Towards better traffic volume estimation: Tackling both underdetermined and non-equilibrium problems via a correlation-adaptive graph convolution network

4 authors:

Tong Nie
Tongji University
4 PUBLICATIONS **4** CITATIONS

SEE PROFILE

Guoyang Qin
Tongji University
20 PUBLICATIONS **80** CITATIONS

SEE PROFILE

Yunpeng Wang
University of Illinois, Urbana-Champaign
16 PUBLICATIONS **240** CITATIONS

SEE PROFILE

Jian Sun
Tongji University
204 PUBLICATIONS **4,003** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

bicycle simulation View project

Data Pond View project

Towards better traffic volume estimation: Tackling both underdetermined and non-equilibrium problems via a correlation-adaptive graph convolution network

Tong Nie[a], Guoyang Qin[a], Yunpeng Wang[b], Jian Sun[a,*]

[a]*Department of Traffic Engineering & Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China*
[b] *Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control, School of Transportation Science and Engineering, Beihang University, Beijing 100191, China*

Abstract

Traffic volume is an indispensable ingredient to provide fine-grained information for traffic management and control. However, due to limited deployment of traffic sensors, obtaining full-scale volume information is far from easy. Existing works on this topic primarily focus on improving the overall estimation accuracy of a particular method and ignore the underlying challenges of volume estimation, thereby having inferior performances on some critical tasks. This paper studies two key problems with regard to traffic volume estimation: (1) underdetermined traffic flows caused by fully undetected paths that can allow arbitrary volume values without violating the conservation law, where using local side information is insufficient to tackle, and (2) non-equilibrium traffic flows arise when traffic flow vary in density over space and time due to congestion propagation delay, which produce time-shifted volume readings to varying degrees. Here we demonstrate a graph-based deep learning method that offers a data-driven, equation-free and correlation-adaptive approach to tackle the above issues and perform accurate network-wide traffic volume estimation. Particularly, in order to quantify the dynamic and nonlinear speed-volume relationships for the estimation of underdetermined flows, a speed pattern-adaptive adjacent matrix based on graph attention is developed and integrated into the graph convolution process, to capture non-local correlations between volumes. To measure the impacts of non-equilibrium flows, a temporal masked and clipped attention combined with a gated temporal convolution layer is customized to capture time-asynchronous correlations between upstream and downstream sensors under varying impacts of congestion delay. We then evaluate our model on a real-world highway traffic volume dataset and compare it with several benchmark models. It is demonstrated that the proposed model achieves high estimation accuracy even under 20% sensor coverage rate and outperforms other baselines significantly, especially on underdetermined and non-equilibrium flow locations. Furthermore, comprehensive quantitative model analysis are also carried out to justify the model designs.

Keywords: Traffic volume estimation, Graph convolution neural networks, Speed-volume relationship, Spatiotemporal correlation, Attention mechanism, Kriging

1. Introduction

Traffic volumes have been widely measured due to the rapid development of traffic sensing technology. As one of the three macroscopic traffic characteristics, volume directly depicts the operation of traffic flows. Real-time traffic volume information at network-wide is one of the main ingredients required to provide fundamental information for both traffic agencies and individuals to make timely decisions. All the while, ever-increasing problems in transportation systems has prompted traffic managers in dire need of obtaining full-scale of volume data for better control and management of traffic. For instance, estimating time-dependent origin-destination flows (Ashok and Ben-Akiva, 2002), formulating proactive traffic control policies (Sirmatel and Geroliminis, 2017; Yildirimoglu et al., 2018), measuring urban traffic emissions (Liu et al., 2019), and calibrating microscopic traffic simulator (Saeedmanesh et al., 2021) entail detailed traffic volume data, and a majority of modern traffic forecasting models are built on complete volume observations.

Obtaining network-wide traffic volume over a period of time is nontrivial due to the limited deployment of expensive traffic volume sensors. Current sensor coverage rate hardly meets the demands of high spatial resolution of volume data. As traffic control strategy is generally performed at the network level (Aboudolas and Geroliminis, 2013), efficient and accurate traffic volume estimation approaches are therefore much-needed for analyzing multi-lane traffic network conditions. Distinct from traffic forecasting, in this paper we study network-wide multi-lane urban highway traffic volume estimation, which aims to predict traffic volume data at current time slot for the whole sensor network based on partial measurements of link traffic. This problem can be

---

viewed as installing virtue volume sensors on road segments without sensors to enhance spatial resolution of link traffic. Fig. 1 gives an example to illustrate the studied problem.
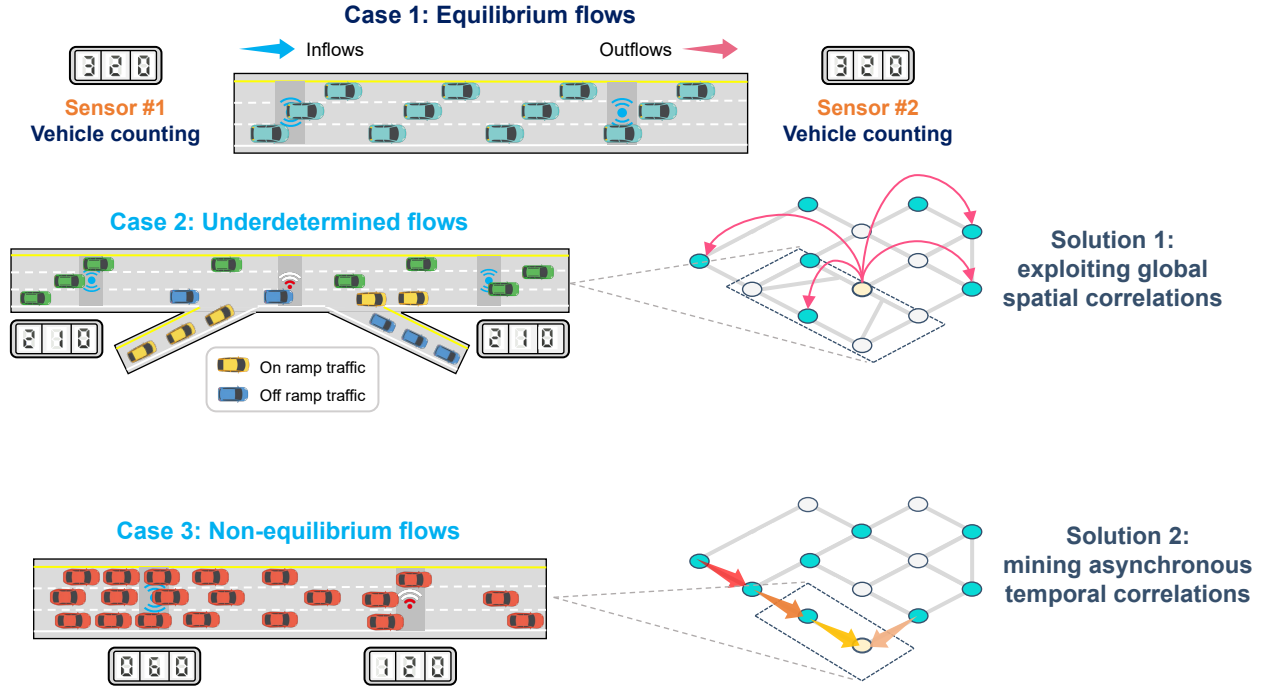

Figure 1: Illustration of highway traffic volume estimation problem. Three typical road segments of the highway network are shown to illustrate the three scenarios studied in this paper. (a) Case of equilibrium and determined traffic flows: the density between sensor #1 and #2 is constant so volumes of sensor #2 are identical to sensor #1. (b) Case of underdetermined traffic flows: due to lack of measurements at on/off ramps, arbitrary volume values may exist between the two sensors, even if the two detectors have the same or close reading. In this case, volumes in the middle position cannot be inferred directly from conservation equation. (c) Case of non-equilibrium traffic flows: influenced by congestion delay and different reactions of vehicles, the outflows are distinct from inflows, resulting in time-shifted readings of sensor #1 and #2. In general, cases (b) and (c) are more intractable because when sensor #1 or #2 is unavailable, their readings deviate from neighborhood values, which the focus of this paper is on. To tackle the two key problems, global spatial correlations and asynchronous temporal correlations are exploited by customizing a GNN model.

In light of the preceding analysis, a collection of pioneering works have been dedicated to tackle this challenge and achieved valuable results (Aslam et al., 2012; Zhan et al., 2016; Meng et al., 2017; Tang et al., 2019; Zhang et al., 2020). However, a majority of existing studies on this topic has focused on primarily validating the capability of a particular model to accurately estimate or forecast real-world traffic volume. The issue discussed much less than the overall accuracy is that of a model's ability of dealing with core challenges of volume estimation. We claim that some of the shortcomings of previous studies are not primarily their lack of overall accuracy, but more so their inferior performances on some critical and challenging scenarios, e.g., estimation of underdetermined and non-equilibrium traffic flows.

To elaborate the above view and emphasize the motivations of this paper, we discuss several inherent issues of network-wide traffic volume estimation that have not been fully solved in the literature:

Modeling semantic relationships between sensors. Estimating sensors' readings for unmeasured locations are typically cast into spatiotemporal traffic kriging in prior works (Appleby et al., 2020; Wu et al., 2021a; Nie et al., 2022b). Most recent approaches that target performing kriging on traffic speeds adopt a simple distance metric or function (e.g., graph Laplacian) to model the pair-wise correlations between sensors (Takeuchi et al., 2017; Lei et al., 2022; Nie et al., 2022b). Nevertheless, the interplay and semantic similarities between traffic volume sensors are much more complex (Saeedmanesh and Geroliminis, 2016) and volume loaded on a network is strongly associated with the underlying environment such as road geometric factors, driving directions and POIs. Besides, the stochastic and non-stationary nature of traffic flow pattern further complicates this problem. Although graph neural network (GNN) has been widely leveraged to perform spatial modeling in existing works (Appleby et al., 2020; Wu et al., 2021a), they are designed for general issues that are essentially different from the traffic estimation problem. Direct usage of an existing kriging model on this task might be less than satisfactory. As a consequence, more advanced spatial relations modeling techniques and a tailored model architecture that can handle complex spatial semantic correlations as well as physical constraints simultaneously in the context of spatiotemporal traffic volume data are much-needed.

Non-local spatial dependence for underdetermined traffic flows. Due to the limited observations and existence of routine choice, not all link flows can be directly inferred from conservation equations (see Fig. 1 case 2). In terms of systems of linear equations, this road network is *underdetermined* (Ng, 2012). Underdetermined

flows can arise on open paths which went fully undetected and could allow arbitrary volume values without violating the local flow conservation law if no further information is provided. For instance, traffic participants usually travel to/from POIs settled in the middle of two adjacent detectors, and these undetected behaviors cause volume records of adjacent sensors large inconsistency, where neighboring information is inadequate for accurate recovery. Fortunately, it is demonstrated that traffic flow interdependence among road segments can exist and evolve in a wider range of network (Wu et al., 2019; Wang et al., 2020b). In this case, how to identify globally-related sensors to can provide referable information for the estimation of underdetermined sensors.

Current GNN based kriging models typically consider locally proximal detectors and aggregates available information only from k-th order neighborhoods (Appleby et al., 2020; Wu et al., 2021a). Although a few techniques are developed for capturing spatial correlations globally on the graph for traffic forecasting models (Wu et al., 2019; Yu et al., 2019a; Bai et al., 2020), the requirement of complete historical data for recognizing hidden patterns renders them unfeasible for traffic volume estimation problem. As traffic speed and traffic volume are linked by traffic density, they usually co-evolve under certain rules. Two road segments with similar road network context and traffic speeds are prone to exhibit similar volume patterns, despite far distance on the graph. Compared to cross-sectional volume that requires careful and laborious vehicle counting, average spot speed or average link speed are much more accessible (Meng et al., 2017). By exploiting shared speed patterns among roads, well-established speed-volume relationships can be utilized to offer hints for estimating underdetermined flows, beyond the description of spatial proximity.

However, modeling speed-volume relationship has been studied for decades yet is far beyond solved. Most prior works assume speed-volume vary by locations linearly, and the correlations of speed patterns approximate to ones of volume patterns (Zhan et al., 2016; Meng et al., 2017; Zhang et al., 2020). In these works, they elaborate specific rules (e.g., Pearson correlation coefficient, or static adjacent graph) to simplify this problem. In practice, nevertheless, these correlations are found to be nonlinear (Zhu et al., 2022), heterogeneous (Ji and Geroliminis, 2012; Ramezani et al., 2015) and dynamically changing over time (Yao et al., 2019). For example, different congestion formation and dissipation patterns across the traffic network could produce heterogeneous macroscopic fundamental diagrams with multiple kinds of speed-volume curves. And different traffic demands during rush hours and off-peak hours lead to various intensities of correlations. Consequently, open questions remain about how to understand the entangled relationships between speed and volume in a transferable, adaptive, and elegant routine without exogenous assumptions on traffic flow characteristics, such as a fundamental diagram (Geroliminis and Daganzo, 2008).

Time-asynchronous correlations for non-equilibrium traffic flows. Intrinsic traffic flow phenomena are strongly associated with high-resolution volume estimation, and few studies go further on this issue. Apart from the inductive bias (or misfit) brought by machine learning models themselves, the behaviors of traffic flow could also affect the estimation accuracy to a large extent. Specifically, if the traffic density between two sensors keeps constant over space and time, then the incoming and outgoing flux/flow of each upstream and downstream sensor pair should be identical or matched at the same time slot, which is called *equilibrium flow* (see Fig. 1 case 1). Otherwise, there exists *non-equilibrium* traffic flows (Zhang, 1998; Zhang et al., 2009) caused by asynchronous response time of drivers, manifesting as the macroscopic congestion propagation of traffic flow. The affect of traffic congestion propagates from upstream segment to downstream segment with a time delay, and this propagation is determined by the road network topology and traffic speed (Ji et al., 2014). Such non-equilibrium flows could make the volume dependence of sensors inactivated within a short time window, or in other words, generating time-asynchronous correlations between upstream and downstream sensors (see Fig. 1 case 3). However, as presented in (Yang et al., 2023), recognizing a reasonable scope to consider the influence of propagation delay is not an easy task. And such delayed temporal correlations often vary by space and time, associating with the current traffic conditions. A prescribed value may be somewhat subjective. Besides, temporal correlations are often ignored or considered as a feature extraction module independent of spatial features in existing kriging or imputation works (Appleby et al., 2020; Wu et al., 2021a; Liang et al., 2022b), which are inadequate for consideration of time-asynchronous correlations.

Motivated by the above research gap, this study contributes to the state-of-the-art with novel perspective and methodologies. A GNN framework is proposed to serve as a data-driven, equation-free, and correlation-adaptive link traffic volume estimator. In summary, our contributions are four-fold:

- We model the traffic volume estimation as a kriging problem and identify two critical problems that affect the volume estimation significantly, i.e., underdetermined and non-equilibrium traffic flows. Two quantitative metrics are developed to measure their impacts.

- To improve the performance on underdetermined flows, we quantify the dynamic and nonlinear speed-volume relationships by proposing a graph attention based speed pattern-adaptive adjacent matrix (SPAM) construction module. After incorporating SPAM into graph convolution, our model can capture both local and non-local spatial relations simultaneously.

- To consider the impacts of non-equilibrium flows, we formulate a temporal masked and localized attention (Tatt) module to learn the potential impacts of previous congested traffic states. Collaborating with a gated

temporal convolution network (TCN), the time-asynchronous correlations between neighborhoods are well captured.

- Experiments are conducted on a real-world highway traffic volume dataset, and results indicate the state-of-the-art performances of our model on both overall accuracy and critical scenarios. Besides, ablation studies, sensitivity analysis, as well as interpretations of sample rate and learned graphs, further justify the rationality of model designs.

The rest of this article is organized as follows. Section 2 briefly reviews related works about traffic state estimation. Section 3 describes basic concepts and formulates the problem. Section 4 introduces our model architecture and details. In Section 5, we evaluate our model on a real-world dataset and ablation studies, sensitivity analysis as well as model interpretations are provided. Section 6 concludes this study and provides future directions.

2. Literature review

This section revisits the status quo about traffic data modeling and state estimation. This review covers three aspects: 1) traffic volume estimation methods based on heterogeneous data fusion; 2) spatiotemporal kriging models and graph neural networks for traffic data modeling; 3) dynamic traffic assignment and link flow observability methods.

2.1. Traffic volume estimation by data fusion

Unlike like traffic state imputation problem centering on estimating missing elements of multivariate time series, traffic volume estimation for locations without volume detectors is more crucial but challenging. In general, there exists two two categories of volume estimation models in the literature. The first is a fully supervised learning treatment (Aslam et al., 2012; Idé et al., 2016; Yi et al., 2019; Zhang et al., 2022), where various kinds of statistic models or machine learning models are established for volume regression with selected covariates. Another paradigm is the semi-supervised learning that graph-based approaches are desirable choices (Meng et al., 2017; Tang et al., 2019; Yu et al., 2019b; Zhang et al., 2020). These pioneering works prove that fusing different sources of traffic data is a viable option for volume estimation (Xing et al., 2022). However, most of them either rely on strong assumption (e.g., specific locations of volume sensors (Rostami-Shahrbabaki et al., 2020), availability of historical average data (Zhang et al., 2022)), or extra data that is laborious to collect (e.g., probe vehicle counts (Aslam et al., 2012), taxi trajectories (Meng et al., 2017; Yu et al., 2019b)), undermining their utility and effectiveness in real-world applications.

Particularly, there are a few works achieve this task by considering speed-volume relationships. It is widely acknowledged that there is a potential link between traffic speed and volume. Therefore, it is rational to attempt to infer volume from speed. To this end, some studies model spatial dependencies of traffic volumes with speed information. Meng et al. (2017) and Zhang et al. (2020) assume that traffic speed and volume co-evolve in a linear relationship and locally adjacent road segments/sensors with similar speed profiles share similar volume patterns. Zhan et al. (2016) describe this relationship by a calibrated fundamental diagram. Mahajan et al. (2022) predicts volumes from speeds directly with a deep learning model. In addition, Zhu et al. (2022) design a speed-volume joint imputation framework by setting a shared feature extraction and learning module, achieving volume reconstruction from speed. However, their method depends on clear correlations between volume and speed that a Greenshields model can be well-fitted. As shown in (Zhu et al., 2022; Mahajan et al., 2022), the estimation performances degrade dramatically when the speed-volume relationships change. To fully exploit the speed-volume relationship and filter the most useful speed information, a data-driven approach is preferable.

2.2. Graph neural networks and low-rank kriging models for network-wide traffic flow modeling

GNNs have driven giant advances in prior works for temporal and spatial analysis. In the literature, GNNs have been broadly adopted for traffic state forecasting (Yu et al., 2017; Li et al., 2017; Yu et al., 2017; Wu et al., 2019; Wang et al., 2020b) and imputation (Ye et al., 2021; Liang et al., 2022a,b; Chen and Chen, 2022). Among these works, Li et al. (2017) developed a diffusion convolution networks that model the diffusion process of traffic flow as random walks and enable ones to build GNNs for traffic flow on directed graph. Wu et al. (2019) further generalized the diffusion convolution model and proposed an adaptive graph learning module. These pioneering studies have a wide spectrum of applications and shed light on the versatility of GNNs to model network-wide traffic flows. Whereas, different from forecasting task with sufficient historical data for model training, volume estimation problem considered in this paper is much more tricky because indirect information is treated as predictor variables for unmeasured locations.

In the sphere of geostatistics, the problem of estimating values of unobserved locations is usually cast as spatial kriging (spatial regression). Most recently, kriging models have aroused great interests in transportation domain (Takeuchi et al., 2017; Wu et al., 2021a,b; Liang et al., 2022a; Lei et al., 2022; Nie et al., 2022b). To

demonstrate the inductive power of GNNs, Wu et al. (2021a) first proposed a kriging convolution network for spatiotemporal traffic speed. Wu et al. (2019) and Liang et al. (2022a) further improve the model performances with more advanced architectures. The key rationale of traffic kriging methods is to design a kernel function to capture spatial relationships between traffic states at different locations. However, almost all of them focus on traffic speeds that can be fully characterized by simple and static graphs. Unfortunately, traffic volume behaves much more complicated than speed that simplified and static rules are not applicable to model the correlations. As traffic volume is strongly associated with the current traffic condition and many other factors, direct usage of a kriging model on traffic volume estimation could produce undesirable results.

2.3. Dynamic traffic assignment and network flow conservation methods for link flow inference

In the area of transportation research, a typical link traffic estimation method falls into the category of traffic assignment. In general, dynamic traffic assignment (DTA) solves the travel times for each time interval of each road segment with user equilibrium assumptions and all-day link flows can also be obtained (Janson, 1991).

Despite more factors can be considered and more detailed information can be produced, DTA has stringent requirements for inputs. On the one hand, a complete picture of traffic demand is needed, but collecting all-day traffic demand is quite difficult (e.g., origin-destination survey). On the other hand, as the computation of DTA contains a path enumeration process (Wang et al., 2020a; Zhang and Qian, 2020), a large-scale vehicle path dataset is sometimes necessary. Therefore, origin-destination matrix estimation (Krishnakumari et al., 2020), or path reconstruction (Yang and Sun, 2015) procedure which are both intractable, could be a prerequisite. In DTA, the relationship of travel speed (time) and link volume is established by calibrating a Bureau of Public Roads (BPR) power function (Merchant and Nemhauser, 1978). This simplified assumption could deviate from the reality. In the meanwhile, as DTA method usually bases on path assignment at the lower level of a bi-level optimization model, these analytical or simulation solutions require high computational burden (Ros-Roca et al., 2022).

There is also another branch of link flow inference research that is of wide interest, i.e., the full link traffic flow observability problem (Ng, 2012). The goal of this problem is to identify the minimum number and combination of link sensors, to infer the link traffic for the whole network, using conservation equation only (Viti et al., 2014; Xu et al., 2016). We would like to clarify that our focus is different from these works. Link flow observability problems place emphasis on the flow distribution of a static network, and the time dimension is ignored. They can select the locations of sensors and usually assume intact observations without noises. The volume estimation is implemented by solving conservation equations and macroscopic simulation is widely used for validation (Salari et al., 2019; Shao et al., 2021).

While we consider more realistic scenarios and data collected in real-world traffic measurement systems, where the conservation equations may be inadequate. We estimate the all-day network-wide link flow series for a long time period in the perspective of kriging and data reconstruction, given the detector already installed. Besides, our work targets at higher-resolution sensor-level measurements and the sensor network, which is a basic unit for modern traffic control and management strategies.

3. Preliminaries and problem descriptions

In this section, we first introduce basic concepts used herein and formulate the traffic volume estimation problem explicitly. Then we present the rationale for how GNN can treat traffic volume estimation as a kriging problem.

3.1. Graph representation of traffic network

In this work, we formulate the network-wide traffic volume estimation task as a graph modeling and reconstruction problem. As traffic flows load on a road network that connects related detectors rather than a 2D grid, we can abstract their relationships using a directed graph. Please note that we use the terms "sensor", "detector", and "node" interchangeably throughout the paper. A sensor network can be modeled as a weighted directed graph $\mathscr{G} = \{\mathscr{V}, \mathscr{E}, \mathbf{A}\}$, where vertices $\mathscr{V}$ denote N sensors or detectors of the network, edges $\mathscr{E}$ indicate the connectivity between nodes and their directions are assigned according to the moving directions of traffic flow. $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the weighted adjacent matrix (weights of edges) describing the correlations of nodes, and it is usually treated as a function of network topology or travel distance.

In the literature, a commonly used distance-based adjacent matrix $\mathbf{A}$ is computed as:

$$a_{ij} = \begin{cases} \exp\left(-\left(\text{dist}(v_i, v_j)/\delta\right)^2\right), & \text{if } \text{dist}(v_i, v_j) \leq \epsilon, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $\text{dist}(\cdot)$ is the travel distance between two sensors, δ is a scaling parameter which can be selected as the standard deviation of distances, and ϵ is the threshold. As we can seen, $\mathbf{A}$ is asymmetric because the sensor graph is directed.

Traffic flow operating on the graph can be recorded as graph signals. We use $\mathbf{X}_v \in \mathbb{R}^{N \times T}$ and $\mathbf{X}_s \in \mathbb{R}^{N \times T}$ to represent traffic volumes and speeds produced at N locations during T time intervals. Accordingly, each element like $x_{v_n}^t$ is the volume of location n at time window t. To obtain a fine-grained spatiotemporal traffic volume estimation, we set the time window size as 5 minute in this work.

3.2. Network-wide traffic volume estimation with speed information

This paper centers on performing spatial estimation of traffic volume at locations without volume detectors and we can discuss this problem in the context of spatial kriging (Nie et al., 2022b). Let $\mathscr{V}^o$ be the locations with volume sensors, $\mathscr{V}^u$ be the unobserved locations, and $n_o = |\mathscr{V}^o|$, $n_u = |\mathscr{V}^u|$ are number of observed/unobserved locations, the goal of traffic volume estimation/kriging is:

$$\begin{aligned}
\widehat{\mathbf{X}}_v &= \mathscr{F}(\mathbf{X}_v^i, \mathbf{X}_s^j, \mathscr{G}),\ i \in \mathscr{V}^o, j \in \mathscr{V}^o \cup \mathscr{V}^u, \\
\text{s.t. } \widehat{\mathbf{X}}_v^i &= \mathbf{X}_v^i,
\end{aligned} \tag{2}$$

$\widehat{\mathbf{X}}_v \in \mathbb{R}^{N \times T}$ is the estimated volume of entire graph and $\mathscr{F}$ is the targeted neural network model. In this paper, we keep the same assumption as in (Meng et al., 2017) and (Zhang et al., 2020), that traffic speed records are supposed to be complete and available for all road segments (sensors). Actually, this assumption is easy to satisfy in reality. On the one hand, a full picture of traffic speeds can be readily measured and recorded by float cars or crowdsourcing vehicles (Liu et al., 2019; Zhang et al., 2020; Yu et al., 2020). Such Lagrangian measurements can account for a large proportion of urban road network, with a high data updating frequency. By conducting a map-matching procedure, the speed data is applicable in practice. On the other hand, recent studies aiming at estimating network-wide traffic speed have demonstrated that full-scale of speeds can be recovered with high accuracy, having limited amount of observations (Lei et al., 2022; Nie et al., 2022b; Wu et al., 2021b; Liang et al., 2022a). Given partially observed traffic volume and speed, high-quality and complete speed data can be estimated at first, using an already established spatiotemporal speed kriging model.

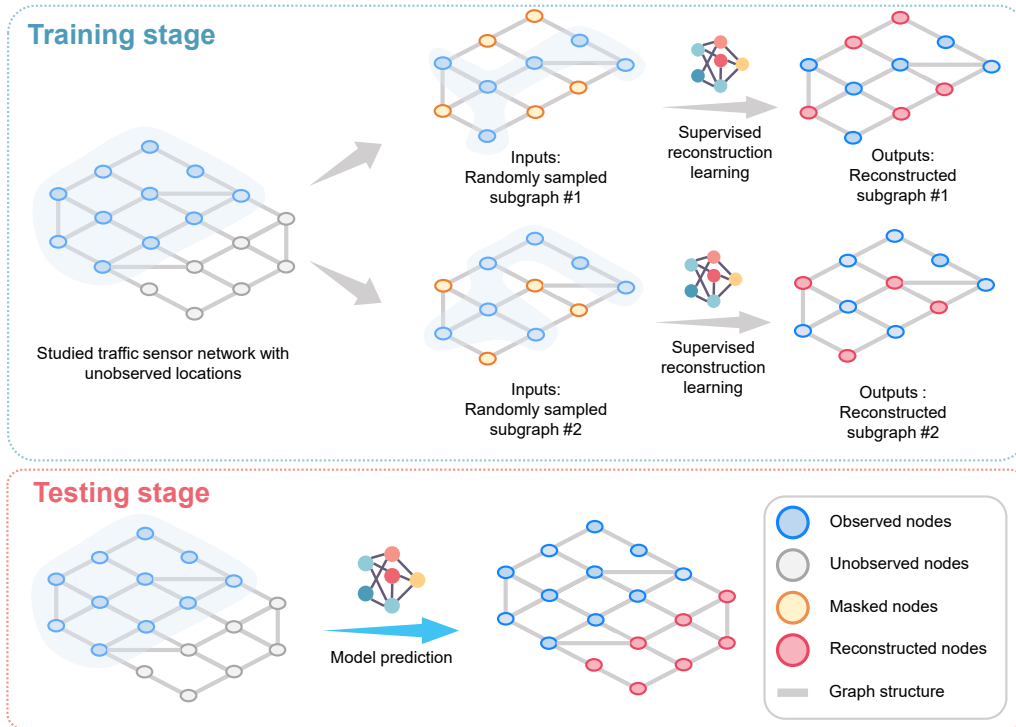### 3.3. Treating volume estimation as kriging

Figure 2: Explanation of the working process of volume kriging model. Only observed graph information (data and location) is provided during training stage. By randomly sampling different sub-graphs, the model can inductively learn to reconstruct the whole graph and generalize to unseen nodes. A trained model can make predictions at newly added locations, even if they do not appear in the training graph.

To solve Eq. (2), in this work we develop a GNN kriging architecture to handle highway traffic volume estimation problem, and our model belongs to the category of kriging convolution networks (KCN) (Appleby et al., 2020; Wu et al., 2021a). Typically, KCN is used for spatial modeling of non-cumulative data, such as temperature and traffic speed. To explain the rationality of kriging on traffic volume, in what follows we will first demonstrate how our model can be used for volume estimation under kriging settings.

Generally, volume estimation problem can be solved by a regression method in which some selected independent variables such as historical traffic measurements, traffic demand, network features, as well as time information, are used to fit the target volume value (Aslam et al., 2012; Idé et al., 2016; Zhang et al., 2022).

$$\mathbf{X}^i = \mathrm{AGG}(\{\mathbf{x}_j : j \in \mathcal{N}_i\} \cup \{\mathbf{z}_j : j \in \mathcal{F}_i\}),$$
$$\mathbf{Y}^i = \mathrm{REG}(\mathbf{X}^i, \mathcal{G}), \tag{3}$$

where $\mathrm{AGG}(\cdot)$ denotes a parameterized function that aggregates all available information from the neighborhood $\mathcal{N}_i$ and relevant features $\mathcal{F}_i$ of node i. respectively. Then combining with the graph information $\mathcal{G}$, $\mathrm{REG}(\cdot)$ produces the target volume values $\mathbf{Y}^i$.

To use as few external features as possible, which are usually laborious to collect, e.g., full-scale traffic demand, the input volume value can become as the only feature to consider, and the label is also the true volume value, problem (3) turns into the following:

$$\min_{\Theta} \sum_{i \in \mathscr{V}} \mathrm{LOSSFUNC}(\mathbf{X}^i, \hat{\mathbf{X}}^i),$$
$$\hat{\mathbf{X}}^i = \mathrm{GNN}(\mathbf{X}, \mathbf{M}, \mathcal{G}|\Theta), \tag{4}$$

where Θ is the model parameter needs to learn, $\mathrm{LOSSFUNC}(\cdot)$ measures the discrepancy of estimated and true values. Here, GNN serves as the predictive function to produce estimation. By this supervised learning setting, GNN learns the message aggregation mechanism on the graph and can generalize to new node if its location is given. We can see that problem (4) has equivalent form to problem (3), with specific input features. In this work, we only choose the historical traffic speed data and network geometries as external features, which are readily available for most urban road networks.

One can find that in this volume kriging problem, the labels are true volume data for all locations, which are also the inputs to the GNN model. However, directly feed the labels into model training will produce a trivial model without learning anything. Therefore, a random sampling matrix $\mathbf{M}$ is introduced to generate diverse training samples, see Fig. 2. In this sense, this problem resembles a graph reconstruction learning process.

A trained or calibrated volume estimator should have the ability to predict the label of new test data point at any location using the observed training labels directly. Fortunately, the inductive power of GNN makes it possible (Wu et al., 2021a). Besides, the training of GNN in Eq. (4) allows missing values of the observations, and real-time implementation is also feasible.

Although treating volume estimation as kriging is reasonable, however, three key characteristics of highway traffic volume estimation problems hinder the direct usage of a KCN on this topic. In short, traffic flow characteristics and underdetermined problem will affect $\mathrm{AGG}(\cdot)$, e.g., aggregating locally proximal information is insufficient when neighbor sensors vary greatly in space, and non-equilibrium phenomenon will impact $\mathrm{REG}(\cdot)$, e.g., the inputs and labels are time-shifted on temporal dimension. In the rest of this section, we will describe how these problems affect the estimation performance and how our model designs tackle such challenges.

4. Methodology

In this section, we first develop two metrics that are used to evaluate the undetermined and non-equilibrium problems of highway traffic volume estimation. In what follows, we elaborate the proposed **S**patio**T**emporal **C**orrelation-**A**daptive **G**raph **C**onvolution **N**etworks (STCAGCN) model. The core concept of STCAGCN is to make use of the redundancy of sensor networks and rich correlations of traffic volumes, i.e., unobserved volumes can be estimated by exploiting spatial relationships from both globally or locally correlated sensors, and time-asynchronous correlations from previous traffic states.

Fig. 3 shows the holistic architecture of STCAGCN. It consists of two basic components: (1) spatial information aggregation block for alleviating underdetermined problem with learned speed-volume relationships. (2) time-asynchronous correlations extraction block for coping with non-equilibrium flows.

4.1. Measuring undetermined and non-equilibrium phenomena

As introduced in section 1, we attribute the core challenges of volume estimation to underdetermined and non-equilibrium traffic flows. To examine the existence and evaluate the degree of these phenomena, we first develop the spatial smoothness and temporal alignment metrics.

A common belief is that adjacent sensors report similar traffic volumes because flows are supposed to be conserved if vehicles only traverse between them. This can be recognized as a spatial stationary property of traffic network (Zhang et al., 2020). However, traffic participants often travel to/from middle POIs of urban networks or on-off ramps of highway networks. If there are only a few observations, the sensor network can
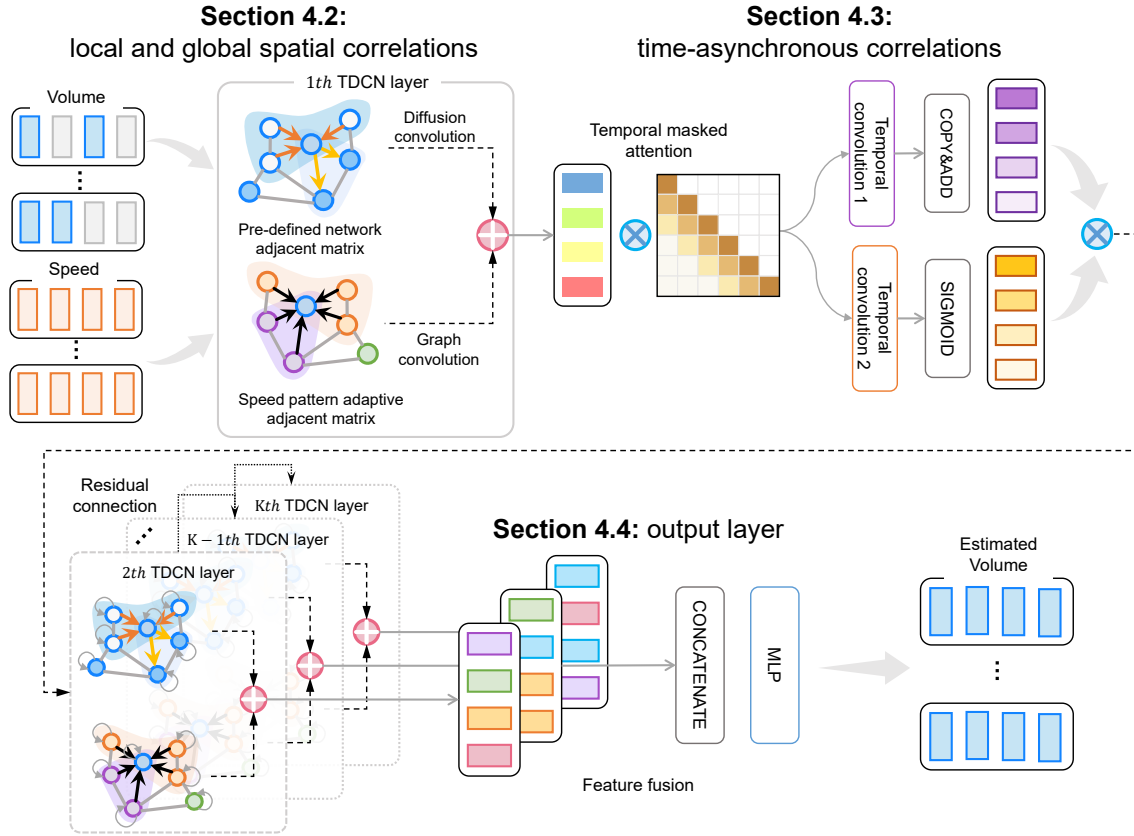
Figure 3: Overall architecture of proposed STCAGCN model. One initial TDCN layer, one Tatt-TCN layer and several stacked TDCN layers are applied. The inputs are masked traffic volume and complete traffic speed data, and STCAGCN reconstructs the complete volume for whole network.

be viewed as a underdetermined linear system and volume values can not be obtained from conservation laws directly. The spatial smooth index is thereby adopted to measure this nonstationary phenomenon.

Following (Zhang et al., 2020), a weighted and directed spatial smoothness index is modified as follows:

Definition 1. *(Weighted and directed spatial smooth index): The weighted and directed spatial smooth index (WDSSI) quantifies the spatial local stationary (or conservation) property of the traffic volume loaded on a road network based on the ground truth volume values and network geometry, which can be formulated as:*

$$\text{WDSSI}(i) = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{\sum_{j \in \mathcal{N}_i} a_{ij} x_{v_j}^t}{\sum_{j \in \mathcal{N}_i} a_{ij}} - x_{v_i}^t \right| / x_{v_i}^t, \ \forall i \in \mathcal{V}, \tag{5}$$

where $\text{WDSSI}(i)$ *is the value for the i-th location of the sensor network as well as the i-th node on the affinity graph, $\mathcal{N}_i$ is the set of neighbors of node i either on inflow or outflow direction.*

WDSSI considers the proximity and observability of a sensor network. If the road segment between a sensor and its adjacent sensors is fully conserved, the WDSSI value should close to zero. In highway or freeway network, there exists a large amount of on/off ramps that are not equipped with sensors. Such positions could cause large discrepancy of the measurements between adjacent sensors, and we use WDSSI to identify the exist of this underdetermined issue.

Dynamic time warping is a widely adopted distance metric for time series data that can be used to compare time series of different lengths and rhythms (Rakthanmanon et al., 2012). Generally speaking, if two time series are not time-aligned, timestep based distance metrics like Euclidean distance are inapplicable. In this case, DTW will automatically warp the time series (i.e., local scaling on the time axis) so that the morphology of the two series is as consistent as possible to obtain the maximum possible similarity. Based on dynamic programming algorithm, the DTW calculation is given as follows.

Definition 2. *(Dynamic time warping distance): Given two time series* $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ *and* $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ *with length n and m, let $d_{ij} = (x_i - y_j)^2$ be the Euclidean distance of two series points, then the accumulated distance $\gamma(i, j)$ can be defined as:*

$$\gamma(i, j) = d_{ij} + \min(\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)), \tag{6}$$

i.e., the sum of the Euclidean distance of points i and j and the cumulative distance of the smallest neighboring elements that can reach this point. After increment of i, j, the final accumulated distance $\gamma(n, m)$ has the best alignment and the accumulated distance along the warping path is minimized.

Under this definition, if congestion delay occurs between two sensors, their recorded signals are supposed to be time-asynchronous and should have small DTW distance but large Euclidean distance. With regard to this, we propose a time alignment indicator (TAI) to examine if two time series from upstream and downstream are time-asynchronous (i.e., non-equilibrium flows occur in the middle road segment). The smaller the TAI is, the higher possibility the target sensor is influenced by traffic propagation delay.

Definition 3. *(Time alignment indicator): The time alignment indicator quantifies the temporal local alignment (or asynchronous correlation) of volume between upper- and lower-stream detectors:*

$$\text{TAI}(i) = \frac{\text{DTW}(i, j)}{\text{EUC}(i, j)}, \; j \in \mathscr{N}_i^u, \tag{7}$$

where $\mathscr{N}_i^u$ is the nearest upstream neighbor of node i, $\text{DTW}(\cdot), \text{EUC}(\cdot)$ are DTW and Euclidean distance matrix respectively.

These two metrics help us identify and understand the occurrence of underdetermined and non-equilibrium problems, elaborate corresponding solutions, and evaluate model performances on each task.

4.2. Modeling both local and global spatial relations for estimation with underdetermined flows

4.2.1. Task-specific diffusion convolution for local spatial dependency

Graph convolution networks (GCNs) have been widely applied in recent spatial modeling works (Li et al., 2017; Yu et al., 2017; Wu et al., 2019; Xue et al., 2022). By aggregating features from neighbor nodes on the graph, GCNs transform graph signals in feature space and model local spatial dependencies in a data-driven routine. For any node on a sensor graph, both inflows and outflows decide its instant traffic state. Therefore, directed nature of real-world traffic networks encourages a model that can deal with asymmetric adjacent matrix (Nie et al., 2022b).

Diffusion convolution network (DCN) proposed by (Li et al., 2017) provides an effective tool for this problem. DCN models the propagation of traffic flow as a finite step diffusion process to facilitate convolution on directional graphs. According to (Wu et al., 2019), for an input graph signal matrix $\mathbf{X} \in \mathbb{R}^{N \times C_{in}}$, the diffusion convolution operator can be generalized into the following form:

$$\mathbf{Y} = \sum_{k=0}^{K} \left(\mathbf{A}_f^k \mathbf{X} \mathbf{W}_{kf} + \mathbf{A}_b^k \mathbf{X} \mathbf{W}_{kb} \right), \tag{8}$$

where $\mathbf{A}_f^k, \mathbf{A}_b^k$ are the forward and backward transition matrix at k-th diffusion step, K is the total steps of diffusion, $\mathbf{W}_{kf}, \mathbf{W}_{bf} \in \mathbb{R}^{C_{in} \times C_{out}}$ are learnable parameters of DCN layer. The bidirectional transition matrices can be obtained from the adjacent matrix of sensor graphs:

$$\mathbf{A}_f = \frac{\mathbf{A}}{\text{rowsum}(\mathbf{A})}, \; \mathbf{A}_b = \frac{\mathbf{A}^\top}{\text{rowsum}(\mathbf{A}^\top)}. \tag{9}$$

$\mathbf{A}_f$ and $\mathbf{A}_b$ are in fact the row-wise normalization of $\mathbf{A}$ and $\mathbf{A}^\top$, in order to sum the weights to one. This diffusion convolution operator can take into account the influence of inflows and outflows simultaneously.

Recall that the propagation of traffic flow is directed and along each travel direction, DCN only considers inflows and outflows of nodes, but ignores the fact that traffic states from opposite moving direction could be quite different. Considering this, we modify the basic DCN with a Task-specific Diffusion Convolution Network (TDCN) in three aspects: diffusion direction, setting of self-loop, and lane-wise normalization. We will demonstrate the effectiveness of these simple but useful modifications in section 5.

As discussed in section 2, DCN has driven tremendous achievements in traffic speed forecasting and kriging applications. Almost all these prior works construct the adjacent matrix $\mathbf{A}$ in Eq. (1) by computing the Euclidean distance or travel distance between sensors in a direction-agnostic way. In other words, they treat all sensors from different driving directions, upstream-downstream segments indiscriminately. This treatment is acceptable for traffic speeds, but could be inadequate for volumes that are strongly associated with the road network characteristics. In this work, we confine the diffusion process to the lane where the detector is located. In other words, TDCN only aggregates neighboring features along specific driving direction. As a consequent, a new asymmetric adjacent matrix $\widetilde{\mathbf{A}}$ can be constructed:

$$\widetilde{a}_{i,j} = \begin{cases} a_{ij}, & \text{if } i \text{ and } j \text{ are in the same driving direction}, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Typically, historical records are important input features for traffic forecasting task, so after fusing available information from neighborhood detectors, the original features/inputs of each sensor itself need to be incorporated and the mixed features are feed into the next GCN layer. Given this, the adjacent matrix $\mathbf{A}$ is supposed to contain a self-loop for each node (Yu et al., 2017). However, with regard to our kriging problem, if we feed the ground truth values of unobserved locations to the neural network during model training stage, this model would not learn any useful information but only outputs the input itself. Hence, TDCN has no access to the inputs of unobserved or masked locations (details about random masking will appear in subsection 4.4.2) in the first layer and the adjacent matrix of first-layer should not have self-loops.

In addition, another important impact factor of traffic volume is the number of lanes. Since cross-sectional volume directly relates with the lane number, a wider road segment with more lanes will record more passing vehicles within a time interval. Rather than using lane number as an indicator to predict volume directly, we normalize the input volume data by lane numbers thus per lane volume is used for training and testing.

After considering all these special characteristics of this task, we customize the TDCN layer as following:

$$
\begin{aligned}
\mathbf{H}^1 &= \sum_{k=0}^{K} \left(\widetilde{\mathbf{A}}_{f0}^k \mathbf{X} \mathbf{W}_{kf0} + \widetilde{\mathbf{A}}_{b0}^k \mathbf{X} \mathbf{W}_{kb0} \right) \oslash \mathbf{e}, \; l = 0 \\
\mathbf{H}^{l+1} &= \mathrm{a} \left(\sum_{k=0}^{K} \left(\widetilde{\mathbf{A}}_{f}^k \mathbf{H}^l \mathbf{W}_{kf} + \widetilde{\mathbf{A}}_{b}^k \mathbf{H}^l \mathbf{W}_{kb} \right) \right), \; l \geq 1
\end{aligned}
\tag{11}
$$

where $\mathbf{e} \in \mathbb{R}^N$ denotes a vector whose element is the lane number of each sensor, $\oslash$ denotes the element-wise division, $\widetilde{\mathbf{A}}_{f0}, \widetilde{\mathbf{A}}_{b0}$ are adjacent matrices without self-loop, $\mathbf{H}^l$ is the output spatial features of l-th DCN layer, and $\mathrm{a}(\cdot)$ denotes the nonlinear activation function, e.g., ReLU function. By stacking several TDCN layers, GCN can capture multi-hop spatial correlations on the graph and the features of each node contain information from neighbor nodes.

In practice, to encourage parallel computation and mini-batch treatment, the implementation of TDCN can be extended to high-dimensional tensors. For input batch of graph signal embedding $\mathcal{H} \in \mathbb{R}^{B \times N \times T \times C_{in}}$ with batch size B, number of nodes N, sequence length T and feature dimensions C_{in}, Eq. (11) can be expressed as:

$$
\mathcal{H}^{l+1} = \mathrm{a} \left(\sum_{k=0}^{K} \widetilde{\mathcal{A}}_{f}^k \otimes \mathcal{H}^l \times_4 \mathbf{W}_{kf} + \widetilde{\mathcal{A}}_{b}^k \otimes \mathcal{H}^l \times_4 \mathbf{W}_{kb} \right), \; l \geq 1
\tag{12}
$$

where $\otimes$ is the batch-wise matrix multiplication operation, $\times_4$ denotes the mode-4 tensor-matrix product.

As can be seen, Eq. (11) only aggregates localized spatial information and suitable for dealing with determined flow locations where neighborhood sensors can provide sufficient information. When confront with the underdetermined scenario, observations in the proximity of target sensor become less informative.

4.2.2. Quantifying speed-volume relationships by speed pattern-adaptive graph construction module
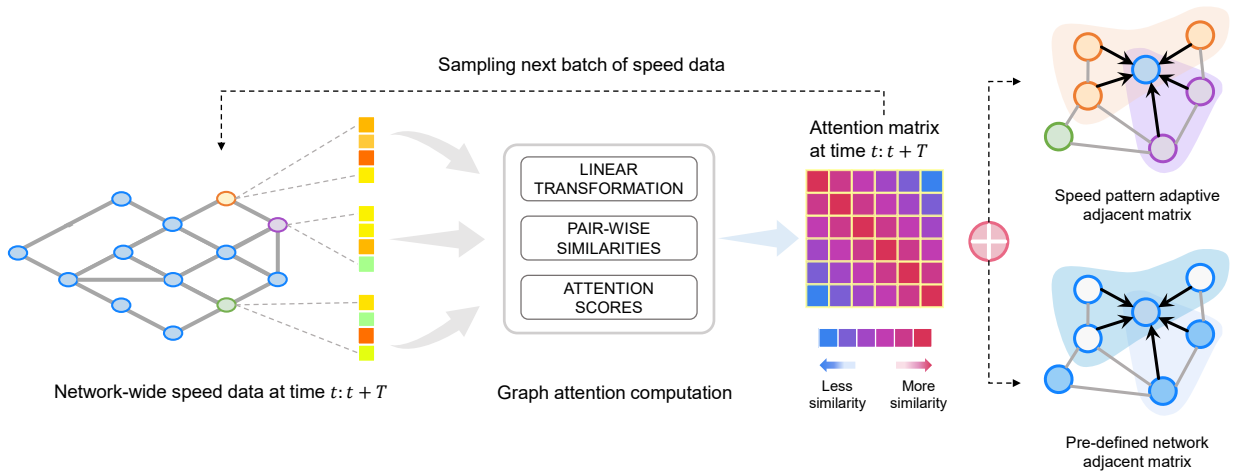

Figure 4: Mechanism of speed pattern adaptive matrix construction module. SPAM is constructed for each input speed batch, thereby changing over time.

A shortcoming not to be overlooked of DCN is that it highly depends on the pre-defined adjacent matrix (based on distance metric or correlation analysis). The pre-defined graph could be subjective and cannot fully mirror the underlying semantic similarity of items. Previous works have confirmed that the specification of

adjacent matrix has significant impacts on the performances of GCN (Wu et al., 2019). To alleviate this restriction, many studies about traffic forecasting resort to the internal similarities of data and learn a semantic graph as a complementary (Wang et al., 2020b; Bai et al., 2020; Liang et al., 2022b). However, the adaptively learned graph in these works are acquired from historical input data and can not be generalized to our case that no historical data is available for unmeasured locations. Therefore, external data source like traffic speeds need to be incorporated as an auxiliary. As stated in section 3, in this paper we assume the historical speed records are available for all road segments. These speed information provide opportunities for revealing the underlying correlations of traffic volume.

A natural question comes up: how we use the underlying relationship between traffic speed and volume to help the estimation of underdetermined volume? Modeling the relationship between speed and volume is a longstanding but open question in traffic studies (Weng and Meng, 2011; Fu et al., 2020). There are some pioneering works introduce the usage of simplified speed-volume relationship to help estimate the network-wide volume (Zhan et al., 2016; Meng et al., 2017; Zhang et al., 2020; Zhu et al., 2022), and their approaches can be divided into two categories: 1) adopting a deterministic model to describe this relationship; 2) estimating volume from speed directly. Both of these two considerations rely on a well-fitted speed-volume model (such as well-defined fundamental diagrams (Geroliminis and Sun, 2011)) and the estimation accuracy is significantly affected by the variance of speed-volume scatter. Moreover, as highway traffic states form a high-dimensional nonlinear dynamical system, with time-evolving spatial patterns (Wang and Sun, 2023), these kinds of static and linear methods fail to reveal the dynamically varying, non-exclusive, nonlinear speed-volume relationships.

To circumvent these restrictions, we propose a Speed Pattern Adaptive adjacent Matrix (SPAM) learning module to quantify the underlying speed-volume correlations directly from data, which is shown in Fig. 4. Specifically, inspired by the graph attention (GAT) architecture (Veličković et al., 2017), SPAM first computes the mutual importance of each node's speed features, and then mapping the similarity of speed features to the similarity of volumes via end-to-end stochastic gradient descent, discarding all structural and prior information about the sensor graph.

Given the input traffic speed for all sensors $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}, \mathbf{x}_i \in \mathbb{R}^D$, where D is the feature dimension, N is the node numbers, SPAM first computes the query vectors $\mathbf{q}$ and key vectors $\mathbf{k}$ by linear transform:

$$\begin{aligned} \mathbf{q} &= \mathbf{W}^q \mathbf{x}_i, \\ \mathbf{k} &= \mathbf{W}^k \mathbf{x}_j. \end{aligned} \tag{13}$$

Then, an attention scoring function $\text{att} : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is used to calculate the attention scores, which represent the similarity or importance weight of node j's feature on node i. Here, we select the scaled dot product (Vaswani et al., 2017) as the attention function. Followed by a normalization function, the attention scores are computed as:

$$\alpha_{ij} = \text{Softmax}(\text{att}(\mathbf{q}, \mathbf{k})) = \text{Softmax}(\frac{\mathbf{q} \cdot \mathbf{k}^{\mathsf{T}}}{\sqrt{d_k}}), \tag{14}$$

where d_k is a scaling constant.

Eq. (14) in effect computes the pair-wise similarity of each transformed speed vector and normalize it, making sure the sum is one. In SPAM, we utilize a shared weight matrix $\mathbf{W}$ to convert the queries and keys to the same space, then apply a LeakyReLU activation function to add nonlinearity. And we use the product of $\mathcal{L}$-2 norm as the scaling constant. Formally, given the speed data $\mathbf{X}_s$ of all sensors, the SPAM $\mathbf{A}_{spa}$ as well as the attention weights matrix can be computed by:

$$a_{ij}^{spa} = \frac{\exp(\text{LeakyReLU}((\mathbf{W}\mathbf{x}_{\mathbf{s}i})^{\mathsf{T}} \cdot \mathbf{W}\mathbf{x}_{\mathbf{s}j})/(\|\mathbf{W}\mathbf{x}_{\mathbf{s}i}\|\|\mathbf{W}\mathbf{x}_{\mathbf{s}i}\|))}{\sum_{j'} \exp(\text{LeakyReLU}((\mathbf{W}\mathbf{x}_{\mathbf{s}i})^{\mathsf{T}} \cdot \mathbf{W}\mathbf{x}_{\mathbf{s}j'})/(\|\mathbf{W}\mathbf{x}_{\mathbf{s}i}\|\|\mathbf{W}\mathbf{x}_{\mathbf{s}i}\|))}, \tag{15}$$

where $\|\cdot\|$ is the $\mathcal{L}$-2 norm of vector. One can find that this attention matrix $\mathbf{A}_{spa}$ indicates the similarity between features $\mathbf{x}_{\mathbf{s}i}$ and $\mathbf{x}_{\mathbf{s}j}$, and it ranges from 0 to 1 with summation to 1, which can also be regarded as an adjacent matrix.

By developing SPAM, we provide a calibration-free way to infer the inter-dependence of volume from speed data automatically. In STCAGCN, we compute SPAM for each input speed batch and thereby constructing dynamic graphs over time. And the activation function enables SPAM to achieve nonlinear mapping from speed correlations to volume correlations. Additionally, different from pre-defined graphs that are highly localized, SPAM connects globally related nodes adaptively, even they are remote in the sensor graph.

As SPAM is obtained from the speed data of all sensors, for batch forms of input, the produced SPAMs change over time with size $\mathcal{A}_{spa} \in \mathbb{R}^{B \times N \times N}$. Combining SPAM with TDCN, the final spatial feature extractor at l-th ($l \geq 1$) layer can be formulated as:

$$\mathcal{H}^{l+1} = \mathrm{a}\left(\sum_{k=0}^{K} \widetilde{\mathcal{A}}_f^k \otimes \mathcal{H}^l \times_4 \mathbf{W}_{kf} + \widetilde{\mathcal{A}}_b^k \otimes \mathcal{H}^l \times_4 \mathbf{W}_{kb} + \mathcal{A}_{spa}^k \otimes \mathcal{H}^l \times_4 \mathbf{W}_{spa}\right). \tag{16}$$

In Eq. (16), TDCN performs forward and backward diffusion convolution to capture localized correlations, while SPAM conducts single graph convolution to extract global correlations. Through stochastic gradient descent, $\mathbf{A}_{spa}$ is trained to filter most relevant features on the whole graph, thereby enlarging the receptive field of GCN. In this way, we inject the speed information to STCAGCN adaptively and can capture multiresolution spatial dependencies in a unified module. Graph convolution is performed on each time slot of $\mathcal{H}[:, t, :, :]$, which can be parallelized across time.

By this design, STCAGCN not only can correlate locally adjacent sensors, but also globally relevant sensors with similar traffic patterns. When dealing with underdetermined flow locations where local conservation is less effective, global information provides a useful complementary. When dealing with determined flows, local information plays a dominating role and proper features are aggregated by TDCN and feed into the next layer. Distinct from predicting volumes directly from speeds, SPAM does not assume a well-developed speed-volume scatter, or fundamental diagram, which is a model-free routine.

4.3. Capturing time-asynchronous correlations by gated temporal convolution and temporal masked self-attention

Traffic volumes are significantly affected by the formation and dissipation of traffic congestion. If traffic flows operate in free-flow speeds and the distance between upstream and downstream sensors is close enough that all vehicles can pass through within the aggregating time window, or in the case traffic is operating in an equilibrium state, it is expected that the inflows and outflows are the same at any time intervals and the volume profiles of them should be identical. While when congestion occurs, vehicles could show different react times and some vehicles can not reach downstream sensor within the time window. This non-equilibrium phenomenon is strongly associated with the distance, traffic density and speed. In this case, we claim that the volume of target sensor at current time interval has time-asynchronous correlations with the volume of neighborhood sensors at previous time steps. Most recently, Yang et al. (2023) also notice this phenomenon in traffic forecasting task and suppose traffic propagation could affect other road segments with a flow propagation delay that depends on traffic speed and network topology. As a result, flow dependencies of some roads may be inactive during the current time window due to traffic propagation delay.
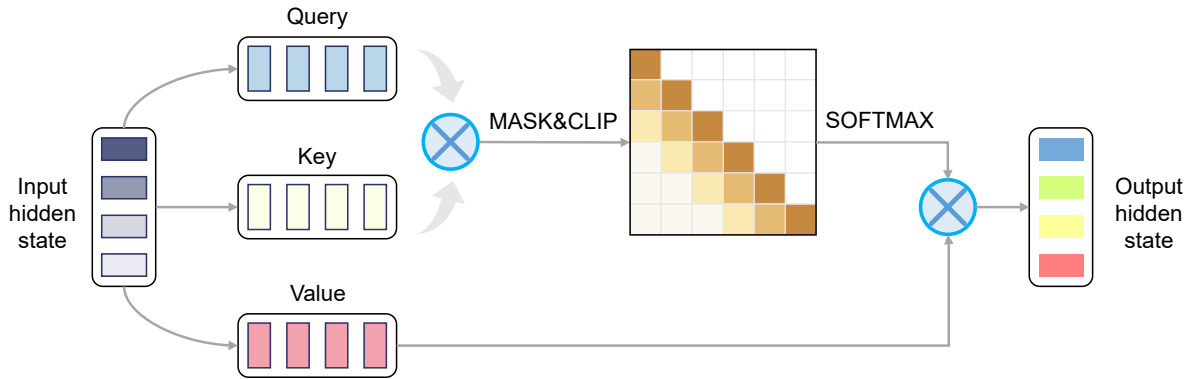

Figure 5: Mechanism of temporal masked self-attention module. Each input temporal embedding is weighed by attention scores.

As discussed in (Yang et al., 2023), obtaining a reasonable time window to consider the impact of congestion delay is a non-trivial task. Besides, traffic state correlates with its previous states and such correlations vary with the traffic status at different times, e.g., during rush hours, the influence of congestion could exist a few hours. In order to capture the the propagation dynamics of traffic congestion and exploit time-asynchronous correlations adaptively to tackle the non-equilibrium phenomenon, we customize a masked and localized temporal self-attention module (Tatt), to dynamically obtain temporal embedding. Attention mechanism has been successfully applied in sequence and time series data modeling (Zheng et al., 2020; Hao et al., 2020; Wang et al., 2020b). Given a temporal hidden state slice $\mathbf{Z} = \mathcal{H}[b, n, :, :] \in \mathbb{R}^{T \times C}$ produced by TDCN layer, we have:

$$\mathbf{E}_{att} = \frac{\mathbf{ZW}_t \cdot (\mathbf{ZW}_t)^{\mathsf{T}}}{\sqrt{d_t}} \tag{17}$$

where $\mathbf{E}_{att}$ is the unnormalized attention weight matrix, and a shared parameter $\mathbf{W}_t$ is used for projection of both key, value and query matrices. To avoid numerical overflow of Softmax function, we set the scale parameter as $\sqrt{d_t} = \|\mathbf{ZW}_t\|_F^2$.

After that, a time index masking matrix $\mathbf{M}_t$ is constructed to consider the casualty of time series, i.e., we take into account the impacts of previous time steps and avoid the future states affect the current state (Zheng et al., 2020), which is given by:

$$m_{ij} = \begin{cases} 0, & \text{if } t_i \leq t_j, \\ -\infty, & \text{otherwise.} \end{cases} \tag{18}$$

By adding $\mathbf{M}_t$, we can extract the lower triangular part of $\mathbf{E}_{att}$. Finally, the attention weighted hidden state is computed by applying a clip operation, normalization and weighted aggregation in sequence:

$$\mathbf{Z}_{att} = \text{SOFTMAX}\left(\text{TOPk}(\mathbf{E}_{att} + \mathbf{M}_t)\right) \cdot (\mathbf{Z}\mathbf{W}_t), \tag{19}$$

in which $\text{TOPk}(\cdot)$ is a attention clip operation that only the largest k weights are kept, to ensure the model pay more attention on most relevant time slots and adapt for temporal features that have marginal influences on current time slot, e.g., free-flow intervals. Fig. 5 illustrates the workflow of Tatt layer. In vector form, Eq. (19) can also be formulated as:

$$\mathbf{z}_i = \sum_{j \in \mathcal{N}_{t_i} \cap \mathcal{N}_{k_i}} e_{ij}(\mathbf{W}_t \mathbf{z}_j), \tag{20}$$

where $\mathcal{N}_{t_i}$ is the set consists of time points before t_i, and $\mathcal{N}_{k_i}$ denotes the set contains the largest k values in i-th row of $\mathbf{E}_{att}$. $\mathbf{z}_i$ is the aggregated representation of hidden state at time point i, considering potential influences of previous traffic states.

Furthermore, we also employ the gated temporal convolution network (TCN) after Tatt layer to ensure the consistent and shift-invariant traits of the estimaed traffic time series. Some recent works use dilated causal convolution (Wu et al., 2019; Liang et al., 2022a), recurrent neural network (Li et al., 2017) models to capture long-term temporal dependence for traffic forecasting. Consider the fact that we do not focus on modeling complex temporal dependencies in a long time range, because traffic congestion propagation between upstream and downstream detectors happens within a short time window. Therefore, TCN with localized receptive field is suitable for this purpose.

Besides, a gated linear unit (GLU) (Dauphin et al., 2017) is used to screen features of previous time steps that can pass through TCN layer. For an input temporal embedding $\mathbf{Z}_{att} \in \mathbb{R}^{T \times C}$ produced by temporal attention layer, a gated 1-D temporal convolution with kernel size k_t and residual connection is performed along the time dimension:

$$\text{GatedTCN}(\mathbf{Z}) = (\boldsymbol{\theta}_1 \star \mathbf{Z} + \mathbf{b}_1 + \mathbf{Z}) \odot \sigma(\boldsymbol{\theta}_2 \star \mathbf{H} + \mathbf{b}_2) \tag{21}$$

where $\star$ denotes the convolution operator, $\odot$ denotes the Hadmard product, $\sigma(\cdot)$ is the Sigmoid gate function which controls the ratio of temporal features flow to the following layer, $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ and $\mathbf{b}_1, \mathbf{b}_2$ are learnable parameters. Note that we use a zero-padding technique to keep the input time length unchanged. For each node, TCN assumes data in each time point only correlates with previous time points within the length of convolution kernel k_t. Owing to this non-recursive computation way, Eq. (24) can also be parallelized across all nodes, which makes TCN more lightweight than recurrent neural network.

Both Eq. (19) and (21) can also be calculated in the form of tensors. Given the output of TDCN layer $\mathcal{H} \in \mathbb{R}^{B \times N \times T \times C}$, we first permute it to $\widetilde{\mathcal{H}} \in \mathbb{R}^{B \times T \times N \times C}$ and the following equations process the temporal features $\mathcal{T}$:

$$\mathcal{E}_{att} = \frac{\text{UNFOLD}(\widetilde{\mathcal{H}} \times_4 \mathbf{W}_t) \otimes \text{UNFOLD}(\widetilde{\mathcal{H}} \times_4 \mathbf{W}_t)^{\mathsf{T}}}{\sqrt{d_t}}, \tag{22}$$

$$\widetilde{\mathcal{H}}_{att} = \text{SOFTMAX}\left(\text{TOPk}(\mathcal{E}_{att} + \mathcal{M}_t)\right) \cdot (\widetilde{\mathcal{H}} \times_4 \mathbf{W}_t), \tag{23}$$

$$\mathcal{T} = (\Theta_1 \star \widetilde{\mathcal{H}}_{att} + \mathcal{B}_1) \odot \sigma(\Theta_2 \star \widetilde{\mathcal{H}}_{att} + \mathcal{B}_2) \in \mathbb{R}^{\widetilde{\mathcal{H}} \in \mathbb{R}^{B \times T \times N \times C_{out}}}, \tag{24}$$

where $\text{UNFOLD}(\cdot): \mathbb{R}^{B \times T \times N \times C} \to \mathbb{R}^{B \times T \times NC}$ is a linear transform to change the order of a tensor.

In STCAGCN, we apply a single Tatt-TCN layer after the first TDCN layer without self-loops, so that the aggregated spatial features of neighborhood nodes are feed into Tatt-TCN layer and it extracts temporal features from most correlated previous time points for each node. Then the obtained temporal embedding containing the information of neighboring nodes at previous time slots are feed into next TDCN layer with self-loops. By this means, each sensor could have access to time-asynchronous observations with neighbor nodes in an adaptive manner.

4.4. Model implementations

4.4.1. Output layer with feature concatenation and residual connection

To alleviate the over-smoothing problem of stacked GNNs (Chen et al., 2020) and accelerate the back propagation of gradient, we add a residual connection (He et al., 2016) at the end of each TDCN block, i.e.,

$$\mathcal{H}^{l+1} = \mathcal{H}^l + \text{TDCN}(\mathcal{H}^l). \tag{25}$$

Besides, to fully utilize the multi-scale spatiotemporal features produced by TDCN and Tatt-TCN, we further employ a feature concatenation layer after the last TDCN layer to integrate these features. Finally, a linear fusion layer is used to produce the output:

$$\widehat{\mathcal{X}}_v = \text{CONCAT}(\mathcal{H}^0, \mathcal{H}^1, \ldots, \mathcal{H}^m) \times_4 \mathbf{W}_o + \mathcal{B}_o, \tag{26}$$

where $\mathbf{W}_o \in \mathbb{R}^{mC_{out} \times C_{out}}$ and $\mathcal{B}_o \in \mathbb{R}^{C_{out}}$ are parameters of the fusion layer, m is the number of TDCN layers, and $\text{CONCAT}(\cdot)$ denotes the data concatenation operation along the last dimension.

4.4.2. Training strategy and loss function

As our volume estimation problem can be viewed as a special kriging problem, the objective is to reconstruct the manually masked sub-graphs and generalize to unknown sensor locations during testing. Therefore, the reconstruction loss is computed on the whole input volume data, including both observed and masked nodes, to benefit the generalization ability :

$$\mathcal{L}_r = \|\mathcal{X}_v - \widehat{\mathcal{X}}_v\|_1, \tag{27}$$

where $\|\cdot\|_1$ is the $\mathcal{L}_1$ norm.

We observe that the adjacent matrix in reality is sparse and a small number of connected edges play a dominant role, i.e., a node's volume value is related with a fraction of other nodes'. And the node values usually vary smoothly across adjacent nodes. To control the sparsity and smoothness of the learned graph, we apply a graph Laplacian (a.k.a., graph Dirichlet energy (Belkin and Niyogi, 2001)) regularization on the training graph using SPAM $\mathbf{A}_{spa}$.

Specifically, given the volume signals $\mathcal{X}_v$, the graph Laplacian is computed on all batch inputs as follows:

$$\mathcal{L}_g = \mathcal{D}(\mathbf{A}_{spa}, \mathcal{X}_v) = \frac{1}{BNT} \sum_b \sum_{j,j'} a_{jj'}^{spa} \|\mathbf{x}_{\mathbf{s}j} - \mathbf{x}_{\mathbf{s}j'}\|_2^2 = \sum_b \mathrm{Tr}(\mathcal{X}_v[b,:,:,:]\mathbf{L}\mathcal{X}_v[b,:,:,:]^\mathsf{T}), \tag{28}$$

where $\mathbf{L} = \mathrm{Diag}(\sum_j a_{ij}^{spa}) - \mathbf{A}_{spa}$ is the graph Laplacian matrix, $\mathrm{Tr}(\cdot)$ computes the trace of a matrix, and $\|\cdot\|_2$ represents the Frobenius norm. It is observed that minimizing $\mathcal{D}(\mathbf{A}_{spa}, \mathcal{X}_v)$ encourages connected nodes of the learned graph to share similar embedding, leading to a smooth and sparse graph.

Then, the total loss function controlled by a parameter λ is:

$$\mathcal{L} = \mathcal{L}_r + \lambda \mathcal{L}_g. \tag{29}$$

Following (Wu et al., 2021a), the training strategy is in fact a random sub-graphs masking and reconstructing process (as discussed in section 3.3). The input is a series of manually masked graph signals, and output is the complete graph signals. Locations in $\mathcal{V}^u$ are not exposed in the training stage and only used for testing. By this setting, STCAGCN is trained to reconstruct incomplete graphs and generalize to unseen nodes (not in the constructed graph during training). We provide the pseudo code of model training in Algorithm 1. Owing to this inductive training manner, the proposed model has the ability to perform real-time volume kriging given instant measurements, which is a clear advantage over matrix/tensor based methods (Lei et al., 2022; Nie et al., 2022a).

Algorithm 1: Training sample generation and model update for STCAGCN training

Input: Training volume data from n_o observed locations during T time steps $\mathbf{X}_v \in \mathbb{R}^{n_o \times T}$, speed data from n_o observed locations during the same period $\mathbf{X}_s \in \mathbb{R}^{n_o \times T}$, adjacent matrix $\mathbf{A}_f$, $\mathbf{A}_b$, sequence length L, hidden dimension D, kernel size k_t, diffusion steps K, batch size B, number of batches N_b, number of masked nodes n_m, maximum iterations I.

Output: Trained model $\mathscr{F}^I$ and learned graph $\mathbf{A}_{spa}$.

1   initialize the model as $\mathscr{F}^0 = \mathrm{STCAGCN}(L, D, k_t, K)$;
2   **for** $i = 1 : I$ **do**
3   **for** $n = 1 : N_b$ **do**
      `// Split sample batches along time dimension`
4   randomly generate B individual time points $T_r = [t_{r1}, t_{r2}, \ldots, t_{rB}]$ within $[0, T-h]$ to split samples for each batch;
      `// Generate input data tensors`
5   initialize $\mathcal{X}_v \in \mathbb{R}^{B \times n_o \times L}, \mathcal{X}_s \in \mathbb{R}^{B \times n_o \times L}$ and $\mathcal{M} \in \mathbb{R}^{B \times n_o \times L}$ as zeros;
6   **for** $b = 1 : B$ **do**
7   prepare training volume data with sequence length L: $\mathcal{X}_v[b,:,:] \leftarrow \mathbf{X}_s[:, T_r[b] : T_r[b] + L]$;
8   prepare training speed data with sequence length L: $\mathcal{X}_s[b,:,:] \leftarrow \mathbf{X}_v[:, T_r[b] : T_r[b] + L]$;
9   randomly sample n_m masking index $M_r \leftarrow [m_{n_1}, m_{n_2}, \ldots, m_{n_m}]$ within $[0, n_o]$ and set $\mathcal{M}[b, M_r, :] \leftarrow 0$
10   construct masked input volume tensor as: $\mathcal{X}_{\text{input}} \leftarrow \mathcal{M} * \mathcal{X}_v$;
      `// Model training via stochastic gradient descent`
11   forward computation: $\widehat{\mathcal{X}}, \mathbf{A}_{spa} \leftarrow \mathrm{STCAGCN}(\mathcal{X}_{\text{input}}, \mathcal{X}_s, \mathbf{A}_f, \mathbf{A}_b)$;
12   compute loss function by Eq. (29);
13   loss backward and model parameters updating;

---

5. Case study

In this section, we evaluate the proposed STCAGCN model on a publicly available highway traffic volume dataset and provide comprehensive result analysis and discussions. All experiments are carried out on a Windows 10 platform with Intel Core i7-12700KF CPU and an NVIDIA GeForce RTX 3080 GPU. **The trained models and PyTorch implementations are publicly available in our GitHub repository:** https://github.com/tongnie/GNN4Flow.

5.1. Data preparation and experiment setup

5.1.1. Highway traffic volume dataset

To verify our model, we conduct case studies using PeMS-Bay traffic volume data collected from the Performance Measurement System of California [1]. We collect this traffic volume data from 325 loop sensors in the San Francisco bay area, ranging from January to March 2018 with a 5-min time interval. Note that due to failures of some volume sensors, original PeMS data suffers from a proportion of missing values and data imputation is already conducted for users. Data in the selected time span has a relatively high observation rate (about 80%) that is desired to perform experiments in this paper. The speed data within the same period, network topology and road background information (including lane number and driving direction) are also prepared.

5.1.2. Analysis of traffic flow phenomenon of PeMS data

Recall that we interpret the challenges of volume estimation as underdetermined and non-equilibrium problems, which are caused by undetected movements and traffic congestion delay respectively. To reveal these phenomena on PeMS-Bay dataset, we first calculate the WDSSI using Eqs. (5). The percentage of road segments/sensors with different WDSSI values are shown in Fig. 6 (a).
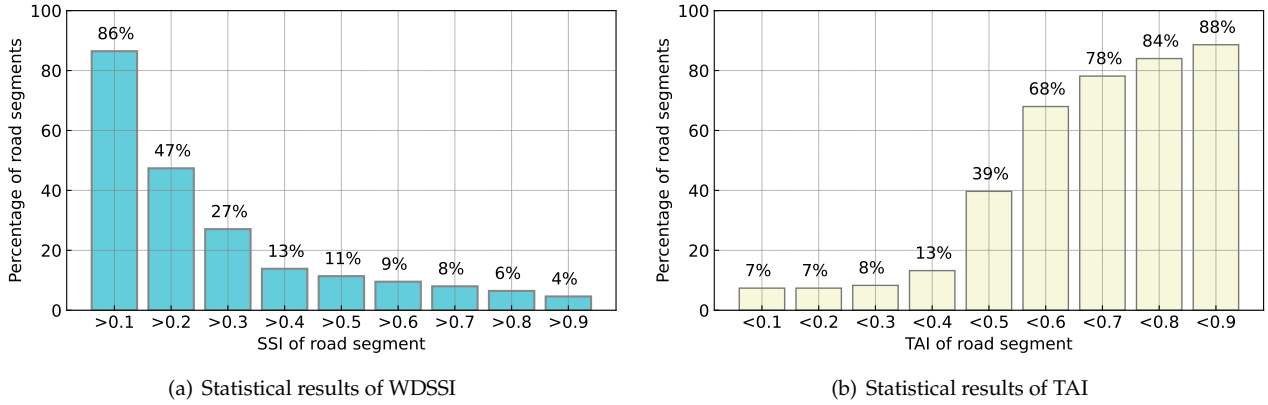

| (a) Statistical results of WDSSI | (b) Statistical results of TAI |

Figure 6: Histogram of WDSSI and TAI of sensors in PeMS-Bay data set.

As can be seen, about 30% of sensors have WDSSI larger than 0.3, which could cause high estimation error due to the unmeasured drive-in and drive-out between sensors (Zhang et al., 2020). We further compute the TAI of those road segments with WDSSI smaller than 0.3 (can be approximated as closed road segments) using Eq. (7) to examine the influence of traffic propagation. Fig. 6 (b) reports that the TAI of nearly 40% of sensors is less than 0.5, and we can regard them as the results of being affected by non-equilibrium traffic flows.

We will show that these small proportions of sensors could account for the majority of estimation errors. In order to demonstrate the effectiveness of our model design, these two metrics are used to distinguish between two types of estimation errors in the following experiments.

5.1.3. Baseline models and experiment setup

We compare our STCAGCN model with several baseline models, including the benchmark KNN method, several state-of-the-art deep learning and matrix/tensor models, to justify our model designs. Brief descriptions of each method are displayed as follows.

- **KNN**: K-nearest neighborhoods method is the benchmark for spatial kriging and it estimates unobserved volumes by simply averaging neighborhoods' values. By comparing with KNN, we could show the significance of modeling time-asynchronous and non-local correlations.

---

[1]from https://pems.dot.ca.gov/

- **ST-SSL (Meng et al., 2017)**: Graph based semi-supervised learning model for city-wide traffic volume estimation. This is a pioneering work for estimating city-wide traffic volume using speed information as supplements, which formulates a semi-supervised learning model.

- **TGMF-S (Zhang et al., 2020)**: Temporal geometric matrix factorization with speed Pearson coefficients. This model uses graph Laplacian and one-order Toeplitz matrix as spatial and temporal regularization. Speed information is used to calculate Pearson coefficients and construct a static adjacent matrix.

- **IGNNK (Wu et al., 2021a)**: Inductive graph neural networks for traffic speed kriging. By replacing the adjacent matrix with a speed-based adjacent matrix, this model can also perform volume kriging.

- **STAR (Liang et al., 2022a)**: This model modifies IGNNK by adding dilated causal convolution layers for temporal features and external attention modules for spatial features.

- **SGET (Nie et al., 2022b)**: Spatiotemporal graph embedded low-rank tensor completion model. Both spatiotemporal correlations and low-rankness are modeled in a unified tensor nuclear norm minimization framework.

Main hyper-parameters in STCAGCN are specified as follows. For spatial module, the diffusion step is set to 1, hidden dimension is set to 128, and the number of TDCN layers is 5. For temporal module, we set the kernel size as 3, the sequence length as 24, and the hidden dimension is also 128. In the training stage, the input batch size is 32, and we choose the Adam optimizer with a fixed learning rate being 5e-4. In the loss function, λ is set to 1e-4. We train STCAGCN for 300 epochs with early stopping to avoid over-fitting. To give a fair comparison, the pre-defined adjacent matrices in IGNNK, STAR and SGET are calculated by speed Pearson coefficients, which are the same as in TGMF-S.

We first compare our model with baselines under 50% sensor coverage rate, i.e., 160 sensors are randomly selected as observed locations used for model training, and the remaining 165 sensors are used only for evaluation during testing. In this scenario, we set the randomly masked nodes in Algorithm 1 as 80, which equals to 50% of the observed nodes. This consistency between training masked rate and testing coverage rate is essential to lean an unbiased model (discussed in subsection 5.3.3).

Three quantitative metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Weighted Mean Absolute Percentage Error (WMAPE) in Eq. (30) are used to evaluate the estimation results. As the range of traffic volume values varies widely, WMAPE is applied to alleviate the influence of extremes.

$$
\text{MAE} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} |x_{ij} - \hat{x}_{ij}|, \ \text{RMSE} = \sqrt{\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} (x_{ij} - \hat{x}_{ij})^2},
$$
$$
\text{MAPE} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} |\frac{x_{ij} - \hat{x}_{ij}}{x_{ij}}|, \ \text{WMAPE} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{T} |x_{ij} - \hat{x}_{ij}|}{\sum_{i=1}^{N} \sum_{j=1}^{T} x_{ij}}.
$$
(30)

5.2. Volume estimation results and model comparison

5.2.1. Estimation results

Tab. 1 reports the volume estimation performances of STCAGCN and other baselines under 50% spatial coverage rate. As described in subsection 5.1.2, we first split these results into sensor groups caused by different reasons. Specifically, three types of road segments can be categorized: 1) *MAPE_udt/WMAPE_udt*: errors of sensors with WDSSI higher than 0.4, which can be viewed as underdetermined parts and account for about 17% of the test sets; 2) *MAPE_dt_eq/WMAPE_dt_eq*: errors of sensors with WDSSI lower than 0.4 but TAI larger than 0.5, which are those stationary segments with minor influence of propagation delay and account for about 46% of the test sets; 3) *MAPE_dt_neq/WMAPE_dt_neq*: errors of sensors with lower WDSSI and smaller TAI, which are space-stationary but time-asynchronous parts and account for about 37% of the test sets. Their respective positions are shown in Fig. 7.

Observing the distributions of two types of sensors and speeds, an interesting finding is that underdetermined flows generally occur at the interchange of freeways, where vehicles frequently access to other expressways via on/off ramps. Besides, as expected, non-equilibrium phenomenon widely exists in congested road segments. These observations justify the use of two metric in section 4.1 to quantify the two special challenges in this work.

By comparing three types of errors longitudinally, we can see that type 1 error is the largest, followed by type 3 error. This result clearly verify our assumption that a small number of underdetermined or non-equilibrium road segments dominate most of the estimation errors, thereby confirming the necessity of customizing models targeting these two phenomena. As enough accurate information can be borrowed from neighborhoods directly, it is natural that type 2 error is lowest.

By comparing different models horizontally, several noteworthy observations are stated as follows:

(a) Sensors with underdetermined and non-equilibrium flows (only data in test are displayed).
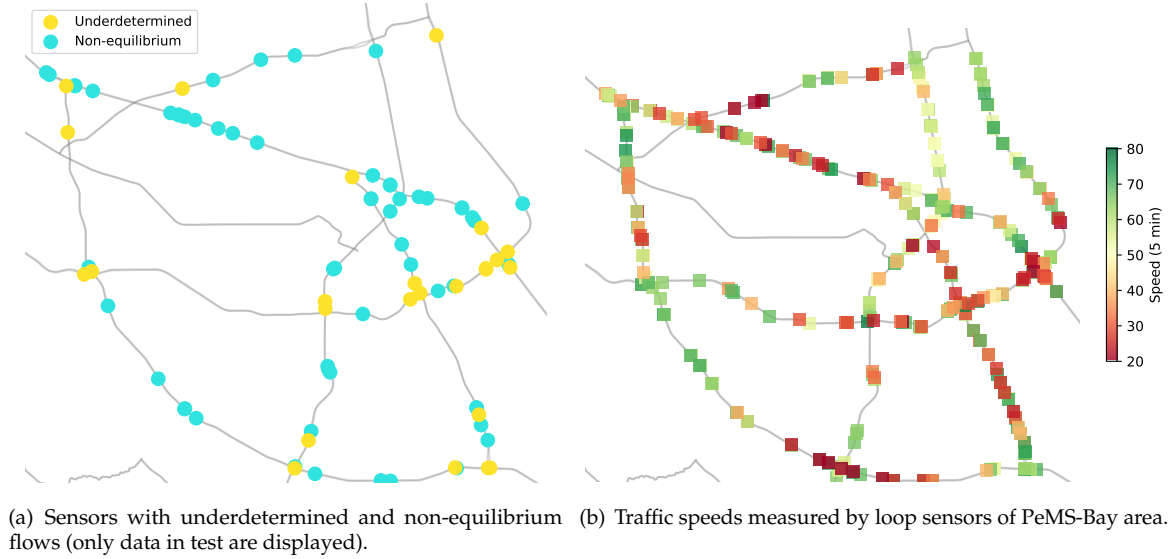(b) Traffic speeds measured by loop sensors of PeMS-Bay area.

Figure 7: Locations of two kinds of sensors and traffic states on PeMS highway network.

Table 1: Volume estimation results under 50% coverage rate of different methods.

	STCAGCN	KNN	TGMF-S	ST-SSL	IGNNK	STAR	SGET
MAE	**43.83**	53.83	49.28	52.78	49.43	50.09	51.88
RMSE	**65.99**	77.73	74.28	82.10	72.77	74.89	77.51
MAPE	**30.13%**	46.51%	39.05%	40.69%	38.25%	38.59%	44.41%
WMAPE	**19.59%**	24.17%	22.12%	23.71%	22.09%	22.38%	23.38%
MAPE_udt	**58.85%**	113.98%	99.47%	95.96%	72.15%	71.20%	101.56%
WMAPE_udt	**30.90%**	56.67%	48.55%	51.65%	34.56%	36.17%	50.66%
MAPE_dt_eq	**22.41%**	27.76%	23.89%	25.95%	24.78%	25.97%	24.36%
WMAPE_dt_eq	**16.37%**	18.90%	17.02%	17.76%	18.04%	18.35%	17.53%
MAPE_dt_neq	**26.74%**	39.27%	31.79%	33.13%	39.58%	39.46%	42.98%
WMAPE_dt_neq	**19.80%**	22.23%	20.81%	25.07%	22.97%	22.79%	29.83%

Best results are bold marked.

(1) STCAGCN outperforms all baselines consistently with significant improvements. Compared with the second best model IGNNK, our model is improved by more than 8 percentage points with respect to MAPE.
(2) Due to the design of SPAM, STCAGCN reduces MAPE_udt/WMAPE_udt dramatically. With the adaptive injection of speed information, STCAGCN can borrow information from globally related sensors on the whole network to fix the spatial underdetermined problem.
(3) All models achieve desired results on MAPE_dt_eq/WMAPE_dt_eq because the volumes of this portion of sensors can be effectively recovered by neighborhood values. Notice that STCAGCN can also reduce this part of errors since the TDCN module can filter the most appropriate neighbor sensors according to the directed nature of traffic flow.
(4) By considering time-asynchronous correlations caused by propagation delay, STCAGCN can increase accuracy of non-equilibrium sensors significantly.

To conclude, our model achieves state-of-the-art in two aspects: STCAGCN performs better than other baselines on the simplest tasks, while at the same time can achieve satisfactory performances on tricky tasks that other models are less effective.

To display the proportion of each type of error, the stacked bar chart is given in Fig 8. Each error value is weighted by their percentage, i.e., 17%, 46% and 37% for MAPE_udt, MAPE_dt_eq, and MAPE_dt_neq respectively, and the sum of them equals to the overall MAPE in Tab. 1. From this figure, we can see that underdetermined and non-equilibrium errors account for more than half of the MAPE errors, which emphasizes the necessity of optimizing these parts. Moreover, it can be intuitively observed that STCAGCN reduces both underdetermined and non-equilibrium errors significantly, compared with other baselines.

We further visualize the ground truth volume values and STCAGCN estimations using the testing data in Fig. 9. As we can see, traffic volumes in this district are relatively high and many road segments have volume values higher than 400 within 5 min. Therefore, traffic congestion could have obvious implications for the propagation
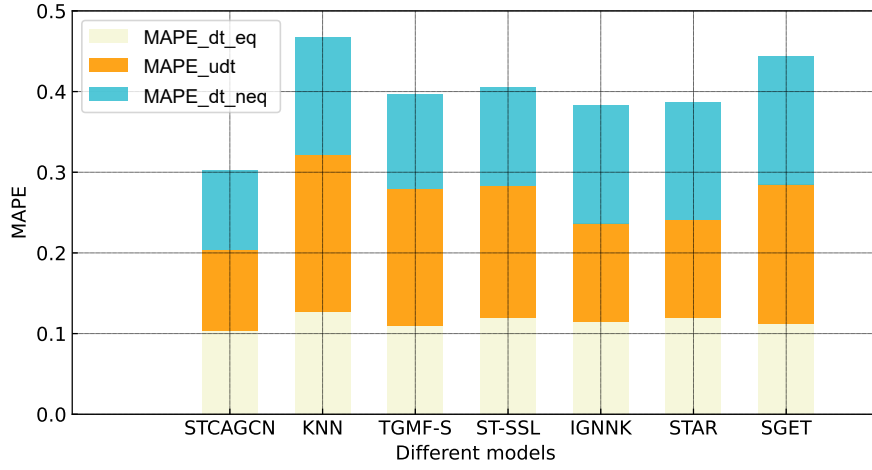
Figure 8: Proportion of different types of errors for each model.

of traffic flow. STCAGCN can produce traffic volume estimations quite close to the true values even 50% of the observations are unavailable, which intuitively confirms the effectiveness of our model.
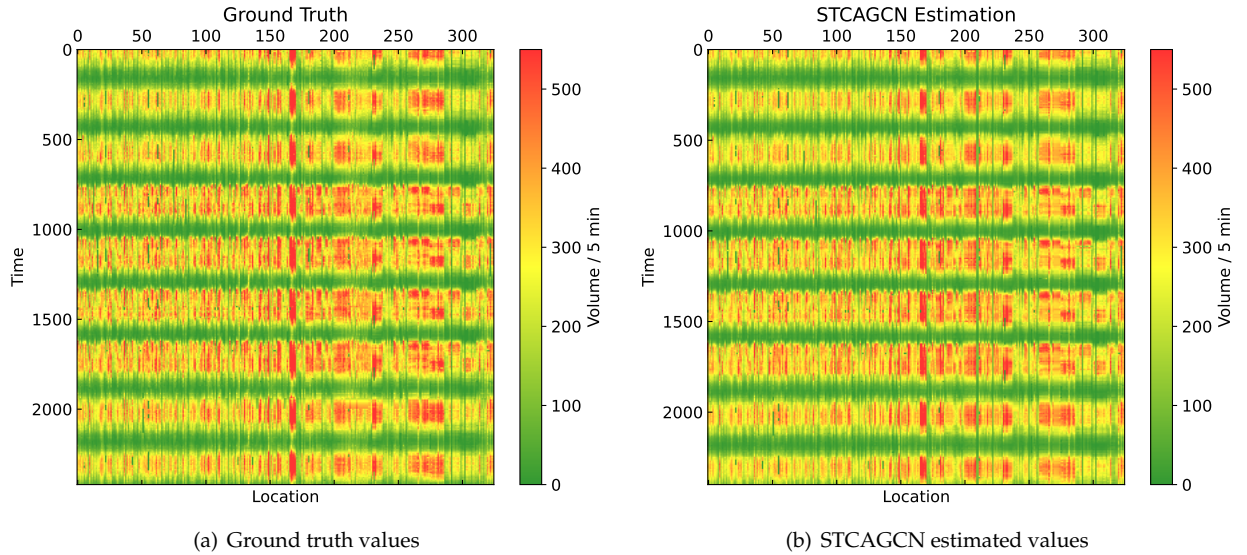

(a) Ground truth values

(b) STCAGCN estimated values

Figure 9: Ground truth traffic volumes versus STCAGCN estimated traffic volumes of PeMS data.

5.2.2. Model performances under lower coverage rates

In order to examine the model effectiveness under lower sensor coverage rate, we set the missing rate ranging from 60%, 70% to 80%, to compare different models. Tab. 2 gives the evaluation metrics in total and the RMSE results are also shown in Fig. 10.

With lower coverage rate, less redundancy of the sensor network can be used so that the estimation performances of all models degrade. In spite of this, STCAGCN is more robust than competing models with minimal increase in errors. Under 80% missing rate (20% coverage rate), the proposed model is the only one with a MAPE error of less than 40%. These findings further confirm the superiority of the proposed model.

5.3. Model analysis

5.3.1. Ablation studies

To inspect the significance of each component of the proposed model, in this part we conduct ablation studies to compare the performances of model variations under 50% coverage rate. Totally, we examine the following three variations of the original STCAGCN model by removing or replacing specific modules:

- **STCAGCN w/o SPAM**: We remove the speed pattern adaptive adjacent matrix construction module in STCAGCN. A static speed Pearson coefficients matrix is used instead.

Table 2: Model comparison under lower coverage rates.

|		STCAGCN	KNN	TGMF-S	ST-SSL	IGNNK	STAR	SGET
60%	MAE	**45.99**	52.22	50.61	57.72	52.66	51.94	52.56
	RMSE	**67.57**	77.44	76.12	92.03	74.46	76.78	77.94
	MAPE	**32.83%**	43.95%	39.00%	39.08%	48.47%	38.23%	45.14%
	WMAPE	**20.50%**	23.47%	22.74%	25.95%	23.55%	23.23%	23.70%
70%	MAE	**46.37**	55.24	51.79	58.51	55.65	53.71	54.05
	RMSE	**68.02**	80.58	77.80	92.12	77.37	78.37	79.40
	MAPE	**31.18%**	44.85%	38.60%	38.16%	50.39%	47.62%	46.73%
	WMAPE	**20.74%**	24.74%	23.20%	26.22%	24.81%	23.94%	24.29%
80%	MAE	**48.51**	63.91	56.02	70.92	66.64	59.23	56.77
	RMSE	**69.55**	88.67	83.02	107.07	90.63	84.09	81.26
	MAPE	**36.38%**	54.12%	40.60%	44.85%	72.16%	65.42%	55.82%
	WMAPE	**21.61%**	28.61%	25.08%	31.76%	29.69%	26.39%	25.50%

Best results are bold marked.
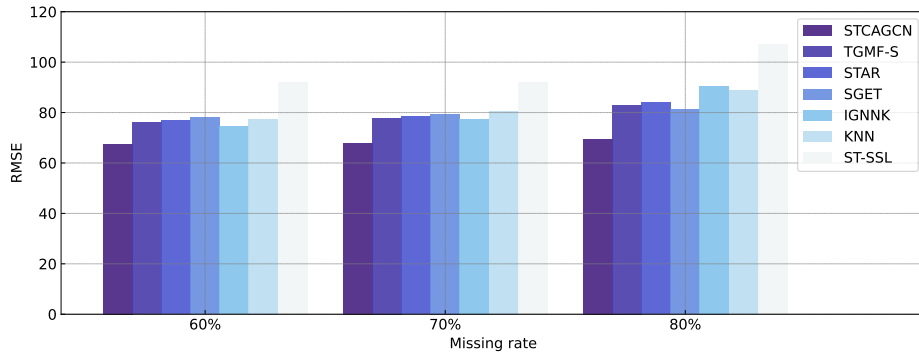
Figure 10: Model performances under lower coverage rate.

- **STCAGCN w/o Tatt-TCN**: STCAGCN without temporal attention and convolution layer. In this case, input time series are treated as individual samples and only spatial features are extracted.

- **STCAGCN w/o TDCN**: We replace the direction-specific diffusion convolution with a symmetric graph spectral convolution operator in (Yu et al., 2017).

Table 3: Ablation studies.

	w/o Tatt-TCN	w/o SPAM	w/o TDCN
MAE	46.21	51.56	46.77
RMSE	67.75	76.36	69.30
MAPE	34.00%	49.00%	35.35%
WMAPE	20.52%	23.04%	20.90%
MAPE_udt	66.34%	87.51%	68.98%
WMAPE_udt	33.04%	38.23%	33.22%
MAPE_dt_eq	25.76%	29.20%	27.09%
WMAPE_dt_eq	17.47%	18.22%	18.08%
MAPE_dt_neq	30.21%	56.03%	30.43%
WMAPE_dt_neq	20.39%	23.98%	20.25%

Results of different variations are shown in Tab. 3. By comparing results of the integral STCAGCN in Tab. 1, it is observed that the performances of all variations are degraded, indicating that each component contributes the overall improvements of STCAGCN model remarkably.

After removing Tatt-TCN layer, the errors at time-asynchronous locations have an obvious increase, showing that time-asynchronous correlations are indispensable factors of traffic volume estimation in congested area. The ablation study on TDCN manifests that the direction-specific, topology-related characteristics of traffic flow

should be considered and incorporated into the computation process of graph convolution and the increments of MAPE_udt and MAPE_dt_eq verify this finding. Moreover, it is notable that without SPAM, the accuracy decreases sharply on all three types of errors. This indicates that a static speed adjacent matrix is incapable to fully reflect the dynamic, time-varying and non-linear speed-volume relationships. Therefore, defining a speed adjacent matrix subjectively could lead to a biased model with inferior prediction performance.

5.3.2. Sensitivity studies on hyper-parameters

To investigate the affects of hyper-parameters on estimation accuracy, we conduct model sensitivity studies on three main hyper-parameters of STCAGCN, including the diffusion step K, number of TCN layers, and hidden dimension. Hidden dimension is the most common parameter in deep learning model and we set the hidden dimension of TCN, TDCN and output layer to the same value. Diffusion step determines the order of neighbors that GCN can aggregate information, which can be viewed as the spatial feaure receptive field. In addition, in our STCAGCN we apply a single TCN layer after the first TDCN layer to capture temporal correlations in a localized window. The more the TCN layers, the longer temporal dependence is extracted.

Fig. 11 shows the estimation MAE under 50% missing rate with different settings. First, it can be seen that with the increment of hidden dimension, the estimation error first drops sharply but then increases. A larger hidden dimension enables the GNN model to have more parameters to learn, thereby enhancing the fitting ability. While at the same time, the risk of over-fitting and additional learning complexity may undermine the capacity. Second, with the increase of K, the MAE grows gradually. Enlarging K can incorporates irrelevant information from higher-order neighbors and cause over-smoothed node representations. Third, Fig. 11 (c) shows that stacking more TCN layers has no benefits for estimation results. A rational explanation is that a series of stacked TCN layers are used to capture long-term temporal dependence, while as stated in section 4.3, the influence of propagation delay happens within a short time period.
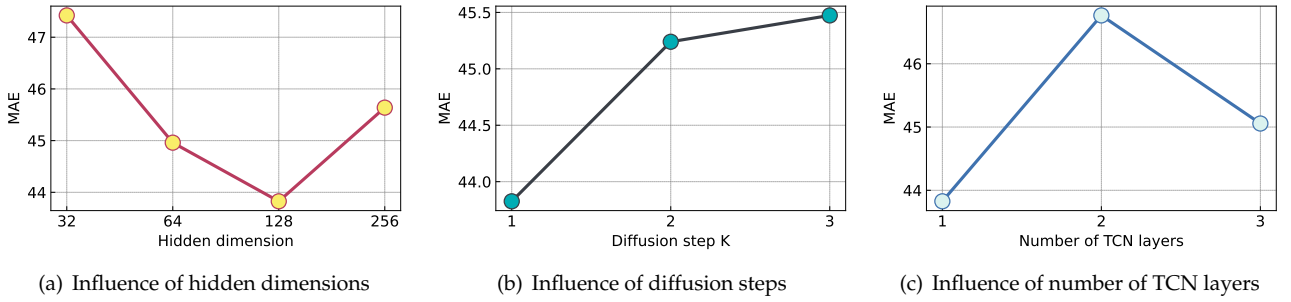

(a) Influence of hidden dimensions   (b) Influence of diffusion steps   (c) Influence of number of TCN layers

Figure 11: STCAGCN performances under different hyper-parameter settings.

5.3.3. Influence of different sample rates during model training and testing

(a) Model trained with 50% masking rate tests on different missing rates scenarios   (b) Models trained with different masking rates test on 50% missing rate scenario
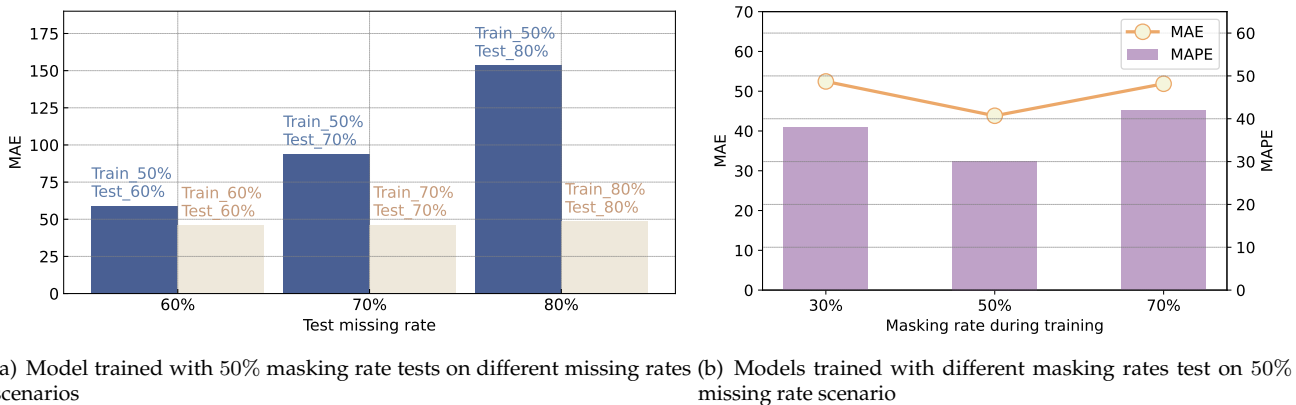
Figure 12: Influence of different masking and missing rate during training and testing.

Masking rate is an important hyper-parameter during model training. This parameter controls what proportions of sub-graphs need to be reconstructed during training. We notice that recent works about traffic data imputation or kriging (Wu et al., 2021a,b; Liang et al., 2022a) usually treat it as a general hyper-parameter and ignore its significance. In contrast, we find that it is essential to keep the training masking rate the same as the testing missing rate, in order to train a model with better generalization ability and make the training stage more stable with less fluctuation.

To give evidence of this proposition, we inspect the performances of models trained with different number of masked nodes and tested on different missing rates scenarios. Fig. 12 (a) displays the errors of the model trained at 50% masking rate but tested at different missing rates, and the model trained at the same masking rate as the tested missing rate. All these models with different masking rate have obviously worse performances than models with the same masking rate as missing rate. Fig. 12 (b) shows the model trained at various masking rates while tested at 50% missing rate and model trained at 50% has the best result.

Both of the two figures reveal the importance of making the masking rate close to the missing rate of the whole sensor graph, thus keeping the distributions of masked sub-graph and actually observed graph as close as possible. We suppose that this setting could alleviate the covariate shift problem of deep learning models (Ioffe and Szegedy, 2015).

5.3.4. Interpretation of learned graphs

To demonstrate that SPAM truly leans meaningful patterns, we visualize the learned SPAM at different moment in Fig. 13 and more examples are given in Appendix A. Please note that we remove the self-loop (diagonal) for better visualization and the values of SPAM are normalized to one. From these figures, we can observe that only a small proportion of nodes have relative high values, which means the correlation graphs of traffic volume is sparse in reality. In this sense, speed information is not always helpful for volume estimation and filtering valid speed data to construct speed-volume relationships in an adaptive way is thus important.

We further take sensors #23, #31 and #36 at $8:00 - 10:00$ as an example and their locations, speed, and volume (per lane) profiles are visualized on the road network in Fig. 14. From Fig. 13 (a), the attention score (correlation) between #23 and #36 is much higher than that between #23 and #31. At the same time, we can observe similar speed-volume patterns of sensors #23 and #36 in Fig. 14, even though they are not proximal on the road network. By seizing the speed similarity, the volume values of #23 are highly informative for #36, and vice versa. Moreover, despite sensors #31 and #36 are pretty close in distance, their volume series show very different patterns. This inconsistency phenomenon prompts us to seek for non-local correlations that can overcome the deficiency of typical graph convolution.

Therefore, these observations suggest that the SPAM successfully captures non-local correlations between sensors by dynamically learning the speed-volume relationships. Compared with static speed adjacent used in previous works, our SPAM is agnostic to the form of speed-volume function and robust to misleading information, which is much more flexible and reliable.

(a) SPAM at $8:00 - 10:00$

(b) SPAM at $10:00 - 12:00$

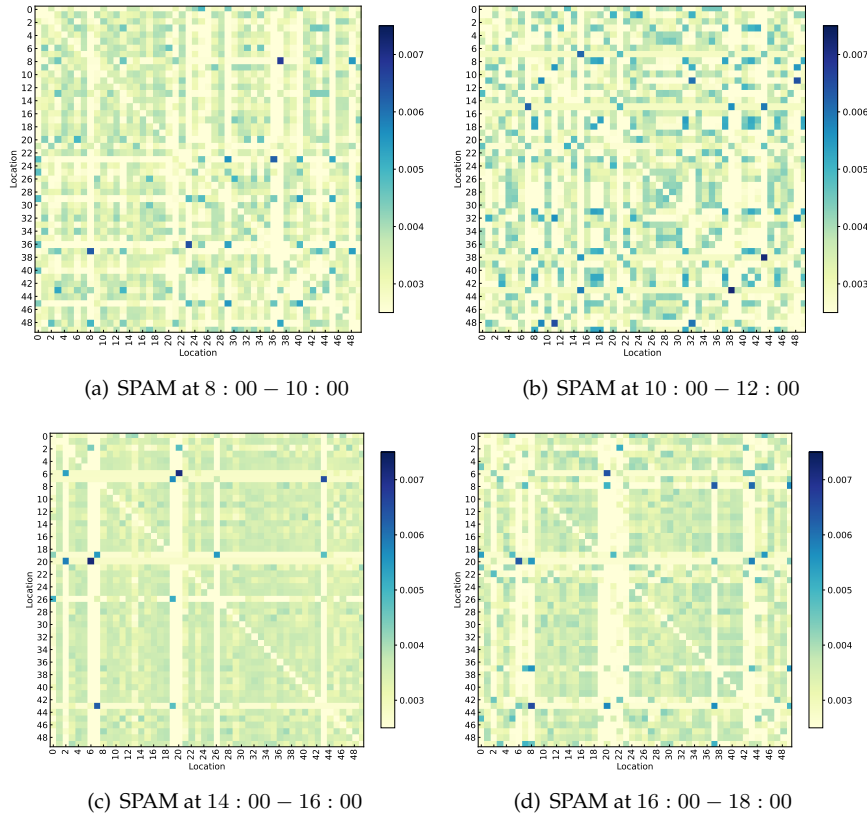(c) SPAM at $14:00 - 16:00$

(d) SPAM at $16:00 - 18:00$

Figure 13: SPAM at different time point on test set. Note that only the first 50 sensors are displayed for better viewing.
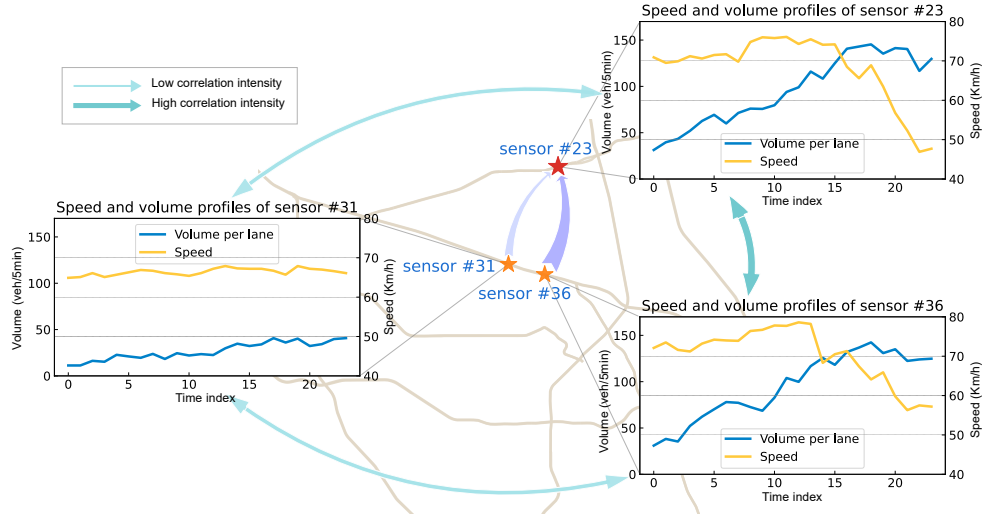
Figure 14: Illustration of the non-local correlations learned by SPAM. In this figure, wider arrows with darker color denotes higher intensity of correlation.

6. Conclusion and future directions

Acquiring fine-grained spatiotemporal traffic volume data is of great challenge due to the limited deployment of traffic detectors. The underdetermined and non-equilibrium phenomena further complicate the requirements of utilizing available observations. To tackle these issues, in this paper we elaborate a correlation-adaptive graph convolution network named STCAGCN that features both non-local spatial dependence and time-asynchronous correlation for highway traffic volume estimation under sparse sensor coverage. To consider the impacts of road geometry and propagation direction, we develop a task-specific diffusion convolution module to model the semantic similarities between sensors. To leverage the dynamic, nonlinear and time-varying speed-volume relationships for enhancing the estimation of underdetermined flows, we propose a graph attention based speed pattern-adaptive learning module to encode the speed information as a complement. To deal with non-equilibrium flows caused by congestion propagation, we develop a temporal masked self-attention model combined with a localized temporal convolution layer to exploit time-asynchronous correlations adaptively. Detailed case studies and model analysis are conducted on a real-world highway traffic volume dataset, and several conclusions can be drawn as follows:

(1) Results compared with other baseline models including state-of-the-art deep learning and tensor learning methods on different observation conditions, illustrate obvious superiority of our STCAGCN model on both overall accuracy and critical scenarios.

(2) Model ablation studies further justify the rationality of model designs. Especially, SPAM captures the global correlations of volume to provide complementary side information for underdetermined flows. And Tatt-TCN can seize the time-shifted correlations under varying impacts of congestion delay.

(3) As our model achieves graph reconstruction based on randomly masked sub-graphs, random sampling plays a key role in model generalization capability. Studies on sampling rate indicate the necessity of keeping the masking rate in training stage and missing rate in testing stage the same.

(4) It is well known that there exists complicated mapping relationships between traffic speed and volume. Our model captures this subtle connection well and interpretations of learned spatial correlation graphs reveal the globally shared volume patterns by extracting from the underlying speed-volume relationships.

In addition, there are still some directions require future efforts. First, how to achieve joint estimation of incomplete traffic speed and volume is worth trying. A possible solution is to impute the incomplete speed data at the same time, with speed reconstruction errors included in the loss function. Second, as collecting link volume data of urban road network is much more difficult, further efforts are needed to conduct case studies on urban networks.

Appendix A.

We give more examples about the learned SPAM in the following figure 15.

Acknowledgement

(a) SPAM of test sample #19   (b) SPAM of test sample #21   (c) SPAM of test sample #24   (d) SPAM of test sample #25

(e) SPAM of test sample #31   (f) SPAM of test sample #32   (g) SPAM of test sample #35   (h) SPAM of test sample #39
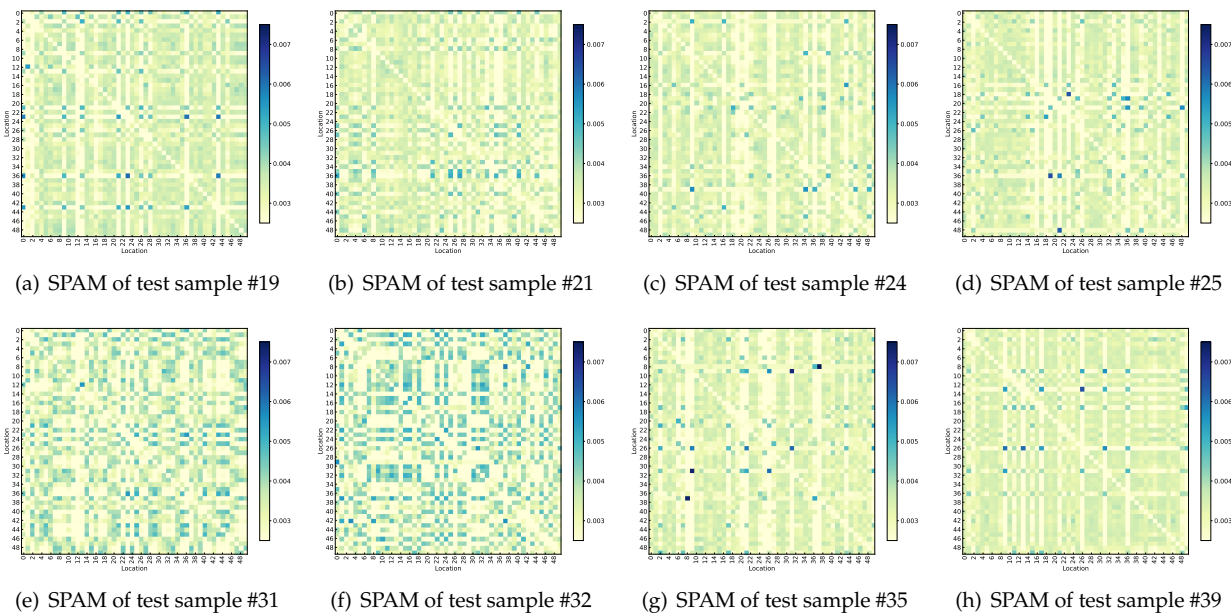
Figure 15: More examples of the learned SPAM on test samples.

References

Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. Transportation Research Part B: Methodological 55, 265–281.

Appleby, G., Liu, L., Liu, L.-P., 2020. Kriging convolutional networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. pp. 3187–3194.

Ashok, K., Ben-Akiva, M. E., 2002. Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows. Transportation science 36 (2), 184–198.

Aslam, J., Lim, S., Pan, X., Rus, D., 2012. City-scale traffic estimation from a roving sensor network. In: Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems. pp. 141–154.

Bai, L., Yao, L., Li, C., Wang, X., Wang, C., 2020. Adaptive graph convolutional recurrent network for traffic forecasting. Advances in neural information processing systems 33, 17804–17815.

Belkin, M., Niyogi, P., 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. Advances in neural information processing systems 14.

Chen, D., Lin, Y., Li, W., Li, P., Zhou, J., Sun, X., 2020. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. pp. 3438–3445.

Chen, Y., Chen, X. M., 2022. A novel reinforced dynamic graph convolutional network model with data imputation for network-wide traffic flow prediction. Transportation Research Part C: Emerging Technologies 143, 103820.

Dauphin, Y. N., Fan, A., Auli, M., Grangier, D., 2017. Language modeling with gated convolutional networks. In: International conference on machine learning. PMLR, pp. 933–941.

Fu, H., Wang, Y., Tang, X., Zheng, N., Geroliminis, N., 2020. Empirical analysis of large-scale multimodal traffic with multi-sensor data. Transportation Research Part C: Emerging Technologies 118, 102725.

Geroliminis, N., Daganzo, C. F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological 42 (9), 759–770.

Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transportation Research Part B: Methodological 45 (3), 605–617.

Hao, H., Wang, Y., Xia, Y., Zhao, J., Shen, F., 2020. Temporal convolutional attention-based network for sequence modeling. arXiv preprint arXiv:2002.12530.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778.

Idé, T., Katsuki, T., Morimura, T., Morris, R., 2016. City-wide traffic flow estimation from a limited number of low-quality cameras. IEEE Transactions on Intelligent Transportation Systems 18 (4), 950–959.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. PMLR, pp. 448–456.

Janson, B. N., 1991. Dynamic traffic assignment for urban road networks. Transportation Research Part B: Methodological 25 (2-3), 143–161.

Ji, Y., Geroliminis, N., 2012. On the spatial partitioning of urban transportation networks. Transportation Research Part B: Methodological 46 (10), 1639–1656.

Ji, Y., Luo, J., Geroliminis, N., 2014. Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. Transportation Research Record 2422 (1), 1–11.

Krishnakumari, P., Van Lint, H., Djukic, T., Cats, O., 2020. A data driven method for od matrix estimation. Transportation Research Part C: Emerging Technologies 113, 38–56.

Lei, M., Labbe, A., Wu, Y., Sun, L., 2022. Bayesian kernelized matrix factorization for spatiotemporal traffic data imputation and kriging. IEEE Transactions on Intelligent Transportation Systems, 1–13.

Li, Y., Yu, R., Shahabi, C., Liu, Y., 2017. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. arXiv preprint arXiv:1707.01926.

Liang, W., Li, Y., Xie, K., Zhang, D., Li, K.-C., Souri, A., Li, K., 2022a. Spatial-temporal aware inductive graph neural network for c-its data recovery. IEEE Transactions on Intelligent Transportation Systems.

Liang, Y., Zhao, Z., Sun, L., 2022b. Memory-augmented dynamic graph convolution networks for traffic data imputation with diverse missing patterns. Transportation Research Part C: Emerging Technologies 143, 103826.

Liu, J., Han, K., Chen, X. M., Ong, G. P., 2019. Spatial-temporal inference of urban traffic emissions based on taxi trajectories and multi-source urban data. Transportation Research Part C: Emerging Technologies 106, 145–165.

Mahajan, V., Cantelmo, G., Rothfeld, R., Antoniou, C., 2022. Predicting network flows from speeds using open data and transfer learning. IET Intelligent Transport Systems.

Meng, C., Yi, X., Su, L., Gao, J., Zheng, Y., 2017. City-wide traffic volume inference with loop detector data and taxi trajectories. In: Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 1–10.

Merchant, D. K., Nemhauser, G. L., 1978. A model and an algorithm for the dynamic traffic assignment problems. Transportation science 12 (3), 183–199.

Ng, M., 2012. Synergistic sensor location for link flow inference without path enumeration: A node-based approach. Transportation Research Part B: Methodological 46 (6), 781–788.

Nie, T., Qin, G., Sun, J., 2022a. Truncated tensor schatten p-norm based approach for spatiotemporal traffic data imputation with complicated missing patterns. Transportation Research Part C: Emerging Technologies 141, 103737.

Nie, T., Qin, G., Wang, Y., Sun, J., 2022b. Correlating sparse sensing for network-wide traffic speed estimation: An integrated graph tensor-based kriging approach. arXiv preprint arXiv:2210.11780.

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., Keogh, E., 2012. Searching and mining trillions of time series subsequences under dynamic time warping. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 262–270.

Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. Transportation Research Part B: Methodological 74, 1–19.

Ros-Roca, X., Montero, L., Barceló, J., Nökel, K., Gentile, G., 2022. A practical approach to assignment-free dynamic origin–destination matrix estimation problem. Transportation Research Part C: Emerging Technologies 134, 103477.

Rostami-Shahrbabaki, M., Safavi, A. A., Papageorgiou, M., Setoodeh, P., Papamichail, I., 2020. State estimation in urban traffic networks: A two-layer approach. Transportation Research Part C: Emerging Technologies 115, 102616.

Saeedmanesh, M., Geroliminis, N., 2016. Clustering of heterogeneous networks with directional flows based on "snake" similarities. Transportation Research Part B: Methodological 91, 250–269.

Saeedmanesh, M., Kouvelas, A., Geroliminis, N., 2021. An extended kalman filter approach for real-time state estimation in multi-region mfd urban networks. Transportation Research Part C: Emerging Technologies 132, 103384.

Salari, M., Kattan, L., Lam, W. H., Lo, H., Esfeh, M. A., 2019. Optimization of traffic sensor location for complete link flow observability in traffic network considering sensor failure. Transportation Research Part B: Methodological 121, 216–251.

Shao, M., Xie, C., Sun, L., 2021. Optimization of network sensor location for full link flow observability considering sensor measurement error. Transportation Research Part C: Emerging Technologies 133, 103460.

Sirmatel, I. I., Geroliminis, N., 2017. Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance. IEEE Transactions on Intelligent Transportation Systems 19 (4), 1112–1121.

Takeuchi, K., Kashima, H., Ueda, N., 2017. Autoregressive tensor factorization for spatio-temporal predictions. In: 2017 IEEE International Conference on Data Mining (ICDM). IEEE, pp. 1105–1110.

Tang, X., Gong, B., Yu, Y., Yao, H., Li, Y., Xie, H., Wang, X., 2019. Joint modeling of dense and incomplete trajectories for citywide traffic volume inference. In: The World Wide Web Conference. pp. 1806–1817.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., 2017. Graph attention networks. arXiv preprint arXiv:1710.10903.

Viti, F., Rinaldi, M., Corman, F., Tampère, C. M., 2014. Assessing partial observability in network sensor location problems. Transportation research part B: methodological 70, 65–89.

Wang, P., Lai, J., Huang, Z., Tan, Q., Lin, T., 2020a. Estimating traffic flow in large road networks based on multi-source traffic data. IEEE Transactions on Intelligent Transportation Systems 22 (9), 5672–5683.

Wang, X., Guan, X., Cao, J., Zhang, N., Wu, H., 2020b. Forecast network-wide traffic states for multiple steps ahead: A deep learning approach considering dynamic non-local spatial correlation and non-stationary temporal dependency. Transportation Research Part C: Emerging Technologies 119, 102763.

Wang, X., Sun, L., 2023. Anti-circulant dynamic mode decomposition with sparsity-promoting for highway traffic dynamics analysis. arXiv preprint arXiv:2302.07108.

Weng, J., Meng, Q., 2011. Modeling speed-flow relationship and merging behavior in work zone merging areas. Transportation research part C: emerging technologies 19 (6), 985–996.

Wu, Y., Zhuang, D., Labbe, A., Sun, L., 2021a. Inductive graph neural networks for spatiotemporal kriging. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. pp. 4478–4485.

Wu, Y., Zhuang, D., Lei, M., Labbe, A., Sun, L., 2021b. Spatial aggregation and temporal convolution networks for real-time kriging. arXiv preprint arXiv:2109.12144.

Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C., 2019. Graph wavenet for deep spatial-temporal graph modeling. arXiv preprint arXiv:1906.00121.

Xing, J., Wu, W., Cheng, Q., Liu, R., 2022. Traffic state estimation of urban road networks by multi-source data fusion: Review and new insights. Physica A: Statistical Mechanics and its Applications, 127079.

Xu, X., Lo, H. K., Chen, A., Castillo, E., 2016. Robust network sensor location for complete link flow observability under uncertainty. Transportation Research Part B: Methodological 88, 1–20.

Xue, J., Jiang, N., Liang, S., Pang, Q., Yabe, T., Ukkusuri, S. V., Ma, J., 2022. Quantifying the spatial homogeneity of urban road networks via graph neural networks. Nature Machine Intelligence 4 (3), 246–257.

Yang, H., Du, L., Zhang, G., Ma, T., 2023. A traffic flow dependency and dynamics based deep learning aided approach for network-wide traffic speed propagation prediction. Transportation Research Part B: Methodological 167, 99–117.

Yang, J., Sun, J., 2015. Vehicle path reconstruction using automatic vehicle identification data: An integrated particle filter and path flow estimator. Transportation Research Part C: Emerging Technologies 58, 107–126.

Yao, H., Tang, X., Wei, H., Zheng, G., Li, Z., 2019. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In: Proceedings of the AAAI conference on artificial intelligence. Vol. 33. pp. 5668–5675.

Ye, Y., Zhang, S., Yu, J. J., 2021. Spatial-temporal traffic data imputation via graph attention convolutional network. In: International Conference on Artificial Neural Networks. Springer, pp. 241–252.

Yi, X., Duan, Z., Li, T., Li, T., Zhang, J., Zheng, Y., 2019. Citytraffic: Modeling citywide traffic via neural memorization and generalization approach. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. pp. 2665–2671.

Yildirimoglu, M., Sirmatel, I. I., Geroliminis, N., 2018. Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance. Transportation Research Part B: Methodological 118, 106–123.

Yu, B., Li, M., Zhang, J., Zhu, Z., 2019a. 3d graph convolutional networks with temporal graphs: A spatial information free framework for traffic forecasting. arXiv preprint arXiv:1903.00919.

Yu, B., Yin, H., Zhu, Z., 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.

Yu, J., Stettler, M. E., Angeloudis, P., Hu, S., Chen, X. M., 2020. Urban network-wide traffic speed estimation with massive ride-sourcing gps traces. Transportation Research Part C: Emerging Technologies 112, 136–152.

Yu, Y., Tang, X., Yao, H., Yi, X., Li, Z., 2019b. Citywide traffic volume inference with surveillance camera records. IEEE Transactions on Big Data 7 (6), 900–912.

Zhan, X., Zheng, Y., Yi, X., Ukkusuri, S. V., 2016. Citywide traffic volume estimation using trajectory data. IEEE Transactions on Knowledge and Data Engineering 29 (2), 272–285.

Zhang, H. M., 1998. A theory of nonequilibrium traffic flow. Transportation Research Part B: Methodological 32 (7), 485–498.

Zhang, P., Qian, S., 2020. Path-based system optimal dynamic traffic assignment: A subgradient approach. Transportation Research Part B: Methodological 134, 41–63.

Zhang, P., Wong, S., Dai, S., 2009. A conserved higher-order anisotropic traffic flow model: description of equilibrium and non-equilibrium flows. Transportation Research Part B: Methodological 43 (5), 562–574.

Zhang, Y., Cheng, Q., Liu, Y., Liu, Z., 2022. Full-scale spatio-temporal traffic flow estimation for city-wide networks: a transfer learning based approach. Transportmetrica B: Transport Dynamics, 1–27.

Zhang, Z., Li, M., Lin, X., Wang, Y., 2020. Network-wide traffic flow estimation with insufficient volume detection and crowdsourcing data. Transportation Research Part C: Emerging Technologies 121, 102870.

Zheng, C., Fan, X., Wang, C., Qi, J., 2020. Gman: A graph multi-attention network for traffic prediction. In: Proceedings of the AAAI conference on artificial intelligence. Vol. 34. pp. 1234–1241.

Zhu, Y., Wang, J., Wang, J., He, Z., 2022. Multitask neural tensor factorization for road traffic speed-volume correlation pattern learning and joint imputation. IEEE Transactions on Intelligent Transportation Systems.