

For a graph $G = \{G_1, \ldots, G_u, \ldots, G_k\}$ with k subgraphs, the intra-partition correlation $W_{t(G_u, G_u)}$ of the u-th subgraph at t is defined as

$$W_{t(G_u, G_u)} = \sum_{v_i \in G_u, v_j \in G_u} S_{t(i,j)} \tag{15}$$

where $W_{t(G_u, G_u)}$ is the summation of similarity associated with all vertexes in G_u.

The inter-partition correlation $W_{t\left(G_u, \tilde{G}_u\right)}$ of the u-th subgraph at t is defined as

$$W_{t\left(G_u, \tilde{G}_u\right)} = \sum_{v_i \in G_u, v_j \notin G_u} S_{t(i,j)} \tag{16}$$

where $W_{t\left(G_u, \tilde{G}_u\right)}$ is the summation of correlation associated with all the vertexes at one end in G_u and at other end in subgraphs except G_u.

The minimization of the objective function is obtained by transforming eigenvalues and eigenvectors. Assuming the graph network G is divided by k connected subgraphs. Defining $\mathbf{1} \in \mathbf{R}^m$ as a vector with each of its element value is 1. c_u denotes the index vector of G_u. If $v_i \in G_u, c_u(i) = 1$, and $c_u(i) = 0$ otherwise, as shown in Eq. (17). The degree matrix at t is represented by $\mathbf{D}$, which could be given by Eqs. (18–19).

$$c_u(i) = \begin{cases} 1, & v_i \in G_u \\ 0, & v_i \notin G_u \end{cases} \tag{17}$$

$$\mathbf{D} = diag\{d_i\} \tag{18}$$

$$d_i = \sum_{j=1}^{m} S_{t(i,j)} \tag{19}$$

Applying the indicator vectors, the formulation $\alpha - Cut(G)$ can be simplified by replacing $W_t(G_u, V)$ by $\mathbf{1}^T \mathbf{D} c_u$, $W_t(G_u, G_u)$ by $c_u^T \mathbf{S} c_u$, $W_t(V, V)$ by $\mathbf{1}^T \mathbf{D} \mathbf{1}$, and $|G_u|$ by $c_u^T c_u$. The conversion process can be expressed by Eq. (20).

$$\alpha - Cut(G) = \sum_{u=1}^{k} \left(\frac{\mathbf{1}^T \mathbf{D} c_u}{\mathbf{1}^T \mathbf{D} \mathbf{1}} \times \frac{\mathbf{1}^T \mathbf{D} c_u}{c_u^T c_u} - \frac{c_u^T \mathbf{S} c_u}{c_u^T c_u} \right)$$
$$= \sum_{u=1}^{k} \frac{c_u^T \mathbf{P} c_u}{c_u^T c_u} \tag{20}$$

$$\mathbf{P} = \frac{\left(\mathbf{1}^T \mathbf{D}\right)^T \left(\mathbf{1}^T \mathbf{D}\right)}{\mathbf{1}^T \mathbf{D} \mathbf{1}} - \mathbf{S} \tag{21}$$

The derived matrix $\mathbf{P}$ is defined as the $\alpha - Cut$ matrix. As the objective function needs to be minimized, the optimization problem can be solved by setting the derivative to 0. The smallest eigenvalues (Lagrange multiplier) $\{\lambda_1 \leq \cdots \leq \lambda_k\}$ of $\mathbf{P}$ and corresponding eigenvectors $\{y_1, \ldots, y_k\}$ are computed. Thus, the result is given by Eq. (22).

$$\min \alpha - Cut(G) = y_1^T \mathbf{P} y_1 + \cdots + y_k^T \mathbf{P} y_k = \lambda_1 + \cdots + \lambda_k \tag{22}$$

Equation (22) arranges all the eigenvalues of the P matrix in positive order, then extracts the eigenvectors corresponding to the top smallest eigenvalues to form the matrix and finally performs the slow coherency algorithm for partition.

Minimizing the objective function precisely is a NP-complete problem; however, the discrete value can be approximated affectively by solving the partition matrix. Consequently, we apply the slow coherency theory introduced in the next section.

3.3 α-Cut to slow Coherency theory (α-SC)

Clustering is considered as an effective method to solve the problem of partitioning of urban traffic sub-region. In [19], K-means has been found to be a traditional algorithm in which the location determination of initial center has great influence on the partitioning results. Accordingly, slow coherency theory is a clustering algorithm that groups variables by solving the correlation between each row feature vector and the corresponding reference row according to the feature matrix [33]. Slow coherency theory theoretically solves the problem of identifying the weakest correlation between links, avoiding the influence of initialization values on clustering results. Previous studies have shown that Gaussian elimination on the feature matrix can be used to determine sub-regions with homogeneity [20]. A traffic sub-region initial partitioning algorithm (α-SC) based on the slow coherency theory is proposed here (Algorithm 1). The number of sub-regions can be artificially set to adapt to the actual application scenario.

Obtaining k smallest eigenvalues $\{\lambda_1 \leq \cdots \leq \lambda_k\}$ of matrix P and corresponding set of eigenvectors $\{y_1, \ldots, y_k\}$ (lines 1–8). Then, the feature matrix Y should be row-normalized (lines 9–10). The steps of initial partitioning are listed as follows (lines 11–24):

1. Several initial variables are set, including vertexes $V = \{v_1, \ldots, v_m\}$ and sub-region set $G = \{G_1, \ldots, G_k\}$. The updating status of V and G are recorded in Gaussian elimination procedure;
2. When the condition $l \leq k$ is satisfied, a feature sub-matrix $Y^{(l)} = Y(l : m, l : k)$ is obtained. In other words, $Y^{(l)}$ is consisting of rows $l : m$ and columns $l : k$ of Y, $i = 1, 2, \ldots, m-l+1$, $j = 1, 2, \ldots, k-l+1$. Otherwise, it turns to step (4);

Algorithm 1 α−SC Initial Partitioning Procedure

1: $S \leftarrow$ similarity matrix of G
2: $D \leftarrow$ degree matrix of G at t, where $d_i = \sum_{j=1}^{m} S_{t(i,j)}$
3: $V \leftarrow$ set of links
4: $k \leftarrow$ number of desired sub-regions
5: $P \leftarrow \left(\frac{(1^T D)^T (1^T D)}{1^T D 1} - S \right)$
6: $\cup_{i=1}^{m} \{(\lambda_i, y_i)\} \leftarrow$ get eigenvalue and eigenvector pairs of P
7: sort eigenvalues λ_i to have $\lambda_1 \leq \cdots \leq \lambda_m$
8: select eigenvalues $\{\lambda_1, \ldots, \lambda_k\}$ and corresponding eigenvectors $\{y_1, \ldots, y_k\}$
9: obtain feature matrix $Y_{m \times k} = [y_1, \ldots, y_k]$
10: $Y \leftarrow$ rows of Y is normalized; // eigenvector of the reference row
11: **for** $l = 1 \rightarrow k$ **do**
12:     $Y^{(l)} = Y(l : m, l : k)$
13:     $v_{r+l-1}, G_{c+l-1} \leftarrow$ row and column of the max $\{Y^{(l)}\}$ in Y
14:     $Y(l, :) \leftrightarrow Y(r + l - 1, :)$
15:     $Y(:, l) \leftrightarrow Y(:, c + l - 1)$
16:     $y_{ij} = y_{ij} - y_{lj} \cdot (y_{il}/y_{ll})$, $i = l+1, \ldots, m$, $j = l, \ldots, k$; // Gaussian elimination
17: **end for**
18: reference variable $v_l \leftarrow v_{r+l-1}$, reference matrix $Y_1(l, :) \leftarrow Y(v_l, :)$
19: partition matrix $L_p = Y Y_1^{-1}$
20: **for** $i = 1 \rightarrow m$ **do**
21:     $L_p(i, u) = \max L_p(i, :)$
22:     $v_i \in G_u$
23: **end for**
24: **return** $G = \{G_1, \ldots, G_u, \ldots, G_k\}$; // sub-regions when desired number equals to k

---

3. Obtaining the row number r and column number c of the maximum element in $Y^{(l)}$, and the element v_{r+l-1} is considered as the reference variable of the sub-region G_{c+l-1}. The l−th row is exchanged with the $(r + l - 1)$−th row, and the l−th column is exchanged with the $(c + l - 1)$−th column in Y. Similarly, the l−th element is exchanged with the $(r + l - 1)$−th element in V, and the l−th element is exchanged with the $(c + l - 1)$−th element in G. The Gaussian elimination method is applied to $Y^{(l)}$, $y_{ij} = y_{ij} - y_{lj}(y_{il}/y_{ll})$, $i = l + 1, \ldots, m$, $j = l, \ldots, k$. The variable l is updated to $l + 1$, and then it turns to step (2);
4. According to the reference variables $\{v_1, \ldots, v_l, \ldots, v_k\}$, the nonsingular reference matrix Y_1 is established with $Y_1(l, :) = Y(y_l, :)$. The partition matrix is calculated by function $L_p = Y Y_1^{-1}$;
5. Each column element of the i−th row in L_p indicates the correlation of the vertex v_i to each sub-region. The column u corresponding to the maximum value is searched in the i−th row in L_p,

$L_p(i, u) = \max L_p(i, j)$, $i = 1, \ldots, m$, $j = 1, \ldots, k$. Accordingly, there is a maximum correlation between v_i and G_u, and v_i can be arranged in G_u.

In the α-Cut algorithm, the eigen-decomposition task is generally done in $O(m^3)$ time and $O(m^2)$ time for sparse matrices. The application of Gaussian elimination to slow coherency theory to extract the traffic sub-regions costs $O(km^3)$, where k is the number of desired sub-regions.

Remark 2 In this section, we proposed α-SC algorithm for the initial partition of sub-regions based on the Moran's I index to extract the homogeneous sub-regions in the road network.

3.4 Fine tuning of the sub-region boundary

After the initial partitioning process, the uniform components can be extracted from the heterogeneous network; thus, the first two goals (i.e., extraction of the homogeneous sub-regions and the sub-region coverage constrain for the network) are achieved. This section introduces the proposed method to establish the smooth boundary between sub-regions (the 3rd goal). We propose a slow coherency algorithm aiming at partitioning the weakest correlative sub-regions based on spectral graph theory. The sub-region exported from initial partitioning may appear disconnected. Here, a fine turning procedure aiming at minimizing the variance of the link saturation degree is developed. This procedure defines a membership degree concept to control the expansion trend of clustering stable block and reduce the interference of the boundary links on the partition quality, which expands the α-SC to the α-SC-FT algorithm.

The fine-tuning procedure is used to update the allocation results for the discrete links that leave the largest connected area in the sub-region. Clustering after the initial partitioning is shown in Fig. 4. Vertexes of the same color represent the same sub-region they belong to. In order to maintain the compactness inside the sub-region, the largest connected area in the sub-region can be regarded as the core without changing the partitioning result. The decision variables in this fine-turning step are only the discrete links within the sub-region. Here, the discrete links include the isolated links and the small connected areas away from the core, which correspond to the links containing the filler in the left
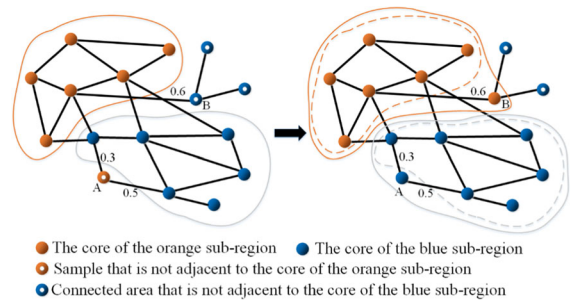

Fig. 4 Schematic diagram for before and after the fine-turning procedure. (Color figure online)

of Fig. 4. This procedure greatly reduces the computational complexity. Now, discrete links do not belong to any sub-region, and the coverage of the sub-regions for the network is less than 100% at this stage.

Remark 3 Given a graph $G = \{G_1, \ldots, G_k\}$, the spatial distance from the discrete link v_i to G_u is measured by the shortest path length $r(v_i, G_u) = l$. The parent link of v_i is defined as the adjacent link v_j that satisfies $r(v_j, G_u) = l-1$. If multiple links meet the condition, the link with the largest $w_{t(i,j)}$ is selected as the parent link. The children of the discrete link v_i is defined as the set of adjacent links $\{v_j | a_{(i,j)} = 1\}$ that have v_i as their parent link.

One feature of fine turning step is to explicitly implement the spatial connection constraint, and this is conducted through a heuristic algorithm. The core within the sub-regions is identified by the initial partitioning procedure. Among them, there only have one maximum connected area in each sub-region. For a discrete link v_i, at least one ordered tree is always found to connect v_i to sub-regions. We define a membership function to measure the probability to which link v_i belongs to the sub-region G_u. The function is affected by two factors: (i) the weight $W_{t(i,G_u)}$ between v_i and G_u; (ii) the geographic location of v_i and G_u. This step preferentially reassigns the noise links with high membership degree. The membership degree $d^i_{G_u}$ of v_i in G_u is calculated as follows:

$$d^i_{G_u} = \frac{1}{|N_i|} \max_{v_j \in G_u} \{w_{t(i,j)}\} \cdot a_{(i,j)} \tag{23}$$

$$|N_i| = \sum_{r(v_i, G_u)=1, G_u \in G} a_{(i,G_u)} \tag{24}$$

where $|N_i|$ represents the number of sub-regions adjacent to v_i. The range for $d_{G_u}^i$ is [0, 1]. A large membership value indicates that the sub-regions homogeneity will be enhanced if the link is assigned in the sub-region. The membership function is only for the discrete link v_i that satisfy constraint $r(v_i, G_u) = 1$ with at least one sub-region G_u, which explicitly implements the sub-region connection constraint.

Algorithm 2 Fine Tuning of the Sub-region Boundary

1: $G = \{G_1, \ldots, G_u, G_b, \ldots, G_k\}$: initial partitioning results
2: $V = \{v_1, \ldots, v_i, v_j, \ldots, v_m\}$: set of vertexes (links)
3: $\hat{V} = \{v_1, \ldots, v_i, v_j, \ldots, v_f\}$: set of discrete vertexes (links), $\hat{V} \subset V$
4: $s1 = \emptyset$: initialize stack
5: **repeat**
6: 　**if** $\hat{V} \neq \emptyset$ **then**
7: 　　**for all** $v_i \in \hat{V}$ **do**
8: 　　　$d_{Gu} = \frac{1}{|N_i|} \max_{v_j \in G_u} \{W_{t(i,j)}\} \cdot a_{(i,j)}$;// compute membership degree
9: 　　　**if** $d_{G_u}^i > 0 \, (G_u \subset G)$ **then**
10: 　　　　$v_i \in G_u \left(d_{G_u}^i > d_{G_b}^i, G_b \subset G \&\& G_b \neq G_u\right)$; // assign the isolated vertexes and the boundary vertexes in discrete connected areas
11: 　　　　$\hat{V} = \hat{V} \cap s1$;
12: 　　　　push v_i into $s1$;// update set of discrete vertexes
13: 　　　**end if**
14: 　　**end for**
15: 　**end if**
16: **until** $\hat{V} = \emptyset$
17: $G \leftarrow$ results after fine tuning of the sub-region boundary
18: **return** $G = \{G_1, \ldots, G_k\}$

The pseudocode of the fine turning procedure can be embodied by Algorithm 2. The algorithm utilizes stack $s1$ to store and traverse discrete links. For a single discrete link v_i, it is directly connected to at least one sub-region. The membership degrees of v_i for all adjacent sub-regions are calculated separately, and v_i is assigned to the sub-region G_u with the constraint $d_{G_u}^i = \max \left\{d_{G_b}^i | b = 1, \ldots, k\right\}$. Similarly, the boundary links of the discrete connected area may first be assigned to sub-region with the highest membership degree (lines 1–10). At this moment, all arranged links are removed from $\hat{V}$ and then pushed into $s1$ (lines 11–12). The above steps used repeatedly can traverse the children links of the boundary links in the discrete connected area, and further traverse their children (lines 13–15). This iterative process continues until all the discrete links are assigned to the appropriate sub-region

(lines 16–18). This procedure satisfies the third goal by considering the connectivity constraints when assigning discrete links. How to apply this static partitioning procedure to the dynamic partitioning framework will be introduced in details in the next section.

Example 1 The left of Fig. 4 shows two sub-regions after initial partitioning stage. It can be seen that there are isolated discrete vertexes (the orange vertex A with filler) and connected area (the blue vertexes with filler) in the sub-regions. We assume that the membership degree of A for blue sub-region is $d_{blu}^A = 0.5$, and the membership degree of B for orange sub-region is $d_{ora}^B = 0.6$. As the other discrete vertexes are the children of B, they are not considered in current iterative allocation process. The results of the first round of fine tuning are shown in the right of Fig. 4.

The complexity of the dynamic partitioning algorithm considering the evolution of traffic congestion proposed in this paper remains $O\left(km^3\right)$, where m denotes the number of links to be controlled and adjusted and k denotes the desired number of sub-regions.

Remark 4 In this section, the α-SC-FT algorithm is proposed for enforcing the connectivity and uniformity of sub-regions based on the α-SC algorithm.

3.5 Performance metrics

The partition quality is evaluated from three perspectives. Firstly, the performance of the first goal can be evaluated by the average sub-region correlation intensity index CI_a defined by Eq. (25). The absolute difference in saturation degree of adjacent links is considered as the distance between the links, and then the average absolute difference of all the sub-regions is calculated.

$$CI_a = \frac{1}{k} \sum_{v_i, v_j \in G_u} \frac{|v_i \cdot s - v_j \cdot s| \cdot a_{(i,j)}}{|v_i \cdot s - v_j \cdot s| \cdot a_{(i,j)} + Cut_{(G_u)}}$$
(25)

$$Cut_{(G_u)} = \sum_{v_i \in G_u, v_j \notin Gu} |v_i \cdot s - v_j \cdot s| \cdot a_{(i,j)}$$
(26)

where $Cut_{(G_u)}$ represents the sum of the differences of saturation degree between G_u and its neighborhood. This index indicates the variances between the link saturation degree of inter-partition and intra-partition. A

small value of CI_a denotes the superior partition performance.

Secondly, the performance of the homogeneity of intra-partition and the size of the sub-region can be evaluated by using a normalized total variance TV_n [34]. TV_n is defined as the ratio of the total variance of each sub-region to the entire urban network. A small value of TV_n denotes a better partitioning.

$$TV_n = \frac{\sum_{u=1}^{k} N_{Gu} \times \text{var}(G_u)}{m \times \text{var}(G)} \quad (27)$$

where $G = \cup_{u=1}^{k} G_u$ denotes the traffic network to be researched. N_{Gu}, m denote number of links in G_u and G, respectively. $\text{var}(G_u)$ represents the variance of traffic characteristics in G_u.

Modularity [35] is a criterion defined by Eq. (28) to verify whether a particular partition is meaningful. This index is by far the most used and best-known quality function. It is commonly used in community discovery that takes the connectivity within the community and communication between vertexes into account.

$$Q = \sum_{u=1}^{k} (w_{t(u,u)} - A_u^2) = Tre - \left\| W^2 \right\| \quad (28)$$

$$A_u = \sum_{u=1}^{k} w_{(u,u)} \quad (29)$$

where A_u denotes the summation of the $u-th$ row in weight matrix W. The range of Q is [0, 1), and it close to 1 indicates a homogeneous sub-region structure in the network.

4 Methodology for dynamic updating

Due to the dynamic propagation characteristics of traffic flow in different regions, the complete static partitioning of the traffic network at each time-stamp is a task with high computation time. The target links of the dynamic partitioning model proposed in this paper are only a few high-heterogeneity links, which can improve the partition efficiency of the network and track the spatiotemporal evolution of sub-regions. In this section, a three-step dynamic updating framework shown in Algorithm 3 is proposed based on heuristic method, which considers the evolution of congestion and utilizes the stacks $s2$, $s3$ to store the calculated values. The dynamic updating model mainly includes merging, cutting and adjusting processes.

When different congested areas need to be merged during the diffusion of congestion or the congestion pocket appears inside an unblocked sub-region [18], the dynamic partitioning method should be able to model the merging or cutting procedure of the sub-regions.

In order to capture the diffusion of congestion, TV_n is applied to measure the homogeneity of the network in the sub-region merging process. Based on the partitioning result at $t - 1$, $TV_n\left(t, G_{a,b}\right)$ at time t is calculated after any adjacent sub-region pairs $\left\{G_a^{t-1}, G_b^{t-1}\right\}$ merged into $G_{a,b}^t$ with the largest similarity at t respectively (line 1–6). The adjacent sub-regions corresponding to the minimum TV_n are merged (lines 7–10). The objective of this step is to find the sub-regions of optimal homogeneity, where the weight between G_a and G_b is calculated by Eq. (30). The merging process needs to meet the following conditions:

(i) The size of $G_{a,b}^t$ needs to be smaller than the size of the largest traffic sub-region at $t - 1$. This is to balance the size of the sub-regions, as shown in Eq. (31) (line 8);

(ii) TV_n of the network at time t is smaller than that at time $t - 1$, which is defined by Eq. (32) (line 9);

(iii) Equation (33) insures that the number of sub-regions is updated from k to $k - 1$ (line 10).

$$W_{t(G_a, G_b)} = \sum_{\substack{v_i \in G_a \\ v_j \in G_b}} W_{t(i,j)}, \left(G_a, G_b \subset G^{t-1}\right)$$
$$(30)$$

$$N_{G_{a,b}}^t < \max\left(N_{G_u}^{t-1}\right), G_u = \left\{G_1^{t-1}, \ldots, G_k^{t-1}\right\}$$
$$(31)$$

$$TV_n(t, G) < TV_n(t - 1, G) \quad (32)$$

$$G^t = \left\{G^{t-1} \cap \left\{G_a^{t-1}, G_b^{t-1}\right\}\right\} \cup G_{a,b}^t \quad (33)$$

The process of forming the congested pocket can be captured by using α-SC-FT algorithm in Sect. 3 to partition the target sub-region G_u^{t-1} into G_{u1}^t and G_{u2}^t at time t. Since the variance index calculated by Eq. (34) can easily measure heterogeneity within the sub-region, the variance is applied to determine whether the original sub-regions satisfy the dichotomous condition (lines 11–17). This process needs to meet the following constraints: (i) the size of the newly generated two sub-regions must be larger than the size of the original minimum sub-region, as defined in Eq. (35) (line 18); (ii)

the variance indexes of G_{u1}^t and G_{u2}^t are all smaller than G_u^{t-1} (line 19) as defined in Eq. (36), which indicates the different levels of congestion areas within G_u at t; (iii) the number of sub-regions is updated from k to $k+1$, as in Eq. (37) (line 20).

$$\text{var}\left(G_u^t\right) = \sum_{v_i, v_j \in G_u} ||v_i.s - v_j.s||_2^2 \tag{34}$$

$$\min\left(N_{u1}^t, N_{u2}^t\right) > \min\left(N_a^{t-1}\right), G_a = \left\{G_1^{t-1}, \ldots, G_k^{t-1}\right\} \tag{35}$$

$$\max\left\{\text{var}\left(G_{u1}^t\right), \text{var}\left(G_{u2}^t\right)\right\} < \text{var}\left(G_u^{t-1}\right) \tag{36}$$

$$G^t = \left(G^{t-1} \cap G_u^{t-1}\right) \cup \left(G_{u1}^t, G_{u2}^t\right) \tag{37}$$

Over a short period of time, the traffic characteristics of the link do not change significantly; thus, it is able to realize clustering with high homogeneity at t by adjusting the highly heterogeneous links at $t-1$. Firstly, the link with the closest degree of saturation to the average value within the sub-region is determined as the clustering seed at t, reflecting the overall traffic state of the sub-region. We apply the improved depth-first search algorithm (DFS) to iteratively capture the clustering seed and control the expansion order of the stable blocks within sub-regions (the block is highly homogeneous at time t) (lines 21–24). It is worth noting that the block can only be in the sub-region where the seed vertex is located at $t-1$. Specially, each sub-region contains only one stable block. Because the variance in block has a nonlinear trend with increasing of the number of iterations, further improvement is carried out with a threshold δ artificially set so that the stable block is cut off when it reaches a certain coverage δ and the running time is reduced. The impacts of θ on the partition quality can be found in Sect. 5.

At time t, the evolution process of the DFS-based stable block in G_u is as follows:

1. Given a set of vertexes $V = \left\{v_{u,1}^{t-1}, v_{u,2}^{t-1}, \ldots, v_{u,r}^{t-1}\right\}$ in G_u at $t-1$ where r is the number of vertexes in G_u. The stable block B_u^t of G_u is initialized to empty set at t;
2. Calculate the distance between the degree of saturation and the average value in G_u separately, and the similarity between vertexes can be derived from the model in Sect. 3.1. The vertex v_i corresponding to the minimum distance is treated as the seed vertex,

and then, $B_u^t = B_u^t \cup v_{u,i}^{t-1}$. The cycle-index n is set to 1;

3. When $n \leq r$, the vertex with the maximum similarity $S\_\max$ to B_u^t can be searched in G_u. If there exists a vertex set that satisfies the condition $A_j^{\max} = \left\{v_j | v_j \in \{v_1, \ldots, v_h\}, \left(S_{t(i,1)} = \cdots = S_{t(i,h)} = S\_\max\right)\right\}$, an arbitrary link v_j is selected; otherwise, jumps to (5);
4. If $S\_\max > \delta$, B_u^t is updated to $B_u^t \cup v_j$. Returns to (3); otherwise, jumps to (5);
5. B_u^t is treated as a constant stable block at t and the identification of the stable block is ended.

The elements in the k stable blocks $\left\{B_1^t, \ldots, B_u^t, \ldots, B_k^t\right\}$ are retained as the original allocation results, and the remaining vertexes that exclude the blocks are considered as unallocated vertexes. Due to the set threshold δ, the extracted k stable blocks have the coverage of 0%–100% for G at t. The fine-tuning algorithm mentioned in Sect. 3.4 can be easily adapted to partition unallocated vertexes with the following modifications (lines 25–31):

(i) "discrete vertex (link)" is changed to "unallocated vertex (link)";
(ii) "discrete connected area (link set)" is changed to "unallocated connected area (link set).

The coverage of the sub-regions is 100% when the discrete vertexes are allocated, which satisfies the second criteria mentioned in Sect. 3. The DFS algorithm is applied to identify the stable block abovementioned by constraining the connectivity and homogeneity of the sub-region.

The decision variables in the dynamic updating procedure are the unallocated links excluding the stable blocks (including the isolated links and connected areas away from the block). Thus, the computational complexity of dynamic partitioning is lower than that of static partitioning according to the time correlation of the traffic flow.

Remark 5 In this section, a dynamic partition algorithm of traffic network control sub-regions is proposed, which can effectively identify different sub-regions with different saturation.

5 Case study and results

To evaluate the effectiveness of the proposed dynamic partitioning algorithm, we carried out a case study with

Algorithm 3 Dynamic Updating Procedure

1: $G^{t-1} = \{G_1, \ldots, G_a, G_b, G_u, G_x, \ldots, G_k\}$: static partitioning

2: $V = \{v_1, \ldots, v_i, v_j, \ldots, v_m\}$: set of vertexes (links)

3: $s2 = \emptyset, s3 = \emptyset$: initialize stacks

4: **for all** $G_a^{t-1} \left(W_{t(Ga,Gb)} = \max \left(W_{t(Ga,Gu)}\right)\right), G_u \subset G^{t-1}$ **do**

5:     push $TV_n\left(t, G_{a,b}\right)$ into $s2$

6: **end for**

7: $G_c^t \leftarrow \left\{G_a^{t-1}, G_b^{t-1}\right\} \left(TV_n\left(t, G_{a,b}\right) = \min\{s2\}\right);$// the two sub-regions $\{G_a, G_b\}$ corresponding to the minimum value in $s2$ are merged into G_c

8: **if** $N_{Ga,b}^t < \max\left(N_{Gu}^{t-1}\right), G_u = \left\{G_1^{t-1}, \ldots, G_k^{t-1}\right\}$ **then**

9:     **if** $TV_n(t, G) < TV_n(t-1, G)$ **then**

10:        $G^t = \left\{G^{t-1} \cap \left\{G_a^{t-1}, G_b^{t-1}\right\}\right\} \cup G_c^t;$ //Merging procedure

11:     **else**

12:        $G^t = G^{t-1}$

13:        **for all** $G_u \subset G^{t-1}$ **do**

14:           $\text{var}\left(G_u^t\right) = \sum_{v_i, v_j \in Gu} ||v_i.s - v_j.s||_2^2$

15:           push $\text{var}\left(G_u^t\right)$ into $s3$

16:        **end for**

17:        $G_{u1}^t, G_{u2}^t \leftarrow G_u^t \left(\text{var}\left(G_u^t\right) = \max(s3)\right)$ is divided by α-SC-FT

18:        **if** $\min\left(N_{u1}^t, N_{u2}^t\right) > \min\left(N_x^{t-1}\right), G_x = \left\{G_1^{t-1}, \ldots, G_k^{t-1}\right\}$ **then**

19:           **if** $\max\left\{\text{var}\left(G_{u1}^t\right), \text{var}\left(G_{u2}^t\right)\right\} < \text{var}\left(G_u^{t-1}\right)$ **then**

20:              $G^t = \left(G^{t-1} \cap G_u^{t-1}\right) \cup \left(G_{u1}^t, G_{u2}^t\right);$ // Cutting procedure

21:           **else**

22:              $G^t = G^{t-1}$

23:              **for all** $G_u^{t-1} \subset G^{t-1}$ **do**

24:                 identifies stable blocks at time t

25:                 uses fine tuning process to allocate remaining vertexes

26:              **end for**; // Adjustment process

27:           **end if**

28:        **end if**

29:     **end if**
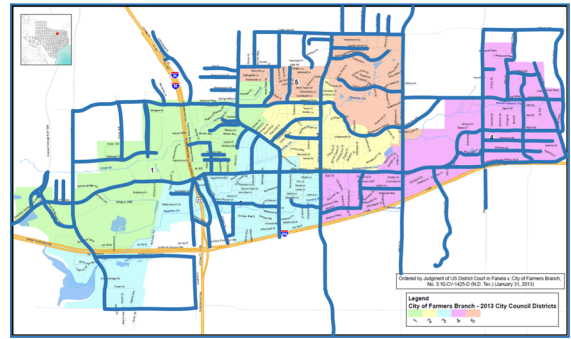
30: **end if**;

31: **return** G^t

Fig. 5 Traffic network of farmers branch. (Color figure online)

5.1 Data preparation

In order to verify the effectiveness of the proposed dynamic partitioning method of the traffic network sub-region, several experiments are conducted by using a real-world urban network data set. The traffic network (Fig. 5) is extracted from the Regional Travel Demand Model of Farmers Branch (North Central Texas Council of Governments, Transportation Department 2014).[1] The network does not have a clear grid structure and the adjacent relationship between road segments is different, which makes the application of the method challenging. The data set includes the average degree of link saturation in Farmers Branch for each working day from October 1, 2014 to October 31, 2014. In the data set, the degree of saturation is taken as the characteristics of the link, which can embody the traffic capacity of the link. In this study, 211 links (including arterial roads and minor arterial roads) of the Farmers Branch are analyzed to make the degree of saturation data be estimated more reliable.

The time period of 00:00–23:15 is divided into 93 time-stamps for 15 min intervals, and n denotes the time intervals. Figure 6 illustrates the spatiotemporal distribution of saturation degree revealing the morning peak ($n = 28$) and evening peak ($n = 69$) clearly. Moreover, different saturation degree values reflect different congestion levels of the network.

To verify the performance of dynamic partitioning, the difference between independent partitioning and static partitioning in terms of CI_a and TV_n is discussed. Independent partitioning is to divide the sub-region at each time stamp, and static partitioning is to divide the sub-region only at the initial moment. In Fig. 7, the
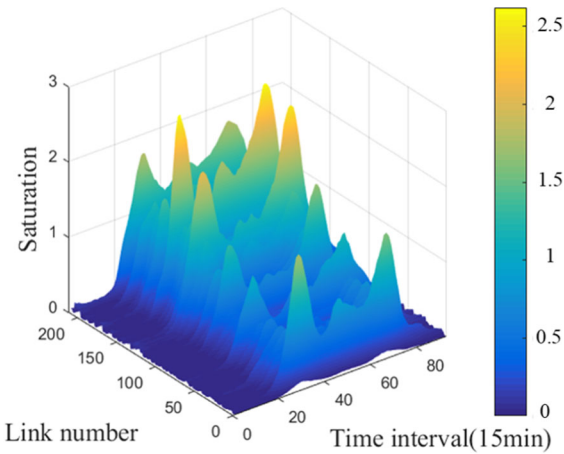
a medium-sized real traffic network. Section 5.2 reports the comparative analysis of the performance of the proposed algorithm with k-way α-Cut [19], spectral clustering and symmetric nonnegative matrix factorization (SNMF) [18]. Section 5.3 presents the performance of dynamic partitioning and the sensitivity analysis of the parameter δ to the stable block.

Fig. 6 Space-time distribution of the degree of saturation. (Color figure online)

Fig. 8 Partitioning quality comparison of α-SC and α-SC-FT. (Color figure online)
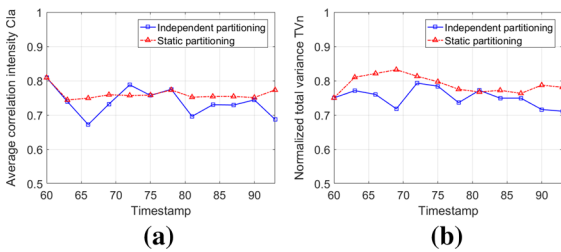

Fig. 7 Efficiency of independent partitioning and static partitioning. (Color figure online)

blue curve describes the independent partitioning quality, while the red curve indicates the static partitioning quality. It is worth noting that the network has better partition performance at each time-stamp by continuously updating partitioning results, which reflects the importance of dynamic partitioning. Since the independent partitioning is executed at each time-stamp to track the development trend of sub-regions and has a high computational complexity, it is necessary to capture the evolution of sub-regions by the proposed dynamic partitioning method.

5.2 Quality of static partitioning

The partitioning quality of α-SC and α-SC-FT applied in the traffic network at $n = 69$ is displayed in Fig. 8. From Fig. 6, it can be seen that the traffic network presents different congestion levels at this time. The main purpose of the fine tuning is to ensure the connectivity within the sub-regions. Since CI_a and TV_n
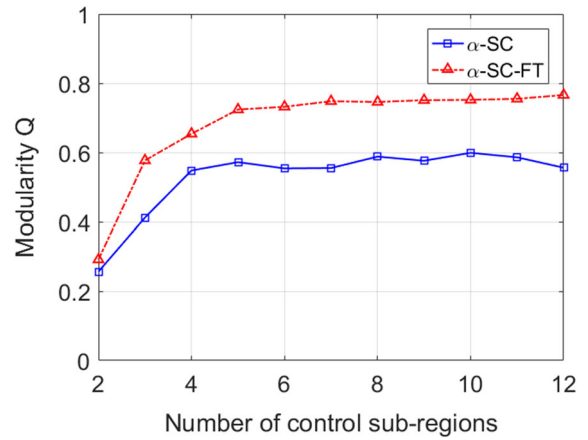
indicators focus on analyzing the homogeneity of subregions, the connectivity performance of α-SC and α-SC-FT is compared using the modularity metrics.

In Fig. 8, the horizontal and vertical axis indicate 2–12 different sub-region number k and modularity Q, respectively. In terms of modularity, the Q value after the fine tuning stage is significantly higher than the initial partitioning under different k, and it can be concluded that this procedure greatly improves the compactness of the sub-regions. This is because the α-SC-FT algorithm fine tunes the sub-region boundaries to alleviate the interference of noise links to the subregions. Moreover, as the parameter k increases, the sub-regions become more homogenous.

The TV_n indicator is used as the baseline, which describes the comparative result from k-way α-Cut, Spectral cluster, SNMF and α-SC-FT with sub-region numbers 2–12 in Fig. 9. The sub-regions corresponding to the minimum TV_n are regarded as the optimal results. The performance of the four algorithms shows an upward trend with the increase of k. The α-SC-FT partitioning scheme outperforms Spectral cluster in most sub-regions except $k = 4$, 7 and 10. And both them are better than the k-way α-Cut and SNMF algorithm, while k-way α-Cut and SNMF have the similar partition performance.

The TV_n values for the four algorithms in the case of the optimal number of traffic sub-regions are shown in Table 2. It can be seen that the optimal sub-region number of k-way α-Cut, Spectral cluster, SNMF and α-SC-FT algorithms are 10, 12, 12 and 8, respectively.
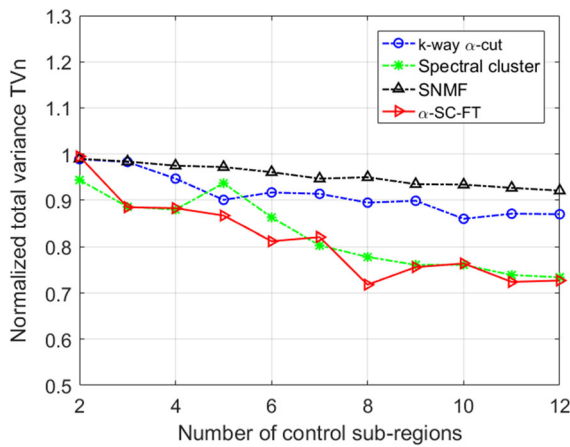
Fig. 9 Quality comparison of different static algorithms. (Color figure online)

Table 2 Overall performance of partition

	k-way α-cut	Spectral cluster	SNMF	α-SC-FT
TVn	0.860	0.734	0.921	0.718
k	10	12	12	8

The TV_n of α-SC-FT algorithm is the smallest, which indicates that a good partition result is guaranteed of the network.

The optimal partitioning result of α-SC-FT algorithm and frequency distribution of saturation degree at $n = 69$ are shown in Fig. 9. The dotted lines in Fig. 10b mark the saturation degree levels where each sub-region has the maximum number of links. The curves of the same color have a high peak at the dotted line, and the number of links corresponding to other degree of saturation is less than the peak value. Thus, it can be seen that the sub-region is relatively homogeneous. In Fig. 10a, due to the connectivity constraints, the algorithm partitions two sub-regions with similar degree of saturation but not geographically adjacent, such as the 5th and 7th sub-region. The links within each sub-region remain connected, so the α-SC-FT algorithm is considered to be superior and can ensure the boundary smoothness of the sub-region.

5.3 Quality of dynamic partitioning

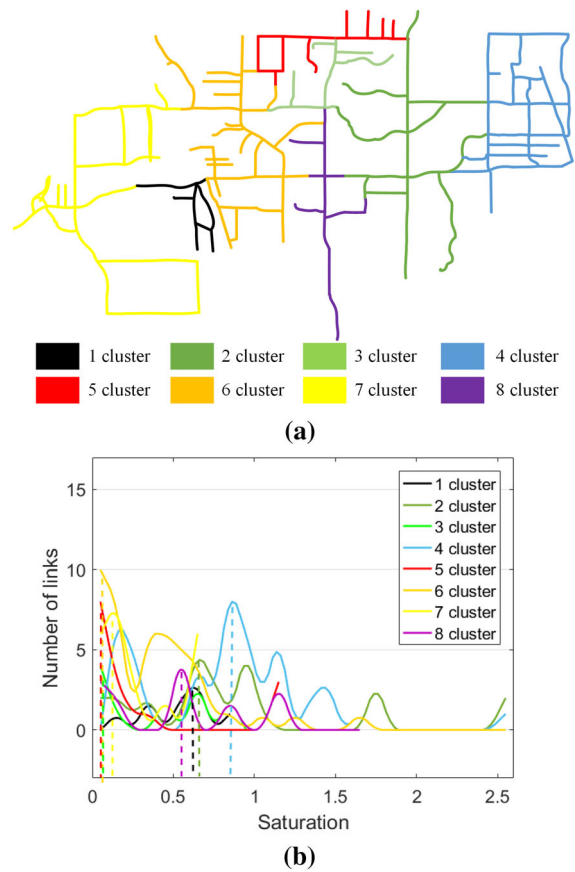Due to the strong spatiotemporal correlation of traffic flow, independent partitioning does not need to be

(a)

(b)

Fig. 10 **a** The partitioning results of α-SC-FT algorithm at $n = 69$; **b** The frequency distribution of degree of saturation. (Color figure online)

performed separately at each time-stamp if the method developed in Sect. 4 runs well. In this section, we will study the dynamic partitioning of the road network during the appearance of congestion pocket at $n = 66$–78. Figure 11a, b, respectively, depicts the dynamic partition performance at different update interval Δ from $n = 66$ to 78. The threshold δ is set to be 0.3. It can be seen that CI_a and TV_n decrease with the shortening of the update interval, which reflects the effectiveness of the dynamic partitioning.

The threshold δ is a key parameter of dynamic updating that directly affects the coverage θ of the stable blocks and the partition quality of the network. As the result of the dynamic partitioning is convincing at the small update interval length Δ (Fig. 11a, b), the sensitivity analysis on different values of δ is performed when $\Delta = 1$ (see Fig. 11c, d). Here, the test values of δ are 0.3, 0.4 and 0.5, respectively, with the hope to
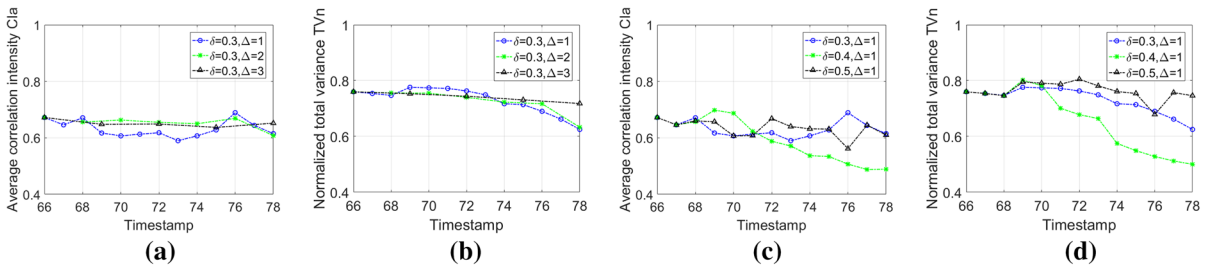
Fig. 11 Sensitivity of the dynamic partitioning to Δ (**a, b**) and to δ (**c, d**). (Color figure online)

Table 3 Merging, cutting and adjusting of sub-regions

Cycle	66	67	68	69	70	71	72
Evolution	–	C	A	A	C	C	A
k	9	10	10	10	11	12	12
Overlap rate	–	–	0.611	0.512			0.616

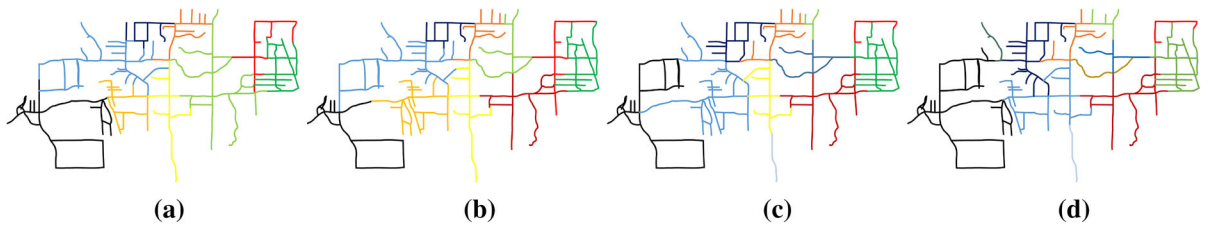| Cycle	73	74	75	76	77	78
Evolution	C	A	A	A	A	C
k	13	13	13	13	13	14
Overlap rate	–	0.768	0.825	0.839	0.853	

Fig. 12 Dynamic partitioning of sub-regions at time-stamps ($n = 66, 68, 70$ and 72). (Color figure online)

provide an optimal value for solving the dynamic partitioning model. All the weights between the links in the stable block are greater than 0.3 when $\delta = 0.3$. In other words, the area with higher homogeneity within the sub-region is preserved. It can be seen that the optimal result can be obtained with $\delta = 0.4$, and when the value of δ is increased or decreased, the result becomes less well. The coverage θ and dynamic partitioning quality will vary with different values of δ. The flexible settings of parameters in this method can be applied to different application scenarios.

We operate dynamic partitioning from $n = 66$, and the optimal number of sub-regions is 9. Three different strategies merging, cutting and adjusting (M, C, A) are performed, respectively, and the smallest objective

functions CI_a and TV_n are applied in dynamic partitioning. Since the strategy of $\Delta = 1$ and $\delta = 0.4$ are convincing for the algorithm, Table 3 shows the evolution results of the sub-regions in this strategy. The coverage θ of the stable blocks for the network is 61.1% at $n = 68$ and shows an increasing trend as the time passes. This result indicates that a congestion pocket appears in the evolution process, and the difference of saturation degree of each area in the network is relatively reduced.

In order to further reveal the evolution of the sub-regions, Fig. 12 visualizes the dynamic partitioning results of $n = 66, 68, 70$ and 72. The average degree of saturation values within the sub-regions is presented in Table 4. During the dynamic partitioning process,

Table 4 Average value of saturation degree in different sub-regions at $n = 66, 68, 70$ and 72

n	Light green	Orange	Light blue	Yellow	Brown	Blue
66	0.758	0.295	0.299	0.556		
68	0.783	0.383	0.360	0.757	0.757	
70	0.518		0.395	0.476	0.860	0.907
72	0.446	0.760	0.308		0.760	1.259
n	Gray	Yellow-green	Dark green	CI_a	TVn	
66				0.672	0.760	
68				0.657	0.746	
70	1.116			0.687	0.782	
72	0.923	0.241	0.566	0.588	0.678	
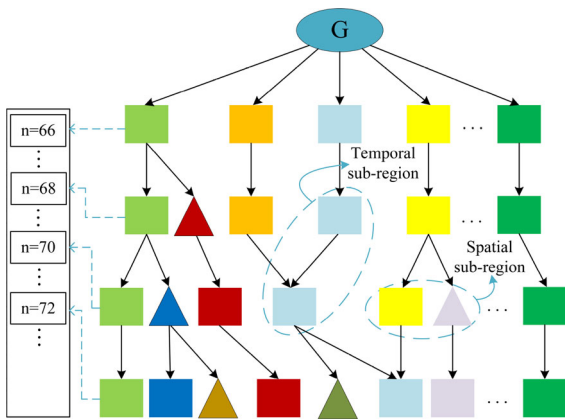
Fig. 13 Spatiotemporal evolution of sub-regions ($n = 66, 68, 70$ and 72). (Color figure online)

triangle represents the newly generated sub-region at that time-stamp.

the algorithm starts with 9 sub-regions at $n = 66$, and the number of sub-regions increases or does not change over time. At $n = 67$, due to the presence of a congestion pocket in the light green mid-saturated sub-region, it is divided into light green (congested) and brown (medium saturated) regions. The adjustment process is performed in the traffic network at $n = 68$. The algorithm partitions the light green sub-region into light green and blue sub-regions at $n = 70$ and then increases the number of sub-regions to 11. It can be observed in Table 4 that the light green region is completely unblocked while the blue one is congested. The CI_a and TV_n are decreased over time, which verify the effectiveness of the dynamic updating algorithm. Figure 13 depicts the spatiotemporal evolution process of sub-regions. The square represents the sub-region that has been generated at the previous time-stamp, and the

6 Conclusions

This paper proposes a dynamic partitioning framework for urban traffic network sub-regions based on Moran's I index and heuristic algorithm. The methodology achieves the spatial partitioning of the network and reveals the evolution of traffic flow on the basis of spatiotemporal correlation. A similarity model between links is constructed, and the spatial autocorrelation characteristics are analyzed based on Moran's I index. Then, the slow coherency theory is applied to the static partitioning framework (α-SC-FT) to solve the identification problem of the weakest correlation between sub-regions. The experiments carried out on the data set of Farmers Branch city show that the algorithm is superior to the existing spatial partitioning method k-way α-Cut. The dynamic updating procedure effectively updates the sub-regions according to time dependence of congestion. This procedure reserves the stable block with higher homogeneity in the original sub-region, and the discrete links are iteratively re-assigned to identify the spatiotemporal propagation of congested pockets. The unique advantages of the proposed two-step dynamic partitioning framework are: (i) a well identification of the sub-region evolution in real traffic network; (ii) the number of traffic sub-regions can be artificially set according to the dynamic characteristics of traffic flow, which can adapt to the actual application scenarios. The numerical analysis of this study shows satisfac-

tory results, including the mean correlation intensity, normalized variance and modularity.

The results of the proposed approach indicate that the critical borders of congestion pockets are properly tracked over time by changing a small percentage of links. When the urban networks with variant demand profiles, a dynamic partitioning coupled with perimeter control strategies might be necessary. Therefore, integrating dynamic partitioning with perimeter control strategies should be a challenging and worthwhile research direction.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

1. Geroliminis, N., Daganzo, C.F.: Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transp. Res. Part B Methodol. **42**(9), 759–770 (2008)

2. Ji, Y., Daamen, W., Hoogendoorn, S., Lanser, S.: Investigatin the shape of the macroscopic fundamental diagram using simulation data. Transp. Res. Rec. **2161**(1), 40–48 (2010)

3. Zhang, L., Garoni, T.M., De, G.J.: A comparative study of macroscopic fundamental diagrams of arterial road networks governed by adaptive traffic signal systems. Transp. Res. Part B Methodol. **49**(2), 1–23 (2013)

4. Zockaie, A., Saberi, M., Saedi, R.: A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: optimal locating of fixed measurement points and sampling of probe trajectories. Transp. Res. Part C Emerg. Technol. **8**(6), 245–262 (2018)

5. Geroliminis, N., Zheng, N., Ampountolas, K.: A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transp. Res. Part C Emerg. Technol. **4**(2), 168–181 (2014)

6. Wada, K., Satsukawa, K., Smith, M., Akamatsu, T.: Network throughput under dynamic user equilibrium: queue spillback, paradox and traffic control. Transp. Res. Part B Methodol. **126**, 391–413 (2019)

7. Geroliminis, N., Sun, J.: Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transp. Res. Part B Methodol. **45**(3), 605–617 (2011)

8. Buisson, C., Ladier, C.: Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transp. Res. Rec. **2124**(1), 127–136 (2009)

9. Zhou, Z., Lin, S., Xi, Y.: A fast network partition method for large-scale urban traffic networks. Contr. Theor. Tech. **11**(3), 359–366 (2013)

10. Anwar, T., Liu, C., Vu, L.H., Leckie, C.: Partitioning road networks using density peak graphs: efficiency vs. accuracy. Inf. Syst. **64**, 22–40 (2017)

11. Loukas, D., Paraskevas, N.: Dynamic partitioning of urban road networks based on their topological and operational characteristics. In: Proc. IEEE Conf. Models and Technologies for ITS, Napoli, Italy, pp. 457–462 (2017)

12. Wang, T.Y., Cao, X., Wang, S.W.: Self-adaptive clustering and load-bandwidth management for uplink enhancement in heterogeneous vehicular networks. IEEE Internet Things J. **6**(3), 5607–5617 (2019)

13. Ji, Y.X., Geroliminis, N.: On the spatial partitioning of urban transportation networks. Transp. Res. Part B Methodol. **46**(10), 1639–1656 (2012)

14. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 888–905 (2000)

15. Ji, Y.X., Luo, J., Geroliminis, N.: Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. Transp. Res. Rec. **2422**(1), 1–11 (2014)

16. Schaeffer, S.E.: Graph clustering. Comput. Sci. Rev. **1**(1), 27–64 (2007)

17. Deng, X., Zhai, J., Lv, T., Yin, L.: Efficient vector influence clustering coefficient based directed community detection method. IEEE Access. **5**(99), 17106–17116 (2017)

18. Saeedmanesh, M., Geroliminis, N.: Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. Transp. Res. Part B Methodol. **105**, 193–211 (2017)

19. Anwar, T., Liu, C., Vu, H.L., Leckie, C.: Spatial partitioning of large urban road networks. In: Proc. EDBT. Athens, 24–28, pp. 343–354 (2014)

20. Yan, D.H., Ling, H., Jordan, M.I.: Fast approximate spectral clustering. In: Proc. Fifteenth Intl Conf. Knowledge Discovery and Data Mining. France, pp. 907–916 (2009)

21. Yang, Y.G., Duan, Q.G., Wu, G.B., Chen, R., Li, L., An, Z., Ni, J.M., Shen, C.: Slow coherency based adaptive controlled islanding scheme of the China Southern Power Grid. In: Proc. Asia-Pac. Power Energy Eng. Conf. Brisbane, QLD, Australia, pp. 1–5 (2015)

22. Xu, N., Shang, P., Kamae, S.: Modeling traffic flow correlation using DFA and DCCA. Nonlinear Dyn. **61**(1–2), 207–216 (2010)

23. Tille, Y., Dickson, M.M., Espa, G., Giuliani, D.: Measuring the spatial balance of a sample: a new measure based on Morans I index. Spat Stat-neth **2**(3), 182–192 (2018)

24. Wagner, M., Zeileis, A.: Heterogeneity and spatial dependence of regional growth in the EU: a recursive partitioning approach. Ger. Econ. Rev. **20**(1), 67–82 (2019)

25. Lopez, C., Krishnakumari, P., Leclercq, L., Chiabaut, N., Van Lint, H.: Spatio-temporal partitioning of transportation network using travel time data. Transp. Res. Rec. **2623**(1), 98–107 (2017)

26. Anwar, T., Liu, C.F., Vu, H.L., Islam, M.I., Sellis, T.: Capturing the spatiotemporal evolution in road traffic networks. IEEE Trans. Knowl. Data Eng. **30**(8), 1426–1439 (2018)

27. Anbaroglu, B., Heydecker, B., Cheng, T.: Spatio-temporal clustering for non-recurrent traffic congestion detection on urban road networks. Transp. Res. Part C Emerg. Technol. **4**(8), 47–65 (2014)

28. Zhu, Y., Zheng, W.X.: Multiple lyapunov functions analysis approach for discrete-time switched piecewise-affine systems under dwell-time constraints. IEEE Trans. Automat. Contr. **65**(5), 2177–2184 (2020)

29. Hajiahmadi, M., Haddad, J., De Schutter, B., Geroliminis, N.: Optimal hybrid perimeter and switching plans control for urban traffic networks. IEEE Trans. Control Syst. Technol. **23**(2), 464–478 (2015)

30. Zhu, Y., Zhong, Z., Basin, M.V., Zhou, D.: A descriptor system approach to stability and stabilization of discrete-time switched PWA systems. IEEE Trans. Automat. Contr. **63**(10), 3456–3463 (2018)

31. Zhou, Z., Lin, S., Xi, Y.: A dynamic network partition method for heterogeneous urban traffic networks. In: Proc. ITSC. USA, pp. 820–825 (2012)

32. Moran, P.A.P.: A test for the serial independence of residuals. Biometrika **37**(1–2), 178–181 (1950)

33. You, H., Vittal, V., Wang, X.: Slow coherency-based islanding. IEEE Trans. Power Syst. **19**(1), 483–491 (2015)

34. Saeedmanesh, M., Geroliminis, N.: Clustering of heterogeneous networks with directional flows based on "Snake" similarities. Transp. Res. Part B Methodol. **9**(1), 250–269 (2016)

35. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E. **69**(2), 026113 (2004)