Received: 23 February 2023 | Accepted: 22 November 2023

DOI: 10.1111/mice.13131

RESEARCH ARTICLE

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

A deep marked graph process model for citywide traffic congestion forecasting

Tong Zhang[1] | **Jianlong Wang**[2] | **Tong Wang**[1] | **Yiwei Pang**[3] | **Peixiao Wang**[4] | **Wangshu Wang**[5]

[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

[2]Changjiang Space Information Technology Engineering Co., Ltd. (Wuhan), Wuhan, China

[3]School of Geodesy and Geomatics, Wuhan University, Wuhan, China

[4]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

[5]Research Unit Cartography, Department of Geodesy and Geoinformation, TU Wien, Vienna, Austria

Correspondence
Tong Zhang, 129 Luoyu Rd., LIESMARS, Wuhan University, Wuhan 430079, China.
Email: zhangt@whu.edu.cn

Wangshu Wang, Department of Geodesy and Geoinformation, TU Wien, Vienna 1040, Austria.
Email: wangshu.wang@tuwien.ac.at

Funding information
National Key R&D Program of China (International Scientific & Technological Cooperation Program), Grant/Award Number: 2019YFE0106500; National Natural Science Foundation of China, Grant/Award Number: 42371470; China National Postdoctoral Support Program for Innovative Talents, Grant/Award Number: BX20230360; Undergraduate Training Programs for Innovation and Entrepreneurship of Wuhan University

Abstract
Forecasting citywide traffic congestion on large road networks has long been a nontrivial research problem due to the challenge of modeling complex evolution patterns of congestion in highly stochastic traffic environments. Arguing that purely data-driven methods may not perform well for congestion forecasting, we propose a deep marked graph process model for predicting the congestion indices and the occurrence time of traffic congestion events for complex signalized road networks. Traffic congestion is considered as a nonrigorous spatiotemporal extreme event. We extend the traditional point process model by integrating a specially designed spatiotemporal graph convolutional network. This hybrid strategy takes advantage of the simple form of the point process model as well as the ability of graph neural networks to emulate the evolution of congestion. Experiments on real-world congestion data sets show that the proposed method outperforms state-of-the-art baseline methods, yielding satisfactory prediction results on a large signalized road network with superior computational efficiency.

1 | INTRODUCTION

Traffic congestion is increasing in major cities around the world, leading to excessive air pollution (Lelieveld et al., 2015) and negative economic impacts (Sweet, 2014). As transportation management authorities seek to alleviate traffic congestion, there is an urgent need to predict the occurrence and propagation of traffic congestion using traffic surveillance data collected by various sensors. Reliable and robust congestion prediction models can provide practical benefits for traffic signal control, traffic flow guidance, as well as road infrastructure performance evaluation and improvement planning (Han et al., 2023; Yao et al., 2023). They are also beneficial for local residents to make informed travel decisions.

In recent years, research on the prediction of future traffic states has proliferated due to the powerful capabilities of modern machine learning methods in capturing complex spatiotemporal traffic dynamics and dependencies (Tedjopurnomo et al., 2022; Wang et al., 2022). Most of these studies aim to predict specific traffic parameters, such as traffic flow, speed, or travel time. As variations in traffic parameters tend to exhibit regular spatial and temporal patterns, the evolution of congestion on signalized urban road networks is highly volatile, making congestion forecasting a challenging task.

This study investigates the problem of short-term traffic congestion prediction for large signalized road networks. Although some attempts have been made (Kumar & Raubal, 2021; Gal-Tzur et al., 2022; Mehdi et al., 2022), significant challenges still remain:

1) Because the traffic environment of signalized road networks is highly stochastic, there are complex factors that cause traffic congestion, including the spatiotemporal variation of traffic demand, the capacity of the roads, the effects of signal control, sudden weather changes, the occurrence of accidents, and road maintenance. It is difficult to build a comprehensive prediction model that explicitly accounts for these factors and the interactions between these possible factors.
2) Traffic congestion is a complex spatiotemporal process with constantly changing spatial coverage and temporal extent. Even at the same location, different congestion events occurring at different times may have different evolutionary patterns with different durations due to spillover effects. It remains a challenge to predict the fine-grained spatiotemporal extent of congestion events over their duration.
3) Predicting citywide congestion poses profound challenges. Traffic congestion propagates across connected road segments and interact with nonlocal traffic, leading to possible cascading failures in the road network. Since the propagation and dissipation of congestion is anisotropic, the evolution of traffic congestion is heterogeneous in space and time.

Data-driven deep learning–based methods have been extensively employed for short-term estimation or prediction of traffic parameters. It seems feasible to apply these established models to predict traffic congestion. Although many existing data-driven models can discover spatiotemporal traffic correlations after being trained on sufficient data samples, traffic congestion does not necessarily recur in the same place and time. Therefore, these vanilla models are difficult to generalize to the entire road network, which manifests heterogeneous congestion development patterns in space and time (Jiang et al., 2023; Zheng et al., 2023). Machine learning–based methods typically require a large number of samples to address this heterogeneity issue, as local road environments (network topology, road layout, time, weather, and traffic flow characteristics) need to be accounted for. On the other hand, complex machine learning models present a significant computational challenge in both training and inference (Pan et al., 2023), as short-term congestion forecasting is a time-critical application.

In light of the above issues, we propose a deep marked graph process (DMGP) model for predicting citywide traffic congestion events on complex signalized road networks. Traffic congestion is considered as a nonrigorous spatiotemporal extreme event and an event-oriented modeling method is developed. By treating congestion as a spatiotemporal event, the event-oriented modeling method can better describe the propagation and dissipation of congestion events throughout its entire life cycle, whereas many previous studies have neglected the evolutionary patterns within individual congestion events. The evolution of congestion events is modeled as spatiotemporal processes occurring in road networks. Instead of using a purely data-driven deep learning model, we adopt a hybrid strategy that takes the simple form of the point process model, in order to avoid including excessive influencing factors in the modeling of congestion evolution. Therefore, all congestion events can be modeled as marked spatiotemporal processes, taking advantage of their general functional forms to improve model robustness, generalizability, and interpretability, all of which are crucial for citywide congestion forecasting, as it is difficult to generalize to a large road network with a limited number of congestion samples.

The proposed approach extends the point process model with the ability to emulate the evolution of congestion using a congestion embedding network that learns expressive spatiotemporal embeddings of historical congestion events with specially designed spatial and temporal graph convolutions, respectively. These embeddings encode the spatial and temporal evolutionary patterns of congestion, which are then integrated with the conditional intensity function of marked spatiotemporal processes to predict spatiotemporal extent of congestion events and their severities. The integration of the congestion embedding network with the marked spatiotemporal process model significantly improves generalizability and accelerates learning efficiency through the use of two key inductive biases, namely, sharing intensity function parameters for citywide forecasting tasks and designing unified congestion embedding framework for the structured representation of congestion events. The contributions of this work can be summarized as follows:

1) A DMGP model is developed for citywide traffic congestion forecasting, being capable of predicting the congestion severity and occurrence time of each congestion event in an end-to-end deep learning framework.
2) A pattern-aware spatiotemporal graph embedding network is developed to generate spatiotemporal embeddings of historical congestion events based on traffic condition and traffic congestion graphs.
3) The traditional one-dimensional temporal point process model is extended to a deep marked spatiotemporal graph process model by integrating the learned congestion embeddings with the conditional intensity function of marked spatiotemporal processes, accounting for stochastic and heterogeneous evolutionary patterns of traffic congestion in urban road networks.

2 | LITERATURE REVIEW AND RELATED WORK

2.1 | Congestion prediction

Earlier studies on congestion prediction attempt to model congestion events using well-established mathematical or microsimulation models (Adeli & Ghosh-Dastidar, 2004; Ghosh-Dastidar & Adeli, 2006; Zhao et al., 2005). However, they usually fail to capture citywide congestion dynamics because strong statistical or mathematical assumptions do not always hold in complex urban traffic environments on a large road network (Ma et al., 2015). Due to the wide availability of traffic surveillance data and the popularity of machine learning algorithms, various data-driven methods have been proposed to predict the occurrence and propagation of traffic congestion (Akhtar & Moridpour, 2021). Several attempts have been made to adapt off-the-shelf deep learning models to congestion prediction tasks (Kumar & Raubal, 2021), such as stacked long short-term memory model (Chen et al., 2016; Mohanty et al., 2020; Yu et al., 2017), deep convolutional neural network (Chen et al., 2018), deep autoencoder (Zhang, Yao, et al., 2019), or the integration of recurrent neural network and other deep learning models (Guo et al., 2021; Ma et al., 2015). While these deep learning models can capture the traffic correlations in road networks to some extent, they are not specifically designed to account for the propagation patterns of traffic congestion and the sparseness of congestion data, thereby yielding inferior prediction results when traffic dynamics are highly stochastic.

Recently, a few studies have explicitly incorporated traffic patterns into deep learning models to developed pattern-aware spatiotemporal prediction models (Di et al., 2019; Leiser & Yildirimoglu, 2021; Zheng et al., 2023). While many previous studies did not explicitly model the evolution of individual congestion events and only predicted congestion based on regular time intervals (Kumar & Raubal, 2021), some researchers have considered congestion as spatiotemporally propagating events and have attempted to perform fine-grained congestion forecasting using graph embedding (Sun et al., 2022; Wang et al., 2023) or point process models (Zhu et al., 2022).

2.2 | Neural point process models

Temporal point processes provide a powerful mathematical framework for modeling time-dependent event data based on parametric intensity functions (Daley & Vere-Jones, 2003). Given prior knowledge of a random process, parametric assumptions can be made for a conditional intensity function to describe the probability of an event occurring in the near future given a sequence of historical events. The temporal point process model can naturally predict the occurrence rates of events over time and has therefore been widely used to model and predict rare events such as earthquakes (Ogata, 1998), war (Zammit-Mangion et al., 2012), and crime (Adepeju et al., 2016). The drawback of classical temporal point process models is that their strict parametric assumptions do not hold for some event generative processes. Recently, the integration of point process models and deep learning has experienced a surge of interest, resulting in a flurry of work on neural temporal point processes (NTPPs) (Du et al., 2016; Shchur et al., 2021). By taking the advantage of deep learning to capture complex hidden temporal propagation patterns, NTPPs outperform classical point process models in predictive ability and expressiveness when prior knowledge is scarce. To better capture complex unseen temporal patterns, Bae et al. (2023) propose to formulate temporal point processes as attentive neural processes based on a meta-learning framework. NTPPs can be further extended as neural spatiotemporal point processes, which can use deep neural networks (Zhou et al., 2022; Zhu et al., 2020) or neural ordinary differential equations (Chen et al., 2021) to parameterize the joint event distribution in both space and time. Similar to our work, Zhu et al. (2022) use spatiotemporal point processes to predict congestion events. However, their study was conducted on highway data collected from a limited number of sensors whereas this study aims to predict citywide congestion for a large urban signalized road network where traffic dynamics are much more complex and stochastic.

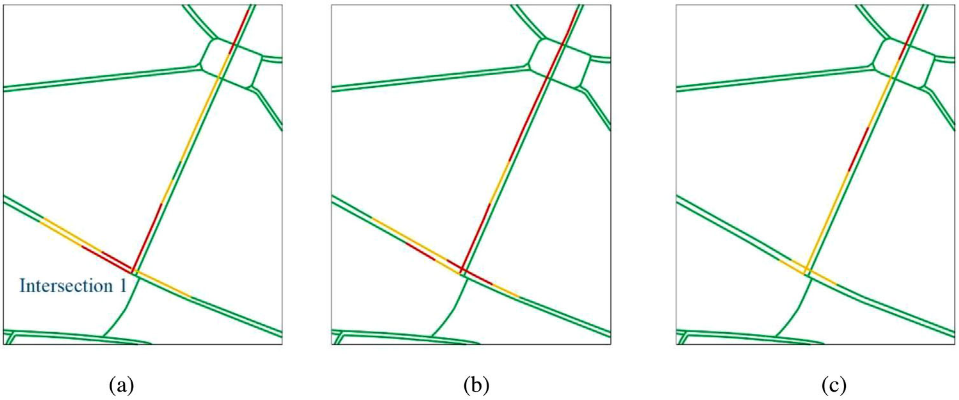

FIGURE 1 Propagation of traffic congestion on an urban road network.

2.3 | Traffic prediction using deep neural networks

Short-term prediction of various traffic parameters (flow, speed, travel time etc.) has long been a well-researched topic in the transportation and computer science communities (Vlahogianni et al., 2014). The use of neural networks for traffic pattern analysis and short-term traffic prediction has a substantial literature (Jiang & Adeli, 2004; Smith & Demetsky, 1994; Yun et al., 1998). Over the past decade, the application of deep learning technique in traffic prediction has gained momentum. Earlier studies directly applied off-the-shelf deep learning models to predict specific traffic parameters, such as stacked autoencoders (Lv et al., 2015) and convolutional neural networks (Ma et al., 2017). To capture spatiotemporal correlations hidden in traffic dynamics, hybrid models have been developed that integrate multiple deep learning modules (Yin et al., 2022), such as deep residual learning (Zhang et al., 2017), hybrid recurrent convolutional networks (Ke et al., 2017; Lin et al., 2020), attention convolutional neural networks (Liu et al., 2018), and deep meta-learning (Pan et al., 2019). Efficient ensemble deep learning models have been proposed to predict citywide traffic parameters, yielding robust and accurate prediction results (Liu et al., 2020; Zhang, Zhou, et al., 2020).

Since transportation networks can be intuitively modeled as graphs, graph neural networks have been widely used for citywide traffic forecasting in recent years (Jiang & Luo, 2022; Rahmani et al., 2023). Intensive efforts have been made to capture the spatiotemporal dependency or correlation between road intersections or segments with specially designed graph convolutional networks (GCNs), such as, spatiotemporal GCN (Yu et al., 2018), spatial–temporal graph inception residual network (Zhang, Cheng, et al., 2019), temporal GCN (Zhao et al., 2020), spatiotemporal gated graph attention network (Tang & Zeng, 2022), time-aware multipersistence spatiosupra GCN (Chen et al., 2022), and graph neural rough differential equations (Choi & Park, 2023). We argue that predicting traffic congestion is more challenging than predicting regular traffic parameters because the latter can exploit the regularities of traffic dynamics whereas the occurrence and evolution of congestion are more volatile. In addition, traffic congestion is a relatively rare extreme event that provides much fewer training samples than traffic parameter data (e.g., speed or flow), making it difficult for GCN models to generalize. Therefore, the GCN models used in traffic parameter prediction may not perform well in traffic congestion prediction because congestion should be considered as a complete event to model its propagation and dissipation patterns (Zhu et al., 2022). Special attention should be paid to these issues.

3 | METHODOLOGY

3.1 | Motivation and definitions

The propagation of traffic congestion in an urban road network is constrained by signalized intersections, exhibiting strong anisotropic evolution patterns in different directions. When modeling citywide traffic congestion, we need to consider nonlocal spatiotemporal traffic correlations because congestion usually propagates further into the road network and interacts with traffic in upstream road links.

As shown in Figure 1, a congestion event is constrained by traffic from surrounding intersections or adjacent saturated road segments and evolves between signalized intersections, exhibiting complex evolution patterns. In Figure 1a, it is shown that a congestion event occurs at intersection 1 at 18:20, while some road links in the northeast direction also start to experience congestion.

Figure 1b shows that the congestion on the northeastern road segments becomes more severe, leading to a further increase in congestion that develops into a gridlock state on these links at 18:25. Figure 1c shows that at 18:30, as the congestion at intersection 1 eases, the gridlock state on the northeastern road segments begins to dissipate and slow-moving traffic can be observed. The propagation and evolution of a congestion event throughout its entire life cycle is a complex, nonrigorous spatiotemporal dynamic process that interacts with the changing traffic dynamics in the neighborhood and sometimes with other congestion events. These interactions give rise to complex spatiotemporal correlations between congestion events and the surrounding road environment, making accurate predictions difficult. This observation inspires us to develop a spatiotemporal process model to account for the anisotropy of congestion propagation in space and time.

Definition 1. Traffic Congestion Event. A traffic congestion event ce describes the occurrence and persistence of congestion states in certain road segments in a road network $\mathcal{G}$, whose nodes $\boldsymbol{N}$ and edges $\boldsymbol{E}$ correspond to intersections and road segments, respectively. A marked congestion event can be formally defined as $ce_e = (b_e,\ t_e,\ m_e)$, where b_e is a Boolean value indicating whether the road segment e is congested or not. t_e is the occurrence time of the congestion event, and $m_e \in \mathbb{M}$ ($\mathbb{M} \in \mathbb{R}^d$ represents the domain of marks) is the marker vector recording other semantic information of the event, such as the duration or congestion severity index. Then, traffic congestion events can be modeled as marked spatial–temporal processes on $\mathcal{G} \times T$ (let $T > 0$ be a time horizon).

Definition 2. Traffic Congestion Event Sequence. Based on the above definition, the sequence of historical traffic congestion events on road segment e can be denoted as $\mathcal{H}_t\ (e) = \ \{b_{t'}, t', m_{t'} | t' < t\}$, which is a realization of a marked spatiotemporal process. Note that events in a sequence may have variable lengths of duration.

The distribution of traffic congestion events on road segment e can be characterized by the conditional intensity function,

$$\lambda^*[t|\mathcal{H}_t\,(e)] \ = \lim_{\Delta t \to 0} \frac{\boldsymbol{Pr}\{t_e \in [t,\ t + \Delta t]\,|\mathcal{H}_t\,(e)\}}{\Delta t} \quad (1)$$

The conditional intensity function describes the expected probability of the occurrence of a congestion event on road segment e at time t, conditional on a given historical event sequence $\mathcal{H}_t(e)$.

Definition 3. Traffic Condition Graph. A traffic condition graph $(\mathcal{G}_{\boldsymbol{d}})_{t \in \{1,\,2,\,...,\,T\}} = (\boldsymbol{V_d},\ \boldsymbol{E_d})_t$ is constructed based on observed traffic surveillance data at time t to encode traffic features on the road network. A node $v_d \in \boldsymbol{V_d}$ ($|\boldsymbol{V_d}| \ = \ N$) corresponds to the road segment between two signalized intersections. An edge $e_d \in \boldsymbol{E}_d$ describes the connectivity between nodes. $(\boldsymbol{A_d})_t \in \mathbb{R}^{N \times N}$ is the adjacency matrix of $(\mathcal{G}_{\boldsymbol{d}})_t$. We use $\boldsymbol{x}_t(v_d) \in \mathbb{R}^F$ to denote the F-dimensional traffic feature vector of node v_d at time t (e.g., traffic flow and average speed). The traffic condition of the entire road network at time t can be represented as $\boldsymbol{X}_t\ (\mathcal{G}_{\boldsymbol{d}}) = [\boldsymbol{x}_t(v_1), \boldsymbol{x}_t(v_2), \dots, \boldsymbol{x}_t(v_N)] \in \mathbb{R}^{N \times F}$.

Definition 4. Traffic Congestion Graph. A traffic congestion graph $(\mathcal{G}_{\boldsymbol{g}})_{t \in \{1,\,2,\,...,\,T\}} = (\boldsymbol{V_g},\ \boldsymbol{E_g})_{t \in \{1,\,2,\,...,\,T\}}$ is constructed to represent the traffic congestion events that occur in the studied road network. At time t, the node set $(\boldsymbol{V_g})_t$ contains the road segments that are being congested. $(\boldsymbol{E_g})_t$ is the set of edges describing the potential connectivity of the congested road segments at time t. If the distance between two congested road segments is less than a threshold, we consider them to be "connected" in the traffic congestion graph. Unlike the traffic condition graph, the adjacency matrix of the traffic congestion graph $\boldsymbol{A_g}(t)$ is not static because it is used to represent time-varying correlations between congested nodes. Each element of $\boldsymbol{A_g}(t)$ represents the strength of such a correlation, which is learnable. Note that i and j are not necessarily directly connected. In other words, if i and j are associated with two different congestion events, each element of $a_g(i, j)$ quantifies the correlation between the two congestion events. A feature matrix $\boldsymbol{X_t}(\mathcal{G}_{\boldsymbol{g}}) \in \mathbb{R}^{d+1}$ describes the associated features of each node in $(\mathcal{G}_{\boldsymbol{g}})_t$, including the occurrence time and the marked values. Overall, the traffic congestion graph represents a subgraph of the traffic condition graph while focusing on the representation of congestion events and their evolution.

The predicting problem for the citywide traffic congestion event sequence $\boldsymbol{CE}$ can be defined as follows:

$$\boldsymbol{CE}_{t \to\ t+T}E = \varphi\left(\cup_{e=1}^{|E\}}\,[b_e\,(t)\,, m_e\,(t)]\,\middle|\,\mathcal{G}_{\boldsymbol{d}}, \mathcal{G}_{\boldsymbol{g}}, \cup_{e=1}^{|E\}}\mathcal{H}_t\,(e)\right] \quad (2)$$

where $t{+}T$ is the prediction horizon, and $\cup_{e\,=\,1}^{|E\}}\mathcal{H}_t(e)$ represents the entire historical event sequence for all road segments during the time interval $[0,\ t]$. The prediction task deals with a multitask learning problem, that is, we simultaneously predict the occurrence of congestion (i.e., b_e) and the marked values of congestion (i.e., m_e) for each segment e on the road network $\mathcal{G}$, based on two virtual graphs, $\mathcal{G}_{\boldsymbol{d}}$ and $\mathcal{G}_{\boldsymbol{g}}$.

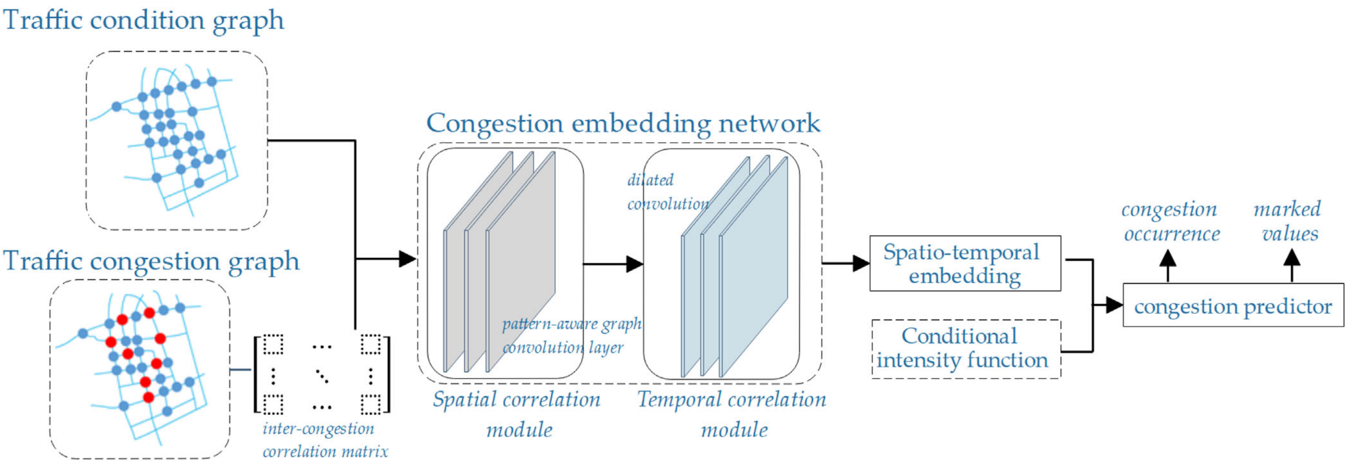

FIGURE 2 Workflow of the proposed deep marked graph process (DMGP) model.

3.2 | Methodology overview

A novel spatiotemporal graph process model is proposed for citywide traffic congestion prediction. The proposed DMGP is capable of predicting citywide traffic congestion events on complex signalized road networks. The DMGP model is based on the classical spatiotemporal point process model and learns the stochastic evolution and interaction patterns of congestion events using a specially designed spatiotemporal graph embedding network that learns expressive spatiotemporal embeddings of historical congestion events. The embeddings are integrated with the conditional intensity function of marked spatiotemporal processes for predicting congestion across the entire road network.

The proposed model is schematically illustrated in Figure 2. The DMGP model consists of three components: (1) learning the spatiotemporal embedding of traffic congestion events through an integrated graph embedding network, (2) constructing the conditional intensity function for citywide congestion events by incorporating the learned embeddings, and (3) training the DMGP model (or making predictions when performing inference).

The first component is comprised of two steps:

1) Learning the spatial correlations between road segments where congestion occurs. Based on the traffic condition graph and the traffic congestion graph, we develop a graph embedding network to capture spatial correlations between upstream and downstream road segments. Considering the scarcity of congestion event samples, we develop a novel pattern-aware graph propagation convolutional kernel capable of capturing the propagation and evolutionary patterns of congestion across the road network. The kernel learns the structural dependencies of the road network using the traffic condition graph and captures the congestion evolution patterns using the traffic congestion graph, respectively. For each node of the traffic condition graph $\mathcal{G}_d$, the spatial correlation module generates a spatial embedding vector;
2) Learning the temporal correlations of traffic congestion events. Using the spatial embeddings learned in the previous step as input, this component continues to learn the temporal dependencies between the nodes of the traffic condition graph $\mathcal{G}_d$. Based on a temporal convolutional network (TCN), the temporal module generates the spatiotemporal embeddings of historical congestion events, encoding complex spatiotemporal propagation patterns of historical congestion events.

The conditional intensity function (Equation 1) is extended to accommodate the learned spatiotemporal embeddings, thereby integrating learned graph structure information into the spatiotemporal point process model. This integration can empower the vanilla spatiotemporal point process model with rich knowledge of temporal dependencies and spatial correlations within and between congestion events. Then, according to Equation (2), the modified conditional intensity function can be used to predict whether congestion events will occur or not, and to predict the marker values of all nodes in the traffic condition graph in an autoregressive manner. With the two graphs, the proposed model not only captures local congestion evolution patterns but also describes the non-local interactions between different congestion events. The details of the spatial embedding, temporal embedding, and modified conditional intensity function are described as follows.

3.3 | Embedding spatial correlations using pattern-aware graph convolution kernels

The first step of the proposed method is to learn the spatial correlations between road segments when congestion occurs. Graph convolutions are performed to embed spatial correlations in the traffic condition graph. The spatial embeddings are then fed to the temporal embedding module to generate spatiotemporal embeddings of traffic congestion events. The spatial embedding module is built based on a GCN, which performs global convolution on the entire traffic condition graph instead of local convolution on neighboring nodes, because many congestion events can propagate far away along the road network.

Specifically, a two-layer GCN (Kipf and Welling, 2017) at time t that embeds spatial correlations via two convolutional operations can be defined as,

$$\boldsymbol{H}_t(1) = \boldsymbol{A_d}\boldsymbol{X}_t(\mathcal{G}_{\boldsymbol{d}})\boldsymbol{W}^{(0)} \tag{3}$$

$$\boldsymbol{H}_t(l) = \sigma\left[\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}_d\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}_t(l-1)\boldsymbol{W}^{(l-1)}\right] \tag{4}$$

where $\boldsymbol{H}_t(\boldsymbol{l}) \in \mathbb{R}^{N\times z}$ is the spatial embedding of the lth layer. $\boldsymbol{A_d}$ is the adjacency matrix of the traffic condition graph. $\tilde{\boldsymbol{A}}_d = \boldsymbol{A_d} + \boldsymbol{I}_N$ ($\boldsymbol{I}_N$ is the identity matrix). $\boldsymbol{X_t}(\mathcal{G}_{\boldsymbol{d}}) \in \mathbb{R}^{N\times F}$ denotes the traffic feature matrix at time t. $\boldsymbol{W}^{(0)}$ and $\boldsymbol{W}^{(l-1)}$ are the learnable weights. $(\bullet)$ is an activation function. $\tilde{\boldsymbol{D}}$ is the degree matrix of $\mathcal{G}_{\boldsymbol{d}}$.

With these graph convolution layers, we can incorporate topological priors into the embeddings of congestion events using the static adjacency matrix $\boldsymbol{A_d}$. However, the evolution of congestion is not only constrained by the topological structure of the road network but also by the propagation patterns of congestion. Inspired by recent pattern-aware event prediction methods (Du et al., 2021; Zhang & Cheng, 2020; Zhu et al., 2020), the proposed model explicitly integrates spatiotemporal propagation patterns into the embeddings of congestion events based on the traffic congestion graph.

Let the spatiotemporal propagation pattern between two congestion events located at nodes i and j be $\rho(i,j) \in [0,1]$, which measures the spatiotemporal dependency between the two congestion events. It is assumed that two congestion events can influence each other if their occurrence times are close and their temporal variations of congestion levels are similar. A congestion event may exert an influence on neighboring upstream road segments and may also affect on some more distant road segments. Therefore, there is no constraint on the spatial dimension when computing intercongestion dependencies. At time t, $\rho_t(i,j)$ can be computed as

$$\rho_t(i,j) = \begin{cases} exp[w_t\left(1-\frac{\Delta t_{i,j}}{\Delta t}\right) + w_m\left(1-\frac{\Delta m_{i,j}}{\Delta m}\right) & \\ 0 \qquad otherwise & \end{cases}$$

$$\text{if } \Delta t_{i,j} < \Delta t \& \Delta m_{i,j} < \Delta m \tag{5}$$

where $w_t \in [0,1]$ and $w_m \in [0,1]$ are learnable weighting parameters used to modulate the effects of time and congestion severity ($w_t + w_m = 1$). Note that w_t and w_m vary with time and pairs of road segments. Δt and Δm are the time and marked value (i.e., congestion severity) thresholds used to determine if two congestions are mutually correlated or not. $\Delta t_{i,j}$ and $\Delta m_{i,j}$ are the differences in the occurrence time and congestion severity between two congestion events i and j. According to Equation (5), large (small) values of $\rho_t(i,j)$ indicate strong (weak) dependencies between i and j.

Based on the definition of the intercongestion dependency $\rho_t(i,j)$, we can construct the time-varying dynamic correlation matrix $\boldsymbol{A_g}(t)$ of the traffic congestion graph using the intercongestion dependency (i.e., $\rho(i,j)$) as the elements of $\boldsymbol{A_g}(t)$. Then the spatiotemporal propagation patterns of all traffic congestion events can be represented by $\boldsymbol{A_g}(t)$. By integrating the time-varying correlation matrix $\boldsymbol{A_g}(t)$, Equation (4) can be rewritten as a pattern-aware graph convolutional layer,

$$\boldsymbol{H}_t(l) = \sigma\left\{[\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}_d\tilde{\boldsymbol{D}}^{-\frac{1}{2}} \odot \boldsymbol{A_g}(t)]\boldsymbol{H}_t(l-1)\boldsymbol{W}^{(l-1)}\right\} \tag{6}$$

where $\odot$ is the Hadamard product. $[\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}_d\tilde{\boldsymbol{D}}^{-\frac{1}{2}} \odot \boldsymbol{A_g}(t)]$ can be considered as a graph convolution kernel that characterizes the propagation of congestion. By integrating correlations between congestion events through graph convolution, the spatial embeddings encode both static topological constraints and dynamic dependencies between congestion events, thereby capturing high-level evolutionary features and improving the generalizability of the prediction model.

3.4 | Capturing sequence correlations using temporal convolution

After obtaining the spatial embeddings from the traffic condition graph, we use temporal convolution to capture the sequence correlations for each node in the traffic congestion graph, as shown in Figure 3. Typical recurrent neural networks are not used due to the gradient

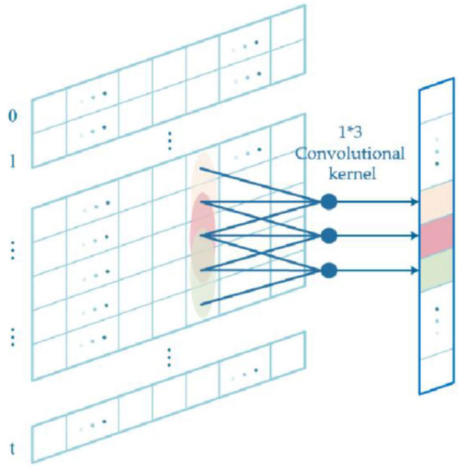

FIGURE 3 Temporal convolution on spatial embeddings.

vanishing/exploding issue when modeling long sequences of congestion events. The temporal embedding is performed based on the architecture of the TCN (Bai et al., 2018). Dilated convolutions are used to extract temporal correlations from a sequence of congestion events at the same node. To facilitate the modeling of frequent congestion that may last for a considerable period of time, we retain the residual blocks in the TCN to enable the construction of a very deep structure.

For each graph node e in $\mathcal{G}$ (i.e., road segment), the dilated convolution takes the spatial embeddings of the previous t historical time steps as input and learns comprehensive spatiotemporal embeddings,

$$\boldsymbol{Z}_{t,e}(l) = Relu\{K(t-1)_{*T}\left[\boldsymbol{H}_{t,e}(l) : ce_{e|0\to t}\right] \tag{7}$$

where $*_T$ denotes the dilated convolution, *Relu* is an activation function, and $K(t{-}1)$ denotes the temporal convolution kernel. $ce_{e\ |0\to t}$ denotes the sequence of event vectors (b_e, t_e, m_e) from time 0 to t. $\boldsymbol{H}_{t,e}(l)$ is the spatial embedding vector for node e at time t after l layers of spatial graph convolutions. [:] represents the concatenation operation.

The temporal convolutions are performed only locally, making it possible for multiple nodes to compute in parallel at the same time. Parallel computing helps to promote the efficiency of training and inference for congestion prediction in large road networks.

3.5 | Integrating spatiotemporal embeddings into the conditional intensity function

After embedding spatial correlations and sequence correlations for each node in the traffic condition graph, we integrate the embeddings with the conditional intensity function to enhance the modeling capability of the vanilla spatiotemporal point process model, which needs to account for the cumulative influences of historical congestion events at the same node as well as congestion occurring at other nodes (e.g., downstream nodes). For each node, its sequence of congestion events is modeled as a marked spatiotemporal point process. Then, the conditional intensity function for each node at time t+1 can be written as

$$\lambda_e^*(b_{(t+1)}, \boldsymbol{m}_{(t+1)}|\boldsymbol{Z}_t) = \exp\{\boldsymbol{W}^{bm}[b_t : \boldsymbol{m}_t] + \boldsymbol{W}^h\boldsymbol{H}_t + \boldsymbol{W}^z\boldsymbol{Z}_t + \boldsymbol{\lambda}_e^{\boldsymbol{0}}(t+1)\} \tag{8}$$

where $\boldsymbol{W}^{bm}, \boldsymbol{W}^h, \boldsymbol{W}^z$ are learnable weighting parameters. b_{t+1} is a Boolean variable indicating whether congestion occurs at time t+1. m_t+1 is a marked value vector of predicted congestion events at time t+1. $\boldsymbol{W}^{bm}[b_t : \boldsymbol{m}_t]$ describes the influences of the events that occurs at time t. $\boldsymbol{H}_t$ encodes the spatial correlations between the studied nodes and other nodes. $\boldsymbol{Z}_t$ represents the cumulative influences of historical congestion events. Note that the influenced nodes may include themselves, neighboring nodes, and possible distant nodes. λ^0 is a constant background rate of congestion at time t+1. In the right-hand terms, e is omitted for brevity. The four weighting terms represent the influences of the current congestion event at e and the spatial correlations to other nodes, as well as the cumulative spatiotemporal influences of congestion events occurring in the vicinity of the considered node.

The future congestion events at each node (i.e., road segment) can be predicted, including the probability of congestion and the marked values.

$$P_{t+1}\left[b_{t+1} = 1|\lambda_e^*(b_{t+1}, \boldsymbol{m}_{t+1}|\boldsymbol{Z}_t)\right] = \frac{\exp[\boldsymbol{W}^{1t}\lambda_e^*(b_{t+1},\ \boldsymbol{m}_{t+1}|\boldsymbol{Z}_t) + \boldsymbol{B}^{1t}]}{\sum_{i=1}^{2}\exp[\boldsymbol{W}^{it}\lambda_e^*(b_{t+1},\ \boldsymbol{m}_{t+1}|\boldsymbol{Z}_t) + \boldsymbol{B}^{it}]} \tag{9}$$

$$\boldsymbol{m}_{t+1} = \boldsymbol{W}^{mt}\ \lambda_e^*(b_{t+1},, \boldsymbol{m}_{t+1}|\boldsymbol{Z}_t) + \boldsymbol{B}^{mt} \tag{10}$$

where $\boldsymbol{W}^{mt}$, $\boldsymbol{W}^{1t}$, and $\boldsymbol{W}^{it}$ are learnable weighting parameters and $\boldsymbol{B}^{1t}$, $\boldsymbol{B}^{it}$, and $\boldsymbol{B}^{mt}$ are bias terms.

3.6 | Model training and inference

Let $\mathcal{L}(\theta_{\boldsymbol{b}}, \theta_{\boldsymbol{m}})$ be the total loss, which is defined as a joint loss of two subtasks:

$$\mathcal{L}(\theta_{\boldsymbol{b}}, \theta_{\boldsymbol{m}}) = \min_{\theta_b,\theta_m}\{\underbrace{logP_{t+1}\left[b_{t+1} = 1\left|\lambda_e^*(b_{t+1},\ \boldsymbol{m}_{t+1}|\boldsymbol{Z}_t)\right.;\theta_{\boldsymbol{b}}\right]}_{\text{predicting congestion occurrence}} + \underbrace{\beta\log f_m\left[\boldsymbol{m}_{t+1}\left|\lambda_e^*(b_{t+1},\ \boldsymbol{m}_{t+1}|\boldsymbol{Z}_t)\right.;\theta_{\boldsymbol{m}}\right]\}}_{\text{predicting marked values}} \tag{11}$$

where β is the weight that regulates the contribution of the two subtasks. $f_m(.)$ is used to compute the cross entropy loss of the marked value predictions. θb and θm denote the learnable parameters of the two subtasks, that is, predicting congestion occurrence and predicting marked values of congestion events, respectively. These parameters are specified in Equations (9) and (10).

The pseudocode for the training algorithm is as follows:

ALGORITHM 1 Model training

Inputs: Traffic condition graph $\mathcal{G}_d$, traffic congestion graph $\mathcal{G}_g$, historical congestion events $\mathcal{H}_t(\boldsymbol{G_g})$, batch size b, number of iteration e, loss weight β

Outputs: optimized θ_b, θ_m

1. initialize θ_b, θ_m
2. shuffle training samples
3. **for** $i \leftarrow 1 \cdots e$:
4. fetch b samples from the training dataset
5. $\boldsymbol{A_g}(t) \leftarrow \{\mathcal{G}_g(t), \mathcal{G}_d(t)\}$ // compute inter-congestion correlation matrix $\boldsymbol{A_g}(t)$ using Equation (5)
6. $\boldsymbol{H_t} \leftarrow \{\boldsymbol{A_g}(t), \mathcal{G}_d(t)\}$ //compute pattern-aware spatial embedding $\boldsymbol{H}_t$ using Equation (6)
7. $\boldsymbol{Z}_t \leftarrow \{\mathcal{G}_g(t), \mathcal{G}_d(t), \mathcal{H}_t(\boldsymbol{G_g}(t))\}$ // computer spatio-temporal embedding $\boldsymbol{Z}_t$ using Equation (7)
8. $\lambda_e^* \leftarrow \{\boldsymbol{H_t}, \boldsymbol{Z}_t\}$ //construct conditional intensity function using Equation (8)
9. $b_{t+1}, \boldsymbol{m}_{t+1} \leftarrow \{\lambda_e^*\}$ // predict congestion occurrence & marked values using Eqs.(9) and (10)
10. $\hat{\theta} = \text{arcmin}\ 1/b_{\theta_b,\theta_m} \sum_{i=1}^{b} (logP(\theta_b) + \beta log f_m(\theta_m))$ // compute loss
11. compute gradient $\nabla\hat{\theta}$
12. $\hat{\theta} \leftarrow \hat{\theta} + \nabla\hat{\theta}$ //update model
13. End

The core of the training procedure (Algorithm 1) is the construction of the conditional intensity functions for each road segment. All input data were standardized to a scale with a mean of 0 and a standard deviation of 1. The Apache Mxnet framework was used to implement the proposed algorithm. A two-layered GCN (Kipf & Welling, 2017) was used to embed spatial correlations. The dimensionality of both the spatial and the comprehensive spatiotemporal embedding $\boldsymbol{Z}_t$ were set to 64. The temporal convolution kernel has a size of 1×3. After graph and temporal convolutions, batch normalization was employed to accelerate the optimization. Sigmoid was used as the activation function. The loss was optimized for 100 iterations by Adam with a learning rate of 0.001 and a batch size of 32. Given historical data for the previous 10 time steps, the task is to predict the occurrence time and severity level of congestion for the entire road network of the studied city for the next five time steps. TCN takes the spatial embeddings of the previous 10 time steps at inputs and learns the comprehensive spatiotemporal embeddings. We performed only one temporal convolution to produce the comprehensive spatiotemporal embeddings of historical congestion events. A grid search was performed using the validation data set to find the best hyperparameter $\beta = 0.1$.

Once the model has been trained, it can be used to predict future congestion, including the occurrence and marked values of each congestion event in the road network. Figure 4 shows how the learned spatiotemporal embeddings of congestion events occurring before time t are integrated with the current embeddings at time t, and how the integrated embeddings are incorporated into the conditional intensity function to predict congestion at time t+1. Note that we separate the embedding of time t from those at time t+1 because it helps to reduce the inference time of the spatiotemporal convolutions and to emphasize the role played by the data of time t. The model autoregressively predicts future congestion after updating the traffic condition/congestion graphs and the historical event sequence.

4 | EXPERIMENTAL RESULTS

4.1 | Data

The study area is located in Xi'an City, China, which has a population of nine million. Figure 5 shows the city's arterial road network, which consists of 2957 road segments. The congestion index (the ratio of free flow speed over actual speed) and average speed data for each road segment were collected from AutoNavi, which is a leading web mapping and navigation company in China, with a granularity of 10 min for the entire year of 2018. Missing data were imputed based on a gated generative adversarial network model (Zhang et al., 2022), which jointly models the spatiotemporal correlations between road segments using a specially designed attentional mechanism and improves the consistency between the imputed data and the overall data distribution through two subdiscriminators. According to a recently released standard ("road congestion evaluation metrics") (TMRI & Tongji, 2020), a congestion event is identified based on the difference between the actual traveling speed and the limited speed. When the average traveling speed falls below the specified speed threshold, a congestion event occurs. Based on the actual speed, congestion events can be categorized into three levels: light, medium, and severe congestion. Congestion event data were collected for all road segments to build historical congestion event sequences. Traffic condition and congestion graphs were then constructed according to the definitions in Section 3.1. Speed data were used to construct traffic feature vectors for the graph nodes.

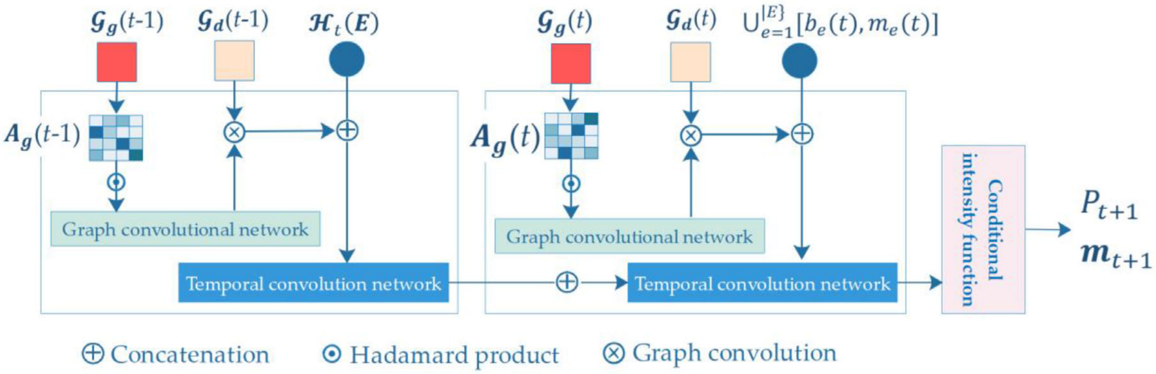

FIGURE 4 Autoregressive prediction of congestion.

FIGURE 5 Road network in Xi'an City. Road segments are rendered in four colors to indicate different levels of congestion (severe congestion, medium congestion, light congestion, and free flow).

All experiments were conducted on a desktop machine with a 3.7 GHz Intel® i7-9700K processor, 32 GB of memory, and an Nvida GeForce RTX®2080 Ti graphics card. The data set was split into training (60%), validation (10%), and test (30%) data sets.

4.2 | Baselines

Five baseline models were selected for comparison. One baseline is a classical time series model, two baselines are point process models, and the remaining two are typical spatiotemporal GCNs.

Details of these baselines are as follows:

1. autoregressive integrated moving average (ARIMA; Box and Jenkins, 1970), which is a classical time-series model;
2. self-attentive Hawkes process (SAHP; Zhang, Lipani, et al., 2020), which uses a self-attention mechanism to integrate historical event information into the conditional intensity function. The attention mechanism captures only temporal evolutionary information, while spatial correlations are not modeled;
3. transformer Hawkes process (THP; Zuo et al., 2020), which integrates the transformer with the Hawkes process to capture both long-term and short-term interevent dependencies;
4. attention-based spatial–temporal graph convolutional network (ASTGCN; Guo et al., 2019), which uses spatiotemporal attention and spatiotemporal graph convolutions to explicitly model the dynamics of traffic flows;
5. dynamic spatial–temporal aware graph neural network (DSTAGNN; Lan et al., 2022), which relies on dynamic spatial–temporal graphs to better capture complex spatial–temporal correlations in road networks through an enhanced multihead attention mechanism and multiscale gated convolution.

In addition to the above baselines, three variants of the proposed model were tested to evaluate the contribution of the different components.

1. DMGP(S): the temporal embedding component is removed. The purpose of testing this variant is to evaluate the benefits of sequence correlation modeling.
2. DMGP(T): the graph convolutional layer is removed. The testing of this variant is to evaluate the contribution of spatial correlation embedding in capturing heterogeneous propagation patterns of congestion;
3. DMGP(G): the time-varying correlation matrix $\boldsymbol{A}_g(t)$ is not included in Equation (6), which means that pattern-aware capability is not realized in the GCN.

TABLE 1 Performance comparison of predicted congestion severity index during weekday peak-hours.

Horizon	Metrics	ARIMA	DSTAGNN	SAHP	THP	ASTGCN	DMGP (S)	DMGP (T)	DMGP (G)	DMGP
1	MAE	0.27	0.25	0.31	0.29	0.27	0.28	0.24	0.21	0.16
	RMSE	0.68	0.53	0.64	0.61	0.56	0.59	0.53	0.45	0.35
	MAPE (%)	15.41	14.05	17.18	16.08	14.80	15.07	13.02	11.11	8.70
2	MAE	0.30	0.26	0.31	0.30	0.29	0.28	0.25	0.22	0.19
	RMSE	0.76	0.53	0.64	0.62	0.57	0.60	0.54	0.49	0.43
	MAPE (%)	16.99	14.43	16.90	16.06	16.07	15.15	13.08	11.38	10.52
3	MAE	0.32	0.26	0.33	0.30	0.29	0.29	0.25	0.23	0.22
	RMSE	0.81	0.54	0.66	0.63	0.57	0.61	0.55	0.51	0.47
	MAPE (%)	18.01	14.64	17.58	16.17	16.20	15.47	13.15	12.09	11.67
4	MAE	0.33	0.27	0.33	0.30	0.30	0.29	0.25	0.25	0.24
	RMSE	0.82	0.55	0.67	0.65	0.57	0.62	0.56	0.55	0.51
	MAPE (%)	18.31	14.82	17.59	16.18	16.24	15.36	13.19	12.98	12.59
5	MAE	0.34	0.28	0.33	0.31	0.29	0.29	0.26	0.25	0.23
	RMSE	0.83	0.56	0.67	0.66	0.57	0.64	0.57	0.54	0.53
	MAPE (%)	18.57	15.43	17.43	16.04	15.68	15.22	13.66	12.84	12.84

TABLE 2 Performance comparison of predicted congestion severity index during weekend peak-hours.

Horizon	Metric	ARIMA	DSTAGNN	SAHP	THP	ASTGCN	DMGP (S)	DMGP (T)	DMGP (G)	DMGP
1	MAE	0.30	0.25	0.32	0.31	0.28	0.30	0.25	0.19	0.15
	RMSE	0.81	0.53	0.67	0.64	0.56	0.62	0.54	0.41	0.33
	MAPE (%)	18.71	13.63	16.71	16.50	15.15	15.74	13.29	10.17	8.27
2	MAE	0.33	0.25	0.32	0.31	0.30	0.30	0.26	0.20	0.18
	RMSE	0.87	0.54	0.68	0.65	0.57	0.63	0.56	0.45	0.40
	MAPE (%)	20.06	13.56	16.63	16.30	16.26	15.66	13.42	10.53	9.84
3	MAE	0.35	0.25	0.33	0.32	0.30	0.31	0.26	0.22	0.20
	RMSE	0.92	0.54	0.70	0.66	0.57	0.64	0.57	0.49	0.44
	MAPE (%)	20.89	13.43	17.09	16.38	16.34	15.92	13.35	11.55	10.86
4	MAE	0.36	0.25	0.33	0.31	0.30	0.31	0.26	0.25	0.22
	RMSE	0.94	0.55	0.71	0.68	0.58	0.66	0.57	0.53	0.48
	MAPE (%)	21.25	13.31	16.90	16.39	16.38	15.78	13.25	12.69	11.79
5	MAE	0.37	0.26	0.34	0.32	0.29	0.31	0.27	0.23	0.23
	RMSE	0.95	0.56	0.72	0.71	0.58	0.69	0.58	0.53	0.52
	MAPE (%)	21.60	13.82	16.94	16.20	15.74	15.77	13.75	11.53	12.09

4.3 | Results

4.3.1 | Quantitative evaluation

The following performance metrics were used in the experiments.

4.3.2 | Congestion severity index

The congestion severity index is predicted as the marked value of the proposed model. To measure the prediction performance, we used three commonly used error measures, including the mean average error (MAE), root mean square error (RMSE), and mean average percentage error (MAPE).

4.3.3 | Occurrence time of congestion events

Predicting whether a congestion event will occur at a given time can be considered as a binary classification problem. Therefore, the F1 score and the area under the receiver operating characteristic curve (AUC) were used as

TABLE 3 Performance comparison of predicted occurrence times during weekday peak-hours.

		Prediction horizon				
		1	2	3	4	5
F1 score	ARIMA	0.88	0.87	0.86	0.86	0.86
	DSTAGNN	0.89	0.89	0.89	0.88	0.88
	SAHP	0.65	0.64	0.59	0.57	0.59
	THP	0.73	0.71	0.70	0.68	0.67
	ASTGCN	0.89	0.89	0.87	0.87	0.85
	DMGP(S)	0.74	0.72	0.71	0.68	0.67
	DMGP(T)	0.87	0.87	0.86	0.86	0.85
	DMGP(G)	0.89	0.88	0.87	0.85	0.86
	DMGP	0.92	0.90	0.88	0.87	0.86
AUC	ARIMA	0.85	0.84	0.83	0.83	0.83
	DSTAGNN	0.85	0.85	0.84	0.84	0.83
	SAHP	0.70	0.70	0.65	0.65	0.67
	THP	0.76	0.75	0.75	0.74	0.73
	ASTGCN	0.88	0.88	0.87	0.86	0.83
	DMGP(S)	0.78	0.77	0.76	0.74	0.73
	DMGP(T)	0.85	0.84	0.84	0.84	0.83
	DMGP(G)	0.87	0.85	0.84	0.81	0.84
	DMGP	0.91	0.88	0.87	0.85	0.84

TABLE 4 Performance comparison of predicted occurrence times during weekend peak-hours.

		Prediction horizon				
		1	2	3	4	5
F1 score	ARIMA	0.86	0.86	0.85	0.85	0.84
	DSTAGNN	0.89	0.90	0.89	0.89	0.89
	SAHP	0.75	0.76	0.74	0.73	0.76
	THP	0.76	0.74	0.72	0.69	0.66
	ASTGCN	0.87	0.87	0.86	0.86	0.85
	DMGP(S)	0.76	0.74	0.73	0.70	0.67
	DMGP(T)	0.89	0.89	0.88	0.88	0.87
	DMGP(G)	0.91	0.90	0.89	0.87	0.88
	DMGP	0.93	0.91	0.90	0.89	0.88
AUC	ARIMA	0.83	0.82	0.82	0.81	0.81
	DSTAGNN	0.86	0.86	0.86	0.86	0.84
	SAHP	0.72	0.72	0.67	0.67	0.69
	THP	0.79	0.78	0.77	0.75	0.73
	ASTGCN	0.85	0.85	0.85	0.84	0.84
	DMGP(S)	0.79	0.78	0.78	0.76	0.74
	DMGP(T)	0.88	0.88	0.87	0.87	0.86
	DMGP(G)	0.90	0.88	0.87	0.85	0.86
	DMGP	0.92	0.90	0.89	0.87	0.86

performance evaluation criteria. The values of both metrics range from 0 to 1. High values of F1 score and AUC indicate better performance. Since the traffic data were collected at a granularity of 10 min, the length of the prediction horizon is a multiple of 10 min, that is, from 10 to 50 min.

The first test was to evaluate the performance of the proposed prediction model against the ground truth data for two different scenarios: weekday peak-hours and weekend peak-hours. The results of the next five time steps, including the marked congestion severity indices and occurrence times, are reported in Tables 1–4, which indicate that (1) the proposed DMGP model provides the best MAE, RSME, and MAPE metrics for congestion severity indices in all scenarios; (2) the proposed model outperforms the baselines and its variants in predicting the occurrence times in most cases; (3) the overall performance of the weekend peak-hours is superior to that of the weekday peak-hours, which can probably be attributed to the large variability of traffic flow on weekdays; and (4) the DSTAGNN model has the best performance among all the baselines due to its ability to capture dynamic spatiotemporal correlations in the traffic data. ARIMA fails to capture complex congestion patterns and performed the worst in predicting congestion severity indices. The worst data-driven model is SAHP, which relies only on the attention mechanism for temporal modeling and neglects historical temporal correlations as well as spatial correlations. THP, which uses an attention mechanism to model both spatial and temporal correlations, outperforms SAHP. However, it is inferior to ASTGCN because it does not account for the topological structure of the road network as ASTGCN does.

For all models in Tables 1 and 2, the errors for the short-term predictions (horizon = 1 or 2) are generally lower than those for the medium- and long-term forecasts (horizon = 3, 4 or 5), which can be expected since traffic states with high forecasting horizons are challenging to model. Compared to the weekday peak-hours, the DMGP model performs slightly better for the weekend peak-hours, probably because extremely heavy congestion, which is challenging to predict, is less likely to occur on weekends.

We further compare the error distributions of DMGP and two competitive baselines for all road segments. As Figure 6 shows, the error distribution of DMGP is consistently smaller than that of the two baselines for different predictive horizons, demonstrating its robustness under different traffic conditions.

As shown in Tables 3 and 4, the predictions results of the congestion occurrence time are similar to those of the congestion severity index: The accuracies of all the models gradually decrease with the increase of the prediction horizon, and the results for weekdays are slightly worse

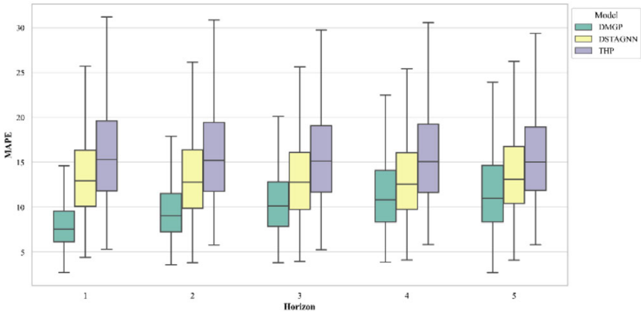

FIGURE 6 Distributions of prediction errors for the congestion severity index during the weekend peak-hours.

TABLE 5 Performance comparison of predicted congestion severity indices on weekdays at different congestion levels (MAPE, %).

	Light			Medium			severe		
	1	3	5	1	3	5	1	3	5
THP	10.60	10.42	10.31	14.08	13.80	13.31	28.18	28.99	30.05
ASTGCN	13.74	15.36	14.79	13.20	14.78	14.09	24.20	24.85	23.96
DSTAGNN	12.78	12.92	13.41	15.36	15.89	16.79	31.05	30.68	29.34
DMGP(S)	10.63	10.83	10.83	12.09	12.45	11.43	26.11	26.47	27.71
DMGP(T)	8.21	8.54	9.02	13.01	13.00	13.67	22.59	23.37	23.73
DMGP(G)	7.47	7.14	7.32	10.94	11.59	11.39	19.93	23.41	23.24
DMGP	6.47	8.25	8.33	8.56	11.40	12.69	14.89	20.66	23.90

than those for weekends. The proposed DMGP model and the two spatiotemporal graph neural network models (i.e., DSTAGNN and ASTGCN) outperform the two Hawkes Process models (i.e., SAHP and THP) by a large margin, indicating that graph neural networks are quite effective in the representation of congestion events. In some cases, DMGP is slightly inferior to DSTAGNN. However, the spatiotemporal attention blocks of DSTAGCNN are much more complex and computationally expensive than the congestion embedding network of DMGP. Interestingly, ARIMA yields competitive results compared to advanced graph neural network models, suggesting that the occurrence of congestion may exhibit some regular temporal patterns.

In the second test, the performance of DMGP was evaluated for three congestion severity levels on weekdays, as shown in Tables 5–6. It can be observed that (1) the prediction performance degrades as the congestion becomes more severe and rarer, making it more difficult to capture long-range spatiotemporal correlations between rare events; (2) the performance gains of the proposed DMGP over the compared baselines under light congestion are not as significant as the gains under moderate and severe congestion, indicating that DMGP can quickly adapt to large changes in traffic conditions and effectively use the traffic evolution information from the limited samples of moderate and severe congestion events; and (3) both spatial

TABLE 6 Performance comparison of predicted congestion occurrence times on weekdays at different congestion levels (AUC).

	Light			Medium			Severe		
	1	3	5	1	3	5	1	3	5
THP	0.69	0.68	0.68	0.71	0.70	0.68	0.75	0.79	0.64
ASTGCN	0.86	0.85	0.85	0.82	0.81	0.79	0.85	0.83	0.81
DSTAGNN	0.76	0.76	0.75	0.75	0.74	0.73	0.68	0.70	0.73
DMGP(S)	0.70	0.69	0.66	0.69	0.68	0.65	0.76	0.81	0.60
DMGP(T)	0.82	0.81	0.80	0.80	0.79	0.78	0.84	0.83	0.82
DMGP(G)	0.83	0.83	0.83	0.81	0.79	0.79	0.85	0.85	0.82
DMGP	0.85	0.83	0.83	0.86	0.81	0.79	0.88	0.85	0.82

and temporal correlations are indispensable for congestion prediction, as evidenced by the poor performance of THP and DMGP(S), which only consider temporal correlations between congestion events.

4.3.4 | Ablation studies

Ablation studies were conducted to evaluate the benefits of the different components of the proposed model. According to the results presented in Tables 1–6, the variant model DMGP(S) achieves the worst performance among all the variants and is sometimes even inferior to some of the baselines because it only captures the spatial correlations of congestion events without accounting for the temporal dependencies, which are critical for modeling congestion events. DMGP(T) achieves better performance than DMGP(S), suggesting that temporal embeddings play a more important role than spatial embeddings when modeling historical congestion events. Compared to DMGP(T), the DMGP(G) model yields a higher prediction accuracy because it can better capture the changes of congestion patterns via the integration of the time-varying dynamic correlation matrix $\boldsymbol{A}_{\boldsymbol{g}}(t)$ in the pattern-aware graph convolution layer.

As shown in the above results, the temporal embedding component, the graph convolutional layer, and the time-varying correlation matrix are essential to improve the predictive performance of the proposed DMGP model. The full model leverages a pattern-aware spatiotemporal GCN to capture stochastic and heterogeneous propagation patterns of congestion based on a traffic condition graph and a traffic congestion graph, thus addressing the challenges of short-term traffic congestion forecasting for large signalized road networks.

4.3.5 | Qualitative results

Qualitative prediction results of the proposed model can be presented using the road network of Xi'an City. From Figure 7, it can be observed that the congestion events

FIGURE 7 Examples of predicted congestion during the morning peak hour (7:30–9:00 a.m.) in Xi'an City. First row: ground truth; second and third rows: the difference between the congestion severity level predicted by deep marked graph process (DMGP)/attention-based spatial–temporal graph convolutional network (ASTGCN) and the ground truth.

predicted by the proposed model evolve closely to the ground truth. More importantly, it can be observed that the proposed model performs consistently well under different congestions severity levels, demonstrating its robustness to different traffic conditions. As shown in Figure 7e, the proposed DMGP model slightly underestimates the severity of congestion on some heavily congested roads. The prediction errors of ASTGCN is significantly larger than that of DMGP under different traffic conditions throughout the city, (Figure 7g–i).

The experimental results show that the proposed prediction model can capture realistic evolution patterns of congestion better than the compared baselines, which may be due to the integration of spatiotemporal correlations into the point process model. The integration of learned spatiotemporal embeddings with the conditional intensity function accounts for both temporal dependencies of historical congestion events and the spatial correlations between neighboring roads. With a limited number of congestion events in the training data set, the proposed model still manages to produce good prediction results and demonstrates better generalization than the baselines through the integration of the point process framework and the spatiotemporal GCN, which naturally impose some forms of prior bias and relational inductive bias (Battaglia et al., 2018). Our model is a two-task learning scheme that uses shared parameters and features to jointly optimize the prediction of congestion occurrence and severity. The proposed model predicts both the occurrence time and the severity index, which facilitates tracking and analyzing the evolution of congestion events.

We also evaluated the time cost of the proposed model. In our experiments, the training time cost of the DMGP model was about 7217 s (slightly more than 2 h), while a typical test (weekday peak-hours) took about 4.0 s, suggesting that the DMGP model can be scaled up to large road networks.

5 | CONCLUSION

This paper has introduced a DMGP model for predicting the congestion severity indices and the occurrence time of traffic congestion events for large urban signalized road networks. The proposed prediction approach integrates a spatiotemporal convolutional graph network with the conventional point process model, accounting for stochastic and heterogeneous evolutionary patterns of traffic congestion. Our model outperforms existing baselines and achieves superior prediction results and computational efficiency on a real-world traffic congestion data set. Congestion management is critical for intelligent transportation systems. The proposed congestion prediction model can be practically used to support both real-world advanced traffic management and traveller information systems. The proposed model can make relatively accurate predictions of the occurrence time and severity of congestion on a citywide scale, thereby facilitating responsive and intelligent control of traffic signals. Congestion information can also be used to estimate travel times online, thus helping to find personalized optimal routes. The model also helps to understand the underlying driving factors of recurrent congestion, which is also important for transportation management authorities to take appropriate measures to improve the management of traffic supply and demand, such as promoting the use of public transportation, car-pooling, high-occupancy vehicle lanes, and developing optimal pricing policies.

However, the practical application of the proposed model may face data challenges as high-quality citywide traffic data may not be available for all road segments. Therefore, error-tolerant prediction methods are worth of scrutiny in the future. We will continue to improve the dynamic GCN to better capture the evolution and interaction patterns of congestion events. More advanced methods will be explored to integrate congestion embeddings with the point process model, such as neural dynamic classification (Rafiei & Adeli, 2017) or dynamic ensemble learning (Alam et al., 2020), which are computationally efficient and excel at modeling traffic dynamics and capturing volatile congestion patterns. Weather and event information can be collected and incorporated into the proposed model to enhance predictive performance.

ACKNOWLEDGMENTS

The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

REFERENCES

Adeli, H., & Ghosh-Dastidar, S. (2004). Mesoscopic-wavelet freeway work zone flow and congestion feature extraction model. *Journal of Transportation Engineering*, *130*(1), 94–103.

Adepeju, M., Rosser, G., & Cheng, T. (2016). Novel evaluation metrics for sparse spatio-temporal point process hotspot predictions: A crime case study. *International Journal of Geographical Information Science*, *30*(11), 2133–2154.

Akhtar, M., & Moridpour, S. (2021). A review of traffic congestion prediction using artificial intelligence. *Journal of Advanced Transportation*, *2021*, 1–18.

Alam, K., Siddique, N., & Adeli, H. (2020). A dynamic ensemble learning algorithm for neural networks. *Neural Computing with Applications*, *32*(10), 8675–8690.

Bae, W., Ahmed, M., Tung, F., & Oliveira, G. (2023). Meta temporal point processes. In *Eleventh international conference on learning representations (ICLR)*, May 1-5, 2023, Kigali, Rwanda. *arXiv preprint arXiv:2301.12023v1*.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint *arXiv:1803.01271*.

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., & Pascanu, R. (2018). Relational inductive biases, deep learning, and graph networks. arXiv preprint *arXiv:1806.01261*.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

Chen, M., Yu, X., & Liu, Y. (2018). PCNN: Deep convolutional networks for short-term traffic congestion prediction. *IEEE Transactions on Intelligent Transportation Systems*, *19*(11), 3550–3559.

Chen, R., Amos, B., & Nickel, M. (2021). Neural spatio-temporal point processes. In *Ninth international conference on learning representations (ICLR)*, May 3–7, 2021, Virtual Event.

Chen, Y., Segovia-Dominguez, I., Coskunuzer, B., & Gel, Y. (2022). Tamp-S2GCNets: Coupling time-aware multipersistence knowledge representation with spatio-supra graph convolutional networks for time-series forecasting. In *Tenth international conference on learning representations (ICLR)*, April 25–29, 2022, Virtual Event.

Chen, Y., Lv, Y., Li, Z., & Wang, F. (2016). Long short-term memory model for traffic congestion prediction with online open data. In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)*, Rio de Janeiro, Brazil, November 1–4, 2016.

Choi, J., & Park, N. (2023). Graph neural rough differential equations for traffic forecasting. *ACM Transactions on Intelligent Systems and Technology*, *14*(4), Article No. 74, 1–27.

Daley, D., & Vere-Jones, D. (2003). *An introduction to the theory of point processes. Vol. I. Probability and its applications*. Springer-Verlag.

Di, X., Xiao, Y., Zhu, C., Deng, Y., Zhao, Q., & Rao, X. (2019). Traffic congestion prediction by spatiotemporal propagation patterns. In *20th IEEE international conference on mobile data management (MDM)*, June 10–13, 2019, Hong Kong, China.

Du, N., Dai, H., Trivedi, R., Upadhyay, U., Geomez-Rodriguez, M., & Song, L. (2016). Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* Association for Computing Machinery, New York, United States. (pp. 1555–1564).

Du, H., Zhou, Y., Ma, Y., & Wang, S. (2021). Astrologer: Exploiting graph neural Hawkes process for event propagation prediction with spatio-temporal characteristics. *Knowledge-Based Systems*, *228*, 107247.

Gal-Tzur, A., Bekhor, S., & Barsky, Y. (2022). Feature engineering methodology for congestion forecasting. *Journal of Traffic and Transportation Engineering (English edition)*, *9*(6), 1055–1068.

Ghosh-Dastidar, S., & Adeli, H. (2006). Neural network-wavelet microsimulation model for delay and queue length estimation at freeway work zones. *Journal of Transportation Engineering*, *132*(4), 331–341.

Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 922–929.

Guo, J., Liu, Y., Yang, Q., Wang, Y., & Fang, S. (2021). GPS-based citywide traffic congestion forecasting using CNN-RNN and C3D hybrid model. *Transportmetrica A: Transport Science*, *17*(2), 190–211.

Han, Y., Lee, H., & Kim, Y. (2023). Extensible prototype learning for real-time traffic signal control. *Computer-Aided Civil and Infrastructure Engineering*, *38*(9), 1181–1198.

Jiang, X., & Adeli, H. (2004). Wavelet packets-autocorrelation function analysis of traffic flow. *Computer-Aided Civil and Infrastructure Engineering*, *19*(5), 324–337.

Jiang, W., & Luo, J. (2022). Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, *207*, 117921.

Jiang, S., Keyvan-Ekbatani, M., & Ngoduy, D. (2023). Partitioning of urban networks with polycentric congestion pattern for traffic management policies: Identifying protected networks. *Computer-Aided Civil and Infrastructure Engineering*, *38*(4), 508–527.

Ke, J., Zheng, H., Yang, H., & Chen, X. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C*, *85*, 591–608.

Kipf, T., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Fifth international conference on learning representations (ICLR)*, April 24–26, 2017, Toulon, France. arXiv preprint arXiv: 1609.02907v4.

Kumar, N., & Raubal, M. (2021). Applications of deep learning in congestion detection, prediction and alleviation: A survey. *Transportation Research Part C*, *133*, 103432.

Lan, S., Ma, Y., Huang, W., Wang, W., Yang, H., & Li, P. (2022). DSTAGNN: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., & Sabato, S. (Eds.). In *Proceedings of the 39th international conference on machine learning* (Vol. *162*, pp. 11906–11917) JMLR.org..

Leiser, N., & Yildirimoglu, M. (2021). Incorporating congestion patterns into spatio-temporal deep learning algorithms. *Transportmetrica B: Transport Dynamics*, *9*(1), 622–640.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, *525*, 367–371.

Lin, Z., Li, M., Zheng, Z., Cheng, Y., & Yuan, C. (2020). Self-attention convlstm for spatiotemporal prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(7), 11531–11538.

Liu, Q., Wang, B., & Zhu, Y. (2018). Short-term traffic speed forecasting based on attention convolutional neural network for arterials. *Computer-Aided Civil and Infrastructure Engineering*, *33*(11), 999–1016.

Liu, Y., Liu, Z., Vu, H., & Lyu, C. (2020). A spatio-temporal ensemble metghod for lage-scale traffic state prediction. *Computer-Aided Civil and Infrastructure Engineering*, *33*(1), 26–44.

Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. (2015). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, *16*(2), 865–873.

Ma, X., Yu, H., Wang, Y., & Wang, Y. (2015). Large-scale transportation network congestion evolution prediction using deep learning theory. *PLoS ONE*, *10*(3), e0119044.

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, *17*, 818.

Mehdi, M. Z., Kammoun, H. M., Benayed, N. G., Sellami, D., & Masmoudi, A. D. (2022). Entropy-based traffic flow labeling for CNN-based traffic congestion prediction from meta-parameters. *IEEE Access*, *10*, 16123–16133.

Mohanty, S., Pozdnukhov, A., & Cassidy, M. (2020). Region-wide congestion prediction and control using deep learning. *Transportation Research Part C*, *116*, 102624.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402.

Pan, Y., Liang, Y., Wang, W., Yu, Y., Zheng, Y., & Zhang, J. (2019). Urban traffic prediction from spatio-temporal data using deep meta learning. In *The 25th ACM SIGKDD conference on knowledge discovery & data mining (KDD'19)*, August 4–8, 2019, Anchorage, AK, USA.

Pan, Z., Sharma, A., Hu, J., Liu, Z., Li, A., Liu, H., Huang, M., & Geng, T. (2023). Ising-traffic: Using Ising machine learning to predict traffic congestion under uncertainty. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(8), 9354–9363.

Rafiei, M., & Adeli, H. (2017). A new neural dynamic classification algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(12), 3074–3083.

Rahmani, S., Baghbani, A., Bouguila, N., & Patterson, Z. (2023). Graph neural networks for intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, *24*(8), 8846–8885.

Shchur, O., Tyurkmen, A., Januschowski, T., & Günnemann, S. (2021). Neural temporal point processes: A review. Zhou, Z.-H. (Ed.). In *Proceedings of the thirtieth international joint conference on artificial intelligence (IJCAI-21)* (pp. 4585–4593).

Smith, B., & Demetsky, M. (1994). Short-term traffic flow prediction: Neural network approach. *Transportation Research Record*, *1453*, 98–104.

Sun, Y., Jiang, G., Lam, S., & He, P. (2022). Learning traffic network embeddings for predicting congestion propagation. *IEEE Transactions on Intelligent Transportation Systems*, *23*(8), 11591–11604. https://doi.org/10.1109/TITS.2021.3105445

Sweet, M. (2014). Traffic congestion's economic impacts: Evidence from US metropolitan regions. *Urban Studies*, *51*, 2088–2110.

Tang, J., & Zeng, J. (2022). Spatiotemporal gated graph attention network for urban traffic flow prediction based on license plate recognition data. *Computer-Aided Civil and Infrastructure Engineering*, *37*, 3–23.

Tedjopurnomo, D., Bao, Z., Zheng, B., Choudhury, F., & Qin, A. (2022). A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. *IEEE Transactions on Knowledge and Data Engineering*, *34*(4), 1544–1561.

Traffic Management Research Institute of the Ministry of Public Safety (TMRI), & Tongji University. (2020). *Road traffic congestion evaluation methods (GAT 115–2020)*.

Vlahogianni, E., Karlaftis, M., & Golias, J. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C*, *43*, 3–19.

Wang, P., Zhang, T., Zheng, Y., & Hu, T. (2022). A multi-view bidirectional spatiotemporal graph network for urban traffic flow imputation. *International Journal of Geographical Information Science*, *36*(6), 1231–1257.

Wang, X., Zeng, R., Zou, F., Liao, Y., & Huang, F. (2023). STTF: An efficient transformer model for traffic congestion prediction. *International Journal of Computational Intelligent Systems*, *16*, Article No. 2.

Yao, Y., Zhou, Y., Xia, J., Wang, X., Li, X., & Ran, B. (2023). Rolling-horizon-based strategy of fully cooperative traffic under signalized intersections. *Computer-Aided Civil and Infrastructure Engineering*, *38*(4), 454–469.

Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., & Yin, B. (2022). Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Transactions on Intelligent Transportation Systems*, *23*(6), 4927–4943.

Yu, R., Li, Y., Shahabi, C., Demiryurek, U., & Liu, Y. (2017). Deep learning: a generic approach for extreme condition traffic forecasting. Chawla, N. & Wang, W. (Eds.). In *Proceedings of the 2017 SIAM international conference on data mining (SDM)* (pp. 777–785). Society for Industrial and Applied Mathematics (SIAM).

Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. Lang, J. (Ed.). In *Proceedings of the 27th international joint conference on artificial intelligence (IJCAI)* (pp. 3634–3640). AAAI Press.

Yun, S., Namkoong, S., Rho, J., Shin, S., & Choi, J. (1998). A performance evaluation of neural network models in traffic volume forecasting. *Mathematical Computer Modelling*, *27*(9-11), 293–310.

Zammit-Mangion, A., Dewar, M., Kadirkamanathan, V., & Sanguinetti, G. (2012). Point process modelling of the Afghan War Diary. *Proceedings of the National Academy of Science*, *109*(31), 12414–12419.

Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flow prediction. Singh, S. & Markovitch, S. (Eds.). In *Proceedings of the thirty-first AAAI conference on artificial intelligence (AAAI-17)*. AAAI Press.

Zhang, S., Yao, Y., Hu, J., Zhao, Y., Li, S., & Hu, J. (2019). Deep autoencoder neural networks for short-term traffic congestion prediction of transportation networks. *Sensors*, *19*, 2229.

Zhang, S., Zhou, L., Chen, X., Zhang, L., Li, L., & Li, M. (2020). Network-wide traffic speed forecasting: 3D convolutional neural network with ensemble empirical mode decomposition. *Computer-Aided Civil and Infrastructure Engineering*, *35*(10), 1132–1147.

Zhang, Y., Cheng, T., & Ren, Y. (2019). A graph deep learning method for short-term traffic forecasting on large road networks. *Computer-Aided Civil and Infrastructure Engineering*, *34*(10), 877–896.

Zhang, Y., & Cheng, T. (2020). Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. *Computers, Environment and Urban Systems*, *79*, 101403.

Zhang, Q., Lipani, A., Kirnap, O., & Yilmaz, E. (2020). Self-attentive Hawkes process. Daumé, H. & Singh, A. (Eds.). In *Proceedings of the 37th international conference on machine learning* (Vol. 119). JMLR.org. 11183–11193.

Zhang, Y., Wang, S., Chen, B., Cao, J., & Huang, Z. (2021). TrafficGAN: Network-scale deep traffic prediction with generative adversarial nets. *IEEE Transactions on Intelligent Transportation Systems*, *22*(1), 219–230.

Zhang, T., Wang, J., & Liu, J. (2022). A gated generative adversarial approach for imputing missing values on signalized road networks. *IEEE Transactions on Intelligent Transportation Systems*, *23*(8), 12144–12160.

Zhao, L., Lai, Y., Park, K., & Ye, N. (2005). Onset of traffic congestion in complex networks. *Physical Review E*, *71*, 026125.

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2020). T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, *21*(9), 3848–3858.

Zheng, W., Yang, H., Cai, J., Wang, P., Jiang, X., Du, S., Wang, Y., & Wang, Z. (2023). Integrated the traffic science with representation learning for city-wide network congestion prediction. *Information Fusion*, *99*, 101837.

Zhou, Z., Yang, X., Rossi, R., Zhao, H., & Yu, R. (2022). Neural point process for learning spatiotemporal event dynamics. Firoozi, R., Mehr, N., Yel, E., Antonova, R., Bohg, J., Schwager, M., & Kochenderfer, M. (Eds.). In *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, PMLR 168:777-789. Machine Learning Research Press. arXiv preprint arXiv:2112.06351.

Zhu, S., Li, S., Peng, Z., & Xie, Y. (2020). Interpretable deep generative spatio-temporal point processes. In *Proceedings of the NeurIPS, workshop AI for earth sciences*.

Zhu, S., Ding, R., Zhang, M., Van Hentenryck, P., & Xie, Y. (2022). Spatio-temporal point processes with attention for traffic congestion event modeling. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 7298–7309. https://doi.org/10.1109/TITS.2021.3068139

Zuo, S., Jiang, H., Li, Z., Zhao, T., & Zha, H. (2020). Transformer Hawkes process. *Proceedings of the 37th International Conference on Machine Learning*, *119*, 11692–11702.

How to cite this article: Zhang, T., Wang, J., Wang, T., Pang, Y., Wang, P., & Wang, W. (2024). A deep marked graph process model for citywide traffic congestion forecasting. *Computer-Aided Civil and Infrastructure Engineering*, *39,* 1180–1196. https://doi.org/10.1111/mice.13131