

Motif discovery based traffic pattern mining in attributed road networks

Guojiang Shen^a, Difeng Zhu^a, Jingjing Chen^b, Xiangjie Kong^{a,*}

^a College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

^b Zhijiang College, Zhejiang University of Technology, Shaoxing 312000, China



ARTICLE INFO

Article history:

Received 31 October 2021
Received in revised form 7 May 2022
Accepted 9 May 2022
Available online 16 May 2022

Keywords:

Traffic pattern
Graph clustering
Motif
Attributed networks
Intelligent transportation systems

ABSTRACT

With the development of intelligent transportation systems, clustering methods are now being adopted for traffic pattern recognition to discover the time-varying laws in road networks; this had attracted significant attention from the industry and academia over the past decades. Existing methods mainly focus on the mobility pattern and spatiotemporal dimension, ignoring the complex relationships among these segments in road networks. The main issues can be divided into two categories: deep integration of the structural and attribute information; global spatial dependencies for clustering structural properties. To address these issues, a clustering method for motif-based attributed road networks is proposed. A higher-order connectivity model based on motif discovery is designed, and a weighted matrix of adjacent segments is defined in the road networks. Moreover, a clustering model for motif-based attributed road networks is constructed, considering the joint relationship between node structure and features. In this study, a set of experiments were conducted on two real-world datasets. The results indicated that the performance of the proposed method is superior to that of the state-of-the-art methods.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

With the rapid development of information technology, city operations are producing massive amounts of information on a daily basis. Road traffic information, which contains the records of state changes and equipment operations, is usually dynamic, multi-source, and continuous. Leveraging complex traffic datasets to draw meaningful conclusions is a hot research topic in industry and academia.

Data analysis has become a core branch in intelligent transportation systems (ITS) [1], and traffic pattern recognition is one of the most important research topics. Traffic patterns present variations in transportation networks, including the topological construction of road networks, vehicle trajectory [2], human mobility [3], and other dimensions, and attempt to discover similar weeks, days, or hours within a day that have similar traffic attribute (i.e., traffic index or speed) [4]. It aims to achieve the following: (1) traffic operators can have a temporal plan for operations, such as publishing traffic information in advance, adjusting traffic signal control in a timely manner, and discriminating between recurrent congestion and outliers [5], and (2) urban travelers can adjust routes and plan according to the traffic rules [6].

With the boom of emerging technologies such as fifth-generation networks [7] and edge computing [8], mobile networks have become the mainstream of research because of the considerable accuracy resulting from their frequently updated information [9]. Nevertheless, researchers have paid significant attention to clustering methods for the mining of traffic patterns [10,11]. Almanna et al. [4] modeled the matching problems between two disjoint sets of agents and used a multi-objective consensus clustering algorithm to perform a spatial analysis of the state of urban traffic congestion. Garcia-Rodenas et al. [11] converted traffic flow data into a pseudo-covariance matrix to collect the dynamic correlation between road links, and obtained daily traffic pattern recognition results using a control strategy repository via the k -means algorithm. Yang et al. [12] developed a mobility pattern model to cluster the taxi original destination point data using the density-based spatial clustering of applications with noise (DBSCAN) algorithm and obtained different mobility traffic patterns. Analyzing the characteristics of human mobility are beneficial for understanding traffic patterns and improving transportation services [13]. To fulfill the requirement of understanding human mobility patterns, Wang et al. [2] constructed the k -nearest neighbor-based Internet of Vehicles to achieve dynamic trajectory clustering. Huang et al. [14] utilized the DBSCAN method to identify frequently visited places and quantified the spatial-temporal entropy rate to measure the regularity of private cars' mobility. These methods have achieved

* Corresponding author.

E-mail address: xjkong@ieee.org (X. Kong).

state-of-the-art performance in terms of mobility patterns and spatiotemporal dimensions.

Identifying the differing characteristics of road networks is essential for analyzing similar traffic patterns [15]. Most existing methods are concerned with data features, without the engagement of complex relationships among these segments in road networks. As shown in Fig. 1, the upper panel denotes a road network, where each node represents road segment, solid lines with arrows represent the connectivity of the vehicle's flow, and the lower panel shows clustered results on traffic attributes, such as traffic index and speed. It does not utilize segment structures to find clusters from traffic data. Recently, several methods [16–18] have been proposed to integrate structural properties into clustering analysis. Despite the empirical success of the aforementioned techniques, two issues of traffic clustering analysis via attributed networks [19] still exist:

- Low accuracy owing to the heterogeneity of the structure and attribute information. The structural and attribute information, describing nodes from two perspectives, is essentially heterogeneous [20,21]. The non-linear relationship between the segments in road networks is not considered, and these methods fail to effectively fuse the structure and attribute information of road segments. The merging of these two aspects of road segments remains an issue for traffic pattern recognition.
- Lower-order connectivity of road network structures. Road networks have long-distance neighborhood relations; therefore it may not be adequate to improve clustering performance by utilizing neighbors within a few hops of each node [22]. However, most of these methods neglect interactions among road segments from a structural perspective. The integration of higher-order connectivity patterns into road networks is a new challenge.

Network representation learning extracts potential information by representing the nodes in a network as a low-dimensional space [23]. Attributed networks are a form of network representation. They fuse the node attributes and topological structures [24] and present various structure patterns with average degree, finiteness connectivity, and spatial pattern correlation [25, 26]. Thus, road networks are real-life examples of attributed networks. Traffic patterns can be considered not only for traffic attributes but also for the underlying physical topology to evaluate better clustering accuracy. Motifs are non-isomorphic connected sub-structures that frequently occur in attributed networks, in which the number of nodes is greater than or equal to three [27]. As motifs consider higher-order correlations of organizational structures, they are suitable for capturing global information to improve the performance of structural feature aggregation [28,29].

Motivated by this, a traffic pattern clustering method utilizing motif-based attributed road networks (PCMAN) is proposed. The method is designed to handle higher-order dependencies and improve the accuracy of traffic pattern recognition. Compared with the current clustering methods, the proposed method has two significant advantages. First, motifs are applied to define road network structures and extract higher-order spatial correlations of traffic information. Second, the consensus graph clustering method is adopted to identify the optimized factorization results through attributed road networks and alleviate the heterogeneity between the structural and attribute properties of the segments. To the best of our knowledge, this study is the first attempt to apply motif-based attributed networks to analyze traffic patterns.

The major contributions of this paper can be summarized as follows:

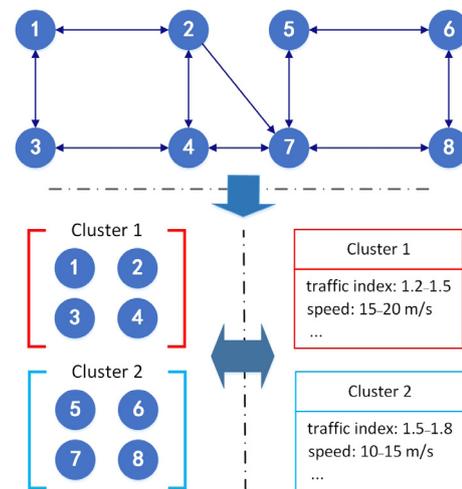


Fig. 1. Example of traffic pattern clustering in a road network.

- The proposed method clusters the joint relationship between node structure and features, and utilizes the k -means algorithm to mine traffic patterns in road networks. Moreover, the joint relationship is determined by considering nonnegative matrix factorization and Karush–Kuhn–Tucker condition optimization.
- A higher-order connectivity model of road networks that utilizes motifs, is presented. It accurately represents the weighted matrix of adjacent road segments and captures the global information of road networks through motif-based search and iterative adjacency matrix string representation.
- Extensive experiments were performed on two real-world datasets to evaluate the performance of the proposed method. The experimental results suggest that the proposed method can achieve the best performance and help improve insights into urban commuting schemes.

The rest of the paper is organized as follows: Section 2 describes the state-of-the-art clustering method. In Section 3, the proposed method is introduced. Section 4 presents the experimental results of the proposed method. Finally, Section 5 concludes the paper.

2. Related work

Clustering, which helps discover the similarity of different data objects, is important for exploring traffic patterns [30]. In this section, related studies are reviewed with regard to traffic pattern clustering and attributed network learning.

2.1. Traffic pattern clustering

Existing studies on traffic pattern clustering have focused on the macroscopic operating state of traffic. Calafate et al. [31] utilized historical data via induction loop detections and then proposed a time-dependent traffic flow model. The obtained traffic patterns provided suitable route recommendations to drivers. Nguyen et al. [32] established a fast-search index based on traffic congestion characteristics and adopted an extended support vector machine (SVM) algorithm for regional congestion pattern classification. Byon et al. [33] developed a real-time pattern recognition model. It is based on an artificial neural network and improved with a geographic information system layer, which was

used for automatic pattern detection on a specific ITS. Zhang et al. [34] estimated the true value of the membership degree of any element in the set by the weighted aggregation operator of the double hesitation fuzzy rough set and proposed a method based on the assessment deviation theories of traffic patterns. However, these methods are designed only to handle data features.

Some graph-based methods consider the strong relations between data structures [35,36] and are feasible for capturing global information to improve the clustering performance [22]. Many graph-based clustering methods aim to leverage graph connectivity patterns and partition data nodes into disjointed groups. Zhang et al. [37] used graph embedding and clustering algorithms to discover hierarchical mobility community structures and identified time-varying traffic behavior patterns based on smart card data and other urban data. Chandra et al. [38] used a weighted dynamic geometric graph to represent the proximity of road agent behaviors and proposed a regularization algorithm based on spectral clustering to improve trajectory prediction accuracy. Ou et al. [39] developed a fuzzy clustering method to identify the spatiotemporal impact areas of incidents in an automated freeway and adopted a clustering algorithm based on graph theory to distinguish the relations of possible multiple independent blocks in a non-recurrent congestion area. However, these methods do not consider the features associated with these road network nodes in graph topological structures.

2.2. Attributed network learning

Recently, convergent development of complex networks has inspired new ideas for graph clustering. Some graph-based clustering methods attempt to consider both node connectivity and data features (i.e., the attributed networks). Wang et al. [40] designed a marginalized graph autoencoder algorithm to integrate node and feature information, and marginalized the corrupted features in a graph autoencoder context to learn graph feature representations. Pan et al. [41] encoded the topological structure and node content to match a prior distribution, and developed two adversarial methods learned with a graph convolutional autoencoder to produce a robust representation in graph clustering and other fields. Bu et al. [42] proposed a graph k -means framework to integrate both node and feature information in social media networks (SMNs) and developed a dynamic game model to explore the evolution of the opinion matrix by discovering the Pareto-optimal community structure. Li et al. [43] developed an embedding graph autoencoder with joint clustering via adjacency sharing (EGAE-JOCAS) model to share the same adjacency within graph convolution layers. It uses node attributes and adjacency to generate adequate representations for joint clustering. Although these methods achieve promising performance in various application contexts, the node attributes and topological structures in a graph fail to fuse effectively owing to the heterogeneity of the two aspects.

Matrix factorization (MF) can approximate the original data matrix by the product of two or more matrices [44] and solve the aforementioned problem (i.e., the heterogeneity of node attributes and topological structures). Nonnegative matrix factorization (NMF) is a typical MF method that has been widely adopted to assign cluster members by leveraging the consensus relationship between node attributes and topological structure [45]. Ma and Dong [46] presented a semi-supervised NMF by incorporating partial information into the topological structure and proved the equivalence of evolutionary modularity density. Meng et al. [47] projected an attributed network into a unified low-dimensional vector space and developed a search result diversification method using NMF to learn node representation. Guo

et al. [48] designed a consensus factorization-based framework to co-cluster networked data and simultaneously aligned networked nodes and features for the best clustering results through NMF. Huang et al. [24] addressed the issue of heterogeneity of structures and attributes using joint NMF and graph optimization. They factorized the established attribute similarity matrix and topological adjacent matrix to detect attributed networks. Considering the dynamic evolution of road networks, these methods are not capable of capturing global graph structures [49] and fail to address the higher-order connectivity patterns of road networks.

To address this problem, in [50], the researchers employed a higher-order organization of complex networks. Motifs are networked sub-structures in a graph and are beneficial for understanding global structural graph principles [30]. Yin et al. [51] proposed a motif-based approximate personalized Pagerank (MAPPR) algorithm to incorporate higher-order clustering information captured by motifs and solve the problem of large-scale real-time changeable networks. Li et al. [52] introduced motif correlation clustering to minimize the cost of clustering errors associated with both edges and higher-order network structures and offered approximation guarantees according to motif size. William et al. [53] captured higher-order graph structures using motif adjacency matrices based on weighted networks and constructed a motif-based model for clustering bipartite weighted networks. Mei et al. [54] designed a graph-based clustering algorithm by exploiting k -core decomposition and motifs (KCoreMotif), and grouped the remaining vertices in the remaining $(k-1)$ -core sub-graphs to obtain the desired clusters of higher-order networks. Several studies indicate that road networks contain complex sub-graph structures, which have certain characteristics of transitivity, interactivity, balance, etc. Therefore, motifs are suitable for higher-order road networks with small sub-graph structures as units. However, motif-based clustering is adopted in medical identification [55], smart meters [56], and other fields and rarely adopted in transportation. Moreover, motifs fail to handle the network heterogeneity of node attributes and topological structures.

Motivated by the aforementioned literature, a clustering method based on an attributed network that combines NMF and motifs, is proposed. In this method, NMF is applied to factorize the established node feature similarity matrix and topological adjacent matrix to obtain a better clustering result of traffic patterns. The performance of the fusion of node topology and attribute similarity has been verified by [57]. In addition, by adopting motifs, the method benefits from capturing higher-order spatial dependencies in the traffic.

3. Methodology

In this section, the proposed PCMAN method is presented to identify the traffic patterns in a road network. Fig. 2 illustrates the PCMAN framework.

The traffic data process generates a topological adjacent matrix (denoted by \mathbf{W}) and a node feature matrix (denoted by \mathbf{F}) for each road segment (node). The motif-based weighted adjacency matrix (denoted by \mathbf{W}_M) is the variant of node adjacent matrix \mathbf{W} using the motif-based search and influence degree weighted calculation method. Simultaneously, node feature matrix was transformed into node feature similarity matrix (denoted by \mathbf{S}) via similarity calculation. NMF was then employed to factorize \mathbf{W}_M and \mathbf{S} , and the common basis matrix (denoted by \mathbf{B}) was obtained when the factorization iteration ended. The outputs of the different basis matrix components were clustered by road segments to obtain the traffic pattern results. The key notations used in this study are listed in Table 1.

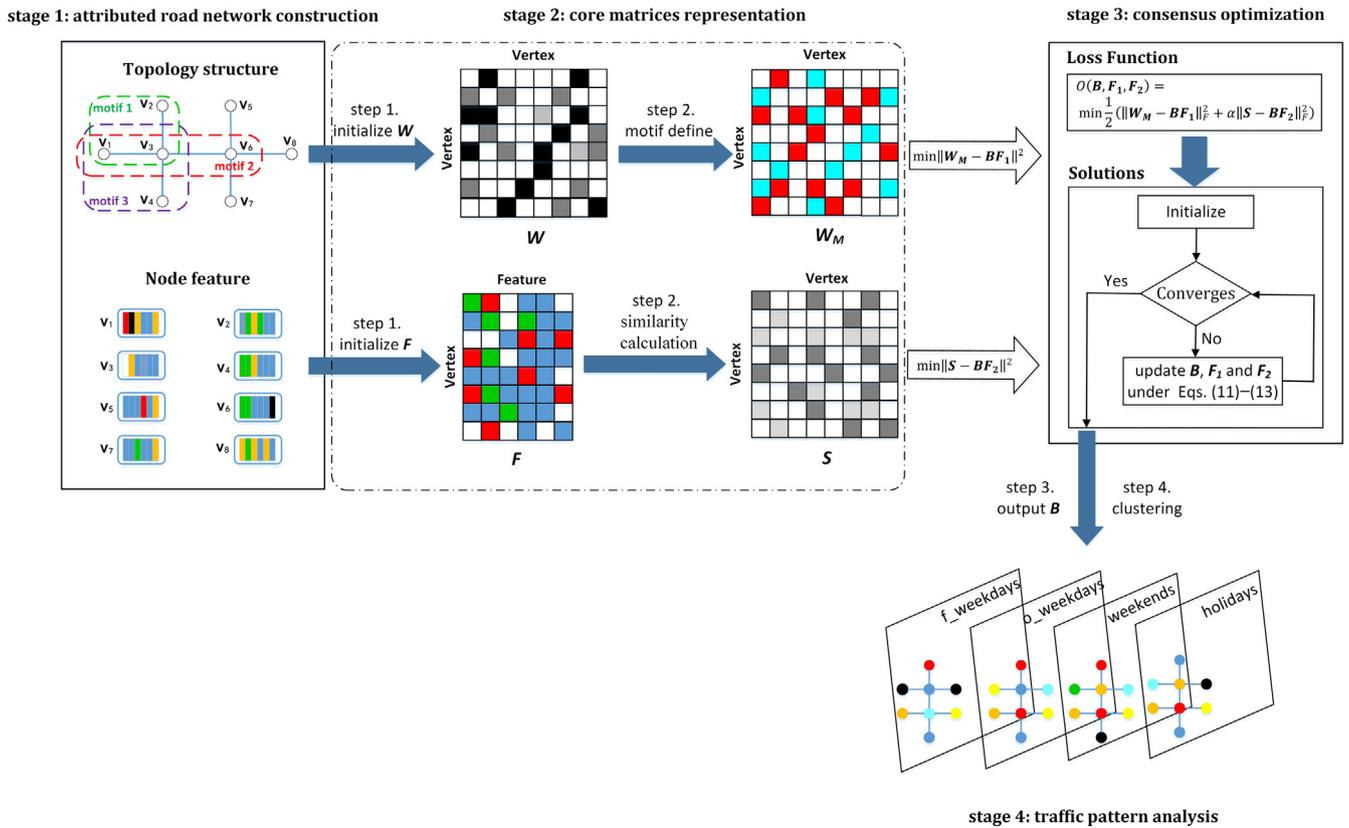


Fig. 2. Overview of the proposed PCMAN Framework. PCMAN performs co-clustering by considering information from two aspects: road network topological structures and node features. PCMAN factorizes each of them and then explores the consensus of their factorized values by using loss function to achieve optimal clustering results. The acronyms “f_weekdays” and “o_weekdays” represent the first and last weekdays, and other weekdays, respectively.

Table 1

Notations used in PCMAN.

Symbols	Definition and description
G	A directed graph
V	Node set of G
E	Edge set of G
\mathbf{W}_M	The motif-based weighted adjacency matrix
$w_{ij,k}$	The weighted value of adjacent nodes v_i and v_j
F	Node feature matrix
S	Node feature similarity matrix
B	The common basis matrix
F_1	The feature matrix for \mathbf{W}_M
F_2	The feature matrix for S
α	Parameter for S and F_2
\bar{S}_k	The average of the silhouette measures

3.1. Problem formulation

A directed graph $G = (V, E, \mathbf{W}_M)$ with m nodes is used to describe a road network, where node $v_i \in V$ denotes road segment, edge $(v_i, v_j) \in E$ denotes the directed connection from node v_i to node v_j and $\mathbf{W}_M \in \mathbb{R}^{N \times N}$ represents a motif-based weighted adjacency matrix. When there is an edge from node v_i to node v_j , \mathbf{W}_M is the sum of the weights on the edge degree involved by motifs; otherwise, \mathbf{W}_M is zero.

To obtain the node feature similarity matrix, node features and node feature similarities must be considered. Node feature matrices $F \in \mathbb{R}^{N \times d}$ are collected to characterize the network node features, where d is the number of features in a node. Node feature similarity matrices $S \in \mathbb{R}^{N \times N}$ are used to explore clustered results for both nodes and features. For simplicity, a linear relationship $S_{ij} = \langle f_i, f_j \rangle$ is used, where $f_i \in F$ is a vector

representation of the i th node across all features and $S_{ij} = \langle f_i, f_j \rangle$ is the similarity degree of f_i and f_j .

3.2. Motif-based higher-order road network structural representation

Because the higher-order connectivity structures presented in road networks have their own biases, triangular motifs are selected as the research object in combination with traffic theory. Fig. 3 presents all the triangular motifs used [50], in which different motifs convey different interactive patterns.

Owing to the directional characteristics of road networks, the five motif types shown in Fig. 3 are selected to represent higher-order structures in road networks. M_1 , M_5 , M_8 , M_9 , and M_{10} represent the ring, detour, diverging, two-hop and converging structures, respectively. Fig. 4 illustrates an example. The road links in local areas form three motif types (two-hop, converging and diverging), the ring motif-based type, and the detour motif-based type in Fig. 4 (a) - Fig. 4 (c), respectively. The five types of two-hop, converging, diverging, ring, and detour are shown in Fig. 4 (d). Considering the different influences of motifs in road networks, a three-step strategy was adopted to express higher-order structures.

3.2.1. Motif-based graph search

To avoid repeatedly counting motifs in road networks, a breadth-first search on the graphs is adopted.

First, ID numbers are performed on all nodes in road networks to form an ordered queue $Q_v = \{v_1, v_2, \dots, v_n\}$, and the first node v_1 is considered the starting node of the motif-based search to form a sub-queue $Q_1 = \{v_1\}$. Next, the second layer enters with a tree structure, and the remaining $n - 1$ nodes are added to Q_1 in

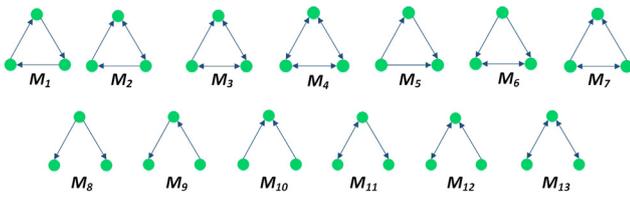


Fig. 3. Triangular motifs.

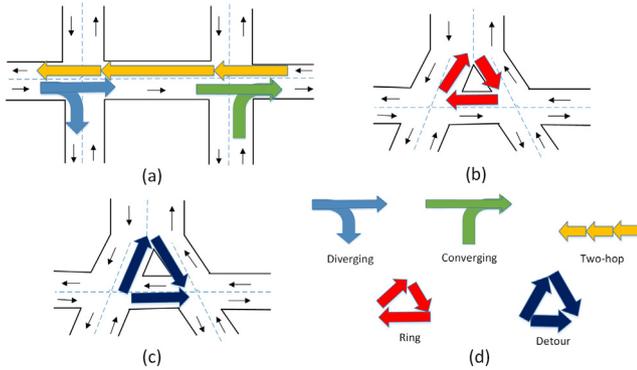


Fig. 4. Higher-order connectivity patterns in a road network.

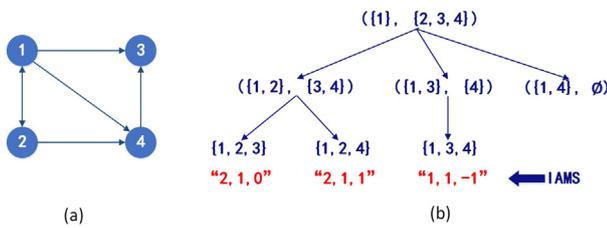


Fig. 5. Illustration for motif-based search and IAMS. All the motifs are at the leaf level with IAMS labeled beneath.

sequence to form $n - 1$ sub-queues $\{v_1, v_2\}, \{v_1, v_3\}, \dots, \{v_1, v_n\}$. If the two nodes of the sub-queue are not adjacent, the sub-queue is deleted. Third, because the number of nodes in the motif-based road networks is three, the tree structure enters the third layer to complete the motif-based search, i.e., $n - 1$ subqueues $\{v_1, v_2\}, \{v_1, v_3\}, \dots, \{v_1, v_n\}$ and the remaining nodes corresponding to Q_v are all added in sequence to form $\frac{n^2-3n+2}{2}$ new sub-queues $\{v_1, v_2, v_3\}, \{v_1, v_2, v_4\}, \dots, \{v_1, v_3, v_4\}, \dots, \{v_1, v_{n-1}, v_n\}$. If the third node of the sub-queue is not adjacent to the previous two nodes, the sub-queue is deleted. In the same way, the second node v_2 is the starting search node. When the previous node v_1 is removed, the aforementioned search method is also adopted to determine the sub-queue with three nodes. It continues until the last node v_n , and then the motif-based search ends.

Fig. 5 is an example of motif-based search: (a) is a directed graph of a road network with four nodes, and (b) is the motif search based on (a). According to the breadth-first search on graphs, node {1} is used as the starting search node, three sub-queues $\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}$ are obtained, and the number of motifs is determined.

3.2.2. Motif-based type determination

According to the breadth-first search on graphs in the first step, the number of motifs in road networks can be determined; however, the specific types of motifs cannot be distinguished. Thus, the iterative adjacency matrix string (IAMS) representation method [28] was adopted to define motif-based identification

and determine motif types. IAMS is defined as follows: if there is a motif with three nodes, 0 indicates that the current node has no adjacent relationship with the previous nodes, 1 indicates that the previous nodes have a directed relationship with the current node, -1 indicates that the current node has a directed relationship with the previous nodes, and 2 represents the bidirectional relationship between the current node and the previous nodes. The IAMS corresponding to the three motifs in Fig. 5(b) is represented as “2,1,0”, “2,1,1”, and “1,1,-1”.

3.2.3. Motif-based weighted adjacency matrix representation

According to the previous two steps, all road network motifs are classified according to their types, and the IAMS values of motifs are normalized corresponding to each pair of adjacent nodes v_i and v_j .

$$w_{ij,k} = \frac{\text{count}(\text{iams}_k)}{|\text{set}(\text{motif})|}, \quad (1)$$

where $w_{ij,k}$ is the weighted value of the adjacent nodes v_i and v_j belonging to the k th motif, $\text{count}(\text{iams}_k)$ is the number of IAMS values iams_k in all motifs, and $\text{set}(\text{motif})$ is the set of all motifs.

The motif-based weighted adjacency matrix is expressed as $\mathbf{W}_M \in \mathbb{R}^{N \times N}$:

$$(\mathbf{W}_M)_{ij} = \sum_{k=1}^n w_{ij,k}, \quad (2)$$

where $(\mathbf{W}_M)_{ij}$ is the motif-based weighted value of the adjacent nodes v_i and v_j , and n is the number of $\text{set}(\text{motif})$.

3.3. Motif-based attributed road network pattern clustering

Because the attribute consistency of nodes corresponds to the graph topological structures in the feature similarity matrix [22], a node feature similarity matrix \mathbf{S} is constructed to alleviate the heterogeneity between structure and attribute information of road networks.

Given the attributed road network G , NMF approximates the motif-based weighted adjacency matrix \mathbf{W}_M and node feature similarity matrix \mathbf{S} by the product of three nonnegative low-rank matrices \mathbf{B} , \mathbf{F}_1 , and \mathbf{F}_2 such that

$$\mathbf{W}_M \approx \mathbf{B}\mathbf{F}_1, \mathbf{S} \approx \mathbf{B}\mathbf{F}_2, \mathbf{B} \geq 0, \mathbf{F}_1 \geq 0, \mathbf{F}_2 \geq 0 \quad (3)$$

Eq. (3) can be solved by minimizing the l_2 norm of the approximation, that is,

$$\min \|\mathbf{W}_M - \mathbf{B}\mathbf{F}_1\|^2, \mathbf{B} \geq 0, \mathbf{F}_1 \geq 0 \quad (4)$$

$$\min \|\mathbf{S} - \mathbf{B}\mathbf{F}_2\|^2, \mathbf{B} \geq 0, \mathbf{F}_2 \geq 0, \quad (5)$$

where \mathbf{B} is the common basis matrix, \mathbf{F}_1 and \mathbf{F}_2 are feature matrices for \mathbf{W}_M and \mathbf{S} , respectively, and $\|\cdot\|^2$ is the l_2 norm.

According to Eqs. (3)–(5), the loss function based on the joint decomposition \mathbf{W}_M and \mathbf{S} is expressed as follows:

$$O(\mathbf{B}, \mathbf{F}_1, \mathbf{F}_2) = \min \frac{1}{2} (\|\mathbf{W}_M - \mathbf{B}\mathbf{F}_1\|_F^2 + \alpha \|\mathbf{S} - \mathbf{B}\mathbf{F}_2\|_F^2), \quad (6)$$

where $O(\mathbf{B}, \mathbf{F}_1, \mathbf{F}_2)$ is the loss function, $\|\cdot\|_F$ is the Frobenius norm of matrices, and α is the parameter for \mathbf{S} and \mathbf{F}_2 .

By expressing the aforementioned loss function as a matrix trace function, the objective function represented by the trace function can be obtained as follows:

$$L = \|\mathbf{B}\mathbf{F}_1\|_F^2 + \|\mathbf{B}\mathbf{F}_2\|_F^2 - 2\text{Tr}(\mathbf{B}^T \mathbf{W}_M \mathbf{F}_1^T) - 2\alpha \text{Tr}(\mathbf{B}^T \mathbf{S} \mathbf{F}_2^T) + \|\mathbf{W}_M\|^2 + \alpha \|\mathbf{S}\|^2, \quad (7)$$

where $\text{Tr}(\cdot)$ is the trace function, and \mathbf{B}^T , \mathbf{F}_1^T and \mathbf{F}_2^T are the transpose matrices of the low-rank matrices \mathbf{B} , \mathbf{F}_1 and \mathbf{F}_2 , respectively.

As $\|\mathbf{W}_M\|^2$ and $\|\mathbf{S}\|^2$ are constant, the three components of \mathbf{B} , \mathbf{F}_1 and \mathbf{F}_2 in Eq. (7) can be derived as follows:

$$\frac{\partial L}{\partial \mathbf{B}} = -(\mathbf{W}_M \mathbf{F}_1^T + \alpha \mathbf{S} \mathbf{F}_2^T) + \mathbf{B} \mathbf{F}_1 \mathbf{F}_1^T + \mathbf{B} \mathbf{F}_2 \mathbf{F}_2^T \quad (8)$$

$$\frac{\partial L}{\partial \mathbf{F}_1} = \mathbf{B}^T \mathbf{B} \mathbf{F}_1 - \mathbf{B}^T \mathbf{W}_M \quad (9)$$

$$\frac{\partial L}{\partial \mathbf{F}_2} = \mathbf{B}^T \mathbf{B} \mathbf{F}_2 - \alpha \mathbf{B}^T \mathbf{S} \quad (10)$$

By setting the partial derivatives in Eqs. (8)–(10) as zero, the updated rules for \mathbf{B} , \mathbf{F}_1 and \mathbf{F}_2 are calculated according to the Karush–Kuhn–Tucker (KKT) condition [24]:

$$\mathbf{B} = \mathbf{B} \odot \frac{[\mathbf{W}_M \mathbf{F}_1^T + \alpha \mathbf{S} \mathbf{F}_2^T]}{[\mathbf{F}_1 \mathbf{F}_1^T + \mathbf{F}_2 \mathbf{F}_2^T]} \quad (11)$$

$$\mathbf{F}_1 = \mathbf{F}_1 \odot \frac{[\mathbf{B}^T \mathbf{W}_M]}{[\mathbf{B}^T \mathbf{B}]} \quad (12)$$

$$\mathbf{F}_2 = \mathbf{F}_2 \odot \frac{[\alpha \mathbf{B}^T \mathbf{S}]}{[\mathbf{B}^T \mathbf{B}]}, \quad (13)$$

where \odot denotes Hadamard product, $[\]/[\]$ denotes Hadamard division.

Algorithm 1: PCMAN.

Input: $G = (V, E)$, m nodes, n motifs, parameter α

Output: cluster C

- 1: Initialize: \mathbf{F} , \mathbf{S}
 - 2: Initialize: \mathbf{W}_M according to Eqs. (1)–(2)
 - 3: set \mathbf{B} , \mathbf{F}_1 and \mathbf{F}_2 according to Eqs. (3)–(5)
 - 4: **repeat**
 - 5: update $O(\mathbf{B}, \mathbf{F}_1, \mathbf{F}_2)$ according to Eq. (6)
 - 6: $\mathbf{B} = \mathbf{B} \odot \frac{[\mathbf{W}_M \mathbf{F}_1^T + \alpha \mathbf{S} \mathbf{F}_2^T]}{[\mathbf{F}_1 \mathbf{F}_1^T + \mathbf{F}_2 \mathbf{F}_2^T]}$
 - 7: $\mathbf{F}_1 = \mathbf{F}_1 \odot \frac{[\mathbf{B}^T \mathbf{W}_M]}{[\mathbf{B}^T \mathbf{B}]}$
 - 8: $\mathbf{F}_2 = \mathbf{F}_2 \odot \frac{[\alpha \mathbf{B}^T \mathbf{S}]}{[\mathbf{B}^T \mathbf{B}]}$
 - 9: **until** the minimum of $O(\mathbf{B}, \mathbf{F}_1, \mathbf{F}_2)$ is reached
 - 10: obtain C clustered by \mathbf{B} using k -means
 - 11: **return** C
-

According to Eqs. (11)–(13), the calculated \mathbf{B} , \mathbf{F}_1 , and \mathbf{F}_2 values are substituted into Eq. (6) for repeated iterations until the value of the loss function reaches 0.001. Finally, the common basis matrix \mathbf{B} is calculated.

3.4. Example of PCMAN

As shown in Fig. 2, the calculating process is presented as follows:

According to the topological structure, the node adjacent matrix \mathbf{W} is expressed:

$$\mathbf{W} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Considering edge (v_1, v_3) as an example, in which the involved motifs are $v_1 \rightarrow v_3 \rightarrow v_2$, $v_1 \rightarrow v_3 \leftarrow v_2$, $v_1 \rightarrow v_3 \rightarrow v_6$,

Table 2
Records of original data.

Segment ID	Time	Traffic index
2271	2018-09-01 00:00	1.1322
2271	2018-09-01 00:10	1.2019
2271	2018-09-01 00:20	1.1394
2271	2018-09-01 00:30	1.2076
...
2271	2018-09-10 23:30	1.1019
2271	2018-09-10 23:40	1.0918
2271	2018-09-10 23:50	1.1797

$v_1 \rightarrow v_3 \leftarrow v_6$, $v_1 \rightarrow v_3 \rightarrow v_4$, and $v_1 \rightarrow v_3 \leftarrow v_4$, thus the number of motifs is 6. The number of $set(motif)$ is 24; thus, $(\mathbf{W}_M)_{13}$ is 0.25. According to Eqs. (1)–(2), the motif-based weighted adjacency matrix \mathbf{W}_M can be numerically solved.

The node feature matrix \mathbf{F} corresponds to the nodes in v_1, v_2, \dots, v_8 . It is directly constructed using the traffic index of six timeslots and is expressed as follows:

$$\mathbf{F} = \begin{pmatrix} 1.4 & 1.5 & 1.3 & 1.2 & 1.2 & 1.3 \\ 1.2 & 1.1 & 1.3 & 1.1 & 1.2 & 1.2 \\ 1.0 & 1.3 & 1.2 & 1.2 & 1.2 & 1.2 \\ 1.1 & 1.1 & 1.3 & 1.2 & 1.2 & 1.2 \\ 1.2 & 1.2 & 1.2 & 1.4 & 1.2 & 1.3 \\ 1.1 & 1.1 & 1.2 & 1.2 & 1.2 & 1.5 \\ 1.2 & 1.2 & 1.1 & 1.2 & 1.2 & 1.3 \\ 1.3 & 1.1 & 1.3 & 1.2 & 1.3 & 1.2 \end{pmatrix}$$

The node feature similarity matrix \mathbf{S} is obtained by the similarity degree method described in Section 3.1:

$$\mathbf{S} = \begin{pmatrix} 1.00 & 0.75 & 0.79 & 0.70 & 0.73 & 0.56 & 0.74 & 0.63 \\ 0.75 & 1.00 & 0.57 & 0.93 & 0.52 & 0.47 & 0.31 & 0.75 \\ 0.79 & 0.57 & 1.00 & 0.72 & 0.70 & 0.39 & 0.41 & 0.48 \\ 0.70 & 0.93 & 0.72 & 1.00 & 0.68 & 0.52 & 0.30 & 0.73 \\ 0.73 & 0.52 & 0.70 & 0.68 & 1.00 & 0.67 & 0.80 & 0.57 \\ 0.56 & 0.47 & 0.39 & 0.52 & 0.67 & 1.00 & 0.65 & 0.41 \\ 0.74 & 0.31 & 0.41 & 0.30 & 0.80 & 0.65 & 1.00 & 0.41 \\ 0.63 & 0.75 & 0.48 & 0.73 & 0.57 & 0.41 & 0.41 & 1.00 \end{pmatrix}$$

According to Eqs. (3)–(6), the feature matrices are updated until convergence (parameter α is set to 0.01). Then the feature matrices \mathbf{F}_1 and \mathbf{F}_2 and the common basis matrix \mathbf{B} are given.

The common basis matrix \mathbf{B} is performed by k -means algorithm, and outputs the clustered results.

The algorithm of PCMAN is illustrated in Algorithm 1.

4. Experiments

4.1. Data preparation

To evaluate the proposed method, two different datasets containing traffic index were used: (1) the dataset in Chengdu, China, the time span is from September to October 2018, and (2) the dataset in Shenzhen, China, the time span is from April to May 2018. Some of the records in the datasets are listed in Table 2. All the datasets were downloaded from <https://outreach.didichuxing.com>.

Considering that each road segment has different characteristics including length, number of lanes, speed limitation, and geographic location, the traffic index is a standardized quantitative indicator to measure traffic congestion. The traffic index can be expressed as follows:

$$C_i^t = \frac{v_i^{max} - v_i^t}{v_i^{max} - v_i^{min}}, \quad (14)$$

Table 3

Statistics of road networks, where $|V|$ and $|E|$ denote the number of nodes and edges, respectively.

Dataset	$ V $	$ E $	Attribute	Cluster
Chengdu	74	164	1776	7
Shenzhen	79	157	1896	8

where C_i^t is the traffic index of road segment i during timeslot t , v_i^t is the average speed, v_i^{max} and v_i^{min} are the maximum and minimum speeds corresponding to road segment i in the historical dataset, respectively. The lower the average speed, the larger the traffic index and the more congested the traffic, and vice versa.

From the perspective of traffic congestion, most of the time on a road is smooth, and congestion generally occurs only during the morning and evening rush hours. For fine-grained traffic congestion, two datasets were filtered in rush hours (7 am–9 am and 5 pm–7 pm), and aggregated into 24 timeslots (10-min. intervals per day). 74 segments in Chengdu and 79 segments in Shenzhen each comprised the road network. Road segments were selected as networked nodes, in which the traffic index at different timeslots was selected as the node features. To meet the requirements of testing, the two datasets were grouped into more clusters empirically, and the information is summarized in Table 3.

Therefore, the node feature matrix in the Chengdu road network $F_c \in \mathbb{R}^{74d \times 24}$ and that in the Shenzhen road network $F_s \in \mathbb{R}^{79d \times 24}$ are generated. Notably, d is the number of days in each dataset.

All experiments are compiled and tested based on Python 3.8 and TensorFlow 2.7.0.

4.2. Experimental settings

4.2.1. Baselines

The proposed method was compared with state-of-the-art methods for traffic pattern clustering.

- **ARGE** [41]: This is an adversarially regularized graph autoencoder for graph clustering with both node features and topology.
- **ARVGE** [41]: This is an adversarially regularized variational graph autoencoder for graph clustering with both topological and content information.
- **GK-means** [42]: This is a graph k -means framework that integrates node and feature information in SMNs and explores the evolution of the opinion matrix.
- **EGAE-JOCAS** [43]: This is an embedding graph autoencoder with joint clustering via adjacency sharing that utilizes node attributes and adjacency to generate adequate representations for joint clustering.
- **PCMAN**: The proposed method defines the weighted matrix of adjacent segments by motifs and uses attributed networks to discover traffic patterns.
- **PCAN**: This is a comparison for **PCMAN**, which uses attributed networks to discover traffic patterns.

4.2.2. Evaluation metrics

Three widely used performance measures were employed to evaluate performance: clustering accuracy (Acc), normalized mutual information (NMI), and macro F1-score (F1). A better result should lead to higher values for all metrics.

$$Acc = \frac{\sum_{i=1}^r \delta(a_i, \text{map}(b_i))}{r} \quad (15)$$

$$NMI = \frac{I(A, B)}{\sqrt{H(A)H(B)}} \quad (16)$$

$$F1 = \frac{2}{s} \sum_{j=1}^s \frac{P_j \times R_j}{P_j + R_j} \quad (17)$$

$$P_j = \frac{TP_j}{TP_j + FP_j} \quad (18)$$

$$R_j = \frac{TP_j}{TP_j + FN_j} \quad (19)$$

where A and B are the clustered results and ground truths, respectively; a_i and b_i are the individuals corresponding to A and B , respectively; $\delta(x, y) = 1$ if $x = y$, otherwise $\delta(x, y) = 0$; $\text{map}(b_i)$ is the best mapping function that permutes clustering labels to match ground truths; $H(A)$ and $H(B)$ are the entropies of A and B , respectively; $I(A, B)$ is the mutual information between A and B ; TP stands for the true value of the positive sample being the same as the predicted value, FP stands for the true value of the negative sample being the same as the predicted value, and FN stands for the true value of the negative sample being different from the predicted value; r and s are the number of clustering and classes, respectively.

4.2.3. Parameter selection

According to Eq. (6), parameter α determines the tradeoff between the topological structure and node features. The PCMAN is trained using parameter α ranging from 0.01 to 0.05.

4.3. Results

The proposed PCMAN method is compared with state-of-the-art methods. These comparisons belong to two categories of tasks: different parameters of the PCMAN and other compared methods. For a fair comparison, these methods were tested under different numbers of clusters.

4.3.1. Parametric study

According to Fig. 6(a) - Fig. 6(c), as the cluster number $K \leq 4$, PCMAN achieves the best performance when α is 0.03, and the best performance is achieved when α is 0.04. According to Fig. 6(d) - Fig. 6(f), as the cluster number $K \leq 5$, PCMAN achieves the best performance when α is 0.03; By contrary, the best performance is achieved when α is 0.04. A possible explanation for this result could be that: as the value of K decreases, PCMAN reaches a good balance between the topological structure and node features when α is 0.03. As the value of K increases, PCMAN achieves a tradeoff among the various attributed road networks when α is 0.04. Therefore, in a subsequent study, the best performance of the PCMAN was provided according to the aforementioned parameter selection.

4.3.2. Comparative results on baselines

Tables 4–9 show the evaluation results for the Chengdu and Shenzhen datasets. For each compared method on all three evaluation metrics, the final performance scores (avg.) were obtained by averaging the scores of different clusters. The experimental results are summarized as follows:

- The proposed PCMAN method outperformed the compared methods in all cases. The overall performance ranking of these methods on the two datasets was PCMAN, GK-means, EGAE-JOCAS, ARVGE, PCAN, and ARGE. On the Chengdu dataset, compared with the second best method, that is, GK-means, PCMAN achieves 3.1%, 12.4%, and 5.3% improvements in Acc, NMI, and F1, respectively. In contrast to GK-means on the Shenzhen dataset, PCMAN achieves 4.3%, 8.5%, and 3.9% improvements in Acc, NMI, and F1, respectively.

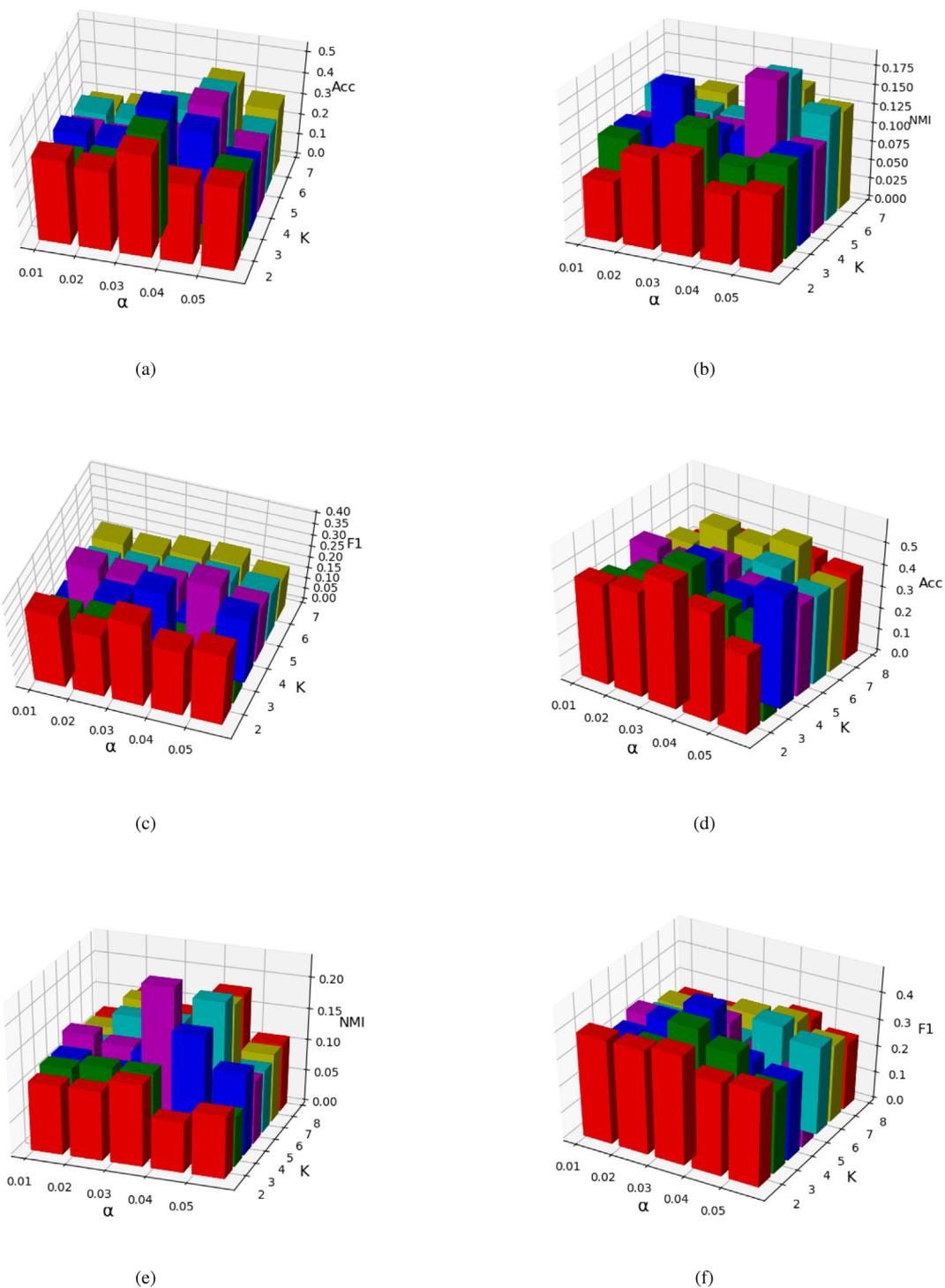


Fig. 6. Performance of PCMAN with combinations of α and K values. (a)-(c) show the performance of Acc, NMI, and F1 on the Chengdu dataset, respectively. (e)-(f) show the performance of Acc, NMI, and F1 on the Shenzhen dataset, respectively.

Table 4
Clustering performance on the Chengdu dataset (Acc). Bold is used to highlight the best result.

K	2	3	4	5	6	7	avg.
ARGE	0.4922	0.4782	0.4264	0.3651	0.3638	0.2493	0.3959
ARVGE	0.5663	0.4441	0.4301	0.3583	0.3965	0.3297	0.4209
GK-means	0.5594	0.5350	0.4482	0.4003	0.4396	0.4331	0.4693
EGAE-JOCAS	0.5045	0.4556	0.4411	0.3679	0.4182	0.3949	0.4304
PCAN	0.4214	0.4686	0.4618	0.4491	0.3680	0.3520	0.4202
PCMAN	0.5052	0.5136	0.5269	0.4928	0.4515	0.4124	0.4838

Table 5

Clustering performance on the Chengdu dataset (NMI). Bold is used to highlight the best result.

K	2	3	4	5	6	7	avg.
ARGE	0.1106	0.1169	0.0923	0.1254	0.1437	0.1406	0.1216
ARVGE	0.1244	0.1162	0.0919	0.1297	0.1776	0.1759	0.1360
GK-means	0.1438	0.1312	0.1669	0.1606	0.1254	0.1419	0.1450
EGAE-JOCAS	0.1120	0.1414	0.1429	0.1321	0.1237	0.1274	0.1299
PCAN	0.1225	0.1291	0.1551	0.1328	0.1416	0.1534	0.1391
PCMAN	0.1309	0.1496	0.1790	0.1864	0.1820	0.1501	0.1630

Table 6

Clustering performance on the Chengdu dataset (F1). Bold is used to highlight the best result.

K	2	3	4	5	6	7	avg.
ARGE	0.3601	0.2895	0.2261	0.2367	0.2297	0.1815	0.2162
ARVGE	0.3656	0.2659	0.2727	0.2426	0.1955	0.2172	0.2599
GK-means	0.3873	0.3004	0.2683	0.2988	0.2169	0.2467	0.2864
EGAE-JOCAS	0.3509	0.2554	0.3155	0.2975	0.2202	0.2313	0.2785
PCAN	0.3183	0.3337	0.2899	0.2251	0.1715	0.2116	0.2584
PCMAN	0.4009	0.2809	0.3399	0.3190	0.2395	0.2297	0.3017

Table 7

Clustering performance on the Shenzhen dataset (Acc). Bold is used to highlight the best result.

K	2	3	4	5	6	7	8	avg.
ARGE	0.5191	0.4043	0.3574	0.2638	0.2340	0.2723	0.2149	0.3237
ARVGE	0.5279	0.5457	0.4787	0.4191	0.4745	0.4191	0.3957	0.4658
GK-means	0.5962	0.6287	0.5135	0.4403	0.4333	0.4743	0.4278	0.5020
EGAE-JOCAS	0.5407	0.5477	0.5011	0.4513	0.4265	0.4513	0.3899	0.4722
PCAN	0.5435	0.4785	0.4156	0.4392	0.4076	0.4236	0.3899	0.4426
PCMAN	0.5781	0.5928	0.5591	0.5043	0.4806	0.5335	0.4167	0.5236

Table 8

Clustering performance on the Shenzhen dataset (NMI). Bold is used to highlight the best result.

K	2	3	4	5	6	7	8	avg.
ARGE	0.1184	0.1165	0.1176	0.1136	0.1222	0.1509	0.1648	0.1291
ARVGE	0.1434	0.1132	0.1298	0.1149	0.1754	0.1628	0.1825	0.1460
GK-means	0.1197	0.1089	0.1151	0.1913	0.1905	0.2280	0.1878	0.1630
EGAE-JOCAS	0.1210	0.1452	0.1524	0.1503	0.1562	0.1624	0.1555	0.1490
PCAN	0.1163	0.1331	0.1328	0.1631	0.1178	0.1370	0.1371	0.1339
PCMAN	0.1260	0.1342	0.1855	0.2266	0.2014	0.1833	0.1815	0.1769

Table 9

Clustering performance on the Shenzhen dataset (F1). Bold is used to highlight the best result.

K	2	3	4	5	6	7	8	avg.
ARGE	0.4189	0.3835	0.2796	0.2387	0.2816	0.2465	0.2251	0.2963
ARVGE	0.4146	0.4060	0.3376	0.3155	0.3265	0.2653	0.2836	0.3356
GK-means	0.4463	0.4641	0.4198	0.3786	0.3674	0.2810	0.2878	0.3779
EGAE-JOCAS	0.4101	0.4477	0.4385	0.3523	0.3377	0.2764	0.3226	0.3693
PCAN	0.3923	0.4098	0.3909	0.3312	0.3266	0.2881	0.2749	0.3448
PCMAN	0.4117	0.4558	0.4752	0.3903	0.3686	0.3431	0.3043	0.3927

- The superiority of PCMAN in most cases is more pronounced when the cluster number K is four or five. One possible reason is that the traffic states are categorized into four classes by the traffic index, that is, the smooth state, lightly congested state, moderately congested state, and heavily congested state, and the cluster number K conforms to the distribution of traffic flow results in different traffic states.

In terms of the advantages of the compared methods, ARGE and ARVGE leverage the topological structure and node features to construct a graph convolutional network, and ARVGE showed better performance than ARGE. This is because ARVGE employs a variational graph autoencoder in the upper tier. EGAE-JOCAS combines relaxed k -means and spectral clustering to generate preferable embeddings and shares the same adjacency within the graph convolution layers; thus it has an improved performance compared with ARVGE. GK-means employs the idea of dynamic game theory to optimize the local community structure and is derived from a continuously nonnegative convex function to explore

the evolution of the opinion matrix. GK-means has a stronger ability for adversarial learning; therefore, it performs better than the aforementioned three methods.

All the methods adopt the topological structure and node features to improve clustering performance, in which PCMAN is higher than the other five methods concerning the following metrics:

- NMF projects large collections of multi-dimensional traffic data into low-dimensional representations; thus, PCMAN integrates the structure and attribute information of road networks to obtain a better joint relationship through NMF. Thus, NMF can be applied to traffic pattern clustering when traffic data need to exhibit high degrees of tightness between structure and attribute information.
- The PCMAN utilizes motifs to capture global information and extract higher-order spatial correlations of node features in road networks. The PCMAN significantly outperformed the

Table 10
Example of clustered results generated from PCMAN.

Segment ID	Time	Clustered result	Ground
2271	2018-09-01	2	2
2271	2018-09-02	1	1
2271	2018-09-03	2	1
2271	2018-09-04	2	2
...
2271	2018-09-08	1	1
2271	2018-09-09	1	2
2271	2018-09-10	3	3

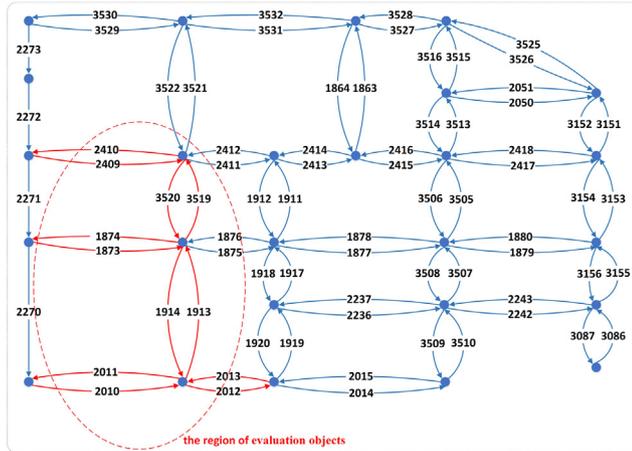


Fig. 7. Topology structure of the Chengdu road network.

compared PCAN method. This indicates that the application of motifs in road networks is feasible and effective.

- By converting the feature matrix into a feature similarity matrix, PCMAN alleviates the heterogeneity between the structure and attribute information of road networks, minimizes the difference between the clustered results and the ground truth, and thereby improves the clustering performance.

The clustered results generated from PCMAN are partially shown in Table 10.

4.4. Analysis

Based on the experimental results, an in-depth traffic pattern analysis of four periods was conducted in the Chengdu road network.

The Chengdu road network is located at 30.66°N ~ 30.73°N, 104.02°E ~ 104.10°E. Fig. 7 represents the network topology, and the region marked by a solid red line with a single arrow is one of the congested areas. To analyze the traffic patterns more meaningfully, 12 segments in the region were used as evaluation objects.

Traffic flow patterns reveal a significant similarity between days with the same date attributes [58]; four periods of pattern clustering in different segments of the road networks are adopted: the first and last weekdays, other weekdays, weekends, and holidays. According to the four periods, the common basis matrix B is categorized into four classes and then re-outputs the clustered results.

Under normal circumstances, the cluster number of traffic states K is optimally determined by the silhouette measure [59], and clustered results are obtained according to the K value.

The average of the silhouette measures of road network nodes \bar{S}_K is expressed as follows:

$$\bar{S}_K = \frac{1}{m} \sum_{a=1}^m Sil(a), \quad (20)$$

where $Sil(a)$ is the silhouette measure of sample A_a and m is the number of road network nodes.

According to the silhouette measure, Fig. 8 indicates that: whether it is morning rush hours or evening rush hours, the Sil of the first and last weekdays reaches the highest value when K is five, and the Sil of the other three periods reaches the highest value when K is four. A possible explanation for this may be that there was a heavy flow of commuters on weekdays, whereas people traveled relatively less on non-weekdays and traffic congestion weakened. In terms of weekdays, traffic jams have become a universal phenomenon during rush hours, and the first and last weekdays will be less deterministic than other weekdays.

According to the best K value in Fig. 8, the pattern similarity is discussed for four periods. In Figs. 9–10, each cluster is associated with a pattern for the probability of traffic states at each road segment. In other words, in four periods, the traffic pattern for each road segment comprises several probabilities of traffic states during morning or evening rush hours.

As it can be observed from Figs. 9–10, the pattern similarity of different segments exists in each period. For example, the patterns of segments 2011 and 1874 in morning rush hours on the first and last weekdays, and the patterns of segments 2013 and 1873 in evening rush hours on holidays are similar. Furthermore, the patterns of segments 3520 and 1873 in the cases of morning rush hours on the first and last weekdays, morning rush hours on weekends, and evening rush hours on other weekdays are similar. When observing closely at the locations of segments 3520 and 1873 (Section of North Station West 2nd Road and North Section of the 1st Ring Road), it is found that they are located close to the Chengdu Railway Middle School. Accordingly, the pattern similarity between these two segments may be linked to the school timetable.

When the patterns of these segments are similar in the morning or evening rush hours of the same period, traffic management agencies will set up the corresponding control strategies to regulate the traffic flow of these similar segments, and prevent the anticipated traffic congestion.

The analysis can provide an improved scheme for urban commuters to understand traffic patterns in road networks, and clarify where imbalances would occur during morning or evening rush hours in four periods.

5. Conclusion

In this paper, a traffic pattern clustering method (PCMAN) is proposed that utilizes motif-based attributed road networks. It incorporates NMF and motifs to address (1) the issue of low accuracy owing to the heterogeneity of the structure and attribute information and (2) the issue of lower-order connectivity of road network structures. Specifically, a higher-order connectivity model is designed to capture the global information of road networks using motif search and IAMS representation. PCMAN then factorizes the motif-based weighted matrix and node feature similarity matrix by considering the Karush–Kuhn–Tucker condition optimization and utilizes the k -means algorithm to obtain clustered results. Experiments were performed on two real-world datasets considering four periods, with morning rush hours from 7:00 to 9:00 am and evening rush hours from 5:00 to 7:00 pm. The results indicated that the performance of the PCMAN is superior to that of other state-of-the-art methods, and the superiority

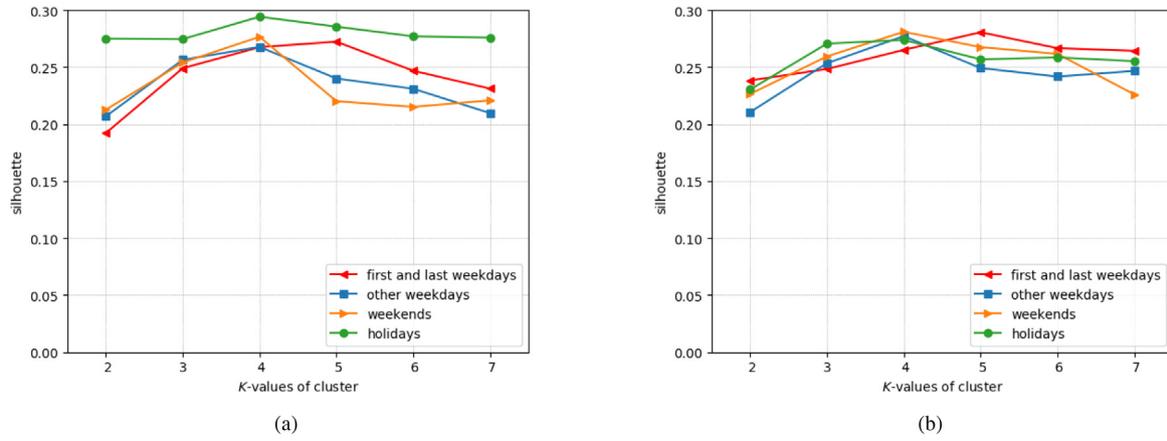


Fig. 8. Silhouette of four periods in morning rush hours (a) and evening rush hours (b).

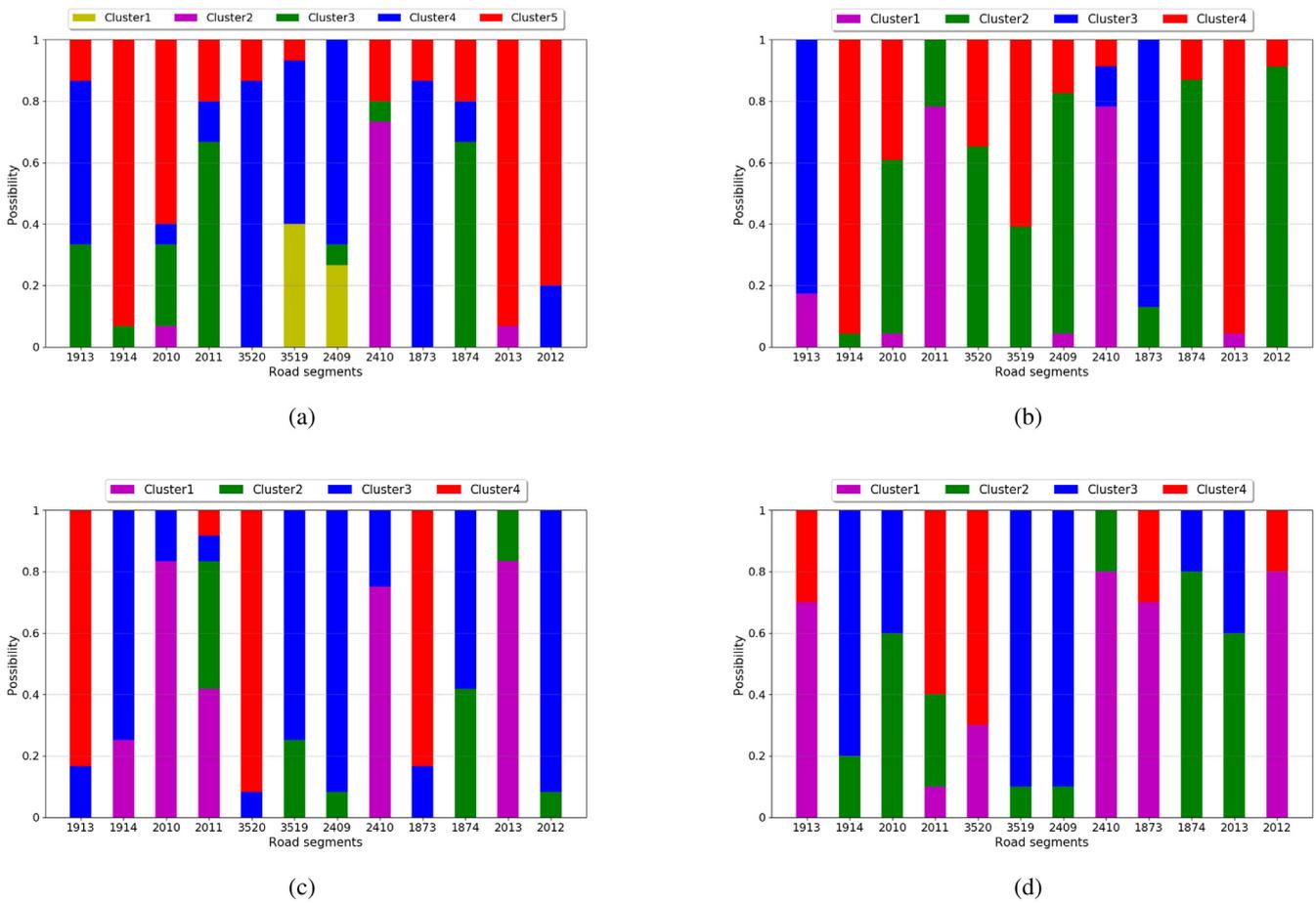


Fig. 9. Pattern of the first and last weekdays (a), other weekdays (b), weekends (c), and holidays (d) in morning rush hours.

of the PCMAN in most cases is more significant when the cluster number K is four or five. Moreover, in the Chengdu road network, the PCMAN is capable of providing insights into the pattern similarity of road segments; for example, the pattern similarity between segments 3520 and 1873 might be linked to the school timetable. The analysis helps urban commuters to better manage transportation modes and alleviate traffic congestion.

The proposed PCMAN method quantitatively evaluated the traffic characteristics and results limited by empirical information sources. Moreover, due to the complexity of heterogeneous information associated with multi-typed components, it is lack

of analyzing the heterogeneity of the structure and attribute information using only traffic index. Future research should focus on overcoming the difficulties of evaluation. It is necessary to use a unified standard to improve the effect of the PCMAN and compare it with other methods. Besides, the attributed road network with more types of node features (traffic flow, average speed, etc.) should be discussed especially those contains continuous attributes. In addition, the application of attributed networks deepens in the proposed method, and it will handle other tasks in ITS, such as traffic prediction and causal discovery of the congestion propagation patterns.

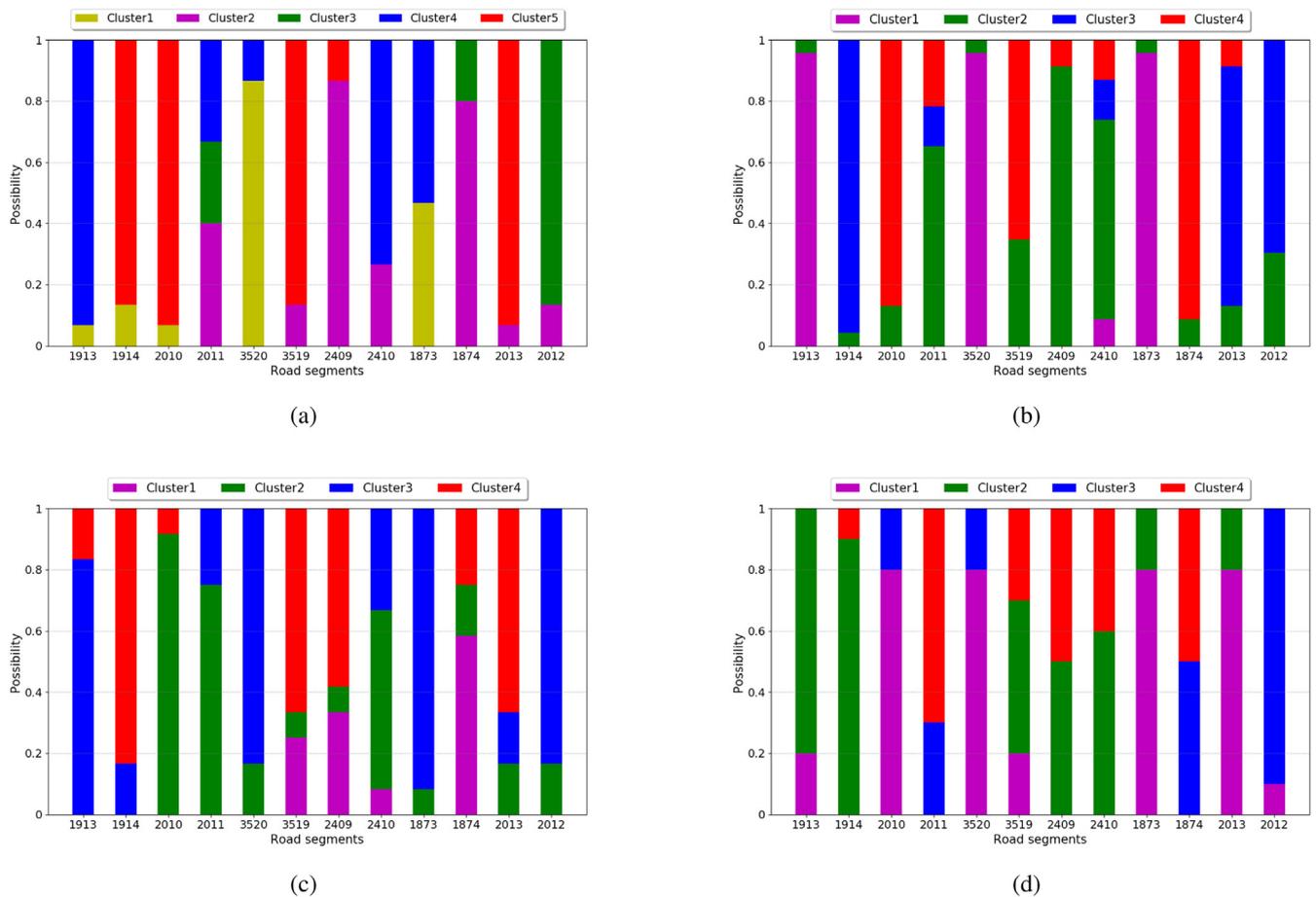


Fig. 10. Pattern of the first and last weekdays (a), other weekdays (b), weekends (c), and holidays (d) in evening rush hours.

CRediT authorship contribution statement

Guojiang Shen: Conceptualization, Methodology, Investigation. **Difeng Zhu:** Methodology, Writing – original draft, Validation, Visualization. **Jingjing Chen:** Writing – review & editing, Validation. **Xiangjie Kong:** Supervision, Project administration, Funding acquisition.

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.

Acknowledgments

This work was supported in part by the “Pioneer” and “Leading Goose” R&D Program of Zhejiang under Grant 2022C01050, in part by the National Natural Science Foundation of China under Grant 62073295 and Grant 62072409, in part by the Zhejiang Provincial Natural Science Foundation under Grant LR21F020003, and in part by the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant RF-B2020001.

References

- [1] L. Zhu, F.R. Yu, Y. Wang, B. Ning, T. Tang, Big data analytics in intelligent transportation systems: A survey, *IEEE Trans. Intell. Trans. Syst.* 20 (1) (2019) 383–398.
- [2] W. Wang, F. Xia, H. Nie, Z. Chen, Z. Gong, X. Kong, W. Wei, Vehicle trajectory clustering based on dynamic representation learning of Internet of vehicles, *IEEE Trans. Intell. Transp. Syst.* 22 (6) (2021) 3567–3576.
- [3] F. Xia, J. Wang, X. Kong, D. Zhang, Z. Wang, Ranking station importance with human mobility patterns using subway network datasets, *IEEE Trans. Intell. Transp. Syst.* 21 (7) (2020) 2840–2852.
- [4] M.H. Almannaa, M. Elhenawy, H.A. Rakha, A novel supervised clustering algorithm for transportation system applications, *IEEE Trans. Intell. Transp. Syst.* 21 (1) (2020) 222–232.
- [5] D. Zhu, G. Shen, D. Liu, J. Chen, Y. Zhang, FCG-ASpredictor: An approach for the prediction of average speed of road segments with floating car GPS data, *Sensors* 19 (22) (2019) 4967.
- [6] T. Nguyen, P. Krishnakumari, S.C. Calvert, H.L. Vu, H. Van Lint, Feature extraction and clustering analysis of highway congestion, *Transp. Res. C* 100 (2019) 238–258.
- [7] X. Kong, K. Wang, M. Hou, X. Hao, F. Xia, A federated learning-based license plate recognition scheme for 5G-enabled internet of vehicles, *IEEE Trans. Ind. Inf.* 17 (12) (2021) 8523–8530.
- [8] X. Kong, K. Wang, S. Wang, X. Wang, X. Jiang, Y. Guo, G. Shen, X. Chen, Q. Ni, Real-time mask identification for COVID-19: An edge computing-based deep learning framework, *IEEE Internet Things J.* 8 (21) (2021) 15929–15938.
- [9] J. Huang, M. Xiao, State of the art on road traffic sensing and learning based on mobile user network log data, *Neurocomputing* 278 (2018) 110–118.
- [10] S. Cheng, F. Lu, P. Peng, Short-term traffic forecasting by mining the non-stationarity of spatiotemporal patterns, *IEEE Trans. Intell. Transp. Syst.* 22 (10) (2021) 6365–6383.
- [11] R. García-Ródenas, M. López-García, M. Sánchez-Rico, An approach to dynamical classification of daily traffic patterns, *Comput.-Aided Civ. Infrastruct. Eng.* 32 (2017) 191–212.
- [12] J. Yang, Y. Sun, B. Shang, L. Wang, J. Zhu, Understanding collective human mobility spatiotemporal patterns on weekdays from taxi origin–destination point data, *Sensors* 19 (12) (2019) 2812.
- [13] Y. Qiao, Z. Si, Y. Zhang, F.B. Abdesslem, Y. Jie, A hybrid Markov-based model for human mobility prediction, *Neurocomputing* 278 (2018) 99–109.
- [14] Y. Huang, Z. Xiao, D. Wang, H. Jiang, D. Wu, Exploring individual travel patterns across private car trajectory data, *IEEE Trans. Intell. Transp. Syst.* 21 (12) (2020) 5036–5050.

- [15] J. Tian, M. Yu, C. Ren, Y. Lei, Network-scape metric analysis: a new approach for the pattern analysis of urban road networks, *Int. J. Geogr. Inf. Sci.* 33 (3) (2019) 537–566.
- [16] E.F. Bernal, A.M. del Rey, P.G. Villardon, Analysis of madrid metro network: From structural to HJ-biplot perspective, *Appl. Sci.-Basel* 10 (16) (2020) 5689.
- [17] Y. Yang, Z. He, Z. Song, X. Fu, J. Wang, Investigation on structural and spatial characteristics of taxi trip trajectory network in Xi'an, China, *Physica A* 506 (2018) 755–766.
- [18] X. Jiang, X. Wen, M. Wu, M. Song, C. Tu, A complex network analysis approach for identifying air traffic congestion based on independent component analysis, *Physica A* 523 (2019) 364–381.
- [19] I.U. Hewapathirana, Change detection in dynamic attributed networks, *Wiley Interdiscip. Rev.-Data Min. Knowl. Dis.* 9 (3) (2019) e1286.
- [20] S. Zhou, J. Bu, Z. Zhang, C. Wang, L. Ma, J. Zhang, Cross multi-type objects clustering in attributed heterogeneous information network, *Knowl.-Based Syst.* 194 (2020) 105458.
- [21] L. Wu, D. Wang, K. Song, S. Feng, Y. Zhang, G. Yu, Dual-view hypergraph neural networks for attributed graph learning, *Knowl.-Based Syst.* 227 (2021) 107185.
- [22] X. Zhang, H. Liu, Q. Li, X. Wu, Attributed graph clustering via adaptive graph convolution, in: *Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI, 2019*, pp. 4327–4333.
- [23] X. Kong, S. Tong, H. Gao, G. Shen, K. Wang, M. Collotta, I. You, S. Das, Mobile edge cooperation optimization for wearable Internet of Things: A network representation-based framework, *IEEE Trans. Ind. Inf.* 17 (7) (2021) 5050–5058.
- [24] Z. Huang, X. Zhong, Q. Wang, M. Gong, X. Ma, Detecting community in attributed networks by dynamically exploring node attributes and topological structure, *Knowl.-Based Syst.* 196 (2020) 105760.
- [25] L. Guo, X. Shi, J. Cao, Turing patterns of gierer-meinhardt model on complex networks, *Nonlinear Dynam.* 105 (1) (2021) 899–909.
- [26] F. Doria, R. Erichsen, M. Gonzalez, F.B. Rodriguez, A. Sanchez, D. Dominguez, Structured patterns retrieval using a metric attractor network: Application to fingerprint recognition, *Physica A* 457 (2016) 424–436.
- [27] D. Zhu, G. Shen, J. Chen, W. Zhou, X. Kong, A higher-order motif based spatio-temporal graph imputation approach for transportation networks, *Wirel. Commun. Mob. Comput.* 2022 (2022) 1702170.
- [28] T. Wang, J. Peng, Q. Peng, Y. Wang, J. Chen, FSM: Fast and scalable network motif discovery for exploring higher-order network organizations, *Methods* 173 (2020) 83–93.
- [29] X. Li, H. Wu, Toward graph classification on structure property using adaptive motif based on graph convolutional network, *J. Supercomput.* 77 (8) (2021) 8767–8786.
- [30] T. Anwar, C. Liu, H.L. Vu, M.S. Islam, T. Sellis, Capturing the spatiotemporal evolution in road traffic networks, *IEEE Trans. Knowl. Data Eng.* 30 (8) (2018) 1426–1439.
- [31] C.T. Calafate, D. Soler, J.C. Cano, P. Manzoni, Traffic management as a service: The traffic flow pattern classification problem, *Math. Probl. Eng.* 2015 (2015) 716598.
- [32] H.N. Nguyen, P. Krishnakumari, H.L. Vu, H. Van Lint, Traffic congestion pattern classification using multi-class SVM, in: *Proceedings of IEEE 19th International Conference on Intelligent Transportation Systems, ITSC, 2016*, pp. 1059–1064.
- [33] Y.J. Byon, J.S. Ha, C.S. Cho, T.Y. Kim, C.Y. Yeun, Real-time transportation mode identification using artificial neural networks enhanced with mode availability layers: A case study in Dubai, *Appl. Sci.* 7 (9) (2017) 923.
- [34] F. Zhang, J. Chen, Y. Zhu, J. Li, Q. Li, Z. Zhuang, A dual hesitant fuzzy rough pattern recognition approach based on deviation theories and its application in urban traffic modes recognition, *Symmetry* 9 (11) (2017) 262.
- [35] X. Han, G. Shen, X. Yang, X. Kong, Congestion recognition for hybrid urban road systems via digraph convolutional network, *Transp. Res. C* 121 (2020) 102877.
- [36] G. Shen, Z. Zhao, X. Kong, GCN2CDD: A commercial district discovery framework via embedding space clustering on graph convolution networks, *IEEE Trans. Ind. Inf.* 18 (1) (2022) 356–364.
- [37] T. Zhang, J. Wang, C. Cui, Y. Li, W. He, Y. Lu, Q. Qiao, Integrating geovisual analytics with machine learning for human mobility pattern discovery, *ISPRS Int. J. Geo-Inf.* 8 (10) (2019) 434.
- [38] R. Chandra, T. Guan, S. Panuganti, T. Mittal, U. Bhattacharya, A. Bera, D. Manocha, Forecasting trajectory and behavior of road-agents using spectral clustering in graph-lstms, *IEEE Robot. Autom. Lett.* 5 (3) (2020) 4882–4890.
- [39] J. Ou, J. Xia, Y. Wang, C. Wang, Z. Lu, A data-driven approach to determining freeway incident impact areas with fuzzy and graph theory-based clustering, *Comput.-Aided Civ. Infrastruct. Eng.* 35 (2) (2020) 178–199.
- [40] C. Wang, S. Pan, G. Long, X. Zhu, J. Jiang, Mgae: Marginalized graph autoencoder for graph clustering, in: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM, 2017*, pp. 889–898.
- [41] S. Pan, R. Hu, G. Long, J. Jiang, L. Yao, C. Zhang, Adversarially regularized graph autoencoder for graph embedding, in: *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI, 2018*, pp. 2609–2615.
- [42] Z. Bu, H.J. Li, C. Zhang, J. Cao, A. Li, Y. Shi, Graph K-means based on leader identification, dynamic game, and opinion dynamics, *IEEE Trans. Knowl. Data Eng.* 32 (7) (2020) 1348–1361.
- [43] X. Li, H. Zhang, R. Zhang, Embedding graph auto-encoder with joint clustering via adjacency sharing, 2020, arXiv preprint arXiv:2002.08643.
- [44] L. Chen, Z. Wu, J. Cao, G. Zhu, Y. Ge, Travel recommendation via fusing multi-auxiliary information into matrix factorization, *ACM Trans. Intell. Syst. Technol. (TIST)* 11 (2) (2020) 22.
- [45] X. Wang, P. Cui, J. Wang, J. Pei, W. Zhu, S. Yang, Community preserving network embedding, in: *Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI, 2017*, pp. 203–209.
- [46] X. Ma, D. Dong, Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks, *IEEE Trans. Knowl. Data Eng.* 29 (5) (2017) 1045–1058.
- [47] Z. Meng, H. Shen, H. Huang, W. Liu, J. Wang, A.K. Sangaiah, Search result diversification on attributed networks via nonnegative matrix factorization, *Inf. Process. Manage.* 54 (6) (2018) 1277–1291.
- [48] T. Guo, S. Pan, X. Zhu, C. Zhang, CFOND: Consensus factorization for co-clustering networked data, *IEEE Trans. Knowl. Data Eng.* 31 (4) (2019) 706–719.
- [49] D. Zhang, M.R. Kabuka, Protein family classification with multi-layer graph convolutional networks, in: *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, BIBM, 2018*, pp. 2390–2393.
- [50] A.R. Benson, D.F. Gleich, J. Leskovec, Higher-order organization of complex networks, *Science* 353 (6295) (2016) 163–166.
- [51] H. Yin, A.R. Benson, J. Leskovec, D.F. Gleich, Local higher-order graph clustering, in: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017*, pp. 555–564.
- [52] P. Li, G.J. Puleo, O. Milenkovic, Motif and hypergraph correlation clustering, *IEEE Trans. Inform. Theory* 66 (5) (2020) 3065–3078.
- [53] W.G. Underwood, A. Elliott, M. Cucuringu, Motif-based spectral clustering of weighted directed networks, 2020, arXiv preprint arXiv:2004.01293.
- [54] G. Mei, J. Tu, L. Xiao, F. Piccialli, KCoreMotif: An efficient graph clustering algorithm for large networks by exploiting k-core decomposition and motifs, 2020, arXiv preprint arXiv:2008.10380.
- [55] C. Sun, Y. Yang, H. Wang, W. Wang, A clustering approach for motif discovery in CHIP-Seq dataset, *Entropy* 21 (8) (2019) 802.
- [56] N.A. Funde, M.M. Dhabu, A. Paramasivam, P.S. Deshpande, Motif-based association rule mining and clustering technique for determining energy usage patterns for smart meter data, *Sustainable Cities Soc.* 46 (2019) 101415.
- [57] Y. Wang, J. Cao, H. Tao, Graph convolutional network with multi-similarity attribute matrices fusion for node classification, *Neural Comput. Appl.* (2021).
- [58] Y. Ren, Z. Hou, I.I. Sirmatel, N. Geroliminis, Data driven model free adaptive iterative learning perimeter control for large-scale urban road networks, *Transp. Res. C* 115 (2020) 102618.
- [59] Q.T. Bui, B. Vo, H.A.N. Do, N.Q.V. Hung, V. Snaesl, F-Mapper: A fuzzy mapper clustering algorithm, *Knowl.-Based Syst.* 189 (2020) 105107.