A Hierarchical Attention Graph Convolutional Network for Traffic Incident Impact Forecasting

Kaiqun Fu*[1], Taoran Ji*[2], Nathan Self[2], Zhiqian Chen[3], and Chang-Tien Lu[2]
[1]*South Dakota State University, Brookings, SD, USA*
[2]*Virginia Tech, Falls Church, VA, USA*
[3]*Mississippi State University, MS, USA*
kaiqun.fu@sdstate.edu, {jtr,nwself}@vt.edu, zchen@cse.msstate.edu, ctlu@vt.edu

2021 IEEE International Conference on Big Data (Big Data) | 978-1-6654-3902-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/BigData52589.2021.9672030

***Abstract*—Predicting the impact of traffic incidents based on traffic sensor data is an essential research topic in the field of Intelligent Transportation Systems (ITS). Tackling the problem of estimating the durations of incidents from their early stages is a challenge due to the variable nature of such incidents and the complex structure of modern road networks. Existing studies on forecasting the incident duration from sensor data are mostly incapable of modeling 1) the spatiotemporal correlations of traffic sensors and arterial roads and 2) the hierarchical topology of the traffic sensor and road networks. In this paper, we propose the Hierarchical Attention-based Spatiotemporal Graph Convolutional Network model (*HastGCN*) to solve the incident duration forecasting problem by formulating the spatiotemporal correlation and traffic patterns on both the sensor level and the road level in their natural hierarchical manner. At the sensor level, we propose a spatiotemporal attention mechanism followed by graph convolutions to model the local correlations and patterns between traffic sensors on the same arterial road. At the road level, a connectivity-aware attention mechanism is designed to learn the global spatial relatedness between each arterial road. Traffic-condition aware graph convolutions are then applied to understand the target incident representation for the incident duration forecasting.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Early detection of non-recurring congestion caused by traffic incidents has become an increasingly important research topic in the field of Intelligent Transportation Systems (ITS). Furthermore, estimation of the duration of such incidents is the natural follow-on problem, especially due to the potential for significant social and economic loss caused by such delays. Indeed, a one-minute reduction in incident duration can produce a 65 USD gain per traffic incident [1]. Due to their natural variability, occurrences of traffic incidents are hard to forecast. But despite this difficulty, the usefulness of such work keeps the problem of forecasting traffic incident duration a primary focus for transportation researchers. To the benefit of such research, over the past decade, there has been a widespread deployment of traffic speed sensors and traffic incident management systems(TIMS) which has made traffic speed and traffic incident records more widely accessible. Thanks to this abundance of traffic data sources, we are able to develop efficient machine learning models to provide accurate estimations of traffic incident duration from the perspective of the natural flow of a traffic incident's lifespan.

*Authors have equal contribution.

The *incident duration* of a traffic incident is quantified by the time elapsed from the incident occurrence until no evidence of the incident remains at the incident scene [2]. A large body of current work [3]–[6] regards the estimation of incident duration as a feature-driven regression task, relying on partial traffic sensor data but ignoring the topology of the road networks. Often, these methods suffer from several drawbacks.

First, **hierarchical structures between traffic sensors and arterial roads are rarely considered**. The topology of modern road networks can be generalized as a connected graph of arterial roads where each arterial road consists of a distance-based correlation graph of traffic sensors deployed along it. This hierarchical structure between traffic sensors and road networks can be applied in designing a hierarchical graph neural network which can improve the performance of predicting the duration of traffic incidents. However, most existing works [3], [7], [8] in the literature on traffic incident duration prediction ignore the advantages of this hierarchy and only consider single arterial roads or road networks as a whole. Second, current methods [1], [4] are **incapable of learning dynamic spatiotemporal feature correlation from traffic sensor data**. The temporal features extracted from the traffic sensors used by regression-based methods cannot represent the spatial correlations of traffic incident duration prediction tasks. Because the duration of a traffic incident is defined and quantified in both the spatial and temporal domain, modulating the spatiotemporal characteristics is essential for precisely estimating duration. Additionally, it is critical to identify the relative importance of spatiotemporal features and the relations among these features. Most of the existing methods do not address these concerns and leave much of this signal on the table. Third, **connectivity between arterial roads and correlations between the traffic sensors are not properly modeled**. In real-world scenarios, the impact of a traffic incident cascades along spatially correlated traffic sensors and interconnected arterial roads. Properly modeling the interconnections of arterial roads becomes integral for predicting the duration of traffic incidents. The existing regression-based and multitask learning-based methods [8], [9], discussed here, offer limited capabilities for modeling spatial correlations between traffic sensors and exploiting the connectivity of arterial roads.

To address these challenges, we propose the **H**ierarchical **A**ttention-based **S**patio**T**emporal **G**raph **C**onvolutional **N**etwork (*HastGCN*) to formulate the spatiotemporal feature correlations and traffic patterns between the traffic sensors and the interconnections of arterial roads separately in their natural hierarchical manner. In particular, because sensor readings during traffic incidents show strong associations in regional traffic sensors, connected arterial roads, and adhesive time slots, we propose multiple attention mechanisms targeting both the spatial and temporal features to learn dynamic spatiotemporal feature correlation from traffic sensor data. Also, because traffic patterns are only available in data transmitted by a closely deployed network of traffic sensors situated along connected arterial roads, we graph the traffic-condition aware convolutional networks to learn the road and incident representations, based on which traffic incident duration forecasting is performed. The main contributions of this paper are summarized as follows:

- **Proposing a novel hierarchical structure for spatiotemporal graph convolutional networks**. We leverage the topology of the road network to model the correlations of the traffic data on multiple levels. Specifically, the sensor level is represented by the spatial relatedness between the traffic sensors on each corridor. The road level is represented by the spatial connections between corridors.
- **Formulating relation-aware, multi-attention mechanisms on spatial and temporal traffic sensor features**. The proposed *HastGCN* model is capable of capturing the dynamical dependencies between spatiotemporal features. In particular, the spatiotemporal attention layer is proposed to identify the density-based correlation between sensors and roads, as well as the inequality of influence of distinct time frames of incidents.
- **Developing a sensor-road traffic-condition aware graph convolutional network to learn road and traffic incident representations**. By considering both the natural spatial connectivity and the invariant traffic conditions (e.g., number of lanes), we propose a traffic-condition aware graph convolutional neural network to deliver the road and incident representation based on the sensor-level and road-level graphs, respectively.

II. Problem Statement

The real-world traffic network $\mathcal{T}$ is a topology which consists of a set of road segments $\mathcal{R}$ and the set of intersections $\mathcal{C}$ which connect them. Intuitively, by modulating this road-level topology, models could make use of road connections for incident duration prediction. However, we argue that, besides the road-level topology, the sensors deployed on each arterial road also construct a topology which contributes a reflection of the real-world traffic situation. Thus, as shown in Fig. 1, we adopt a hierarchical structure which consists of two topologies: a road-level graph (RLG) and a sensor-level graph (SLG).

Definition I: *Road-Level Graph (RLG)*. Naturally, the vanilla road-level topology can be represented mathematically as a graph $\mathcal{G} = (\mathcal{R}, \mathcal{C}, \mathbf{A})$, where $\mathcal{R}$ is the node set which

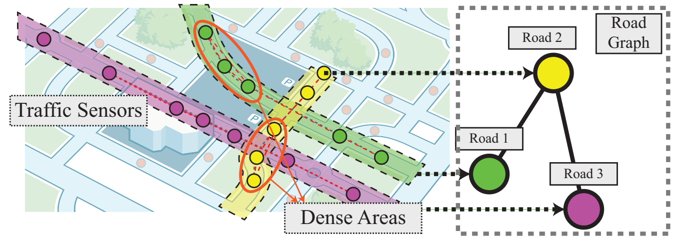


Fig. 1. The Construction of the Hierarchical Graph. The road map on the left-hand side represents the traffic sensors. The green, yellow, and purple nodes represent the traffic sensors on the green, yellow, and purple roads, respectively.

represents the arterial roads in the traffic network $\mathcal{T}$, and $\mathcal{C}$ is edge set which refers the set of intersections $\mathcal{C}$ in the topology $\mathcal{T}$. Each node $r \in \mathcal{R}$ is attached with an attribute vector which is learned based on the sensor-level graph defined below. The square matrix $\mathbf{A} \in \mathbb{R}^{|\mathcal{R}|\times|\mathcal{R}|}$ denotes the adjacency matrix of this graph. In particular, for two arbitrary road segments r_i and r_j, the adjacency matrix $\mathbf{A}$ is formulated as:

$$\mathbf{A}_{i,j} = \begin{cases} 1, & \text{if condition } \varrho(r_i, r_j) \text{ holds,} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where the condition ϱ is a spatial correlation measurement that calculates the spatial relation between the two targets r_i and r_j. These spatial conditions will return True if r_i and r_j *intersect* or *touch* each other.

Definition II: *Sensor-Level Graph (SLG)*. The sensor graph of one road segment r is defined as $\mathcal{G}_r = (\mathcal{V}_r, \mathcal{E}_r, \mathbf{A}_r)$, where $\mathcal{V}_r$ is the node set which represents the n traffic sensors deployed on this road and $\mathcal{E}_r$ refers to the selective pairwise correlation between different sensors. At time τ, the real-time readings of all the sensors on this road compose the attribute matrix $\mathbf{X}_{r,\tau} \in \mathcal{R}^{n\times|F|}$ of graph $\mathcal{G}_r$, where F is the set of traffic features monitored by a sensor. $\mathbf{A}_r \in \mathbb{R}^{n\times n}$ is the adjacency matrix representing the Gaussian weights measured as correlation between pairwise traffic sensors, defined as:

$$\mathbf{A}_{r_{i,j}} = \begin{cases} \exp(-\frac{d_{ij}^2}{\sigma^2}) & i \neq j \text{ and } \exp(-\frac{d_{ij}^2}{\sigma^2}) \geq \epsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

where d_{ij} is the distance between traffic sensors i and j, and σ^2 and ϵ are hyperparameters which together control the distribution and sparsity of matrix $\mathbf{A}_r$.

Assume that we are given a collection of traffic incidents Φ from the traffic incident management system. The duration (impact) of each traffic incident $\phi \in \Phi$ can be characterized and quantified by its *occurrence time* τ_o, *verification time* τ_v and the *restoration time* τ_e (the time when the traffic returns to normal). In this paper, we aim to forecast the duration of the traffic incident defined by $\tau_e - \tau_o$, using a small window of readings around the incident verification time τ_v. In particular, we define the window of size p before τ_v as the *pre-verification window*, and the window of size q following the τ_v as the *post-verification window*. For the pre-verification window of the accident ϕ, the traffic sensor readings for all arterial roads can be formulated as $\boldsymbol{\mathcal{X}}_\phi^- = \{\mathbf{X}_{\tau_v-p+1}, \mathbf{X}_{\tau_v-p+2}, \ldots, \mathbf{X}_{\tau_v}\} \in \mathbb{R}^{|\mathcal{R}|\times p\times n\times |F|}$, where p is the pre-verification window size, and $\mathbf{X}_\tau = \{\mathbf{X}_{1,\tau}, \ldots, \mathbf{X}_{|\mathcal{R}|,\tau}\}$ is the collection of attribute matrix of the sensor graph for each road $r \in \mathcal{R}$. Likewise, the traffic readings during the post-verification window of the incident ϕ can be collected as $\boldsymbol{\mathcal{X}}_\phi^+ = \{\mathbf{X}_{\tau_v+1}, \mathbf{X}_{\tau_v+2}, \ldots, \mathbf{X}_{\tau_v+q}\} \in \mathbb{R}^{|\mathcal{R}|\times q\times n\times |F|}$, where q is the post-verification window size.

Then we can format our problem as: given a traffic incident ϕ which has not yet reached the restoration time τ_e, a small window of size p pre-verification readings $\boldsymbol{\mathcal{X}}_\phi^-$ and a small window of size q post-verification readings $\boldsymbol{\mathcal{X}}_\phi^+$, can we forecast the duration of the traffic incident $y_\phi = \tau_e - \tau_o$ in minutes: $\mathcal{F}(\boldsymbol{\mathcal{X}}_\phi) \to y_\phi$, where $\boldsymbol{\mathcal{X}}_\phi = \{\boldsymbol{\mathcal{X}}_\phi^-; \boldsymbol{\mathcal{X}}_\phi^+\} \in \mathbb{R}^{|\mathcal{R}|\times(p+q)\times n\times|F|}$ is the collection of all the traffic readings during time window $[\tau_v - p + 1, \tau_v + q]$.

III. *HastGCN* Model

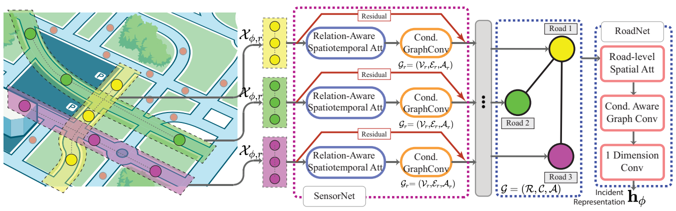


Fig. 2. The framework of the Hierarchical Attention-based Spatiotemporal Graph Convolutional Network (*HastGCN*). The $\boldsymbol{\mathcal{X}}_{\phi,r}$ in the figure represents the road sensor data tensor for the specific road.

This section details the architecture of the *HastGCN* in three major building blocks: *SensorNet*, *RoadNet*, and the incident duration forecasting layer. Aiming to construct an incident representation, the *SensorNet* and *RoadNet* blocks work consecutively to extract, dynamically adjust, and integrate both local (sensor-level) and global (road-level) spatiotemporal features by capturing the spatiotemporal correlations and patterns among traffic sensors and arterial roads. Then, based on the learned incident representation, a prediction layer is used to forecast the incident duration.

A. Model Overview

Recall that for each incident ϕ, the traffic sensor readings are collected as $\boldsymbol{\mathcal{X}}_\phi \in \mathbb{R}^{|\mathcal{R}|\times(p+q)\times n\times|F|}$, *SensorNet* is designed to take $\boldsymbol{\mathcal{X}}_\phi$ as input and learn a representation for each road. In particular, for each road r, we have traffic feature data $\boldsymbol{\mathcal{X}}_{\phi,r} = \{\boldsymbol{\mathcal{X}}_{\phi,r}^-; \boldsymbol{\mathcal{X}}_{\phi,r}^+\} \in \mathbb{R}^{(p+q)\times n\times|F|}$, which consists of the readings of all the sensors on that road, during both the pre- and post-verification windows of that incident. $\boldsymbol{\mathcal{X}}_{\phi,r}$ is first fed into the local spatiotemporal attention layer where both the spatial correlations between the traffic sensors and the importance between different time frames around the verification time are modulated and weighted. The outcome is then fed into the sensor-level convolutional net, which fulfills the graph convolution operations between the sensor attributes and outputs one hidden representation for road r. Next, *RoadNet* modulates the global spatial attention on the road representation learned by *SensorNet*. The output of this is used by the road-level convolutional net, followed by a road-wise one dimensional convolutional layer, to learn the incident representation on the entire road network. By design, our framework allows stacking both *SensorNet* and *RoadNet* multiple times. In order to train the model efficiently and stably, we add residual connections [10] between each stacked module. Finally, a prediction layer is applied to this learned incident representation to forecast the potential incident duration.

B. Relation-Aware Spatiotemporal Attention Layer

In real-world traffic sensor networks, sensors are connected naturally along with the directions of the arterial roads and thus considering them together should enhance the model's generalizability for identifying similar patterns. There are two intuitions behind our proposal of the sensor-level spatial attention: 1) the traffic sensors are not evenly deployed on the arterial roads and 2) the closer the traffic sensors are deployed, the more similar patterns these traffic sensors will share. Based on these two intuitions, we argue that the density of traffic sensors on an arterial road is integral to the spatial attention calculation. Therefore, apply a degree vector $\mathbf{S}_r \in \mathbb{R}^{n\times 1}$ to indicate the density information of the traffic sensors, where $\mathbf{S}_{r_{ij}} = \sum_j \mathbf{A}_{r_{ij}}$. Let $\boldsymbol{\mathcal{X}}_{\phi,r} \in \mathbb{R}^{(p+q)\times n\times|F|}$ be the input raw sensor readings for all the sensors deployed on road r. Inspired by the selective attention mechanism in [11], we propose a density-based sensor attention mechanism to capture the spatial correlations between traffic sensors:

$$\boldsymbol{\alpha} = \mathbf{W}_s \cdot \mathrm{ReLU}\left((\mathbf{W}_{s1}\boldsymbol{\mathcal{X}}_{\phi,r})\mathbf{W}_{s2}(\boldsymbol{\mathcal{X}}_{\phi,r}\mathbf{W}_{s3}) + \mathbf{S}_r\mathbf{W}_d^{\mathrm{T}}\right), \tag{3}$$

$$\bar{\boldsymbol{\alpha}}_{ij} = \frac{\exp(\boldsymbol{\alpha}_{ij})}{\sum_N^{j=1}\exp(\boldsymbol{\alpha}_{ij})}, \hat{\mathbf{X}}_{\phi,r} = \left(\boldsymbol{\mathcal{X}}_{\phi,r}^{\mathrm{T}}\bar{\boldsymbol{\alpha}}\right)^{\mathrm{T}}, \tag{4}$$

where $\bar{\boldsymbol{\alpha}}_{ij}$ represents the attention strength between node i and node j, and $\mathbf{W}_s \in \mathbb{R}^{n\times n}$, $\mathbf{W}_{s1} \in \mathbb{R}^{p+q}$, $\mathbf{W}_{s2} \in \mathbb{R}^{|F|\times(p+q)}$, $\mathbf{W}_{s3} \in \mathbb{R}^{|F|}$, $\mathbf{W}_d \in \mathbb{R}^{1\times n}$ are all learnable parameters. As the sensor-level spatial attention layer modulates the local correlation between sensors on the same road, the road-level spatial attention layer is designed to consider global, road-wise interconnections. In particular, *RoadNet* obtains the spatially integrated and adjusted graph signals $\hat{\mathbf{X}}_r$ by applying the spatial attention layer (Eq. 3 and 4) on the d_s-dimension road representations $\mathbf{X}_r \in \mathbb{R}^{|\mathcal{R}|\times d_s}$ learned by *SensorNet*. The detailed process of generating $\mathbf{X}_r$ is given in Section III-C.

We propose a temporal attention layer, which aims to empower the model with the capacity to dynamically determine the role of each time frame in estimating incident duration. First, the importance of the pre- and post-verification windows is differentiated by a gate mechanism:

$$\gamma = \mathrm{Sigmoid}\left((\mathbf{W}_q\boldsymbol{\mathcal{X}}_{\phi,r}^+\mathbf{W}_f)\mathbf{W}_n\right), \bar{\mathbf{X}}_{\phi,r} = [(1-\gamma)\hat{\mathbf{X}}_{\phi,r}^-; \gamma\hat{\mathbf{X}}_{\phi,r}^+], \tag{5}$$

where $\mathbf{W}_q \in \mathbb{R}^q$, $\mathbf{W}_f \in \mathbb{R}^{|F|}$, $\mathbf{W}_n \in \mathbb{R}^n$ are all learnable matrices. Next, for time frames within each window, the pair-wise correlation is measured. Take the post-verification window as an example, the attention is calculated as:

$$\boldsymbol{\beta}^{+} = \left(\bar{\mathbf{X}}^{+}_{\phi,r}\mathbf{W}^{+}_{t1}\right)\mathbf{W}^{+}_{t2}\left(\mathbf{W}^{+}_{t3}\bar{\mathbf{X}}^{+}_{\phi,r}\right)^{\mathrm{T}}, \tag{6}$$

where $\mathbf{W}^{+}_{t1} \in \mathbb{R}^{|F|}$, $\mathbf{W}^{+}_{t2} \in \mathbb{R}^{n\times|F|}$, and $\mathbf{W}^{+}_{t3} \in \mathbb{R}^{n}$ are all learnable parameters. $\boldsymbol{\beta}^{+}$ is then normalized by the softmax function and directly applied on $\bar{\mathbf{X}}^{+}_{\phi,r}$ to obtain the temporal adjusted attribute matrix $\tilde{\mathbf{X}}^{+}_{\phi,r}$. Finally, by concatenating $\tilde{\mathbf{X}}^{+}_{\phi,r}$ and $\tilde{\mathbf{X}}^{-}_{\phi,r}$ on the temporal, we obtain the temporal attention layer output $\tilde{\mathbf{X}}_{\phi,r} \in \mathbb{R}^{(p+q)\times n\times|F|}$. Note that the temporal attention layer is only used in *SensorNet* since the temporal information is already integrated into the road representation in *RoadNet*.

C. Traffic-condition Aware Graph Convolution

We assume that the traffic patterns which occur during an incident are conducted by the spatial correlations between the sensors (modeled by *SensorNet*) and the interconnections between the arterial roads (modeled by *RoadNet*). To capture the traffic pattern transmission between the sensors and the arterial roads in the traffic networks, we adopt graph convolutional layers to model the entire network. In general settings, the graph signals are dynamically adjusted by the graph convolution which is suitable to the learned representation (e.g., road representation) in our model. However, we argue that, some basic traffic conditions such as the number of lanes also contribute to the understanding of incident and these should, intuitively, be considered in an invariant fashion. To fulfill all the above considerations, we propose a traffic condition-aware graph convolution.

In general, given a graph represented in the spatial domain $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, the graph convolution is implemented by applying linear filters $g(\cdot)$ on the eigenvalue decomposition of the graph's Fourier domain projection, i.e., the normalized Laplacian matrix defined as $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ where $\mathbf{I}$ is the identity matrix and $\mathbf{D}$ is the degree matrix calculated as $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. However, a large scale graph such as the traffic network used in this paper requires expensive computation complexity. As a result, we adopt the m-th polynomial approximation [12] to calculate the graph convolution

$$g_w(\mathbf{L}) * x = \sum_{m=0}^{K} w_m T_m(\mathbf{L}-\mathbf{I})x, \bar{\mathbf{X}}_g = ReLU(g_w(\mathbf{L})\mathbf{X}_g + \zeta_c), \tag{7}$$

where ζ_c denotes real-world traffic condition terms related to graph G, and $\mathbf{X}_g$ is attribute matrix of this graph, $w.$ are learnable parameters, and T_m is the m-th term of the polynomial approximation defined as $T_m(a) = 2xT_{m-1}(a) - T_{m-2}(a), T_0(a) = 1, T_1(a) = a$.

In our current data setting, all traffic sensors deployed on the arterial roads share similar specifications and we therefore omit the ζ_c in *SensorNet*. However, in scenarios where the sensors have different attributes, our framework allows the model to consider specific sensor conditions. For *SensorNet*, the input $\mathbf{X}_g$ for graph convolution (Eq. 7) is $\tilde{\mathbf{X}}_{\phi,r}$ and the output is then flattened to the d_s-dimension vector, used as the road representation. Once every road representation is learned, we obtain $\mathbf{X}_r \in \mathbb{R}^{|\mathcal{R}|\times d_s}$ which is then fed into *RoadNet*. According to the spatial correlations between the traffic sensors in *SensorNet*, we adopt the kernel size $K = 3$.

For *RoadNet*, however, based on the observation that road conditions (e.g. road types and number of lanes) play an important role in predicting traffic incident duration, we introduce a road condition awareness term ζ_c:

$$\zeta_c = ReLU(\rho^{*}\mathbf{W}_c), \tag{8}$$

where $\rho^{*} \in \mathbb{R}^{|\mathcal{R}|\times d_c}$ is the additional road condition information represented by one-hot annotation, and $\mathbf{W}_c \in \mathbb{R}^{d_c\times d_s}$ is a learnable parameter. For *RoadNet*, the input for graph convolution is the output $\hat{\mathbf{X}}_r \in \mathbb{R}^{|\mathcal{R}|\times d_s}$ of the global attention layer. Then, we apply an one dimensional convolution layer to the flattened output of graph convolution $\bar{X}_r$ to obtain the incident representation $\mathbf{h}_\phi$: $\mathbf{h}_\phi = 1DConv(\bar{\mathbf{X}}_r) \in \mathbb{R}^{d_\phi}$.

D. Incident Duration Forecasting Layer

The problem setup for the task of forecasting traffic incident durations implies two constraints on our prediction results: 1) the estimated duration $\hat{y}_\phi$ must be larger than $p+q$ because we aim to use a small window of observation to forecast the entire duration, and 2) the estimated duration must be a positive number. As a result, we modulate these two constraints in our prediction layer:

$$\hat{y}^{-}_{\phi} = p + \text{Softplus}(\mathbf{W}^{-}\mathbf{h}_\phi), \hat{y}^{+}_{\phi} = q + \text{Softplus}(\mathbf{W}^{+}\mathbf{h}_\phi), \tag{9}$$

where $\hat{y}^{-}_{\phi}$ is the estimated duration between the incident occurrence time τ_o and incident verification time τ_v, $\hat{y}^{+}_{\phi}$ is the estimated duration between the verification time τ_v and the time τ_e when the traffic returns to normal. The Softplus layer guarantees that the constraints are satisfied, and $\mathbf{W}^{-} \in \mathbb{R}^{1\times d_h}$ and $\mathbf{W}^{+} \in \mathbb{R}^{1\times d_h}$ are learnable parameters. At last, $\hat{y}_\phi = \hat{y}^{-}_{\phi} + \hat{y}^{+}_{\phi}$ is the estimated duration of incident ϕ.

E. Parameter Learning

In practice, the time period between the verification time and the time when the traffic returns to normal is much longer than the pre-verification period, and thus should play a more important role in the loss calculation. As a result, we use a weighted mean squared error as our loss function:

$$Loss = \sum_{\phi\in\Phi} \epsilon\|\hat{y}^{+}_{\phi} - y^{+}_{\phi}\|^{2}_{2} + (1-\epsilon)\|\hat{y}^{-}_{\phi} - y^{-}_{\phi}\|^{2}_{2}, \tag{10}$$

where the hyperparameter ϵ controls the trade-off between the post-verification duration loss and the pre-verification duration loss.

IV. Experiment

A. Dataset Description and Experiment Setup

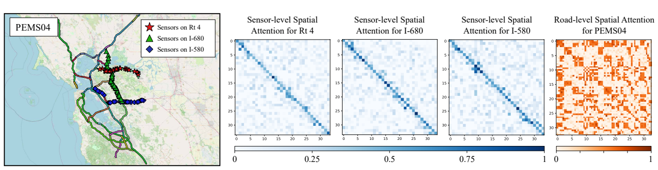
Metrics: To justify the performance of our proposed model on traffic incident duration prediction, we adopt root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). These metrics are widely

TABLE I
TRAFFIC INCIDENT DURATION FORECASTING COMPARISONS (RMSE (MIN), MAE (MIN), MAPE (%))

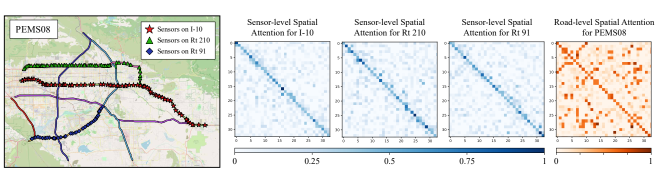
Method	PeMS04			PeMS08		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Ridge	93.6618	74.9687	94.0881	89.0125	70.8411	87.8034
LASSO	90.8160	74.3216	90.4205	76.9342	57.8814	71.1917
SVR	86.4118	70.6930	89.0108	73.5938	54.4448	69.1099
nMTL	87.5108	72.8383	81.5823	55.1493	51.5852	76.2807
TITAN	82.9919	72.1177	79.7070	53.0547	46.9001	68.4498
ASTGCN	68.4086	54.7145	76.8028	39.3745	33.6342	46.2822
GeoMAN	69.3046	61.8778	75.6514	38.7785	31.1322	47.0357
HastGCNr	72.8079	65.7555	75.8787	43.7245	40.7585	59.8090
HastGCNs	70.4996	65.4185	**70.8764**	40.0726	34.8083	59.3436
HastGCN	**66.5545**	**53.3862**	74.5355	**37.7521**	**30.0757**	**46.0682**

utilized in the field of traffic duration prediction studies [3], [8], [13], [14] and reflect upon the predictive performance of the proposed model. The following calculations represent the selected evaluation metrics: $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$, $MAE = \frac{1}{N}\sum_i^N |y_i - \hat{y}_i|$, and $MAPE = \frac{1}{N}\sum_i^N \left|\frac{y_i - \hat{y}_i}{y_i}\right|$ where N is the total number of records; $\boldsymbol{y}$ represents the predicted traffic incident durations as a vector; $\hat{\boldsymbol{y}}$ represents the ground truth value of the corresponding record, also represented as a vector. y_i and $\hat{y}_i$ are the i^{th} predicted result and the i-th ground truth value respectively.

Comparison Methods: To evaluate the performance of our traffic incident duration prediction, five conventional baseline methods are considered in our experiment: ℓ_2 regularized linear regression (ridge regression) [15], ℓ_1 regularized linear regression (LASSO) [9], support vector regression (SVR) [9], naive multi-task learning model (nMTL) [16], and the TITAN model [17]. Two state-of-the-art deep learning methods for traffic flow forecasting are also selected for comparison: ASTGCN [18] and GeoMAN [19]. Due to the different goals of prediction, we change the output to be 1-step prediction in the implementations of these methods.



(a) The sensor-level and road-level spatial attention for Oakland Area



(b) The sensor-level and road-level spatial attention for San Bernardino Area

Fig. 3. Case Studies for Spatial Attention Learning. This figure demonstrates the learned spatial attention parameters at both the local traffic sensor level and the global arterial road level.

B. Performance of Incident Duration Prediction

HastGCN vs. conventional methods with temporal features only. Our model consistently and significantly outperforms the conventional methods (Ridge, LASSO, and SVR) that only consider temporal features. In particular, our model is able to achieve at least 23.0%, 24.5%, and 16.3% improvements on MAE, RMSE, and MAPE, respectively. Furthermore, we observe that the best performing method of the second group also outperforms the "temporal features only" group. These two observations suggest that the models would struggle to modulate the incident duration if only temporal information is available. Also, this experimental result indicates that the temporal attention mechanism is beneficial for identifying and differentiating the importance of different time frames during the observation window during the early stages of an incident.

HastGCN vs. conventional methods with temporal features and spatial constraints. Our model also consistently outperforms the second group of methods that considers temporal features with spatial constraints. In particular, on three different datasets, *HastGCN* can achieve 19.8% to 34.5% improvement on RMSE, 26.0% to 37.8% gain on MAE, and 6.4% to 38.4% improvement on MAPE. Furthermore, we observe that our two ablations (*HastGCN*$_s$ and *HastGCN*$_r$), which considers only *RoadNet* or *SensorNet*, also outperformced this method group in general. We argue that this boost of performance is attributed to both spatiotemporal attention mechanisms and graph convolutional networks. In partular, at the sensor-level, comparing against the spatial constraints used in nMTL and TITAN, our spatial attention layer enables a greater flexibility in modulating sensor-wise correlation and attending only to the important neighboring sensors. Likewise, the temporal attention is more beneficial to model by providing a way to attending the time points during the pre- and post-verification window, comparing to the simple temporal alignment mechanism used in nMTL and TITAN. Besides, we conclude that the graph convolution's capability of modulating the transission of traffic pattern between sensors and roads introduce this advantages in prediction performance.

HastGCN vs. deep learning models. Our proposed model outperforms ASTGCN and GeoMAN on all measures across all datasets. In particular, we observe that, on three different datasets, *HastGCN* can achieve 2.6% to 9.0% improvement on RMSE, 2.4% to 15.5% gain on MAE, and 0.5% to 13.1% improvement on MAPE. Considering that GeoMAN is a spatial-temporal attention based recurrent neural networks, we argue that our model's performance gain can be attributed to the graph convolution layers deployed hierarchically on the sensor-level and graph-level graph. This achievement of performance indicates that, the design of the hierarchical structure and the flexibility it enables in integrating information collected from sensors and roads, are able to improve the accuracy of incident duration predictions. Furthermore, we observe that our model has a more robust performance on all scenarios while GeoMAN and ASTGCN suffer from unstable performance in some cases. For example, on PeMS08, GeoMAN achieves relatively high performance than ASTGCN, but on PeMS04, GeoMAN is inferior to ASTGCN in terms of RMSE and MAE. This can be interpreted as that our hierarchical design could be advantageous in assisting the model to adapt to the complicated and variable nature of the real-world scenarios.

C. Ablation Study

RoadNet Analysis. We first analyze the contribution of the *RoadNet*. In particular, we remove the *RoadNet* module from the *HastGCN*, and create one variant, named *HastGCN*$_s$. Since *SensorNet* can only learn the road representation, in order to predict the duration of the incident, we add one fusion layer to obtain weighted sums of all the roads as the incident representation to our traffic incident prediction layer: $\mathbf{h}_\phi = \sum_{r\in\mathcal{R}} \mathbf{x}_r\mathbf{W}$, where $\mathbf{h}_\phi$ is the incident representation learned by *HastGCN*$_s$, $\mathbf{x}_r \in \mathbb{R}^{d_s}$ is the learned representation for road $r \in \mathcal{R}$, and $\mathbf{W} \in \mathbb{R}^{d_s\times d_\phi}$ are learnable parameters. Note that in this setting, road connectivity information is completely ignored, though a road representation is still learned based on the local sensor readings. The performance is reported in Table I. As expected, we observe that *HastGCN* consistently outperforms *HastGCN*$_s$ under every evaluation case expect for MAPE on PeMS04. This result illustrates that the road network and the connectivity information it contains are critical to the incident duration forecasting problem, which agrees with our intuition.

SensorNet Analysis. Next, we study the contributions of the *SensorNet*. In this ablation test, we create an ablation named *HastGCN*$_r$ by removing the *SensorNet* and keeping only the *RoadNet*. As a result, we use the average of sensor readings of each road as the road representations: $\mathbf{X}_r = \frac{1}{n*k}\sum_\tau\sum_k \mathbf{X}_{\phi,r,\tau,k}$, where $\mathbf{X}_{\phi,r,\tau,k} \in \mathbb{R}^{|F|}$ is the readings of the k-th sensor on road r at time τ. Note that in this setting, the temporal and spatial dependencies of readings of sensors along the road are integrated using the plain average function. The performance is reported in Table I. The fully fledged *HastGCN* consistently outperforms *HastGCN*$_r$ on all datasets. This result demonstrates the importance of allowing the model to dynamically adjust and integrate the raw sensor attributes via considering both the spatial correlation and temporal difference.

V. Conclusion

In this paper, we present a hierarchical attention-based spatiotemporal graph convolution network (*HastGCN*) to predict the impact of traffic incidents from traffic sensor data. This model expands on traditional graph convolutional networks by implementing a hierarchical structure that is capable of modeling graphs with sub-graphs. With this extension, the model considers not only the distances between traffic sensors but also the directions and interconnections of the corridors where the traffic sensors are located. We employ spatial and temporal attention mechanisms that encourage the building of dependencies between traffic sensors and corridor topology and which group the sequences that leverage dynamically learned weights. Our model is evaluated on three real-world traffic datasets collected from the Caltrans Performance Measurement System (PeMS) and the experimental results demonstrate that our model can consistently outperform the existing temporal models.

References

[1] M. W. Adler, J. van Ommeren, and P. Rietveld, "Road congestion and incident duration," *Economics of transportation*, vol. 2, no. 4, pp. 109–118, 2013.

[2] W. Wang, H. Chen, and M. C. Bell, "A review of traffic incident duration analysis [j]," *Communication and Transportati0n Systems Engineering and Information*, vol. 3, p. 022, 2005.

[3] R. Li, F. C. Pereira, and M. E. Ben-Akiva, "Overview of traffic incident duration analysis and prediction," *European Transport Research Review*, vol. 10, no. 2, p. 22, 2018.

[4] S. Boyles, D. Fajardo, and S. T. Waller, "A naive bayesian classifier for incident duration prediction," in *86th Annual Meeting of the Transportation Research Board, Washington, DC*. Citeseer, 2007.

[5] D. Nam and F. Mannering, "An exploratory hazard-based analysis of highway incident duration," *Transportation Research Part A: Policy and Practice*, vol. 34, no. 2, pp. 85–102, 2000.

[6] G. Valenti, M. Lelli, and D. Cucina, "A comparative study of models for the incident duration prediction," *European Transport Research Review*, vol. 2, no. 2, pp. 103–111, 2010.

[7] J. Hu, C. Guo, B. Yang, and C. S. Jensen, "Stochastic weight completion for road networks using graph convolutional networks," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019, pp. 1274–1285.

[8] A. J. Khattak, J. Liu, B. Wali, X. Li, and M. Ng, "Modeling traffic incident duration using quantile regression," *Transportation Research Record*, vol. 2554, no. 1, pp. 139–148, 2016.

[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[11] X. Feng, J. Guo, B. Qin, T. Liu, and Y. Liu, "Effective deep memory networks for distant supervised relation extraction." in *IJCAI*, 2017, pp. 4002–4008.

[12] M. Henaff, J. Bruna, and Y. LeCun, "Deep convolutional networks on graph-structured data," *arXiv preprint arXiv:1506.05163*, 2015.

[13] H. Park, A. Haghani, and X. Zhang, "Interpretation of bayesian neural networks for predicting the duration of detected incidents," *Journal of Intelligent Transportation Systems*, vol. 20, no. 4, pp. 385–400, 2016.

[14] Y. Zou, K. Henrickson, D. Lord, Y. Wang, and K. Xu, "Application of finite mixture models for analysing freeway incident clearance time," *Transportmetrica A: Transport Science*, vol. 12, no. 2, pp. 99–115, 2016.

[15] S. Peeta, J. L. Ramos, and S. Gedela, "Providing real-time traffic advisory and route guidance to manage borman incidents on-line using the hoosier helper program," 2000.

[16] L. Zhao, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan, "Hierarchical incomplete multi-source feature learning for spatiotemporal event forecasting," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 2085–2094.

[17] K. Fu, T. Ji, L. Zhao, and C.-T. Lu, "Titan: A spatiotemporal feature learning framework for traffic incident duration prediction," in *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019, pp. 329–338.

[18] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 922–929.

[19] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "Geoman: Multi-level attention networks for geo-sensory time series prediction." in *IJCAI*, 2018, pp. 3428–3434.