



OPEN ACCESS

EDITED BY
Xingshuo Li,
Nanjing Normal University, China

REVIEWED BY
Dazhong Ma,
Northeastern University, China
Xuguang Hu,
Northeastern University, China

*CORRESPONDENCE
Huaizhi Wang,
wanghz@szu.edu.cn
Ting Wu,
twu920@hotmail.com

SPECIALTY SECTION
This article was submitted to Smart Grids,
a section of the journal Frontiers in Energy
Research

RECEIVED 20 May 2022
ACCEPTED 18 July 2022
PUBLISHED 07 September 2022

CITATION
Kang T, Wang H, Wu T, Peng J and Jiang H
(2022), A pagerank self-attention network
for traffic flow prediction.
Front. Energy Res. 10:948954.
doi: 10.3389/fenrg.2022.948954

COPYRIGHT
© 2022 Kang, Wang, Wu, Peng and Jiang.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A pagerank self-attention network for traffic flow prediction

Ting Kang¹, Huaizhi Wang^{1*}, Ting Wu^{2*}, Jianchun Peng¹ and Hui Jiang³

¹College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China, ²School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, China, ³College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen, China

Traffic information is collected from sensors in the urban road network, and traffic information can be said to be a mapping of people's activities, which are difficult to model as a linear function, so this makes traffic information difficult to be predicted. In other words, traffic information is difficult to build effective models to predict traffic information because of its non-linear characteristics that are difficult to capture. As researchers go deeper, researchers have been able to extract good spatio-temporal features for modern urban road networks. However, it is worth mentioning that most researchers have neglected the importance of models for global potential features under the topology map of urban road networks, yet this global potential feature is very important for traffic prediction. In this paper, we propose a new spatio-temporal graph convolutional network model A Pagerank Self-attention Network (hereafter we abbreviate as PSN) in order to solve this problem based on a full consideration of the urban road network topology features, in which we employ a global spatio-temporal self-attention module to capture the global spatio-temporal features well. and the graph wandering module is used to propagate the spatio-temporal feature information effectively and widely. It is worth mentioning that experiments on two well-known datasets show that our proposed method achieves better prediction results compared to existing baseline methods.

KEYWORDS

traffic forecast, spatio-temporal graph convolutional network, global spatio-temporal self-attention module, graph wandering module, urban road network

1 Introduction

In many developed countries around the world (Tu et al., 2021), severe traffic congestion has caused certain economic losses to them. Traffic congestion brings not only the increase of transportation cost, the decrease of urban transportation efficiency, and the increase of environmental pollution. And it seriously affects people's living standard (Belhadi et al., 2020). With the concern about traffic problems, intelligent transportation in urban road networks has rapidly become an important

research topic. Traffic information is an important data for intelligent system analysis. Taking the famous PEMS traffic dataset as an example, traffic information includes the characteristics of three dimensions: traffic flow, traffic speed, and traffic occupancy. If researchers can predict future traffic data with historical traffic data collected in urban traffic road networks, it can bring positive significance to traffic congestion problems and rational travel planning in cities.

We are inspired by the combination of pagerank and graph convolutional neural network, in order to further improve the improvement of graph convolutional neural network for model prediction, in other words, to improve the over-smoothing problem in graph convolutional neural network. We propose the PNT model to perform prediction.

2 Related work

In the early days of traffic flow forecasting, that is, before the rise of artificial intelligence, most of the traffic forecasting problems were studied using statistical methods. For example, the autoregressive integrated moving average (ARIMA) and its variants were proposed in 1976 and so on (Ahmed and Cook, 1979) (Williams and Hoel, 2003), including spatiotemporal ARIMA (Ding et al., 2011), dynamic spatiotemporal ARIMA (Min et al., 2009), and local spatiotemporal ARIMA (Cheng et al., 2014). However, due to the explosive growth of traffic data in recent years, for traffic forecasting, the previous statistical models can no longer meet the current forecasting needs. Statistical models usually have high computational complexity, but due to the stochastic and nonlinear nature of traffic variables in urban traffic road networks, the prediction effect of statistical models is not satisfactory (Cui et al., 2019). With the development of technology in recent years, the computing power of computers has gradually increased, and machine learning methods have made a splash in various research fields (Wu et al., 2022); Li et al. (2022); Fu (2022), and more and more researchers have started to focus on using machine learning methods for traffic prediction. In the field of traffic prediction, traditional statistical models are also gradually replaced by big data computational intelligence, or machine learning methods (Vlahogianni et al., 2014). Machine learning methods, such as KNN (Van Lint and Van Hinsbergen, 2012), and SVM (Shuman et al., 2013) can model relatively complex data, but they require delicate feature engineering. With the breakthrough of deep learning in natural language processing and image processing, more and more researchers are applying deep learning methods to traffic prediction.

In general, traffic data are the trajectories of road users flowing on urban road networks, so traffic data often exhibit a high degree of spatial and temporal dependence due to

the complexity and regularity of human mobility. Therefore, if the spatio-temporal correlation can be modeled effectively (Tedjopurnomo et al., 2020), then it is possible to obtain accurate traffic prediction results. Although researchers have invested a lot of efforts on traffic prediction (Moreira-Matias et al., 2013) (Zhang et al., 2018), there are still some challenges to be solved.

Challenge 1: It is still difficult to effectively extract the potential global spatio-temporal features of the traffic road network. For example, the locations where some regions in the urban traffic road network are located may present similar traffic patterns due to the intersection of people's habits, even if these two regions are far apart in the urban traffic road network topology graph (Zhang et al., 2017). The traffic similarity between regions is widely present in the urban traffic road network topology, which can also be called as a potential global spatio-temporal feature in the urban traffic road network topology. In order to solve this problem, a global spatio-temporal self-attentive module is designed in our proposed model to focus on the global spatio-temporal features in the urban traffic road network topology and extract them effectively.

Challenge 2: Time-varying traffic data based on the spatial characteristics of urban traffic road networks poses difficulties for traffic prediction. For time-series information such as traffic data, the traffic data on a road depends to a large extent on its past traffic conditions. However, the variation of traffic data is extremely unstable and nonlinear (Zhao et al., 2019). Most previous studies use graph convolution to process traffic data to obtain spatial features, and then use (recursive neural network) RNN (Graves, 2012) and its variants GRU (Van Lint et al., 2002) or LSTM to obtain temporal features to perform traffic prediction. However, it is often limited by the over-smoothing problem caused by the stacking of graph convolution layers, leading to the use of only a few graph convolution layers in the model. For example, the T-GCN (Zhao et al., 2019) model combines two layers of graph convolutional neural network and GRU to extract the spatial and temporal characteristics of road speed respectively. The AST-GCN model (Guo et al., 2019) also applies the graph convolutional neural network to traffic data prediction. According to the results of our experiments, it does not perform as well as our proposed PSN model for predicting the traffic speed of PEMS04 and PEMS08. We believe that this may be due to the fact that the traffic speed features are not sufficiently diffused in the urban traffic network topology. It is well known that the essence of graph convolutional neural network is to aggregate and update the information of neighboring nodes. Therefore, our proposed PSN model is designed with a graph wandering module to spread the traffic speed features extracted by our model sufficiently over the traffic topology. Experimental data show that our model achieves better results on two datasets, PEMS04 and PEMS08.

The main contributions of this paper are as follows. In graph neural networks, the core idea of graph convolutional

neural networks is to aggregate node information using edge information to generate new node representations, but each layer of the graph convolutional neural network can only aggregate and update the first-order nodes around its nodes. If we want to propagate the node information more widely, we can superimpose multiple layers of the graph convolutional neural network, but this will lead to the problem of oversmoothing of the graph convolutional neural network. In order to spread the node information more effectively as well as update it. Inspired by the combination of pagerank and graph neural network, this paper proposes a new PNT model, whose main feature is to bind each sensor node in the urban road network graph with its own traffic information relatively, while using the global spatio-temporal attention module to capture the hidden features behind the traffic information, and then using the GRU framework to filter all the historical data in a certain way, and finally All nodes in the road network graph diffuse the relatively important information effectively. This makes the information of the graph nodes to update themselves effectively and flexibly.

3 Methodology

3.1 Definition

Definition 1: PNT summarizes the spatial information between roads in an urban traffic network by modeling each road as a node in an undirected graph with N roads, which is described in [Figure 1](#). We use $G = \{V, E\}$ for the description, by treating each road as a node and V as the set of this node. where $V = \{V_1, V_2, V_3, \dots, V_{n-1}, V_n\}$, n is the number of nodes. E is a set of edges. We use the adjacency matrix $A \in \mathbb{R}^{N \times N}$ to represent the connectivity between these nodes. Each value in this adjacency matrix is calculated based on the distance of the road network between sensors. The specific description of the elements in the adjacency matrix can be found in ([Guo et al., 2019](#)).

Definition 2: In the urban road network traffic system sensors collect a large amount of traffic information, taking the pems dataset as an example, there are three dimensional features of road flow, occupancy rate and road speed. We can choose any one dimension of traffic information to construct the feature matrix. But in this thesis, we choose road speed in the city to construct our feature matrix $x^{N \times T}$ respectively, and we use the road speed on the city road network to be the attribute feature of the nodes in the city road network, denoted as $x \in \mathbb{R}^{N \times T}$, and T denotes the number of information collected by a sensor at a fixed time interval of a node (for example, in the classical pems dataset we choose both pems04 and pems08 traffic speed data are sampled using a 5 min interval, so we can understand that the sensor sampled a total of T times, each time at an interval of 5 min). $x_t \in \mathbb{R}^{N \times 1}$ is used to represent the speed on each road at moment i . In this article $x_t = Input[t]$.

3.2 PSN model details

In this paper, our prediction task is to extract spatio-temporal features from a period of historical traffic speed information and learn the relevant features to predict the traffic speed information in the future period. The formula is expressed as:

$$\begin{aligned} & [X_{t+1}, X_{t+2}, \dots, X_{t+p-1}, X_{t+p}] \\ & = f(A; (X_{t-n}, X_{t-1}, X_t)) \end{aligned} \quad (1)$$

Here n is the length of the historical series used and p is the length of the time series to be predicted.

In this article, the total PNT model equation is shown below:

$$r_t = \sigma \left(W_r * \begin{bmatrix} GWM \\ h_{t-1} \end{bmatrix} \right) \quad (2)$$

$$z_t = \sigma \left(W_r * \begin{bmatrix} GSSM \\ h_{t-1} \end{bmatrix} \right) \quad (3)$$

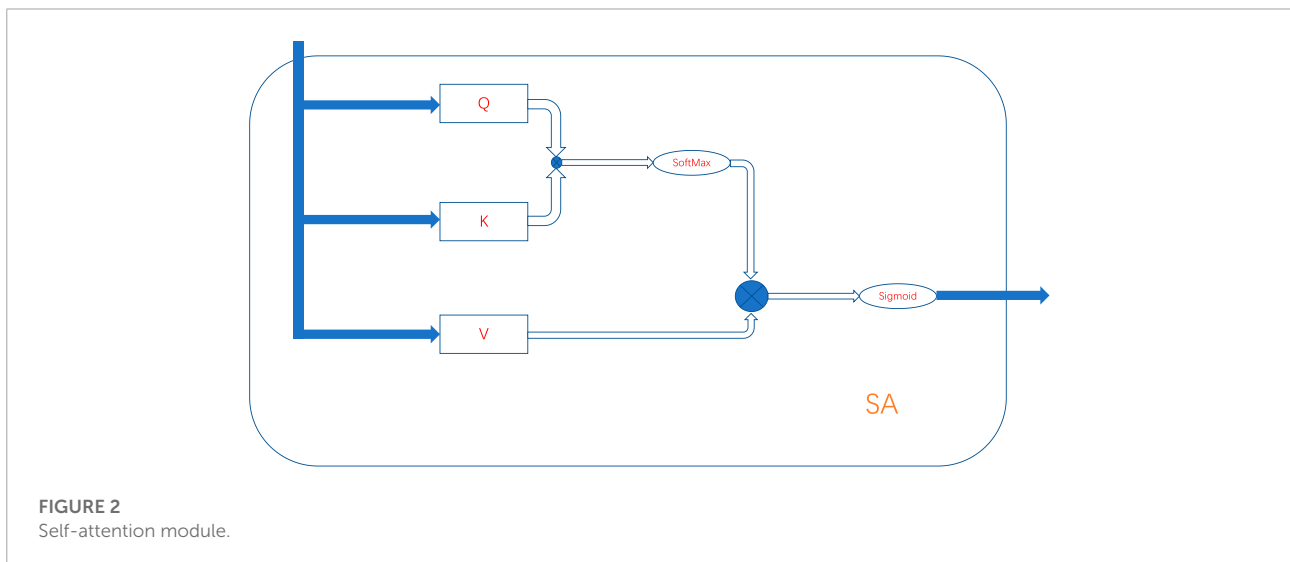
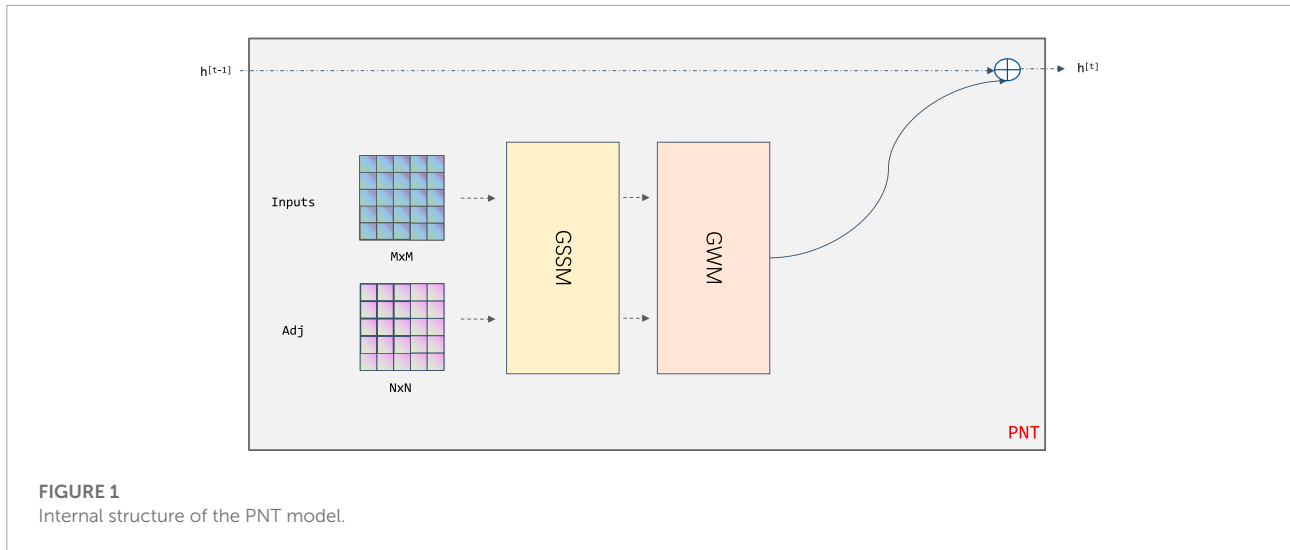
$$\tilde{h}_t = \tanh \left(W_{\tilde{h}} * \begin{bmatrix} GWM \\ r_t * h_{t-1} \end{bmatrix} \right) \quad (4)$$

$$h_t = z_t * h_{t-1} + (1 - z_t) * \tilde{h}_t \quad (5)$$

Briefly, our PNT model consists of two major modules and a final fully connected module, which are the graph wandering module and the global spatio-temporal self-attentive module. We improve the GWM and GSSM combined with GRU to generate a new recursive module. In the following we introduce these two modules of the PNT model separately.

3.2.1 Global spatio-temporal self-attention module:GSSM

For the extraction of global spatio-temporal information, many previous studies have been done. For example, k-hop adjacency matrices ([Zhang et al., 2019](#)) or k-pop Laplace matrices ([Diao et al., 2019](#)) have been used to describe fixed k-hop connections in urban road networks. These ‘‘connectivity’’ approaches have the advantage of considering the location of connections in the urban road network, but unconsciously ignore the global relationships behind the roadways in the traffic network. Generally, if the spatio-temporal features of the road network are simply extracted and the relevant features are neglected, then the training may lead to missing features and affect the prediction results. To solve this problem, we propose a global spatio-temporal self-attentive module ([Vaswani et al., 2017](#)) to extract potential features in urban road networks. As shown in [Figure 2](#) and [Figure 3](#), we are inspired by transformer as well as graph attention, and we use the global spatio-temporal self-attentive module to compute local



and global spatio-temporal correlation information of each node. We use the road speed in the urban road network to construct the query matrix, key matrix and value matrix. In order to alleviate the problems caused by gradient disappearance and gradient explosion in the model, we introduce the residual module to improve the model.

$$Q = x_t * W_q \tag{6}$$

$$K = x_t * W_k \tag{7}$$

$$V = x_t * W_v \tag{8}$$

$$Att = Q * k^{Tr} \tag{9}$$

* in this article all mean matrix multiplication

$$Att'_{ij} = \frac{\exp(Att_{ij})}{\sum_{j=1}^N \exp(Att_{ij})} \tag{10}$$

$$SA = S'_a = \sigma(Att' V) \tag{11}$$

$$GSSM = \frac{(x_t * SA) + x_t}{2} \tag{12}$$

Here, W_q , W_k , and W_v are learnable weight matrices. * denotes matrix multiplication. Tr denotes matrix transpose, and σ denotes Softmax function. Att denotes the measure of

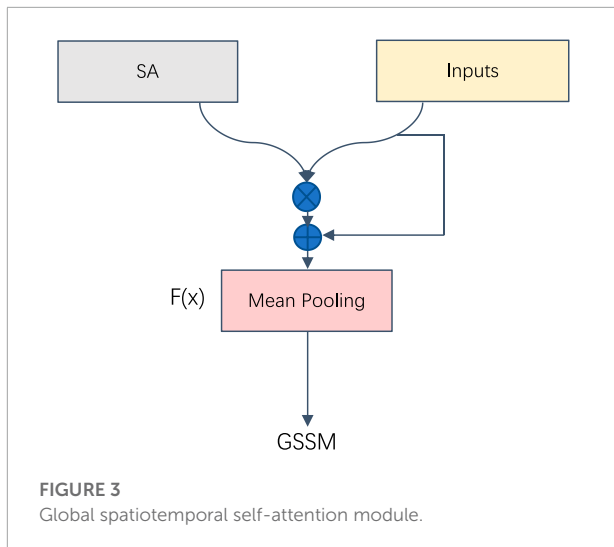


FIGURE 3
Global spatiotemporal self-attention module.

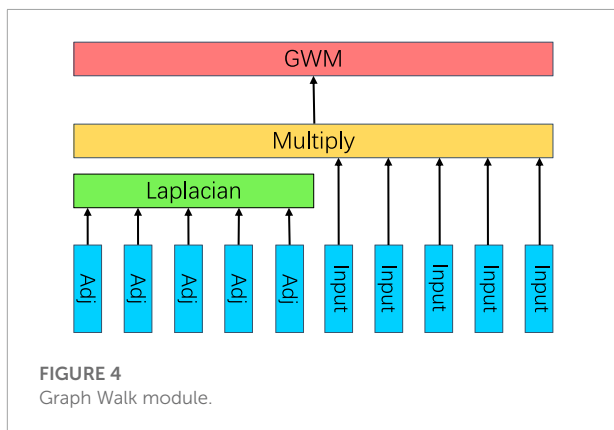


FIGURE 4
Graph Walk module.

similarity between Q and K. Att'_{ij} denotes the coupling strength between road i and road j, which is obtained by performing Softmax calculation on Att . S'_a represents the spatio-temporal coefficient matrix and thus will enable the model to notice the underlying spatio-temporal information behind the traffic speed information.

3.2.2 Graph walking module:GWM

In the field of traffic flow prediction, temporal feature extraction and spatial feature extraction are two key parts. Thanks to the rapid development of hardware devices and deep learning in recent years, deep neural network models have gained attention in the field of traffic prediction because of their excellent results. Generally speaking, to achieve accurate traffic prediction, not only the temporal and spatial features in traffic data have to be extracted effectively, but also their propagation has to be performed efficiently. It is well known that the core idea of graph convolutional neural network is to use the edge information to aggregate the node information to generate new

node representations, and some studies have applied on this basis to use node representations to generate edge representations or graph representations (Yu et al., 2017) to accomplish their tasks. However, graph convolutional neural networks have a limitation that a node can only be influenced by its surrounding first order neighboring nodes. However, we only need to superimpose K such graph convolutional layers to extend the influence of a node to K-order neighbor nodes. But in fact, stacking multiple layers of first-order graph convolution leads to another problem, namely, the oversmoothing problem. By nature, each layer of a graph convolutional network is a special kind of Laplacian smoothing, but Laplacian smoothing makes a point as similar as possible to its surrounding points, with each node's new feature being the mean of its surrounding nodes' special rules. Going back to the road network topology diagram, let's consider each monitoring point in the traffic network as a point. Intuitively, if we want the information of each monitoring point to be fully absorbed by all neighboring nodes for aggregation and update, we need to superimpose N-layer graph convolutional neural network, but it will cause the above-mentioned over-smoothing problem.

Google's founders Larry Page and Sergey Brin invented this technology in 1998 at Stanford University. pagerank was earlier used by Google to reflect the relevance and importance of web pages, and is often used in search engine optimization operations to assess the effectiveness of web optimization as one of the factors. In the early days of search engine development, it was common for people to manually categorize web pages and sort out high quality sites. With more and more web pages, the search engine entered the era of text search, manual classification has clearly can not meet the demand. So pagerank came into existence (Klicpera et al., 2018) proposed to link graph convolutional networks with personalized pagerank in order to solve the over-smoothing problem of graph convolutional networks. Since graph convolutional neural networks have received more attention, previous studies in the literature have usually adopted graph convolutional neural networks or their alternative form of Chebyshev networks to process information based on urban road network monitoring sites, and all of them only used single-digit graph convolutional network superposition (An et al., 2021) to process information from traffic network monitoring sites. However, this may result in each point only being able to utilize a portion of its neighboring nodes' information for aggregation updates. To solve the above problem, we propose a graph wandering module inspired by pagerank to perform aggregated updates of traffic information, which is described in Figure 4.

$$\tilde{A} = A + I \tag{13}$$

$$\tilde{D} = \sum_j \tilde{A}_{ij} \tag{14}$$

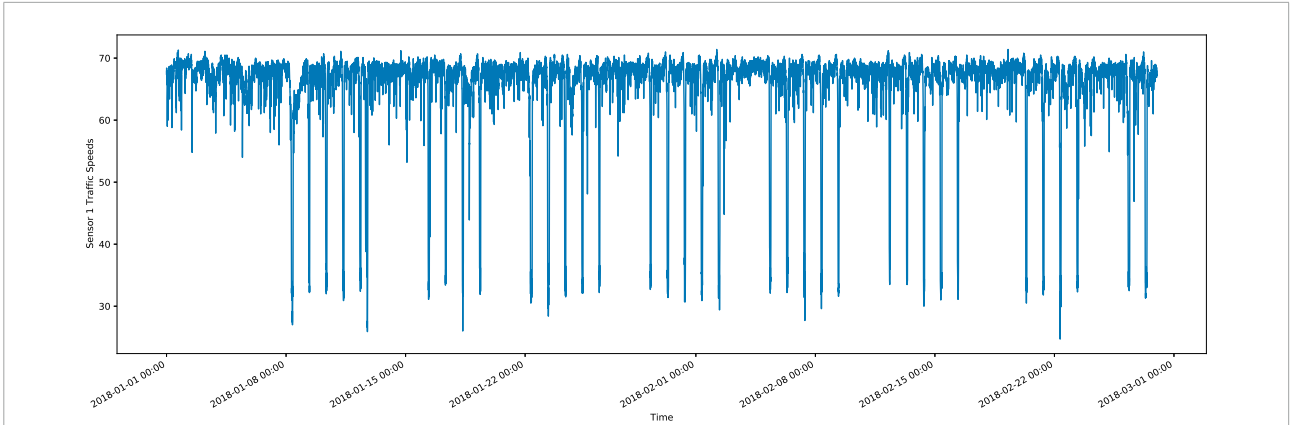


FIGURE 5
PEMS04-sensor 1 traffic speed.

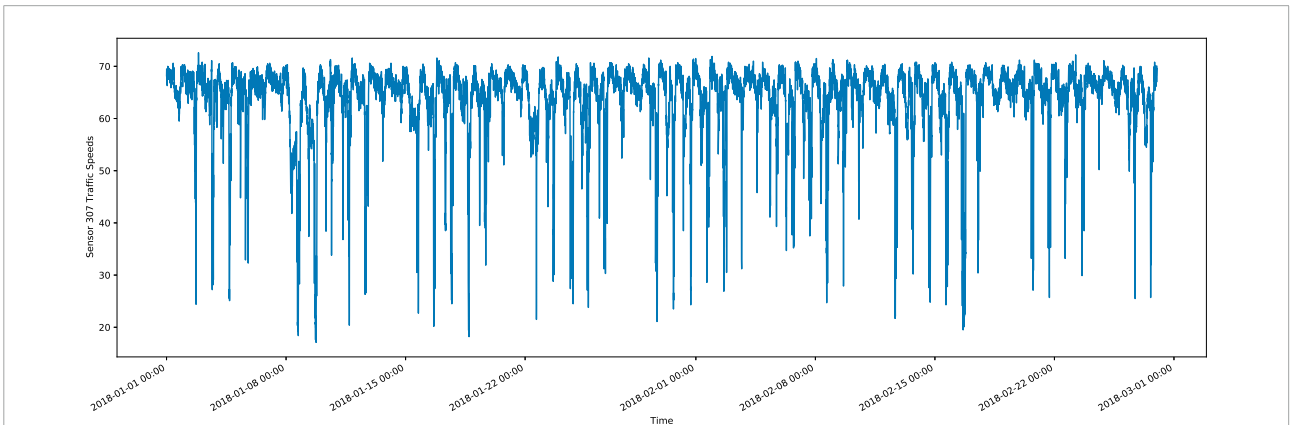


FIGURE 6
PEMS04-sensor 307 traffic speed.

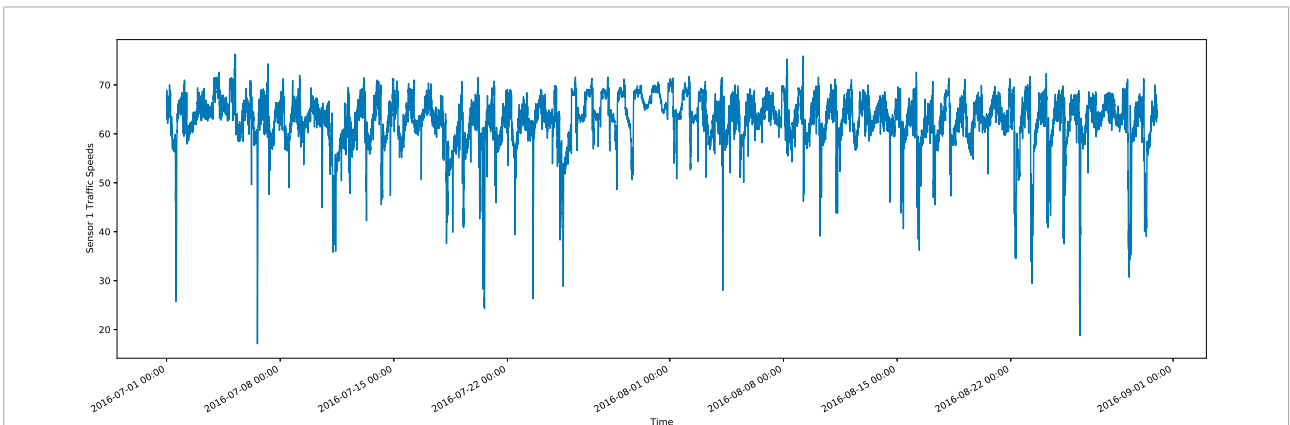


FIGURE 7
PEMS08-sensor 1 traffic speed.

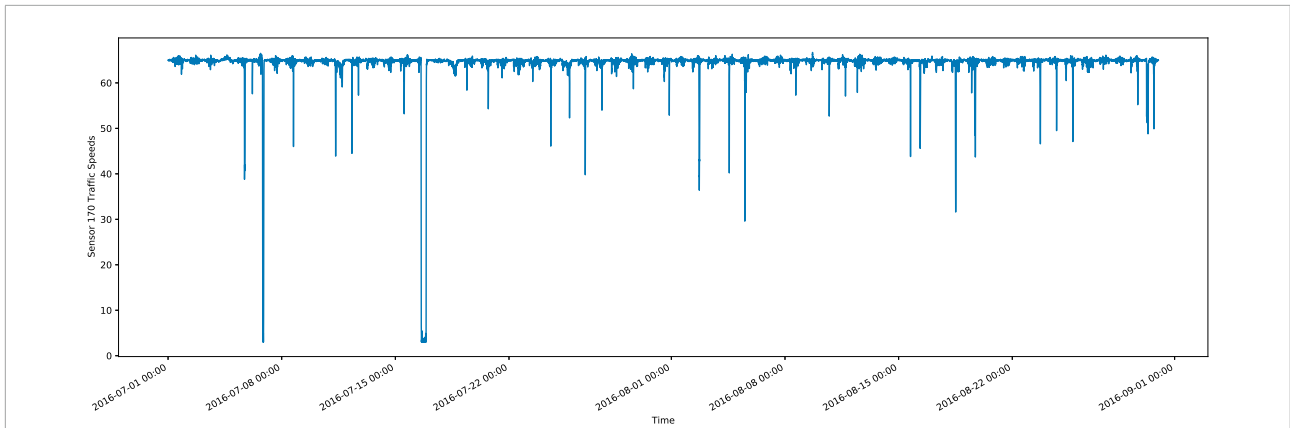


FIGURE 8
PEMS08-sensor 170 traffic speed.

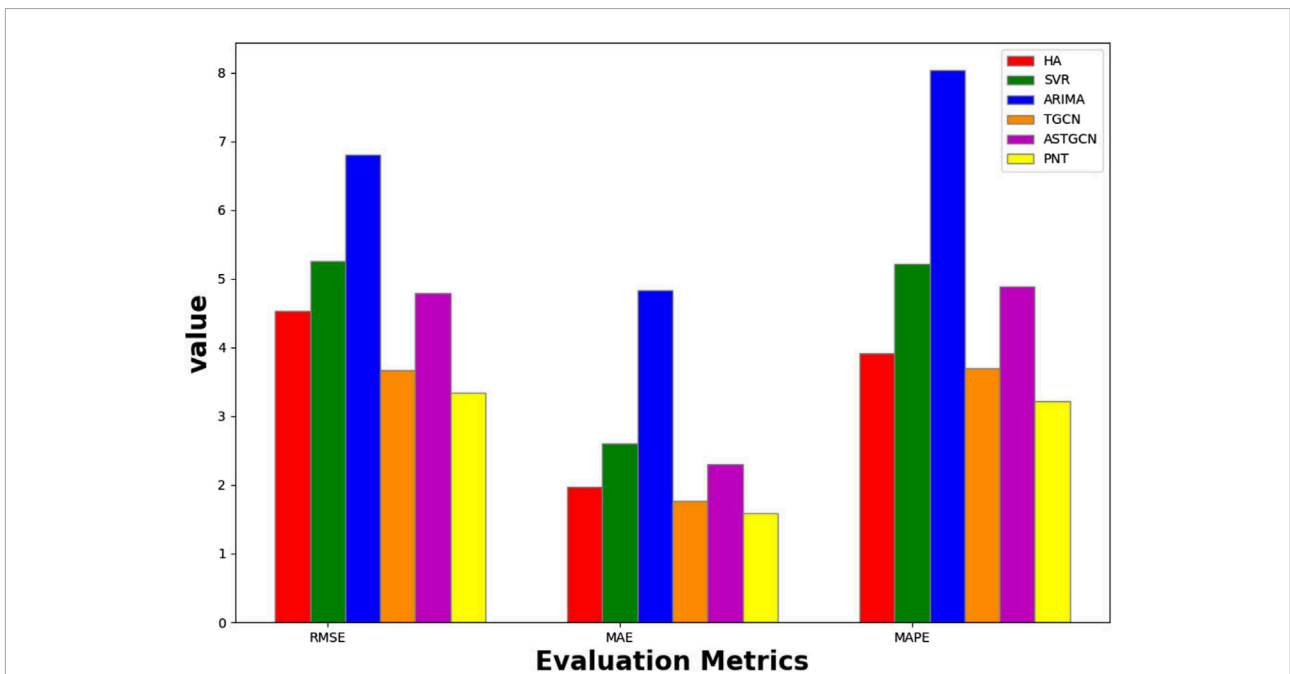


FIGURE 9
pems04-30 min.

$$Laplacian = \hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \tag{15}$$

$$GWM = x_t * Laplacian \tag{16}$$

I is the unit matrix of the corresponding dimension.

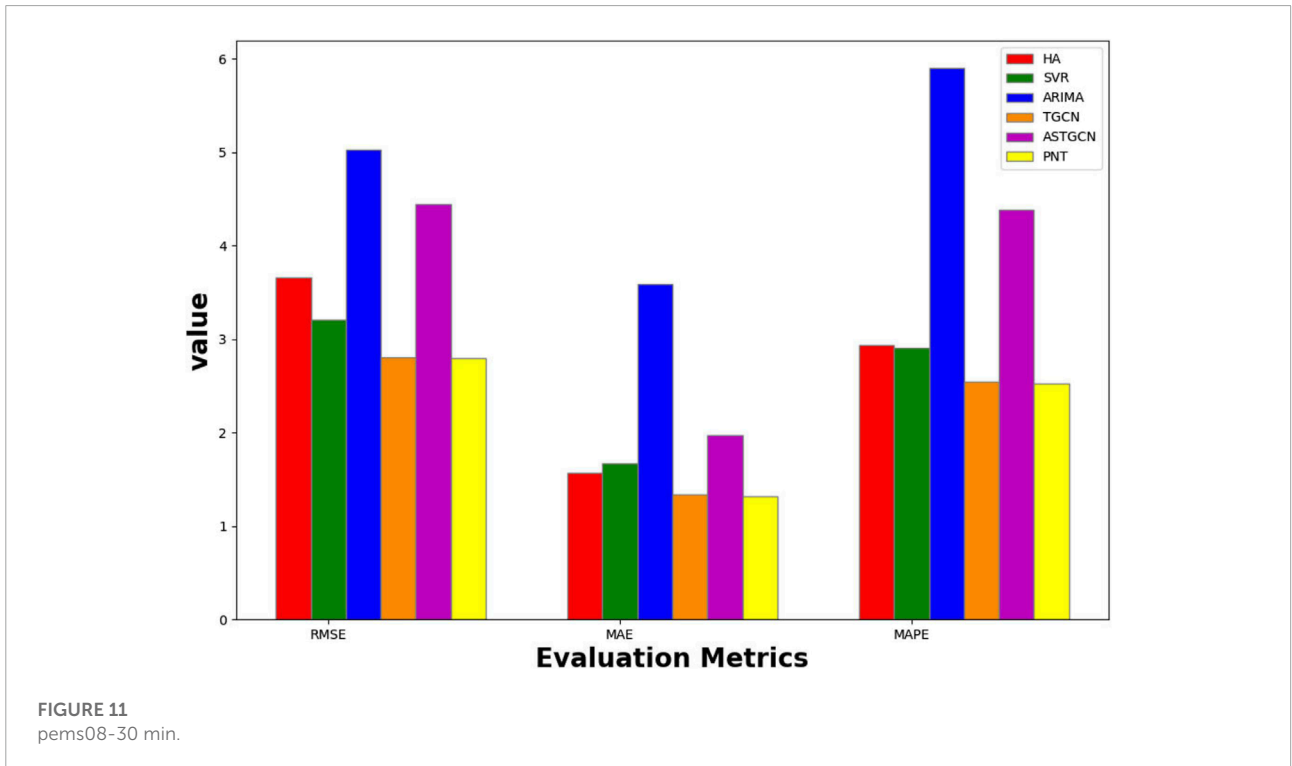
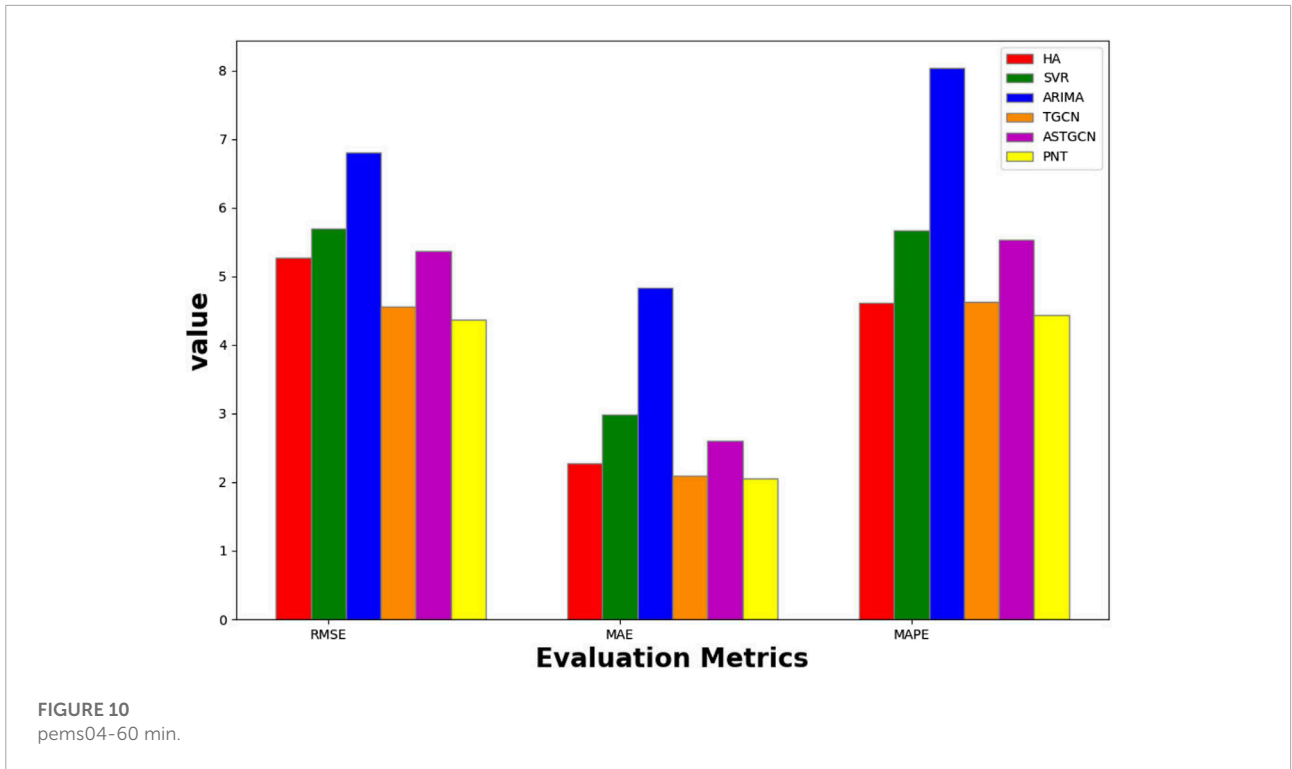
3.2.3 Loss function section

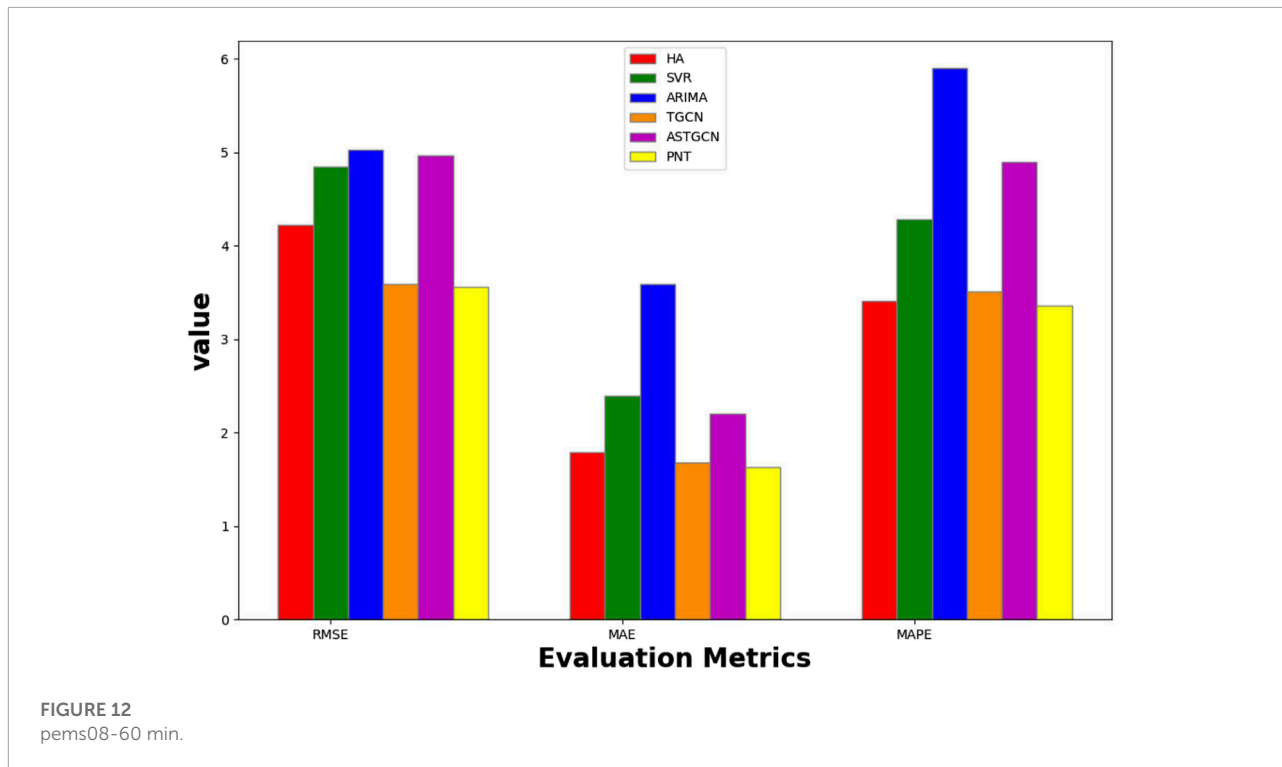
In the PNT model training process, our optimization objective is to minimize the error between the model's prediction

and the actual traffic speed on the road. the loss function of the PNT model uses the following equation:

$$Loss = \|y_t - \hat{y}_t\| + \lambda L_2 \tag{17}$$

We use y_t and \hat{y}_t to represent the real traffic information and the predicted traffic information, respectively. The first term of Eq. is used to minimize the error between the predicted traffic information and the real traffic information, the second term L_2





is the L2 regularization term that helps to avoid the overfitting problem, and λ is the hyperparameter.

4 Experiments

In this section, we describe the experimental setup, including data description, evaluation metrics, parameter settings, and comparison methods.

4.1 Description of experimental data

The experiments of our PNT model are carried out on top of two real highway datasets from the well-known pems in the traffic domain, i.e., pems04 and pems08. The two datasets, pems04 and pems08, are collected in real time on the highway by loop detectors. We selected the traffic speeds from January 1 to 28 February 2018 (speed information collected every 5 minutes) for 307 sensors of pems04 and from July 1 to 31 August 2016 for 170 sensors of pems08, respectively. The whole pems04 data is divided into two major parts in total, one part is the characteristic information of three dimensions, which are the flow, occupancy and speed of the road. The other part is the adjacency matrix composed of the adjacency elements calculated from the adjacency relationship between each sensor, for this part of the information of the adjacency matrix, one can read this paper by ASTGCN [24] for understanding. In our experiments,

we normalize the input traffic speed data to the interval [0,1] and invert it. In addition, we used 80% of the data as the training set and the remaining 20% as the test set and trained the PSN model using the Adam optimizer. We predict the road speed information of these sensors for the next 30 and 60 min.

Next, we give a partial view of the pems04 and pems08 datasets available. From the above, we can see that the pems04 and pems08 datasets are collected from 307 sensors and 170 sensors, respectively. We give the first sensor data and the last sensor data viewable in the pems04 and pems08 datasets, respectively, which are described in Figures 5–8.

4.2 Evaluation metrics

To evaluate the prediction performance of the proposed model, we use the following metrics to evaluate the prediction results.

(1) Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{MN} \sum_{j=1}^M \sum_{i=1}^N (y_i^j - \tilde{y}_i^j)^2} \tag{18}$$

(2) Mean Absolute Error (MAE):

$$MAE = \frac{1}{MN} \sum_{j=1}^M \sum_{i=1}^N |y_i^j - \tilde{y}_i^j| \tag{19}$$

(3) Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{MN} \sum_{j=1}^M \sum_{i=1}^M \left| \frac{y_i^j - \tilde{y}_i^j}{y_i^j} \right| \quad (20)$$

where y_i^j and \tilde{y}_i^j represent the real traffic information and prediction information of the j th time sample on the i th road, respectively. M is the number of time samples; N is the number of roads.

Specifically, RMSE, MAE and MAPE were used to measure the prediction error: the smaller the value is, the better the prediction effect is.

4.3 Model parameters

In our model, we have a number of hyperparameters to determine to ensure that the model will run best. Briefly, the parameters that are relatively important for the PNT model include: learning rate, batch size, and training epoch. In this experiment, we manually set the learning rate to 0.001, the batch size to 32, and the training epoch to 500, and we run the model under NVIDIA-SMI 450.142.00 Driver Version: 450.142.00 CUDA Version: 11.0 python = 3.6. The hardware configuration is three V100 graphics cards. It is worth mentioning that we use a sliding window mechanism for the input data, thus extending the two parameters seq-len and pre-len that vary flexibly according to the prediction task. seq-len indicates the length of the data we obtain from the dataset each time, and pre-len indicates the length of the data we want to predict. For example, in this paper, we make traffic speed predictions for the next 30 min and the next 60 min for the pems04 and pems08 datasets based on seq-len = 12, pre-len = 6 and seq-len = 12, pre-len = 12, respectively, which means we take traffic speed information from the dataset for 1 h at a time to predict the future pre-len time steps (5 min for each time step interval in the pems04 and pems08 datasets). Our experimental results show that the model takes about 3 h to run for 500 training epochs. We currently use runtime to evaluate the computational complexity of algorithms, and we will explore further the computational complexity of algorithms in future work.

4.4 Comparison method

In our experiments, we compare our model with the following methods:

HA: Historical average, in which we calculate the average road speed at the historical moment as the predicted value for the current time step.

SVR: Train the model using historical data, get the relationship between input and output, and then predict future

traffic data for the training model. We use a linear SVR in this model.

ARIMA: differential autoregressive moving average model, The ARIMA model is used to first make the non-stationary data stationary and later to process the stationary data. For more information about ARIMA, please refer to this paper (Gilbert, 2005).

T-GCN: This model combines the graph convolutional network GCN and the Gated Recurrent Unit GRU. In this case, the GCN is used to learn the complex topology to obtain spatial correlation, while the Gated Recurrent Unit is used to learn the dynamics of the traffic data to obtain temporal correlation. For details of the T-GCN model, please refer to this paper (Zhao et al., 2019).

ASTGCN: The ASTGCN model mainly consists of three independent components, by simulating the three temporal dimensional properties of traffic flow separately, i.e., weekly cycle, daily cycle, and recent cycle. Each of these components contains a spatio-temporal attention mechanism and a spatio-temporal convolution. For details of the AST-GCN model, please refer to this paper (Guo et al., 2019).

4.5 Analysis of experimental results

We compared our proposed PSN model with the three baseline methods as well as the T-GCN model and AST-GCN model proposed by previous authors, and the experimental results are shown in Tables 1–4, and the visualization plots are shown in Figures 9–12. By our observation of the PEMS04 table data, we can see that our proposed PSN model achieves the best performance under all evaluation metrics. We can clearly see that the three traditional time series analysis methods, HA, SVR, and ARIMA, do not have good prediction results. This may be due to HA's neglect of temporal features, which leads to poorer prediction results. For ARIMA, because of the structure of its model, it does not capture the nonlinearity and uncertainty of the data well, and the SVR model, although a classical regression algorithm, does not achieve good results for traffic data prediction, which may also be due to the time-varying and nonlinear characteristics of traffic data.

In contrast, deep learning-based methods obtain better prediction results with traditional time-series methods. both T-GCN and AST-GCN models consider the potential temporal and spatial characteristics of the traffic data. However, as the prediction step increases, the prediction accuracy of these two models starts to lag behind that of the PSN model. This is mainly because the propagation of traffic information over the road network topology graph becomes particularly important with the accumulation of time, and the ensuing uncertainty problem is

TABLE 1 PEMS04-30 min

	HA	SVR	ARIMA	T-GCN	ASTGCN	PSN
RMSE	4.5264	5.2506	6.7964	3.6679	4.7900	3.3453
MAE	1.9745	2.5948	4.8356	1.7689	2.3000	1.5840
MAPE (%)	3.91	5.21	8.03	3.70	4.88	3.22

The bold values indicate the optimal value for each of the evaluated metrics.

TABLE 2 PEMS04-60 min.

	HA	SVR	ARIMA	T-GCN	ASTGCN	PSN
RMSE	5.2749	5.6984	6.7964	4.5607	5.3700	4.3651
MAE	2.2712	2.9803	4.8356	2.0996	2.6000	2.0594
MAPE (%)	4.61	5.67	8.03	4.63	5.53	4.43

The bold values indicate the optimal value for each of the evaluated metrics.

TABLE 3 PEMS08-30 min.

	HA	SVR	ARIMA	T-GCN	ASTGCN	PSN
RMSE	3.6596	3.2057	5.0271	2.8036	4.4500	2.7965
MAE	1.5668	1.6711	3.5872	1.3338	1.9701	1.3171
MAPE (%)	2.94	2.91	5.90	2.54	4.38	2.52

The bold values indicate the optimal value for each of the evaluated metrics.

TABLE 4 PEMS08-60 min.

	HA	SVR	ARIMA	T-GCN	ASTGCN	PSN
RMSE	4.2203	4.8524	5.0271	3.5927	4.9710	3.5614
MAE	1.7883	2.3973	3.5872	1.6818	2.2002	1.6321
MAPE (%)	3.41	4.28	5.90	3.51	4.90	3.36

The bold values indicate the optimal value for each of the evaluated metrics.

TABLE 5 PEMS04-30 min.

	Without-GWM	Without-GSSM	PSN
RMSE	3.3572	3.4386	3.3453
MAE	1.5889	1.6221	1.5840
MAPE (%)	3.23	3.28	3.22

The bold values indicate the optimal value for each of the evaluated metrics.

TABLE 6 PEMS04-60 min.

	Without-GWM	Without-GSSM	PSN
RMSE	4.3884	4.5097	4.3651
MAE	2.1413	2.1002	2.0594
MAPE (%)	4.58	4.60	4.43

The bold values indicate the optimal value for each of the evaluated metrics.

TABLE 7 PEMS08-30 min.

	Without-GWM	Without-GSSM	PSN
RMSE	2.8289	2.7863	2.7965
MAE	1.4573	1.3095	1.3171
MAPE (%)	2.76	2.55	2.52

The bold values indicate the optimal value for each of the evaluated metrics.

TABLE 8 PEMS08-60 min.

	Without-GWM	Without-GSSM	PSN
RMSE	3.5799	3.5908	3.5614
MAE	1.6513	1.6617	1.6321
MAPE (%)	3.39	3.37	3.36

The bold values indicate the optimal value for each of the evaluated metrics.

further amplified. However, because our proposed PSN model has better results for the propagation of graph information based on road network topology, it performs better on the dataset.

To further illustrate the role of GWM and GSSM modules in the PNT model. We performed ablation experiments and the experimental results are shown in **Tables 5–8**. In the table, Without-GWM and Without-GSSM represent the PNT model with the GWM module removed and with the GSSM module removed, respectively. We can find from the table that both the GSSM module and the GWM module contribute to the prediction effect of the model. However, we can see that the GSSM module improves the prediction of the model more than the GWM module. We speculate that the global spatio-temporal attention module can capture the hidden information behind the data with spatio-temporal characteristics, but the graph information diffusion performed by the graph wandering module may diffuse some redundant information, which leads to the weak improvement of the GWM module on the prediction of the model.

5 Conclusion and future work

In this paper, we propose a new spatio-temporal attention-based model PNT and apply this to traffic speed prediction. This model is inspired by the pagerank algorithm and proposes a new idea of combining traffic data and graph structure information, which not only avoids the complex operation of graph convolution, but also smoothly aggregates and updates the traffic information based on the graph structure, and achieves a certain traffic prediction accuracy. Our experiments show that our proposed model PNT performs well on both pems04 and pems08 datasets.

In the future, we plan to delve into feature propagation based on topology graphs of urban road networks. Specifically, we will try to explore the impact of other peripheral information (e.g., POIs near roads, road types, etc.) in the urban road network on traffic information.

References

- Ahmed, M. S., and Cook, A. R. (1979). *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*, 722.
- An, J., Guo, L., Liu, W., Fu, Z., Ren, P., Liu, X., et al. (2021). Igagcn: Information geometry and attention-based spatiotemporal graph convolutional networks for traffic flow prediction. *Neural Netw.* 143, 355–367. doi:10.1016/j.neunet.2021.05.035
- Belhadi, A., Djenouri, Y., Djenouri, D., and Lin, J. C.-W. (2020). A recurrent neural network for urban long-term traffic flow forecasting. *Appl. Intell. (Dordr.)* 50, 3252–3265. doi:10.1007/s10489-020-01716-1

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/divanoresia/Traffic>.

Author contributions

TK: Conceptualization, Methodology, validation, Writing—original draft. HW: Methodology, Data curation, Visualization, Software. TW: Writing—Resources, Software, Formal analysis, writing—original draft. JP: Writing—review and editing, Funding acquisition. HJ: Conceptualization, Writing—Supervision.

Funding

This paper has been jointly supported by the National Natural Science Foundation of China (Grant No. 52177102), the Natural Science Foundation of Guangdong Province (Grant No. 2021A1515011685), and the Foundations of Shenzhen Science and Technology Committee (Grant Nos. JCYJ20190808143619749, GJHZ20200731095610032).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Cheng, T., Wang, J., Haworth, J., Heydecker, B., and Chow, A. (2014). A dynamic spatial weight matrix and localized space–time autoregressive integrated moving average for network modeling. *Geogr. Anal.* 46, 75–97. doi:10.1111/gean.12026

- Cui, Z., Henrickson, K., Ke, R., and Wang, Y. (2019). Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Trans. Intell. Transp. Syst.* 21, 4883–4894. doi:10.1109/tits.2019.2950416

- Diao, Z., Wang, X., Zhang, D., Liu, Y., Xie, K., and He, S. (2019). Dynamic spatial-temporal graph convolutional neural networks for traffic

- forecasting. *Proc. AAAI Conf. Artif. Intell.* 33, 890–897. doi:10.1609/aaai.v33i01.3301890
- Ding, Q. Y., Wang, X. F., Zhang, X. Y., and Sun, Z. Q. (2011). “Forecasting traffic volume with space-time arima model,” in *Advanced materials research* (Stafa-Zurich, Switzerland: Trans Tech Publ), Vol. 156, 979–983.
- Fu, X. (2022). Statistical machine learning model for capacitor planning considering uncertainties in photovoltaic power. *Prot. Control Mod. Power Syst.* 7, 5–13. doi:10.1186/s41601-022-00228-z
- Gilbert, K. (2005). An arima supply chain model. *Manag. Sci.* 51, 305–310. doi:10.1287/mnsc.1040.0308
- Graves, A. (2012). Long short-term memory. *Supervised sequence Label. Recurr. neural Netw.* 2012, 37–45.
- Guo, S., Lin, Y., Feng, N., Song, C., and Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proc. AAAI Conf. Artif. Intell.* 33, 922–929. doi:10.1609/aaai.v33i01.3301922
- Klicpera, J., Bojchevski, A., and Günnemann, S. (2018). Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv Prepr. arXiv:1810.05997*.
- Li, D., Zhao, Y., and Zhao, Y. (2022). A dynamic-model-based fault diagnosis method for a wind turbine planetary gearbox using a deep learning network. *Prot. Control Mod. Power Syst.* 7, 22–14. doi:10.1186/s41601-022-00244-z
- Min, X., Hu, J., Chen, Q., Zhang, T., and Zhang, Y. (2009). “Short-term traffic flow forecasting of urban network based on dynamic starima model,” in 2009 12th International IEEE conference on intelligent transportation systems (St. Louis, MO, USA: IEEE), 1–6.
- Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., and Damas, L. (2013). Predicting taxi-passenger demand using streaming data. *IEEE Trans. Intell. Transp. Syst.* 14, 1393–1402. doi:10.1109/tits.2013.2262376
- Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* 30, 83–98. doi:10.1109/msp.2012.2235192
- Tedjopurnomo, D. A., Bao, Z., Zheng, B., Choudhury, F., and Qin, A. (2020). A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. *IEEE Trans. Knowl. Data Eng.* 34, 1544. doi:10.1109/tkde.2020.3001195
- Tu, Y., Lin, S., Qiao, J., and Liu, B. (2021). Deep traffic congestion prediction model based on road segment grouping. *Appl. Intell. (Dordr.)* 51, 8519–8541. doi:10.1007/s10489-020-02152-x
- Van Lint, J., Hoogendoorn, S., and van Zuylen, H. J. (2002). Freeway travel time prediction with state-space neural networks: Modeling state-space dynamics with recurrent neural networks. *Transp. Res. Rec.* 1811, 30–39. doi:10.3141/1811-04
- Van Lint, J., and Van Hinsbergen, C. (2012). Short-term traffic and travel time prediction models. *Artif. Intell. Appl. Crit. Transp. Issues* 22, 22–41.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. neural Inf. Process. Syst.* 30.
- Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we’re going. *Transp. Res. Part C Emerg. Technol.* 43, 3–19. doi:10.1016/j.trc.2014.01.005
- Williams, B. M., and Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *J. Transp. Eng.* 129, 664–672. doi:10.1061/(asce)0733-947x(2003)129:6(664)
- Wu, K., Gu, J., Meng, L., Wen, H., and Ma, J. (2022). An explainable framework for load forecasting of a regional integrated energy system based on coupled features and multi-task learning. *Prot. Control Mod. Power Syst.* 7, 24–14. doi:10.1186/s41601-022-00245-y
- Yu, B., Yin, H., and Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv Prepr. arXiv:1709.04875*.
- Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X., and Li, T. (2018). Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artif. Intell.* 259, 147–166. doi:10.1016/j.artint.2018.03.002
- Zhang, T., Sun, L., Yao, L., and Rong, J. (2017). Impact analysis of land use on traffic congestion using real-time traffic and poi. *J. Adv. Transp.* 2017, 1–8. doi:10.1155/2017/7164790
- Zhang, Z., Li, M., Lin, X., Wang, Y., and He, F. (2019). Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. *Transp. Res. part C Emerg. Technol.* 105, 297–322. doi:10.1016/j.trc.2019.05.039
- Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., et al. (2019). T-Gcn: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* 21, 3848–3858. doi:10.1109/tits.2019.2935152