# Modeling Dynamic Traffic Flow as Visibility Graphs: A Network-Scale Prediction Framework for Lane-Level Traffic Flow Based on LPR Data

Jie Zeng and Jinjun Tang

***Abstract*— Emerging applications in real-time traffic management put forward urgent requirements for lane-level traffic flow prediction. Limited by extremely unstable traffic volumes and heterogeneous spatiotemporal dependencies in urban road networks, network-scale prediction for lane-level traffic flow is still a critical challenge. This study models the dynamic characteristics of lane-level traffic flow as complex networks and proposes a deep learning framework for network-scale prediction. Relying on the visibility graph, we transform the temporal dependence learning task into spatial correlation mining on temporal complex networks. For spatial dependency extraction in urban traffic flows, we establish three topological graphs from traffic, statistical, and semantic perspectives to investigate the static and dynamic correlations. Then, a network-scale traffic volumes prediction model, i.e., spatiotemporal multigraph gated network (STMGG), is proposed to learn spatiotemporal correlations on visibility graphs and spatial topological graphs. This model designs an attention-based gated mechanism to incorporate global features from multigraphs. Additionally, a Seq2Seq structure is integrated to enhance multistep prediction stability. We employ two license plate recognition (LPR) datasets as case studies, and STMGG expresses superiorities over various advanced deep learning models. Meanwhile, an ablation experiment is conducted to evaluate its components, and numerical tests further reveal its impressive inductive learning capability.**

***Index Terms*— Lane-level traffic flow prediction, visibility graph, graph neural network, urban road network, license plate recognition data.**

## I. Introduction

IN RECENT years, the rapid increment of vehicles speeds up the imbalance between road service and travel demand. A series of traffic problems then follow and seriously affect the life satisfaction of urban residents. Nowadays, researchers widely pay attention to intelligent transportation systems (ITS) to counter these challenges. As a pivotal application of ITS [1], short-term traffic flow prediction plays a vital role in dynamic route planning [2], signal optimization [3], and real-time traffic management [4]. Thus, it constantly becomes one of the most crucial topics in traffic/transportation engineering.

From statistical methods to novel deep learning models, this field has developed for more than 40 years, and it has attracted the continued attention of several generations of researchers and engineers. However, limitations and barriers still remain, especially in urban traffic flow prediction. Unlike expressway or freeway road networks, the most typical characteristic of urban roads is signal control at intersections. In this condition, urban traffic flow is constantly interrupted by the widely installed signal control equipment and always expresses extreme fluctuations. Meanwhile, since urban road networks are denser and highly connected, heterogeneous spatiotemporal dependencies exist and increase the difficulty of exploring.

Additionally, emerging technologies in ITS have induced numerous refined applications in traffic systems, such as lane-level traffic management and control. Due to the foundation role in real-time traffic management, the promotion of these emerging scenes requires higher performance in lane-level traffic flow prediction. Meanwhile, previous studies also indicated that lane-level traffic prediction is essential for the control strategy formulation and route guidance of connected automated vehicles (CAVs) [5], [6]. Therefore, the importance of conducting lane-level traffic flow prediction is continually increasing. However, it also faces significant challenges in meeting the accuracy and efficiency requirements of these emerging scenes.

Overall, current barriers in urban traffic flow prediction can be summarized as follows.

- Although several efforts have been applied in lane-level traffic prediction [5], [6], [7], these methods mainly focus on forecasting single or multiple lanes. However, massive lanes with comprehensive correlations exist in urban road networks. Limited by the studied area, these current studies are unsuitable and less practical for real-time applications. Therefore, network-scale prediction of lane-level traffic flow is still a challenge.
- The widely installed traffic detectors are the solid foundation of ITS [8]. Due to the installation and maintenance costs, traffic detectors are always sparsely equipped in urban road networks, resulting in challenges for spatiotemporal dependencies modeling and analysis. Recently, the graph neural network (GNN) has been regarded as a powerful tool in traffic prediction, but

Manuscript received 19 March 2022; revised 25 October 2022; accepted 12 December 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 52172310, in part by the Humanities and Social Sciences Foundation of the Ministry of Education under Grant 21YJCZH147, and in part by the Innovation-Driven Project of Central South University under Grant 2020CX041. The Associate Editor for this article was Y. Lv. *(Corresponding author: Jinjun Tang.)*

The authors are with the Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China (e-mail: zj991130@csu.edu.cn; jinjuntang@csu.edu.cn).

Digital Object Identifier 10.1109/TITS.2022.3231959

it needs a reasonable topological graph in advance. However, the sparse distribution characteristics of fixed detectors make it challenging to transfer detector networks into topological graphs. Additionally, due to complex adjacent relationships and parallel lanes, traditional topological graph construction methods [9] cannot meet the requirements of urban road networks. Hence, how to build reasonable topological graphs to extract spatial dependence in urban road networks still needs further exploration.

- Affected by the signal control at urban intersections, traffic flow is periodically interrupted and expresses strong instability at the short-term scale, especially for lane-level traffic volumes. Existing prediction methods mainly employ deep learning models to learn the temporal dynamics of traffic data, such as the Seq2Seq structure [10], attention mechanism [11], etc. However, time-varying characteristics of the fluctuation are always overlooked. Essentially, it is helpful to identify and investigate the temporal dynamics of traffic flows by modeling their time-varying patterns. Therefore, integrating these characteristics with prediction models might be a potential way to improve prediction performance.

To deal with existing barriers, we propose a spatiotemporal multigraph gated network (STMGG) for lane-level traffic volume prediction from the network scale. In STMGG, both the spatial and temporal dependencies of lane-level traffic volumes are modeled as topological graphs for feature extraction. Specifically, this study first employs the visibility graph [12] to represent temporal dynamics of previous traffic volumes of each lane. Here, nodes in visibility graphs denote traffic volumes at each time step, and GraphSAGE [13] is utilized to capture the temporal evolution patterns. After that, we establish three spatial topological graphs, including two static networks and a dynamic network, from the perspectives of traffic, statistical, and semantic correlations, respectively. In these spatial topological graphs, the average travel time (ATT) and dynamic time warping (DTW) distance [14] are employed for static graph construction, named ATT graph and DTW graph. Meanwhile, this study designs a dynamic adjacency matrix learning module to extract dynamic semantic correlations in the urban road network, named learnable matrix-based graph (LM graph). The proposed STMGG model is composed of STMGG units, which are used to mine the spatial dependence on these three graphs. In each STMGG unit, we employ GraphSAGE as the core aggregator and propose an attention-based gated mechanism to incorporate the global context information from multigraphs. Finally, STMGG units are combined with the Seq2Seq structure to predict future traffic volumes sequentially. The main contributions of this study can be summarized as follows.

- We model the temporal evolution of lane-level traffic volumes as visibility graphs. In this way, the temporal characteristics learning task is transformed into spatial correlation extraction on each temporal complex network. In the constructed visibility graph, *visibility* relationships among nodes are employed to describe the causality relationship and fluctuation phenomenon of traffic volumes.
- We investigate spatial correlations of network-scale traffic flows by integrating static and dynamic spatial dependence. Three topological graphs from different perspectives are established to explore heterogeneous spatial dependence in the urban road network, i.e., traffic engineering, statistical relationship, and dynamic semantic correlation.
- We propose the STMGG model for lane-level traffic flow prediction from the network scale. This Seq2Seq model learns spatiotemporal dependencies in urban road networks from temporal and spatial topological graphs. An attention-based gated mechanism is developed for multigraphs information fusion based on the effective incorporation of multi-head attention and split-attention mechanism.
- Validated on two urban road networks in the Changsha LPR system, China, the proposed STMGG model expresses a superior performance over the advanced baselines. Experiment results further reveal the importance of each component and its inductive learning capability.

The rest of this paper is organized as follows. Existing studies of short-term traffic prediction are briefly reviewed in Section II. We detailly describe the proposed spatiotemporal dependencies modeling methods and STMGG model in Section III. In Section IV, we conduct extensive comparisons for lane-level traffic flow prediction and demonstrate the inductive learning capability of our model. Finally, we summarize this study and point out potential directions for future works in Section V.

## II. Literature Review

In these decades, researchers have proposed numerous methods to improve prediction accuracies, including model-driven and data-driven methods. Overall, the former attempts to construct simulation models to model the evolution process of traffic systems and make predictions [15], while the latter aims to learn the time-varying regularities from real-world traffic data. Due to the rapid increment in available traffic data [16], data-driven methods always express superiority over model-driven [17]. According to prediction modes, existing data-driven prediction methods can be further classified into two categories, i.e., single-point prediction models and network-scale prediction models.

### *A. Single-Point Prediction Models*

In the literature review of this study, the single-point prediction task is defined as: given the previous traffic flow of a single sensor (or road segment), aiming to predict its future states. It is noted that several studies fused previous information of the target sensor with its correlated sensors to predict future traffic states [7], [18]. Since the spatial range of these studies is limited, we also regard them as single-point prediction methods.

In the early stage of this field, traffic engineers and researchers mainly regarded traffic flow as time-series data. Therefore, a lot of statistical methods are applied in short-term traffic flow prediction, including the autoregressive integrated moving average (ARIMA) model [19], [20], Kalman

filter [21], [22], partial least square (PLS) [23], generalized autoregressive conditional heteroscedasticity (GARCH) [24], etc. These statistical methods are convenient to apply and have strong interpretability, but the uncertainty and nonlinear characteristics of traffic systems significantly impact their prediction performance [15]. Meanwhile, in the multistep prediction task, the predicted traffic states of each time step depend on previous time steps, so the prediction errors will accumulate [17].

To overcome these issues, researchers introduced numerous machine learning models in this field, e.g., artificial neural network (ANN) [25], [26], support vector machine (SVM) [27], [28], and k-nearest neighbors (KNN) [29]. Moreover, to further improve prediction accuracy, deep learning models are also favored by scholars. Since short-term traffic flow can be regarded as sequence data, recurrent neural network (RNN) and its variants, i.e., long short-term memory (LSTM) and gated recurrent unit (GRU), are widely used in traffic prediction [30], [31].

Because these single-point prediction methods ignore or cannot fully extract the spatial dependence in the urban road network, their prediction performance is always limited. Meanwhile, applying these models to forecast future traffic states of network-wide traffic flow needs to train and predict each point one by one, leading to a linear increment of computational cost. Therefore, in recent years, network-scale prediction models have become the trend in this field.

### *B. Network-Scale Prediction Models*

Network-scale traffic flow prediction models aim to predict future traffic states of all the nodes simultaneously. Since these prediction methods can fully explore the spatiotemporal dependencies, they always show superiority in both accuracy and efficiency. According to the data structure, existing network-scale prediction methods can be further divided into two categories [32]: grid-based prediction methods and graph-based prediction methods.

Due to its powerful spatial correlation mining capability, convolutional neural networks (CNN) are widely employed in network-scale traffic prediction. The premise of using CNN models in this field is that it is necessary to transform the road network into grid structure data. Thus, many researchers chose to learn urban traffic systems as images [33], [34], [35], where each pixel in the images denotes a region. In this way, the spatial dependence among pixels can be extracted from a series of convolutional operations. To further explore the spatiotemporal dependencies in the grid-based traffic data, Ke et al. [36] integrated the convolutional LSTM (ConvLSTM), LSTM, and CNN for network-scale prediction. Different from these works, Dai et al. [37] defined an arrangement method to transform the highway sensor network into a 2D matrix and designed a CNN-based model for short-term traffic prediction. However, these grid-based models can only be applied to the Euclidean space [15], [38], so the topology structure of road networks is always ignored.

To fill this gap, researchers gradually paid attention to the graph-based models to improve prediction accuracies, including GCN [15], [17], [39], GAT [40], [41], GraphSAGE [42], etc. Among these methods, GCN-based models are the most widely used in this field. Li et al. [9] employed the random walk process to model the spatial dependence in the freeway network and proposed a diffusion convolutional recurrent neural network (DCRNN) for network-scale prediction. Cui et al. [43] proposed a traffic graph convolutional long short-term memory network (TGC-LSTM), in which LSTM and spectral GCN were employed to learn the temporal and spatial dependencies, respectively. After that, they further integrated the graph wavelet operation with the gated recurrent structure for network-scale traffic flow prediction [2]. Lee & Rhee [44] incorporated the spatial correlations from the distance, direction, and positional relationships and then constructed a GCN-based model for prediction. Zhang et al. [8] proposed a graph-based temporal attention model (GTA) to learn the spatiotemporal dependencies in multi-sensor systems.

However, these works mainly employ the physical road network as the topological graph for GNNs and assume spatial dependence never changes. Compared with the freeway, more complex spatial correlations exist in urban road networks due to the higher road density. Therefore, the physical road network cannot fully represent the comprehensive spatial dependence [45]. Meanwhile, since traffic patterns will vary with different time spans [46], the physical road network is challenging to capture the dynamic of the traffic system. Thus, researchers also tried to learn the adaptive topological graphs of road networks and used them for traffic prediction [47], [48].

## III. Methodology

### *A. Problem Formulation*

In general, this study aims to employ the network-scale traffic volumes of previous time steps to sequentially predict their future values at the following time steps. Thus, for a road network with N lanes, the prediction task in this study can be summarized in Eq. (1). Here, $\boldsymbol{x}^t \in \mathbb{R}^{\mathrm{N}}$ denotes the network-scale traffic volumes at step $t$, where N denotes the number of lanes in the road network. Meanwhile, T stands the length of input and output sequences. $\mathcal{F}(\cdot)$ is the mapping function, aiming to forecast future traffic volumes based on previous information. $\hat{\boldsymbol{x}}^{t+i} \in \mathbb{R}^{\mathrm{N}}$ represents the predicted traffic volumes at step $t+i$. Additionally, $\mathbb{G} = \{\mathcal{G}_{\mathrm{ATT}}, \mathcal{G}_{\mathrm{DTW}}\}$ denotes the static topological graph list, where these two graphs denote the ATT graph and DTW graph, respectively.

$$\left\{\hat{\boldsymbol{x}}^{t+1}, \ldots, \hat{\boldsymbol{x}}^{t+\mathrm{T}}\right\} = \mathcal{F}\left(\left\{\boldsymbol{x}^{t-\mathrm{T}+1}, \ldots, \boldsymbol{x}^{t}\right\}; \mathbb{G}\right) \tag{1}$$

### *B. Framework*

This study proposes a spatiotemporal multigraph gated network (STMGG) model for lane-level traffic flow prediction. Fig. 1 shows the framework of the proposed model. According to this figure, we first model the previous traffic volumes of each lane as a visibility graph and employ the GraphSAGE model to extract the temporal dependence. Then, we develop

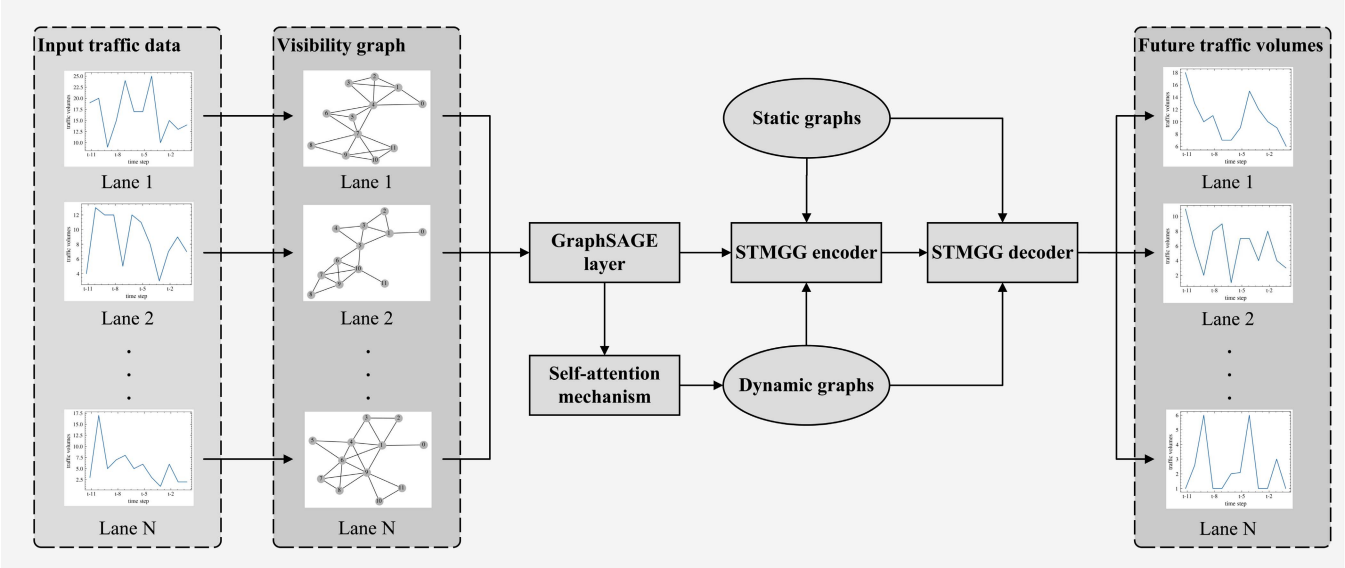



Fig. 1. The framework of the proposed STMGG model.

a dynamic adjacency matrix learning module based on the self-attention mechanism to generate the learnable matrix-based graphs, i.e., LM graph. After that, the STMGG units are applied to incorporate the spatial correlations on the static graphs (i.e., ATT graph and DTW graph) and dynamic graphs (i.e., LM graph). Finally, we integrate the STMGG unit with a GRU-based structure to construct a Seq2Seq prediction framework to enhance multistep prediction performance. Detail descriptions of the proposed STMGG model are introduced in the following subsections.

### C. Temporal Dependence Modeling

It is well known that understanding the temporal dynamics of traffic flow is an essential factor for accurate prediction. Existing studies mainly attempted to learn time-varying characteristics by the deep learning (DL) models directly, e.g., LSTM [30], temporal convolutional network (TCN) [49], attention mechanism [50], etc. However, it is a significant challenge for DL models to understand the laws of the fluctuation phenomenon. During these years, several researchers utilized complex networks to analyze the dynamics and periodicity of traffic flow [51], [52]. Motivated by these works, we attempt to utilize complex networks to extract the time-varying characteristics of lane-level traffic volumes and then employ GNN to explore the temporal dependence.

*1) Temporal Complex Networks Construction:* Complex networks are always regarded as an effective tool for time-series analysis, and numerous methods have been proposed to transform the time-series data into topological graphs [53], e.g., recurrence networks [54], transition networks [55], visibility graphs [12], etc. Since the visibility graph is intuitive and fast computation, we employ this nonparametric method to transform the short-term traffic flow into temporal complex networks. Supposing the temporal dimension of short-term traffic volumes as 12, we display the construction principle of the visibility graph in Fig. 2. Here, each bar denotes traffic

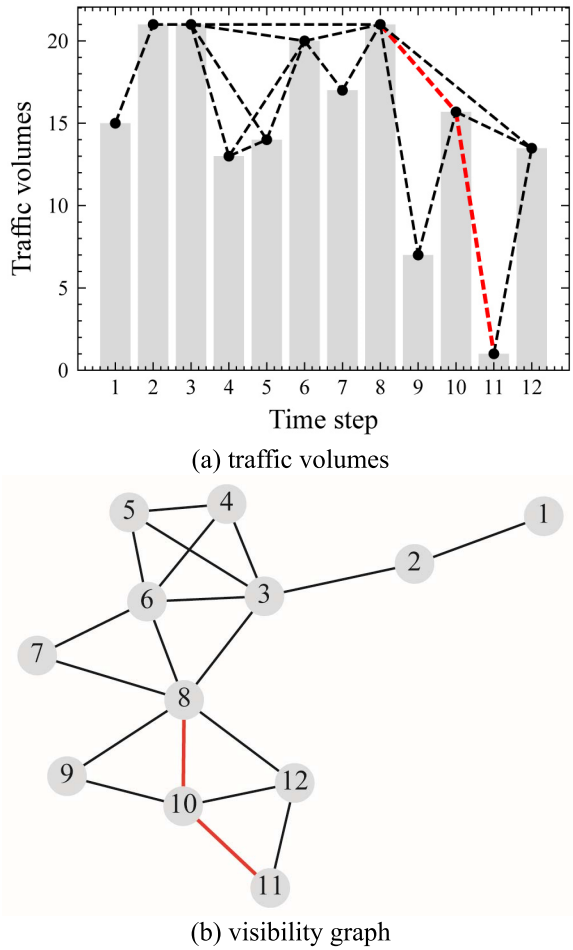



Fig. 2. The construction principle of visibility graph for short-term traffic volumes.

volumes in each 5-min. This construction principle can be summarized as: if traffic volumes at two different time steps can *see* each other, and other bars (i.e., traffic volumes at

other time steps) cannot block the *sight* (e.g., lines in Fig. 2a), we will add an edge to these two steps. Meanwhile, Eq. (2) shows the formal construction process, where $t$ and $v$ denote the time step and corresponding traffic volumes, and **VG** stands for the adjacency matrix of visibility graph of a single lane. It means that if all the data $(t_c, v_c)$ $(t_c \in (t_a, t_b))$ fulfills the constraint in Eq. (2), an undirected edge will be added to node $t_a$ and $t_b$.

$$\mathbf{VG}[i,j] = \begin{cases} 1, & v_c < v_b + \dfrac{t_b - t_c}{t_b - t_a}(v_a - v_b) \\ 0, & \text{else} \end{cases} \tag{2}$$

According to the previous study [56], lines in visibility graphs can be regarded as the maximum limits on the fluctuations of traffic volumes among different intervals. Therefore, this construction principle is employed to determine whether hidden causality exists between the corresponding intervals. For instance, since all the traffic volumes between node 8 and node 10 in Fig. 2a are lower than the red line, we regard there exists causality between these two nodes and add a line to connect them. However, for node 8 and node 11, their *sight* is blocked by node 10, so there is no direct causal relationship between these nodes. Instead, their causality needs to be passed by node 10, i.e., the red lines in Fig. 2.

Meanwhile, fluctuation characteristics of traffic flow are usually caused by random factors. Therefore, the existence of these outliers may dramatically impact prediction performance. According to the collection principle of traffic detectors, outliers always appear with low values due to device failure and data missing. By transforming lane-level traffic volumes into visibility graphs, outliers can obtain lower degrees (e.g., time step 9 and step 11 shown in Fig. 2). As a result, their interactions between other steps are limited, which means the actual time-varying characteristics will be restored. In this way, this vital prior knowledge can be integrated with DL models to adapt temporal dynamics and achieve higher prediction accuracy. Overall, the visibility graph can be considered as a feature engineering to describe the temporal dynamics of lane-level traffic volumes and reduce the impact of outliers.

*2) Temporal Dynamics Extraction:* In the constructed visibility graphs, each node denotes the traffic volume at a specific time step, and the network structure can represent their evolution patterns at the short-term scale. That means we can transform the temporal dependence learning problem into spatial correlation extraction on the constructed visibility graphs. It is natural to consider applying GNNs to solve this problem, but a critical challenge exists for the widely-used GCN model. That is, the constructed visibility graphs rely on actual traffic volumes, so their topological structures are time-varying. However, since GCN is a transductive learning model, it cannot deal with dynamic topological graphs. Many studies have demonstrated the inductive learning ability of the GraphSAGE model, which means it has a strong generalization capability to process graphs with different topological structures. Thus, we employ it to extract the spatial dependence on visibility graphs.

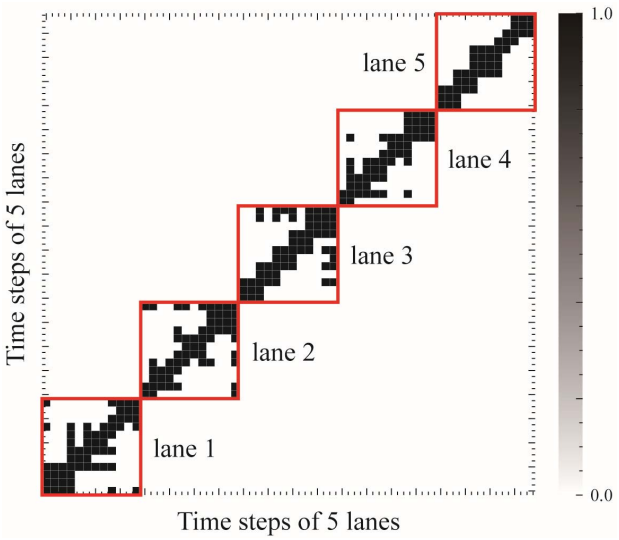



Fig. 3. The batch strategy of visibility graphs for 5 lanes.

The core calculation process of GraphSAGE [13] is summarized in Eq. (3) and (4). For node $v$, it first aggregates information of its neighbors to generate a neighborhood vector $\mathbf{p}^k_{\mathbb{N}(v)}$. As shown in Eq. (4), each node will fuse its neighborhood vector with the current information to update its features. Here, $\mathbf{p}^k_v$ denotes features of node $v$ in layer $k$, and $\mathbb{N}(v)$ presents all the neighbors of node $v$. Meanwhile, $\sigma(\cdot)$ denotes the activate function, which is set as ReLU in this study, and $\mathbf{W}^k$ denotes the learnable weight matrix.

$$\mathbf{p}^k_{\mathbb{N}(v)} = \text{AGGREGATE}_k(\{\mathbf{p}^{k-1}_u, \forall u \in \mathbb{N}(v)\}) \tag{3}$$

$$\mathbf{p}^k_v = \sigma(\mathbf{W}^k \cdot (\{\mathbf{p}^{k-1}_v \bigcup \mathbf{p}^k_{\mathbb{N}(v)}\})) \tag{4}$$

In Eq. (3), AGGREGATE$(\cdot)$ represents the aggregator of GraphSAGE. In the original GraphSAGE model [13], there are three basic aggregator architectures, i.e., mean aggregator, LSTM aggregator, and pooling aggregator. Considering the computing efficiency and GPU occupation, we employ the mean aggregator to implement our model. The mean-based GraphSAGE aggregator is summarized in Eq. (5). From this equation, we can obtain that the GraphSAGE model only relies on features of the neighborhood nodes instead of the whole graph. In other words, when the topological structure changes, nodes can also update their information from new neighbors.

$$\mathbf{p}^k_v = \sigma(\mathbf{W}^k_{\text{mean}} \cdot \text{mean}(\{\mathbf{p}^{k-1}_v \bigcup \{\mathbf{p}^{k-1}_u, \forall u \in \mathbb{N}(v)\}) \tag{5}$$

However, it is noted that if we establish a unique GraphSAGE model for each lane, the training costs will linearly increase with the number of lanes. To solve this problem, we integrate all the visibility graphs at the same period into a larger adjacency matrix, so only one GraphSAGE model is required to extract the spatial correlation. This batch strategy on 5 lanes is displayed in Fig. 3. It indicates that this combining process not only retains the independence of each lane but also improves computational efficiency.

### D. Spatial Dependence Modeling

*1) Spatial Topological Graph Construction:* Spatial correlation is another crucial factor for accurate traffic forecasting [57]. Here, the correlation can be understood in many ways, e.g., physical adjacent relationship, statistical correlation, semantic dependence, etc. Following this opinion, this study constructs three topological graphs to explore the spatial dependence in the urban road network, including two static graphs and a dynamic graph. In these topological graphs, each node denotes a specific lane. To distinguish these graphs from the visibility graphs, we name them spatial topological graphs.

*a) ATT graph:* Since traffic detectors are sparely installed in urban road networks, generating topological graphs directly based on physical adjacent relationships is challenging. Thus, following our previous study [58], we employ the historical average travel time (ATT) matrix to denote the temporal proximities among lanes. Then, the ATT matrix is treated as the distance measure, and the complex network construction algorithm [59] is employed for topological graph construction. The core idea of this algorithm can be considered as an integration of the traditional $k$-nearest and $\varepsilon$- radius method. In this way, lanes with short travel time will be connected, so the constructed ATT graph can realistically represent the physical proximity relationship among lanes. More details of this method can be found in [59] and [60].

*b) DTW graph:* In addition to the proximity relationship, the statistical correlation among lanes is also essential for spatial dependence modeling. Since traffic volumes have typical time-series characteristics, we employ the dynamic time warping (DTW) algorithm [14] to determine the distance instead of the traditional correlation coefficients (e.g., Pearson correlation, Spearman correlation, etc.). After calculating the DTW distance among lane-level traffic volumes, we also treat it as the distance measure and employ the algorithm mentioned above [59] for DTW graph construction.

*c) LM graph:* Since graphs established above are static, they can rarely reflect the temporal evolution of urban traffic flows [32]. Therefore, we further develop a dynamic adjacency matrix learning method to explore the time-varying spatial correlations in urban road networks. The self-attention mechanism is employed to determine correlation coefficients among lane-level traffic flows. Supposing that the output of the above GraphSAGE layer is $\mathbf{p} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{\mathrm{N}}\}$, where $\mathbf{p}_i$ denotes the output features of node $i$, the spatial correlation can be calculated by the following equations.

$$\mathbf{S}[i, j] = \mathbf{p}_i^{\mathrm{T}} \mathbf{p}_j \tag{6}$$

$$\mathbf{R}[i, j] = \frac{\exp(\mathbf{S}[i, j])}{\sum_{k=1}^{\mathrm{N}} \exp(\mathbf{S}[i, k])} \tag{7}$$

Here, $\mathbf{S} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ denotes the spatial correlation matrix, and $\mathbf{R} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ is the normalized correlation matrix. After the Softmax operation in Eq. (7), all the elements in $\mathbf{R}$ are transformed into [0, 1].

As shown in Eq. (8), the $\alpha^{\mathrm{th}}$ quantile of each lane is employed as the threshold to convert the normalized spatial correlation matrix $\mathbf{R}$ into adjacency matrix $\mathbf{D}$. Since $\mathbf{p}$ in Eq. (6) is determined by input data, the learned adjacency matrix $\mathbf{D}$ is adaptive to the actual time-varying characteristics of dynamic traffic flows. Afterwards, the LM graph will be generated from the adjacency matrix $\mathbf{D}$.

$$\mathbf{D}[i, j] = \begin{cases} 1, & \mathbf{R}[i, j] \geq \mathbf{R}_\alpha[i, :] \\ 0, & \mathbf{R}[i, j] < \mathbf{R}_\alpha[i, :] \end{cases} \tag{8}$$

*2) Spatial Correlation Extraction:* Based on the constructed static graphs and dynamic graphs, we propose a deep learning model to explore the spatial correlations of lane-level traffic flow. As illustrated in Fig. 4, we first employ the GraphSAGE model to explore spatial correlations on each spatial topological graph. Then, based on the incorporation of multi-head attention and split-attention mechanism, this study designs a gated module to integrate the global features on these graphs.

In the STMGG unit, we also apply the GraphSAGE model with the mean aggregator to capture the spatial dependence in these spatial topological graphs. Meanwhile, we introduce the multi-head attention mechanism [61] to enhance the training stability. The multi-head mechanism used here can be executed by Eq. (9). It can be regarded as that we construct R unique GraphSAGE model on each graph, and then the prediction results of these independent models are concatenated to output. Hence, the dimension of output features can be written as $\mathbb{R}^{\mathrm{N}\times(\mathrm{C}\times\mathrm{R})}$, where C denotes the output dimension of each GraphSAGE model, and R represents the number of heads.

$$\begin{aligned} \mathbf{p}_v^k &= \|_{r=1}^{\mathrm{R}} \mathbf{p}_v^{r,k} \\ &= \|_{r=1}^{\mathrm{R}} \sigma(\mathbf{W}_{\mathrm{mean}}^{r,k} \cdot \mathrm{mean}(\{\mathbf{P}_v^{r,k-1}\} \\ &\quad \bigcup \{\mathbf{P}_u^{r,k-1}, \forall u \in \mathbb{N}(v)\})) \end{aligned} \tag{9}$$

These three graphs can represent the spatial dependence of network-scale traffic flow from different perspectives. Therefore, understanding the comprehensive spatial correlations by integrating these graphs allows for higher prediction accuracies. Following this opinion, we propose a gated mechanism to incorporate information on these spatial topological graphs. This gated mechanism is based on the integration of split-attention [62] and multi-head attention mechanism. Studies have demonstrated the superiority of the split-attention mechanism in extracting global context information in computer vision, so we utilize it to explore the global features among different graphs. Assuming $\mathbf{p}^{\mathrm{ATT}} \in \mathbb{R}^{\mathrm{N}\times(\mathrm{C}\times\mathrm{R})}$, $\mathbf{p}^{\mathrm{DTW}} \in \mathbb{R}^{\mathrm{N}\times(\mathrm{C}\times\mathrm{R})}$, and $\mathbf{p}^{\mathrm{LM}} \in \mathbb{R}^{\mathrm{N}\times(\mathrm{C}\times\mathrm{R})}$ denote output features of GraphSAGE on the ATT graph, DTW graph, and LM graph, respectively, the calculation procedure of the gated mechanism is summarized in the following equations.

$$\mathbf{p}^{\varepsilon} = \|_{r=1}^{\mathrm{R}} \mathrm{GraphSAGE}_{\varepsilon}^{r}(\mathbf{p}_0, \mathcal{G}_{\varepsilon}), \varepsilon \in \{\mathrm{ATT}, \mathrm{DTW}, \mathrm{LM}\} \tag{10}$$

$$\mathbf{p} = \mathrm{concat}(\{\mathbf{p}^{\mathrm{ATT}}, \mathbf{p}^{\mathrm{DTW}}, \mathbf{p}^{\mathrm{LM}}\}), \mathbf{p} \in \mathbb{R}^{\mathrm{N}\times(\mathrm{C}\times\mathrm{R})\times 3} \tag{11}$$

$$\hat{\mathbf{p}} = \sum_{i=1}^{3} \mathbf{p}[:, :, i] \tag{12}$$

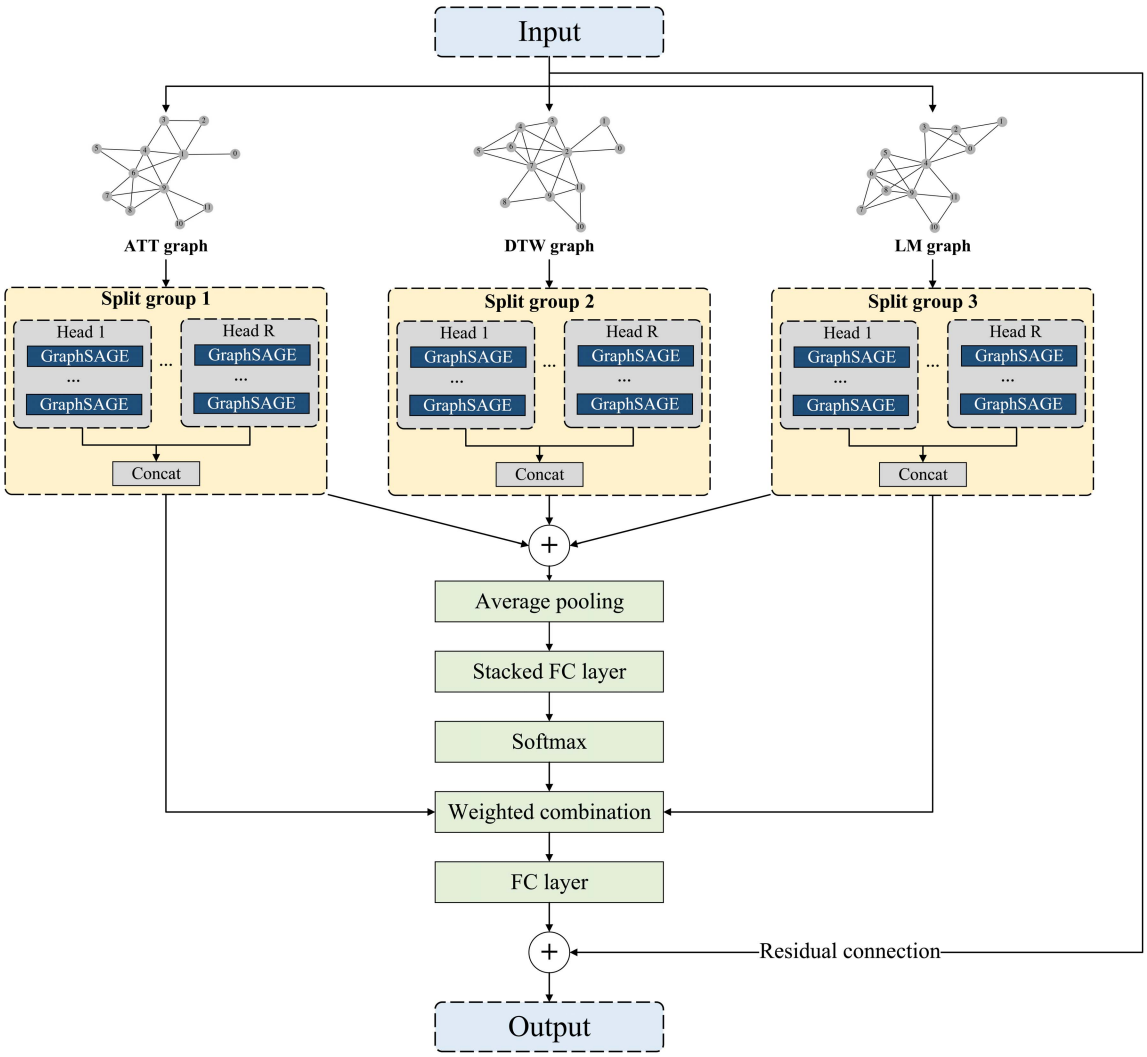



Fig. 4. The spatial correlation mining process of STMGG unit.

$$\mathbf{s} = \text{reshape}(\frac{1}{\text{N}}\sum_{i=1}^{\text{N}}\hat{\mathbf{p}}[i,:]), \mathbf{s} \in \mathbb{R}^{\text{C}\times\text{R}} \tag{13}$$

$$\mathbf{a}[:,j] = \frac{\exp(\xi^j(\mathbf{s}[:,j]))}{\sum_{k=1}^{\text{R}}\exp(\xi^k(\mathbf{s}[:,k]))} \tag{14}$$

$$\mathbf{V} = \mathbf{W}_{\text{out}}(\mathbf{a}\odot\mathbf{p}) + \mathbf{W}_{\text{res}}\mathbf{p}_0 \tag{15}$$

Here, $\mathbf{p}_0$ denotes the input features of the STMGG unit. In Eq. (10), $\varepsilon$ can represent each type of spatial topological graph. Hence, $\mathcal{G}_\varepsilon$, $\text{GraphSAGE}_\varepsilon^r$, and $\mathbf{p}^\varepsilon$ denote the topological graph, GraphSAGE in multi-head attention, and output features of the spatial graph $\varepsilon$, respectively. The abovementioned equations can be viewed as a weighted combination process among different graphs. Detailly, Eq. (12) and (13) denote the global pooling operation, which gathers features across graphs. Eq. (14) aims to calculate the attention weights, and $\xi$ represents two stacked fully connected (FC) layers. In Eq. (15), $\odot$ denotes the Hadamard product, and $\mathbf{W}_{\text{out}}$ is applied to obtain the final output features. Meanwhile, we also utilize the residual connection in Eq. (15) to improve the training stability, where $\mathbf{W}_{\text{res}}$ denotes the weight matrix to keep dimensions equal.

### *E. Seq2Seq Structure*

RNN and its variants have always been regarded as an effective solution to capture temporal dependence [30]. Among the family of RNN, GRU can solve the problem of gradient explosion and reduce the computation cost. To improve the prediction accuracy and stability in the multistep prediction task, we integrate the STMGG unit with GRU to construct a Seq2Seq prediction structure. Fig. 5 illustrates the framework of this Seq2Seq model. In this figure, the blue line denotes the delivery process of hidden states from the encoder to the corresponding decoder. In this way, features at different time steps can interact in the temporal dimension. Meanwhile, this model utilizes the STMGG operation to replace the matrix multiplications in the original GRU. The calculation process is defined in Eq. (16)-(19).

$$\mathbf{R}_t = \text{sigmoid}(\Theta_r \star_g [\mathbf{I}_t, \mathbf{H}_{t-1}] + \mathbf{b}_r) \tag{16}$$

$$\mathbf{Z}_t = \text{sigmoid}(\Theta_z \star_g [\mathbf{I}_t, \mathbf{H}_{t-1}] + \mathbf{b}_z) \tag{17}$$

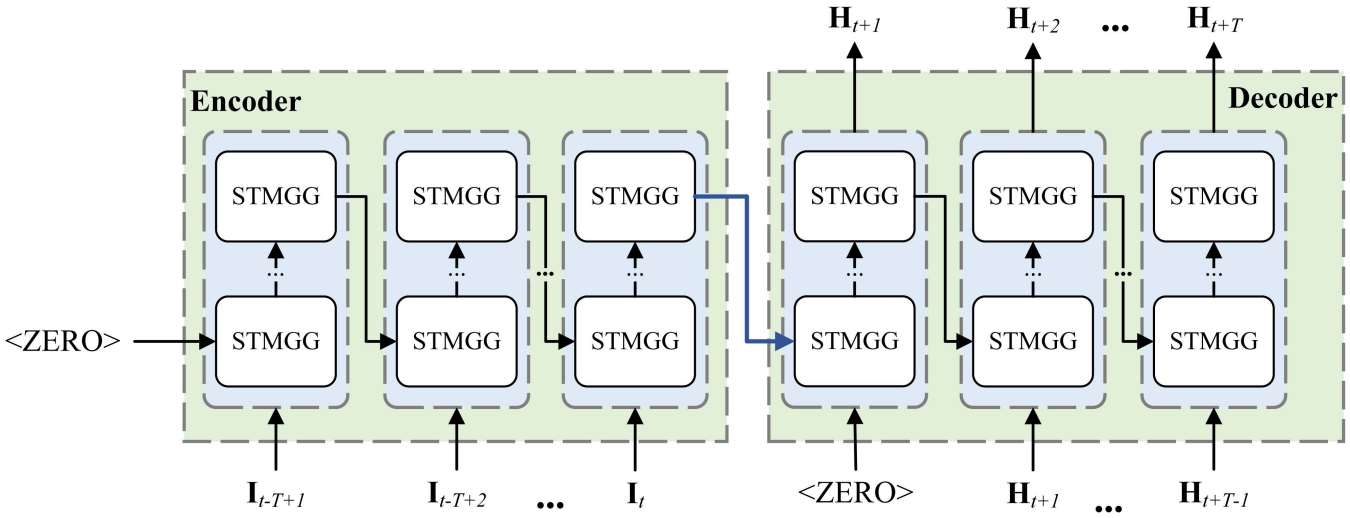



Fig. 5. The framework of the Seq2Seq structure in STMGG model.

$$\hat{\mathbf{H}}_t = \tanh(\Theta_h \star_g [\mathbf{I}_t, (\mathbf{R}_t \odot \mathbf{H}_{t-1})] + \mathbf{b}_h) \tag{18}$$

$$\mathbf{H}_t = \mathbf{Z}_t \odot \mathbf{H}_{t-1} + (1 - \mathbf{Z}_t) \odot \hat{\mathbf{H}}_t \tag{19}$$

In these equations, $\mathbf{R}_t$ and $\mathbf{Z}_t$ denote the reset gate and update gate in the GRU model, respectively. $\mathbf{I}_t$ and $\mathbf{H}_t$ are the input and output features at time step $t$. Specifically, $\mathbf{I}_t$ denotes the hidden features of traffic volumes at time step $t$ encoded by the visibility graph and the first GraphSAGE layer. Meanwhile, $\star_g$ stands for the STMGG operation, and $\Theta$ denotes the parameters for the corresponding STMGG model. Finally, as shown in Eq. (20), we use an FC layer to obtain future traffic volumes.

$$\left\{\hat{\boldsymbol{x}}^{t+1}, \ldots, \hat{\boldsymbol{x}}^{t+\mathrm{T}}\right\} = \mathrm{FC}(\{\mathbf{H}_{t+1}, \ldots, \mathbf{H}_{t+T}\}) \tag{20}$$

## IV. Experiments and Results Analysis

### A. Experiment Settings

*1) Data Description:* To validate the prediction performance of our STMGG model, we introduce two LPR datasets in Changsha city, China, to conduct numerical experiments. The columns of the LPR datasets contain license plate number, collection time, collection address, approach number, and lane number. For privacy protection considerations of drivers, we conduct a hash transformation on the license plate information. These datasets were collected from July 1st to July 31st in 2019, and the spatial distributions of LPR devices are illustrated in Fig. 6. To distinguish these two datasets, we name them LPR-1 and LPR-2, respectively. In total, there are 301 lanes in LPR-1 and 868 lanes in LPR-2. Moreover, we aggregate the traffic volumes every 5 minutes, so 8,928 records are obtained at each lane.

*2) Evaluation Metrics:* Three widely-used metrics are employed to evaluate the prediction performance, including root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). These metrics are defined in the following equations, where N and n denote the number of lanes and data samples, respectively. And in these equations, $\hat{y}_i^j$ represents the predicted traffic volumes of lane $i$ at time $j$, and $y_i^j$ stands for the corresponding ground truth. Existing studies have demonstrated that zero or close-to-zero ground truth data significantly impacts MAPE [63]. Thus, motivated by the previous study [64], we only utilize the actual traffic volumes higher than 5 to calculate this metric.

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n} \cdot \mathrm{N}} \sum_{i=1}^{\mathrm{N}} \sum_{j=1}^{\mathrm{n}} \left(y_i^j - \hat{y}_i^j\right)^2} \tag{21}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{n} \cdot \mathrm{N}} \sum_{i=1}^{\mathrm{N}} \sum_{j=1}^{\mathrm{n}} \left|y_i^j - \hat{y}_i^j\right| \tag{22}$$

$$\mathrm{MAPE} = \frac{1}{\mathrm{n} \cdot \mathrm{N}} \sum_{i=1}^{\mathrm{N}} \sum_{j=1}^{\mathrm{n}} \frac{\left|y_i^j - \hat{y}_i^j\right|}{y_i^j} \times 100\% \tag{23}$$

*3) Implementation Details:* In the static graph construction method, the value of $\lambda$ is set to 3 according to [59], and we set the value of K to 4. Based on these settings, we summarize the structural properties of these graphs in Table I. To keep the network density of static graphs and dynamic graphs comparable, $\alpha$ in Eq. (8) is set to 98%.

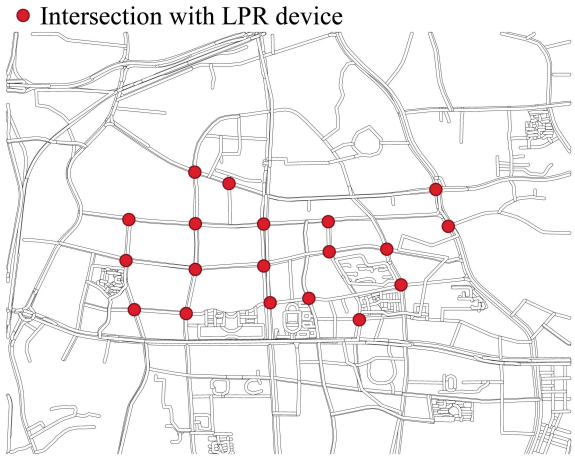
We divide traffic volumes into a training set, a validation set, and a testing set by 70%: 10%: 20%. In the prediction model, we utilize traffic volumes of the previous 12 steps to predict the following 12 steps. The Z-score normalization is applied to the input data and ground truth. The mini-batch training strategy and Adam algorithm [65] with a learning rate $1 \times 10^{-3}$ are employed for model training. Here, the batch sizes of LPR-1 and LPR-2 are set to 16 and 8, respectively. In the multi-head attention mechanism, the number of heads is the same as the number of spatial graphs (i.e., 3). Meanwhile, we stack 2 GraphSAGE layers in each split group and set the number of STMGG unit in each step of encoder or decoder as 1. We employ MAE as the loss function and apply an early-stop strategy according to performance on the validation set to avoid overfitting. Additionally, the hidden dimensions of STMGG are summarized in Table II.

TABLE I
STRUCTURAL PROPERTIES OF THE CONSTRUCTED SPATIAL TOPOLOGICAL GRAPHS

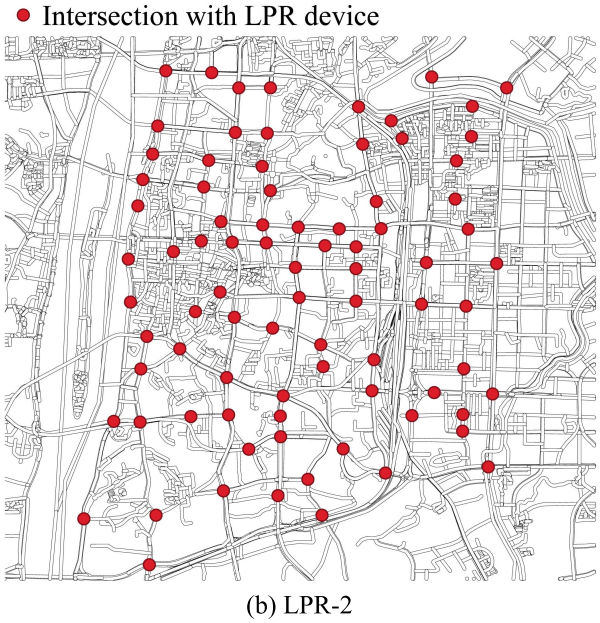
| | LPR-1 | | LPR-2 | |
|---|---|---|---|---|
| | ATT graph | DTW graph | ATT graph | DTW graph |
| Nodes | 301 | 301 | 868 | 868 |
| Edges | 1413 | 1417 | 4058 | 3939 |

TABLE II
HIDDEN DIMENSIONS OF STMGG ON THESE TWO DATASETS

| | GNN in temporal graph | GNN in spatial graph | Inter channel $\xi$ |
|---|---|---|---|
| LPR-1 | 24 | 72 | 16 |
| LPR-2 | 24 | 72 | 16 |




(a) LPR-1




(b) LPR-2

Fig. 6. The spatial distributions of LPR devices in these two datasets.

We employ the open-source deep learning framework *deep graph library* (DGL) [66] and MXNet [67] to implement our STMGG model. The proposed STMGG model is run by Python 3.8.8 on a Windows 10 workstation with one NVIDIA GTX 2080Ti GPU.

### *B. Performance Comparisons*

*1) Baselines:* To validate the superiority of our model, we employ 13 widely-used prediction methods as baselines, including three traditional prediction methods, two tensor-based methods and eight novel GNN-based prediction models. Brief introductions of the selected baselines are listed as follows.

- **HA**. Historical average (HA) is a basic traffic flow prediction method. It utilizes the weighted average of traffic volumes at the corresponding previous intervals for forecasting. In this study, its multistep prediction performance is denoted by 1-step.
- **MLP**. A three-layer multiple layer perceptron (MLP) is employed for short-term traffic flow prediction. According to the grid-search strategy, we set its hidden dimension to 256 and 96 in these two datasets, and ReLU is applied as the activate function.
- **LSTM** [30]. LSTM is a widely-used time series prediction model, and it is effective in capturing temporal dynamics. We also apply a grid-search strategy and set the hidden dimension to 256 and 96, respectively.
- **DCRNN** [9]. DCRNN can be regarded as an early attempt to apply GNN in traffic prediction. It utilizes the random walks and Seq2Seq structure to capture the spatial and temporal correlations.
- **STGCN** [50]. In the spatial-temporal graph convolutional network (STGCN), the graph convolutional operation is employed to extract the spatial dependence. Instead of the RNN-based modules, it utilized the gated convolutional layer to capture the temporal dynamics.
- **GaAN** [68]. The gated attention networks (GaAN) proposed a gated module based on CNN in the multi-head attention mechanism. Meanwhile, it further employed the Seq2Seq structure to explore the long-term dependence in traffic data.
- **Graph-WaveNet** [47]. This model proposed a dynamic topological network learning module and employ the TCN to capture the temporal dynamics.

TABLE III
PREDICTION PERFORMANCE COMPARISON ON LPR-1

| | 3-step | | | 6-step | | | 9-step | | | 12-step | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE |
| HA | 5.078 | 2.518 | 35.532% | 5.078 | 2.518 | 35.532% | 5.078 | 2.518 | 35.532% | 5.078 | 2.518 | 35.532% |
| MLP | 3.096 | 1.967 | 25.153% | 3.315 | 2.109 | 26.665% | 3.482 | 2.191 | 28.025% | 3.736 | 2.342 | 30.076% |
| LSTM | 3.172 | 1.923 | 25.709% | 3.279 | 1.974 | 26.399% | 3.419 | 2.040 | 27.416% | 3.537 | 2.105 | 28.223% |
| STGCN | 3.243 | 1.876 | 26.121% | 3.461 | 1.977 | 27.658% | 3.740 | 2.109 | 29.709% | 4.014 | 2.242 | 31.715% |
| DCRNN | 3.881 | 2.076 | 28.980% | 4.187 | 2.238 | 31.292% | 4.522 | 2.409 | 33.754% | 4.808 | 2.554 | 35.716% |
| GaAN | 3.700 | 2.043 | 28.277% | 4.022 | 2.198 | 30.841% | 4.344 | 2.365 | 33.100% | 4.596 | 2.514 | 34.466% |
| Graph-WaveNet | 2.918 | 1.755 | 24.405% | 3.100 | 1.830 | 25.381% | 3.278 | 1.914 | 26.575% | 3.449 | 2.004 | 27.860% |
| WTMF | 3.700 | 2.034 | 26.673% | 3.945 | 2.096 | 27.656% | 4.125 | 2.148 | 28.544% | 4.274 | 2.195 | 29.330% |
| T-GCN | 3.021 | 1.922 | 24.226% | 3.147 | 1.977 | 24.986% | 3.309 | 2.067 | 25.993% | 3.460 | 2.160 | 27.148% |
| ASTGCN | 3.685 | 2.114 | 29.104% | 3.832 | 2.182 | 30.045% | 3.897 | 2.211 | 30.604% | 3.878 | 2.212 | 30.537% |
| TGC-LSTM | 3.091 | 1.916 | 25.077% | 3.362 | 2.045 | 27.042% | 3.661 | 2.212 | 28.910% | 3.912 | 2.350 | 30.574% |
| STSGCN | 3.019 | 1.804 | 24.766% | 3.150 | 1.865 | 25.644% | 3.276 | 1.927 | 26.566% | 3.399 | 2.002 | 27.617% |
| BTMF | 4.129 | 2.213 | 29.858% | 4.366 | 2.319 | 31.470% | 4.461 | 2.351 | 32.034% | 4.523 | 2.368 | 32.314% |
| STMGG | **2.862** | **1.743** | **24.150%** | **2.975** | **1.784** | **24.959%** | **3.068** | **1.830** | **25.732%** | **3.178** | **1.890** | **26.582%** |

TABLE IV
PREDICTION PERFORMANCE COMPARISON ON LPR-2

| | 3-step | | | 6-step | | | 9-step | | | 12-step | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE |
| HA | 8.103 | 4.173 | 35.196% | 8.103 | 4.173 | 35.196% | 8.103 | 4.173 | 35.196% | 8.103 | 4.173 | 35.196% |
| MLP | 6.526 | 3.778 | 28.685% | 6.604 | 3.788 | 28.836% | 6.741 | 3.894 | 29.640% | 6.878 | 3.973 | 30.590% |
| LSTM | 6.382 | 3.383 | 26.783% | 6.492 | 3.459 | 27.626% | 6.580 | 3.516 | 28.020% | 6.721 | 3.626 | 29.018% |
| STGCN | 6.468 | 3.258 | 26.212% | 7.320 | 3.815 | 30.688% | 8.299 | 4.429 | 35.440% | 8.962 | 5.031 | 40.119% |
| DCRNN | 5.927 | 3.345 | 26.045% | 6.588 | 3.665 | 28.193% | 7.241 | 3.986 | 30.479% | 7.835 | 4.304 | 32.527% |
| GaAN | 5.662 | 3.183 | 25.216% | 6.317 | 3.506 | 27.457% | 6.977 | 3.836 | 29.854% | 7.598 | 4.156 | 32.044% |
| Graph-WaveNet | 5.090 | **2.886** | 23.538% | 5.648 | 3.115 | 25.556% | 6.094 | 3.319 | 27.371% | 6.455 | 3.504 | 28.985% |
| WTMF | 7.010 | 3.453 | 26.635% | 7.352 | 3.553 | 27.545% | 7.465 | 3.616 | 28.109% | 7.542 | 3.674 | 28.666% |
| T-GCN | 5.601 | 3.297 | 25.298% | 5.847 | 3.379 | 25.892% | 6.107 | 3.507 | 26.760% | 6.381 | 3.666 | 27.716% |
| ASTGCN | 6.797 | 3.639 | 29.278% | 7.078 | 3.776 | 30.533% | 7.251 | 3.865 | 31.411% | 7.340 | 3.928 | 32.020% |
| TGC-LSTM | 5.983 | 3.418 | 27.241% | 7.012 | 3.899 | 31.475% | 6.657 | 3.759 | 29.982% | 6.908 | 3.893 | 31.030% |
| STSGCN | 5.423 | 3.012 | 24.511% | 5.784 | 3.134 | 25.466% | 6.071 | 3.243 | 26.289% | 6.315 | 3.363 | 27.314% |
| BTMF | 7.216 | 3.834 | 30.960% | 7.546 | 3.998 | 32.493% | 7.622 | 4.038 | 32.848% | 7.643 | 4.048 | 32.933% |
| STMGG | **5.034** | 2.900 | **23.497%** | **5.478** | **3.050** | **24.691%** | **5.850** | **3.196** | **25.802%** | **6.212** | **3.351** | **27.101%** |

- **WTMF** [69]. In this WTMF model, an alternating minimization scheme is designed to implement low-rank matrix factorization. We employ the open-source code in [70] to run this model.
- **ASTGCN** [71]. Based on the short-term, daily, and weekly dependencies, Guo et al. [71] proposed the attention based spatial-temporal graph convolutional network (ASTGCN) for network-scale traffic prediction.
- **TGC-LSTM** [43]. Considering the traffic significance, Cui et al. [43] proposed a traffic graph convolutional operation and combined it with LSTM to explore the spatiotemporal dependencies.
- **T-GCN** [15]. In this temporal graph convolutional network (T-GCN), GRU is integrated with GCN to enhance the spatiotemporal learning capability.
- **STSGCN** [72]. After the synchronous modeling, Song et al. [72] proposed a spatial-temporal synchronous graph convolutional network (STSGCN) for spatiotemporal traffic prediction.
- **BTMF** [73]. The Bayesian temporal matrix factorization (BTMF) model is also a tensor-based model. It combines the low-rank tensor factorization and VAR for multidimension time-series imputation and prediction.

*2) Comparisons on the Whole Dataset:* In this study, we employ the official codes to implement these GNNs. Since the ATT graph can represent spatial proximity, we utilize it as the adjacency matrix of these models. Table III and Table IV summarize prediction performances of STMGG and the selected baselines at 3-step, 6-step, 9-step, and 12-step, where the best prediction performance is marked in bold.

Here, the prediction accuracies of HA reveal that the unstable and dynamic characteristics in lane-level traffic volumes are challenging to follow, so the traditional HA model cannot achieve acceptable prediction performance. Compared with HA, due to the powerful capability in feature extraction, MLP and LSTM can improve prediction performance. Furthermore, since the topological information is involved in prediction models, GNN-based methods can usually outperform traditional prediction models, and the proposed STMGG model consistently outperforms all the advanced baselines.

We can observe another superiority of STMGG is the stability in multistep prediction. It can consistently achieve

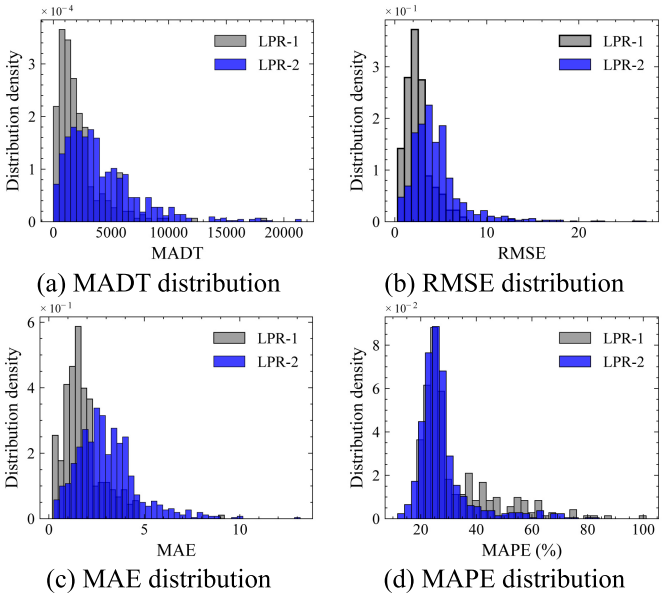



Fig. 7. Distributions of traffic volumes and prediction errors of STMGG among different lanes.

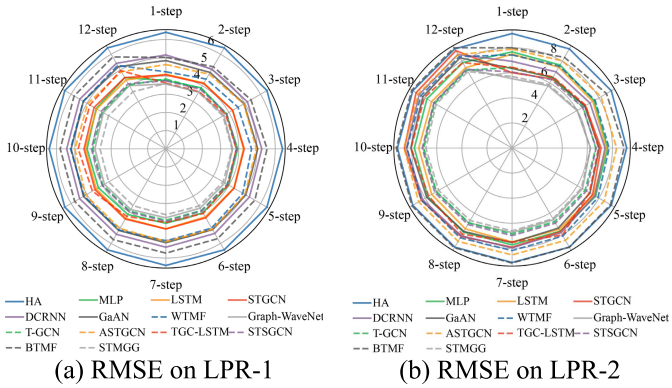



Fig. 8. Prediction performance comparison under rush hours.

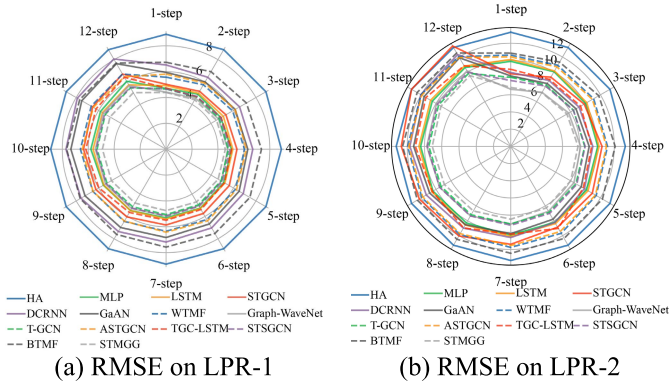



Fig. 9. Prediction performance comparison under top 25% high-volume lanes.

superior performance among baselines in all metrics. Although almost all the baselines involve temporal dependence learning modules (e.g., Seq2Seq, TCN, etc.), the prediction performance of several models (e.g., STGCN, DCRNN, GaAN, etc.) decreases dramatically. But in STMGG, in addition to the Seq2Seq framework, we also transform the temporal dynamics extraction task into spatial correlation modeling on the temporal complex networks. Hence, its capability in temporal dependence learning is significantly enhanced, leading to a more stable prediction performance in multistep prediction.

Interestingly, we can also find that several GNN-based models cannot achieve acceptable (or even comparable) performance to traditional baselines (i.e., MLP and LSTM) on LPR-1. This phenomenon also appeared in the previous study [42], where traditional prediction methods performed similarly or superiorly to GNNs in urban traffic flow prediction. This phenomenon may be that the road network scale and environment are relatively simple on LPR-1, so traditional prediction methods can be adaptive to this situation. Meanwhile, since many GNN-based baselines are proposed for freeway traffic predictions, their spatiotemporal extraction mechanism may be insufficient and unsuitable for urban road networks. However, traditional models are challenging to capture the heterogeneous spatiotemporal dependencies under a complex road network (i.e., LPR-2), so their prediction performance significantly decreases. Therefore, prediction accuracies of almost all the GNN-based models can outperform MLP and LSTM in LPR-2.

Additionally, we further explore deviation distributions for different lanes in Fig. 7, including monthly average day traffic (MADT), RMSE, MAE, and MAPE. Fig. 7a indicates that MADT on most lanes is lower than 10000, and lanes in LPR-2 always have large volumes than LPR-1, especially when MADT is higher than 3000. The reason for this phenomenon is that the LPR-2 dataset locates in the city center, so these intersections always have heavier traffic. Meanwhile, distributions of RMSE and MAE show a similar trend with MADT, i.e., prediction errors are always higher in LPR-2. This is because RMSE and MAE positively correlate with the corresponding ground truth [17]. Conversely, since the MAPE describes the relative errors, this metric is not highly relied on traffic volumes. Therefore, although volume distributions in these datasets are different, their MAPEs still express similar distributions.

*3) Comparisons Under Critical Scenarios:* Traffic congestion usually appears during rush hours and critical intersections, so traffic managers prioritize traffic flow predictions under these extreme conditions. Thus, we further explore prediction accuracies during rush hours and lanes with high volumes. The results are summarized in Fig. 8 and 9, respectively. In this study, 7:30-9:30 and 17:30-19:30 are defined as the rush hours, and we select lanes with the top 25% volumes as the critical lanes.

Under both rush hours and critical lanes, traffic volumes are always higher than in other scenarios. Since RMSE is positively related to the corresponding ground truth, all the RMSE in these figures are higher than that in Table III and Table IV. Compared with prediction performance in the above tables, the RMSE of HA in these figures shows that its performance expresses a sharp decline. It further reveals that traffic volumes under critical scenarios are more unstable, so the

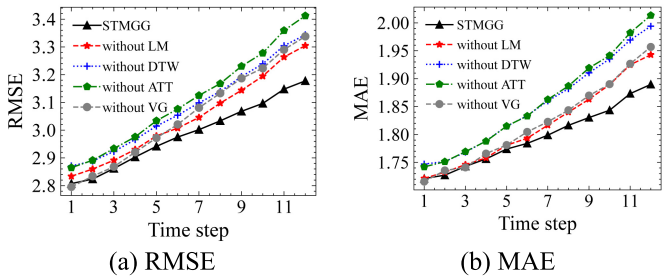



Fig. 10. Ablation experiment of STMGG on LPR-1.

difficulty of accurate predictions significantly increases. Meanwhile, we can observe that although prediction accuracies of STMGG are relatively worse than the previous experiments (i.e., Table III and Table IV), it can also outperform the baselines, especially in multistep prediction.

Furthermore, these figures also indicate that prediction accuracy reduces with the prediction step increase. Besides the prediction accuracies, the standard deviations of STMGG on 12 steps also achieve impressive performance. Significantly, the standard deviation of RMSE under rush hours on LPR-1 reaches 0.014, and this value is just 17.72% of the baselines with the lowest standard deviation (i.e., T-GCN). Therefore, these results reveal that STMGG can maintain both accuracy and stability under critical scenarios.

### C. Ablation Analysis

STMGG comprises several critical components, including visibility graphs and three spatial topological graphs. Therefore, identifying the critical components is essential to reveal why the prediction performance improves and make guidelines for future works. In this subsection, the ablation experiment is conducted to explore the impacts of these components.

To achieve this goal, we remove each component from STMGG and evaluate the corresponding prediction performance, respectively. Fig. 10 shows the stepwise prediction performance of the ablation experiment. The following phenomenon can be identified from these figures.

- We can intuitively find that STMGG always outperforms all the partial models. This phenomenon indicates that integrating these components is helpful in improving prediction accuracies. Furthermore, it demonstrates that STMGG can incorporate in-depth features from different perspectives, which are adaptive to the spatiotemporal evolution patterns of urban road networks.
- The prediction performance of STMGG without VG shows that it achieves the lowest RMSE and MAE at 1-step, even outperforming STMGG. However, with the prediction step increasing, its prediction errors dramatically rise. This phenomenon indicates that introducing VG into short-term traffic flow prediction can significantly improve prediction stability and achieve higher accuracies in multistep prediction.
- From the prediction comparison of three spatial topological graphs, we can see that its prediction performance gains a more significant drop when removing the ATT graph. Thus, it demonstrates that although all these graphs can improve prediction accuracy, the ATT graph plays a more vital role in STMGG than other graphs. Furthermore, we can find that there seems to be no significant difference between the remaining two spatial topological graphs according to RMSE. However, in Fig. 10b, MAE significantly decreases when the DTW graph is removed. Therefore, it indicates that the DTW graph plays a more critical role than the LM graph.

Ablation experiments indicate that introducing each component into the prediction model can further improve prediction performance. To further explore the model interpretation, we display distribution densities of VG, ATT, and DTW distance in Fig. 11.

In Fig. 11a and 11d, the rush hours are also defined as 7:30-9:30 and 17:30-19:30, and the night is set to 23:00-6:00. Meanwhile, we set noontime as 11:00-14:00. According to these two figures, we can observe that visibility graphs express different distribution characteristics at different times. It means that visibility graphs can distinguish traffic patterns among other periods. In this way, STMGG can effectively capture the temporal dynamics to achieve accurate predictions.

Meanwhile, Fig. 11b and 11e also illustrate the distribution comparison between average travel time on the original ATT matrices and ATT graphs. The comparison results indicate that the constructed ATT graphs allow lanes with low travel time to connect, enhancing the spatial correlation from the traffic perspective. Similarly, Fig. 11c and 11f show that the distribution comparisons on DTW distance also hold the same view.

### D. Inductive Learning

Traditional traffic flow prediction methods always encounter limitations in generalization. Specifically, when applying to other datasets, most well-trained models need to tune their parameters by a large amount of data instead of directly making predictions [74]. Sometimes, this data requirement is unrealistic, such as real-time traffic management in the new urban area. Meanwhile, repetitive training may make it too expensive for practical applications.

Inductive learning is a novel field in artificial intelligence. A model with inductive learning capability in traffic prediction means that the well-trained model can be applied to different road networks for forecast without training again. Following this opinion, we further conduct an inductive learning experiment on the STMGG model. For instance, we directly apply the well-trained STMGG model from LPR-1 for network-scale prediction on LPR-2 without training again. The performance comparison is displayed in Fig. 12, where STMGG-i denotes the prediction results of the STMGG model on inductive learning. Here, prediction performance at each step is displayed as a scatter and employed as the metric for these boxplots.

These comparisons show that STMGG-i can achieve acceptable performance on the other dataset without training again. Although indicators of its boxplots seem to be worse than several models, it is worth noting that it can achieve impressive accuracies at 1-step prediction. For instance, its RMSE on

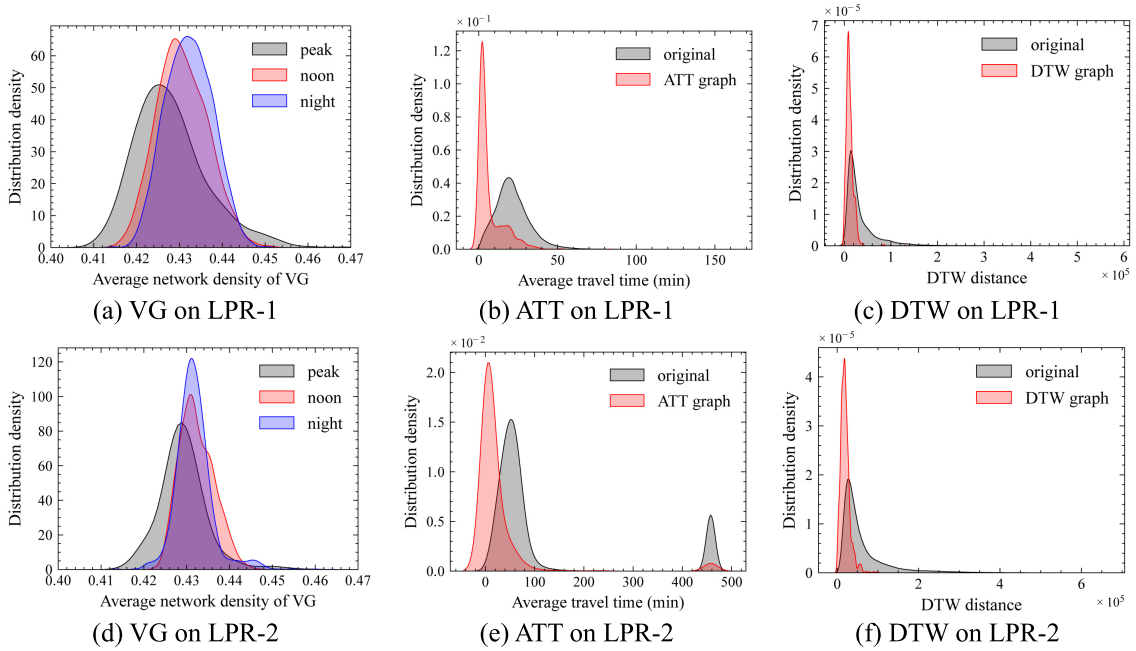



Fig. 11. Traffic features distribution densities in STMGG model.

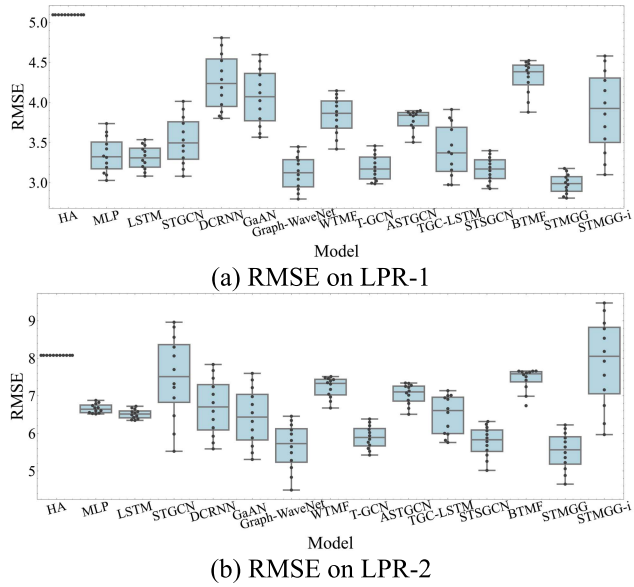



Fig. 12. Prediction performance comparison between baselines and STMGG-i.

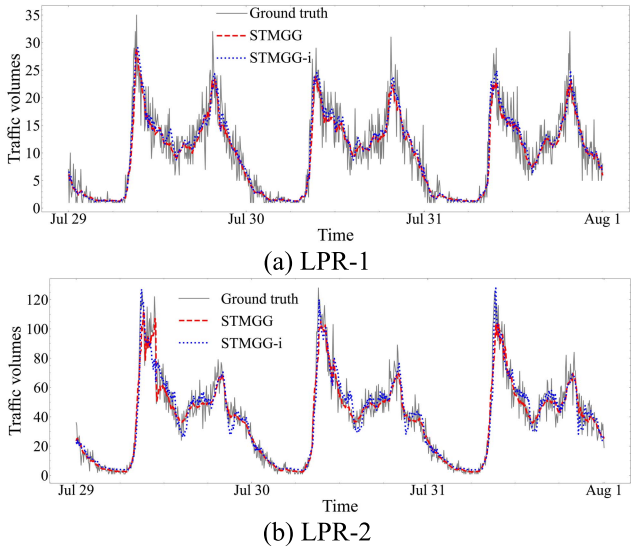



Fig. 13. 1-step prediction results comparison in two LPR datasets.

LPR-1 is lower than several advanced GNN-based models, e.g., ASTGCN, DCRNN, GaAN. Meanwhile, STMGG-i can outperform HA, MLP, LSTM, and ASTGCN on LPR-2. Additionally, an interesting conclusion can be drawn from this figure, i.e., STMGG-i achieves higher accuracies on LPR-1 than LPR-2. Since more lanes exist in LPR-2, the STMGG model can learn more vital and comprehensive knowledge from this dataset. In this way, its generalization is enhanced, leading to a powerful inductive capability. Therefore, even without learning any information from LPR-1, this model can also adapt to traffic situations in this dataset.

Furthermore, the 1-step comparisons of ground truth and prediction results of STMGG-i are illustrated in Fig. 13. Even without training on the target dataset, the STMGG-i model can also capture the temporal evolution of ground truth data effectively. These results demonstrate that the STMGG model has a powerful learning capability and robustness, and its data dependence is insignificant. Since there is no training time for STMGG-i, it is suitable for the 1-step prediction task with a significantly reduced computational cost.

## V. Conclusion

This study proposes a spatiotemporal learning framework, STMGG, to conduct network-scale prediction for lane-level

traffic volumes. Since extreme fluctuation lies in urban traffic flow, we model the temporal dynamics of traffic volumes as visibility graphs to explore the time-varying characteristics. In this way, temporal dependence mining on short-term traffic flows is transformed into spatial correlation learning on visibility graphs. Meanwhile, three spatial topological graphs are constructed from different perspectives to represent the spatial correlations on the urban road network. Based on these spatial topological graphs, we utilize GraphSAGE to explore the spatiotemporal dependencies and develop a gated mechanism to fuse the global contextual information among different graphs. Furthermore, we integrate this model with the Seq2Seq structure to improve the multi-step prediction stability. Validated on two real-world LPR datasets in Changsha, China, the proposed STMGG model expresses superior performance among all the advanced baselines. Additionally, we also conduct an ablation experiment on STMGG and further verify its inductive learning ability.

Due to the critical barriers in data collection, limitations also lie in this study. Overall, the following topics may be worthy of attention for future works.

- Synergy between spatial and temporal correlations. Although this study opens a new perspective for exploring temporal dynamics in short-term traffic flow prediction tasks, the temporal and spatial dependence is still considered separately. It may be a helpful solution by considering the network-scale traffic flow as multivariate time series and transforming the traffic systems into multiplex complex networks, e.g., multiplex visibility graphs [75], multiplex recurrence networks [76], etc.
- Synergy between traffic flows at intersections and road segments. Intersections and road segments can be regarded as the nodes and edges of the traffic network. Thus, future research can consider integrating the node features (e.g., traffic volumes) with the edge features (e.g., speed, travel time, etc.) to achieve simultaneous predictions for both node-level and edge-level traffic states.
- Synergy between different traffic parameters. In addition to the spatiotemporal dependencies, statistical correlations also exist among various traffic parameters (e.g., traffic volumes, speed, density, headway, etc.) [77]. Thus, building multi-task prediction models for different parameters may also be worth trying topic.

## REFERENCES


[1] A. Abdelraouf, M. Abdel-Aty, and J. Yuan, "Utilizing attention-based multi-encoder–decoder neural networks for freeway traffic speed prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11960–11969, Aug. 2022.

[2] Z. Cui, R. Ke, Z. Pu, X. Ma, and Y. Wang, "Learning traffic as a graph: A gated graph wavelet recurrent neural network for network-scale traffic prediction," *Transp. Res. C, Emerg. Technol.*, vol. 115, Jun. 2020, Art. no. 102620.

[3] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[4] A. Dharia and H. Adeli, "Neural network model for rapid forecasting of freeway link travel time," *Eng. Appl. Artif. Intell.*, vol. 16, nos. 7–8, pp. 607–613, 2003.

[5] W. Lu, Y. Rui, and B. Ran, "Lane-level traffic speed forecasting: A novel mixed deep learning model," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3601–3612, Apr. 2022.

[6] W. Lu, Z. Yi, W. Liu, Y. Gu, Y. Rui, and B. Ran, "Efficient deep learning based method for multi-lane speed forecasting: A case study in Beijing," *IET Intell. Transp. Syst.*, vol. 14, no. 14, pp. 2073–2082, Dec. 2020.

[7] Y. Gu, W. Lu, L. Qin, M. Li, and Z. Shao, "Short-term prediction of lane-level traffic speeds: A fusion deep learning model," *Transp. Res. C, Emerg. Technol.*, vol. 106, pp. 1–16, Sep. 2019.

[8] S. Zhang, Y. Guo, P. Zhao, C. Zheng, and X. Chen, "A graph-based temporal attention framework for multi-sensor traffic flow forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7743–7758, Jul. 2022.

[9] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*.

[10] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 4, Sep. 2014, pp. 3104–3112.

[11] L. N. N. Do, H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung, "An effective spatial–temporal attention based neural network for traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 108, pp. 12–28, Nov. 2019.

[12] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, "From time series to complex networks: The visibility graph," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 13, pp. 4972–4975, 2008.

[13] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," 2017, *arXiv:1706.02216*.

[14] D. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. Workshop Knowl. Discovery Databases*, vol. 398, 1994, pp. 359–370.

[15] L. Zhao, Y. Song, C. Zhang, and Y. Liu, "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[16] C. Furtlehner, J.-M. Lasgouttes, A. Attanasi, M. Pezzulla, and G. Gentile, "Short-term forecasting of urban traffic using spatio-temporal Markov field," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 10858–10867, Aug. 2022.

[17] Y. Zhang, T. Cheng, and Y. Ren, "A graph deep learning method for short-term traffic forecasting on large road networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 34, no. 10, pp. 877–896, Oct. 2019.

[18] Z. Li, S. Jiang, L. Li, and Y. Li, "Building sparse models for traffic flow prediction: An empirical comparison between statistical heuristics and geometric heuristics for Bayesian network approaches," *Transportmetrica B, Transp. Dyn.*, vol. 7, no. 1, pp. 107–123, Dec. 2019.

[19] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1678, no. 1, pp. 179–188, Jan. 1999.

[20] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.

[21] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 50–64, Jun. 2014.

[22] Y.-J. Wu, F. Chen, C.-T. Lu, and S. Yang, "Urban traffic flow prediction using a spatio-temporal random effects model," *J. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 282–293, 2016.

[23] W. Li et al., "Short-term traffic state prediction from latent structures: Accuracy vs. efficiency," *Transp. Res. C, Emerg. Technol.*, vol. 111, pp. 72–90. Feb. 2020.

[24] L. Zhao, X. Wen, Y. Wang, and Y. Shao, "A novel hybrid model of ARIMA-MCC and CKDE-GARCH for urban short-term traffic flow prediction," *IET Intell. Transp. Syst.*, vol. 16, no. 2, pp. 206–217, Feb. 2022.

[25] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transp. Res. C, Emerg. Technol.*, vol. 13, no. 3, pp. 211–234, 2005.

[26] H. Liu, C. Yu, H. Wu, Z. Duan, and G. Yan, "A new hybrid ensemble deep reinforcement learning model for wind speed short term forecasting," *Energy*, vol. 202, Jul. 2020, Art. no. 117794.

[27] X. Feng, X. Ling, H. Zheng, Z. Chen, and Y. Xu, "Adaptive multi-kernel SVM with spatial–temporal correlation for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2001–2013, Jun. 2019.

[28] H. Yan, Y. Qi, Q. Ye, and D. Yu, "Robust least squares twin support vector regression with adaptive FOA and PSO for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 14542–14556, Sep. 2022.

[29] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 61–78, May 2015.

[30] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.

[31] C. Ma, Y. Zhao, G. Dai, X. Xu, and S.-C. Wong, "A novel STFSA-CNN-GRU hybrid model for short-term traffic speed prediction," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 2, 2022, doi: 10.1109/TITS.2021.3117835.

[32] H. Yan, X. Ma, and Z. Pu, "Learning dynamic and hierarchical traffic spatiotemporal features with transformer," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 22386–22399, Nov. 2022.

[33] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[34] J. Zhang, Y. Zheng, J. Sun, and D. Qi, "Flow prediction in spatio-temporal networks based on multitask deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 3, pp. 468–478, Mar. 2020.

[35] L. Liu, Z. Qiu, G. Li, Q. Wang, W. Ouyang, and L. Lin, "Contextualized spatial–temporal network for taxi origin-destination demand prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3875–3887, Oct. 2019.

[36] J. Ke, H. Zheng, H. Yang, and X. M. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *J. Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017.

[37] X. Dai et al., "DeepTrend 2.0: A light-weighted multi-scale traffic prediction model using detrending," *Transp. Res. C, Emerg. Technol.*, vol. 103, pp. 142–157, Jun. 2019.

[38] L. Liu, J. Chen, H. Wu, J. Zhen, G. Li, and L. Lin, "Physical-virtual collaboration modeling for intra- and inter-station metro ridership prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3377–3391, Apr. 2022.

[39] Z. Zhang, M. Li, X. Lin, Y. Wang, and F. He, "Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies," *Transp. Res. C, Emerg. Technol.*, vol. 105, pp. 297–322, Oct. 2019.

[40] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 1720–1730.

[41] J. Tan, H. Liu, Y. Li, S. Yin, and C. Yu, "A new ensemble spatio-temporal PM2.5 prediction method based on graph attention recursive networks and reinforcement learning," *Chaos, Solitons Fractals*, vol. 162, Sep. 2022, Art. no. 112405.

[42] J. Liu, G. P. Ong, and X. Chen, "GraphSAGE-based traffic speed forecasting for segment network with sparse data," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1755–1766, Mar. 2022.

[43] Z. Cui, K. Henrickson, R. Ke, and Y. Wang, "Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4883–4894, Nov. 2020.

[44] K. Lee and W. Rhee, "DDP-GCN: Multi-graph convolutional network for spatiotemporal traffic forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 134, Jan. 2022, Art. no. 103466.

[45] J. Ye, J. Zhao, K. Ye, and C. Xu, "How to build a graph-based deep learning architecture in traffic domain: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 3904–3924, May 2022.

[46] Z. Diao, X. Wang, D. Zhang, Y. Liu, K. Xie, and S. He, "Dynamic spatial–temporal graph convolutional neural networks for traffic forecasting," in *Proc. 33rd AAAI Conf. Artif. Intell., 31st Innov. Appl. Artif. Intell. Conf. (IAAI), 9th AAAI Symp. Educ. Adv. Artif. Intell. (EAAI)*, 2019, pp. 890–897.

[47] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for deep spatial–temporal graph modeling," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 1907–1913.

[48] J. J. Q. Yu, "Graph construction for traffic prediction: A data-driven approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15015–15027, Sep. 2022.

[49] G. Ge and W. Yuan, "Short-term traffic speed forecasting based on graph attention temporal convolutional networks," *Neurocomputing*, vol. 410, no. 14, pp. 387–393, Oct. 2020.

[50] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3634–3640.

[51] J. Tang, L. Fang, W. Zhang, Z. Shen, and Y. Wang, "Exploring dynamic property of traffic flow time series in multi-states based on complex networks: Phase space reconstruction versus visibility graph," *Phys. A, Stat. Mech. Appl.*, vol. 450, pp. 635–648, May 2016.

[52] Y. Yan, S. Zhang, J. Tang, and X. Wang, "Understanding characteristics in multivariate traffic flow time series from complex network structure," *Phys. A, Stat. Mech. Appl.*, vol. 477, pp. 149–160, Jul. 2017.

[53] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths, "Complex network approaches to nonlinear time series analysis," *Phys. Rep.*, vol. 787, pp. 1–97, Jan. 2019.

[54] X. Xu, J. Zhang, and M. Small, "Superfamily phenomena and motifs of networks induced from time series," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 50, pp. 19601–19605, 2008.

[55] M. McCullough, M. Small, T. Stemler, and H. H.-C. Iu, "Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 25, no. 5, May 2015, Art. no. 053101.

[56] H. Liu, X. Zhang, and X. Zhang, "Exploring dynamic evolution and fluctuation characteristics of air traffic flow volume time series: A single waypoint case," *Phys. A, Stat. Mech. Appl.*, vol. 503, pp. 560–571, Aug. 2018.

[57] E. Alireza and L. David, "Spatiotemporal short-term traffic forecasting using the network weight matrix and systematic detrending," *Transp. Res. C, Emerg. Technol.*, vol. 104, pp. 38–52, Jul. 2019.

[58] J. Tang and J. Zeng, "Spatiotemporal gated graph attention network for urban traffic flow prediction based on license plate recognition data," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 37, no. 1, pp. 3–23, Jan. 2022.

[59] T. H. Cupertino, J. Huertas, and L. Zhao, "Data clustering using controlled consensus in complex networks," *Neurocomputing*, vol. 118, pp. 132–140, Oct. 2013.

[60] T. C. Silva and L. Zhao, *Machine Learning in Complex Networks*. Cham, Switzerland: Springer, 2016.

[61] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 2017, pp. 5998–6008.

[62] H. Zhang et al., "ResNeSt: Split-attention networks," 2020, *arXiv:2004.08955*.

[63] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecasting*, vol. 32, no. 3, pp. 669–679, Sep. 2016.

[64] J. Tang, J. Liang, F. Liu, J. Hao, and Y. Wang, "Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network," *Transp. Res. C, Emerg. Technol.*, vol. 124, Mar. 2021, Art. no. 102951.

[65] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[66] M. Wang et al., "Deep graph library: A graph-centric, highly-performant package for graph neural networks," 2019, *arXiv:1909.01315*.

[67] T. Chen et al., "MXNet: A flexible and efficient machine learning library for heterogeneous distributed systems," 2015, *arXiv:1512.01274*.

[68] J. Zhang, X. Shi, J. Xie, H. Ma, I. King, and D. Yeung, "GaAN: Gated attention networks for learning on large and spatiotemporal graphs," in *Proc. Conf. Uncertain. Artif. Intell. (UAI)*, Mar. 2018, pp. 339–349.

[69] P. V. Giampouras, A. A. Rontogiannis, and K. D. Koutroumbas, "Alternating iteratively reweighted least squares minimization for low-rank matrix factorization," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 490–503, Jan. 2019.

[70] X. Chen, C. Zhang, X.-L. Zhao, N. Saunier, and L. Sun, "Non-stationary temporal matrix factorization for multivariate time series forecasting," 2022, pp. 1–12, *arXiv:2203.10651*. [Online]. Available: http://arxiv.org/abs/2203.10651

[71] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial–temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, Jul. 2019, vol. 33, no. 1, pp. 922–929.

[72] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial–temporal synchronous graph convolutional networks: A new framework for spatial–temporal network data forecasting," in *Proc. 34th AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 1, pp. 914–921.

[73] X. Chen and L. Sun, "Bayesian temporal factorization for multidimensional time series prediction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4659–4673, Sep. 2022.
[74] A. M. Avila and I. Mezic, "Data-driven analysis and forecasting of highway traffic dynamics," *Nature Commun.*, vol. 11, no. 1, pp. 1–16, Apr. 2020.
[75] L. Lacasa, V. Nicosia, and V. Latora, "Network structure of multivariate time series," *Sci. Rep.*, vol. 5, no. 1, p. 15508, Dec. 2015.
[76] D. Eroglu, N. Marwan, M. Stebich, and J. Kurths, "Multiplex recurrence networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 97, no. 1, pp. 1–9, Jan. 2018.
[77] Y. Zou and Y. Zhang, "A copula-based approach to accommodate the dependence among microscopic traffic variables," *Transp. Res. C, Emerg. Technol.*, vol. 70, pp. 53–68, Sep. 2016.

**Jie Zeng** received the B.E. degree from the School of Traffic and Transportation Engineering, Central South University, Changsha, China, in 2021, where he is currently pursuing the master's degree. His current research interests include traffic flow prediction, spatiotemporal data mining, and intelligent transportation systems.

**Jinjun Tang** received the Ph.D. degree in transportation engineering from the Harbin Institute of Technology, Harbin, China, in 2016. From 2014 to 2016, he was a Visiting Scholar at the Smart Transportation Applications and Research Laboratory (STAR Lab), University of Washington, Seattle, WA, USA. He is currently an Associate Professor with the School of Traffic and Transportation Engineering, Central South University, Changsha, China. He has published more than 50 technical articles in the journal as the first author and corresponding coauthor. His research interests include traffic flow prediction, data mining in the transportation systems, intelligent transportation systems, and transportation modeling.