



Article Hybrid Graph Models for Traffic Prediction

Renyi Chen 🕩 and Huaxiong Yao *

Computer School, Central China Normal University, Wuhan 430079, China; chenrenyi@mails.ccnu.edu.cn * Correspondence: yaohuaxiong@ccnu.edu.cn

Abstract: Obtaining accurate road conditions is crucial for traffic management, dynamic route planning, and intelligent guidance services. The complex spatial correlation and nonlinear temporal dependence pose great challenges to obtaining accurate road conditions. Existing graph-based methods use a static adjacency matrix or a dynamic adjacency matrix to aggregate spatial information between nodes, which cannot fully represent the topological information. In this paper, we propose a Hybrid Graph Model (HGM) for accurate traffic prediction. The HGM constructs a static graph and a dynamic graph to represent the topological information of the traffic network, which is beneficial for mining potential and obvious spatial correlations. The proposed method combines a graph neural network, convolutional neural network, and attention mechanism to jointly extract complex spatial–temporal features. The HGM consists of two different sub-modules, called spatial–temporal attention module and dynamic graph convolutional network, to fuse complex spatial–temporal information. Furthermore, the proposed method designs a novel gated function to adaptively fuse the results from spatial–temporal attention and dynamic graph convolutional network to improve prediction performance. Extensive experiments on two real datasets show that the HGM outperforms comparable state-of-the-art methods.

Keywords: traffic prediction; graph neural network; attention

1. Introduction

Traffic prediction is of great significance for optimizing urban traffic systems, improving traffic efficiency, reducing congestion, and improving environmental quality. Recently, many countries have been committed to developing intelligent transportation systems. As an indispensable part of traffic prediction, obtaining real-time traffic conditions can help people better arrange travel plans and share resources [1–6].

Traffic prediction aims to estimate future road conditions from the historical state recorded by the traffic system [7–10]. A real-time and accurate grasp of road conditions plays an important role in traffic management and resource allocation. However, the complex spatial correlation and nonlinear temporal dependence of the traffic network pose great challenges to obtaining accurate traffic prediction. Although many works have achieved excellent predictive performance, accurate traffic prediction still faces the following challenges, as shown in Figure 1:



Figure 1. Complex spatial-temporal correlations.



Citation: Chen, R.; Yao, H. Hybrid Graph Models for Traffic Prediction. *Appl. Sci.* 2023, *13*, 8673. https:// doi.org/10.3390/app13158673

Academic Editor: Augusto Ferrante

Received: 10 May 2023 Revised: 16 July 2023 Accepted: 24 July 2023 Published: 27 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). (1) Complex spatial correlation. Node *O* received different impacts from adjacent nodes (such as *A*, *B*, *C*, *D*) at different time steps. Capturing the dynamic relationship between nodes is crucial to improving predictive performance.

(2) Nonlinear temporal dependence. Node *B* is influenced by its own node and adjacent nodes at the previous state and the influence weight β_i changes dynamically over time.

Traditional traffic prediction methods such as the autoregressive integral moving average algorithm (ARIMA) [11] or Kalman filter extract time series information of the traffic network based on time stationary assumption. Traditional methods have achieved great success in traffic prediction. However, the flexibility of these methods is limited due to the complex spatial correlations and nonlinear temporal dependencies of the traffic network.

Due to the powerful feature representation ability of deep learning, deep learningbased methods have made breakthroughs in computer vision, natural language processing, and traffic prediction. Deep learning-based methods automatically learn the rich spatialtemporal information of traffic data through deep neural networks. The temporal signal of a traffic network can be regarded as a sequence-to-sequence model, and the temporal interactivity of nodes can be extracted by the 1D convolution operations in traffic prediction. Compared with 1D convolutional neural networks, recurrent neural networks (RNNs) have attracted extensive attention because these methods can model long-range temporal information [12]. However, RNNs learning-based methods are prone to the vanishing gradient, which makes the backbone network unable to effectively train and learn deep semantic information.

In the spatial dimension, convolutional neural networks (CNNs) have excellent feature extraction capabilities on the Euclidean data structure. However, the spatial distribution of the traffic network nodes can be regarded as a non-Euclidean data structure, which leads to the fact that CNNs learning-based methods have not been widely used in traffic prediction. In recent years, graph neural networks (GNNs) have received extensive attention due to their excellent results on non-Euclidean structures [13–16]. The GNNs, which can roughly be divided into the spectral domain and spatial domain, update the state of the current node by aggregating the information of adjacent nodes. However, the adjacency matrix generated based on distance or similarity has several limitations. These methods are subjective and incomplete and cannot represent the potential correlation of the traffic network. For example, as shown in Figure 1, node O represents a commercial area and node D is a residential area. Although node O and node D are far apart in space, they have a strong correlation.

In this paper, we propose a novel Hybrid Graph Model (HGM) to extract rich spatialtemporal information about the traffic network. To make full use of spatial information, the proposed method represents the topological information of the traffic network through a general graph structure. The HGM combines a graph neural network, convolutional neural network, and attention mechanism to jointly extract complex spatial-temporal information through different branches. The proposed method constructs the spatial information of distance-based nodesthrough a large amount of traffic data recorded by sensors, which we call static graph. Then, the HGM aggregates the spatial information of nodes using a graph convolutional neural network and combines attention to simultaneously extract temporal information. In order to mine the potential topological structure of the traffic network, the HGM updates the topological information during the model training process, which is called a dynamic graph, and then uses a 1D convolutional neural network to mine the temporal dependencies between nodes. Furthermore, we design a novel gated function to fuse the results from different branches. The gated function dynamically calculates weights for both components according to the branch results to improve the prediction performance. Experiments on two real-world traffic datasets demonstrate that the HGM achieves state-of-the-art performances. The main contributions of this research are summarized as follows:

(1) The proposed method constructs topological information of the network from different visions to fully represent the topological information. The HGM represents road information through a static graph and a dynamic graph, which is conducive to highlighting the information of adjacent nodes and mining the potential correlation between nodes.

(2) The proposed method uses different components to perform feature extraction on bigraph topology, which can fully capture the dynamic road information, and a gated function is designed for the adaptive fusion of results from different branches.

(3) We evaluate the HGM on two real-world traffic datasets and have demonstrated prediction performance over comparable state-of-the-art methods in both long-term and short-term predictions.

The remainder of this paper is organized as follows: Section 2 reviews the related work. The problem statement of traffic prediction is presented in Section 3. Section 4 details the proposed approach. Section 5 reports the experimental settings, followed by the discussion of experimental results in Section 6. Section 7 briefly concludes this work.

2. Related Work

An intelligent and smart traffic management system is essential for managing the increasing volume of vehicles and the human population in smart cities. Traffic prediction aims to predict the future road state through the historical information recorded by the traffic system. The complex spatial correlation and nonlinear temporal dependence of the traffic network pose great challenges to accurately obtaining future road conditions. Many works have been proposed and have achieved excellent classification performance in traffic prediction.

2.1. Traditional Method

Classic statistical and machine learning-based methods are two major representative works in traffic prediction [17,18]. Compared with statistics-based methods, machine learning-based methods such as support vector machines and autoregressive integral moving averages take into account the complex spatial–temporal information of the traffic network, and these methods achieve better predictive performance than linear models. These methods can effectively alleviate the complex spatial correlation and nonlinear temporal dependence of the traffic network. However, these methods rely on manual feature extraction and cannot effectively extract deep semantic information, and the predictive performance of these methods is limited.

2.2. Deep Learning Method

Since deep learning methods can extract deep semantic information and have powerful feature learning capabilities, deep learning-based methods have made breakthroughs in computer vision, natural language processing, and traffic prediction [19,20]. Convolutional neural networks use convolution operations to aggregate features of adjacent nodes to extract rich spatial-temporal information. Yao et al. [21] and Yu et al. [22] divided the traffic network according to geographic location and then fused the domain information through a convolutional neural network. However, the performance of CNN learning-based methods will drop sharply on non-Euclidean data structures. Due to the irregular distribution of sensors in the traffic network, the nodes of the traffic network are more suitable to be regarded as a non-Euclidean structure. Since the graph-based method can effectively represent the topological information between nodes, it has gradually become an alternative method and has received extensive attention. Zhao et al. [23] and Zhang et al. [24] used a graph to represent road spatial information and demonstrated the effectiveness of the proposed method. Yin et al. [25] proposed to set up different hops and consider the information of long-range adjacent nodes to provide rich spatial information. As shown in Figure 1b, although the distance between node O and node D is large, they have a strong correlation. Previous methods construct the topological information of the network distance-based or similarity-based between nodes, which cannot discover the potential interactivity.

Many works have started to reflect other factors apart from distance information [26,27]. Li et al. [26] modified the distance map by adding additional information, such as inflow/outflow and reachability. Lv et al. [28] considered multi-graph convolutions to provide rich spatial-temporal information for traffic prediction. In order to achieve better prediction performance, Wu et al. [29] proposed to dynamically generate the topology structure between nodes during the model training process to mine the potential information between nodes. However, these methods do not utilize static and dynamic topology information at the same time, which can also lead to information loss. In the temporal dimension, the information of adjacent moments is extracted by means of 1D CNNs or RNNs in the early stage. Guo et al. [30] and Yu et al. [31] adopted CNNs to construct nonlinear temporal dependencies for different time steps. Recurrent neural networks such as long short-term memory or gated recurrent units have been successfully applied to traffic prediction due to their ability to model long-range temporal information. However, recurrent neural networks for sequence learning require iterative training, which may introduce error accumulation and cause a vanishing gradient.

2.3. Attention

The attention mechanism uses limited resources to quickly select high-value information from a large amount of information [32]. Attention-based methods have been widely used and achieved great successes in various tasks, such as natural language processing [33,34], computer vision [35], and speech recognition [36]. To characterize the spatial correlations and temporal dependencies, several works apply attention mechanisms to model spatial-temporal information. Chen et al. [37] applied attention to aggregate spatial information of different scales. Yao et al. [21] used LSTM and attention mechanism to extract temporal information. Zheng et al. [38] and Guo et al. [30] employed attention to capture the relationships between intersequence and intrasequence nodes. Inspired by the above research methods, in order to capture the topology and complex spatial-temporal patterns of the traffic network, the proposed method uses graph convolution and attention mechanisms to model the network structure.

3. Preliminaries

3.1. Traffic Prediction on Road Graphs

In this paper, we define the traffic network as a directed graph G = (V, E, A), where V is a set with |V| = N nodes, E is a set of edges representing the connectivity between the nodes, and $A \in \mathbb{R}^{N \times N}$ denotes the adjacent matrix of the graph representing the proximity between nodes. The intelligent transportation system records the road conditions $X = \{X_0, X_1, \dots, X_t, \dots\}$ of all nodes in the traffic network, where $X_t = \{x_{1,t}, x_{2,t}, \dots, x_{N,t}\} \in \mathbb{R}^{N \times 1}$ represents the road information at time t. The goal of traffic prediction is to predict the future state based on the observed historical state. In this paper, we formulate the traffic prediction problem as finding a function F to predict the next Q time steps based on historical P steps of historical data:

$$\hat{Y} = \{X_{t+1}, X_{t+2,\dots,X_{t+O}}\} = F_{\theta}(X_t, X_{t-1}, \dots, X_{t-P+1})$$
(1)

where θ denotes learnable parameters.

3.2. Graph Neural Network

As shown in Figure 2a, previous studies [21,22] have demonstrated that the CNN learning-based method exhibits excellent performance in grid-structured data (e.g., images, videos). The convolutional neural network is usually used as an effective method for

feature extraction in image processing. Given image $X \in \mathbb{R}^{M \times N}$ and filter $W \in \mathbb{R}^{U \times V}$, the output of convolution operation at position (i, j) can be defined as:

$$y_{i,j} = \sum_{u=1}^{U} \sum_{v=1}^{V} w_{u,v} x_{i-u+1,j-v+1}$$
(2)

As the number of hidden layers of the convolutional neural network increases, the computing resources required by the model also increase sharply. Due to the irregular distribution of sensors, the Euclidean data structure cannot truly reflect road information. Graph neural networks can be roughly divided into spectral-based methods and spatial-based methods. Spectral-based methods define graph convolution operations by introducing filters from the perspective of graph signal processing [39], which are interpreted as removing noise from graph signals. Spatial-based methods [16,40,41] formulate the graph convolution operation as aggregating the information of adjacent nodes, which can roughly be seen as a type of CNN. Spatial-based methods have developed rapidly due to their attractive efficiency, flexibility, and generality.



Figure 2. Convolutional neural network and graph neural network. The black node represents the central node and the light blue node is the adjacent node of the black node.

Thomas et al. [42] introduced a simple yet flexible model f(X, A) for information propagation on graphs and considered GCNs on a graph with a symmetric adjacency matrix A, which can be defined as:

$$H^{l} = f(X, A) = AH^{l-1}W$$
(3)

where $X \in \mathbb{R}^{N \times H}$ denotes the input signals and $W \in \mathbb{R}^{H \times D}$ represents learnable parameters.

4. Methodology

Figure 3 is the framework of the proposed HGM. The proposed method aggregates traffic network information from different visions, which is beneficial for strengthening the shallow information and can dynamically mine potential correlations between nodes. The HGM captures complex spatial correlations and nonlinear temporal dependencies through the spatial-temporal attention (STA) module and the dynamic graph convolution network (DGCN). Specifically, the HGM constructs a distance-based static adjacency matrix to represent topology information and combine spatial-temporal attention to synchronously extract spatial-temporal information in the STA module. In addition, in order to characterize the potential dependencies between nodes, the proposed method randomly generates a dynamic adjacency matrix, which can better represent the topology of the network in the DGCN. The DGCN extracts complex spatial information and nonlinear temporal dependencies at the same time through spatial domain-based graph convolutional neural network and convolutional neural network, respectively. Finally, the HGM uses a gated function to adaptively fuse the results of both branches to improve the prediction performance. Next, we introduce the main components STA module and DGCN in detail.



Figure 3. Model framework.

4.1. STA Module

In this paper, we follow the idea of [38] to design a spatial-temporal attention module, which mainly includes temporal attention and spatial attention, to extract features of historical traffic conditions. In the spatial-temporal attention, the HGM obtains spatial embedding $E_{v_i}^S \in \mathbb{R}^D$ and temporal embedding $E_t^T \in \mathbb{R}^{7+q}$ by using word2vec and one-hot encoding methods, respectively. To obtain both temporal and spatial information, the method obtains the spatial-temporal embedding $STE_{v_i,t} = (E_t^T + E_{v_i}^S) \in \mathbb{R}^{(T_P + T_Q) \times N \times D}$ through linear projection and concatenation operations, where $v_i \in V$, $T \in (t - P + 1, \dots, t, t + 1, \dots, t + Q)$ and D represents the output dimensions of the model. Then, the proposed method uses a multi-head attention to simultaneously extract spatial correlations h_S^A and temporal dependencies h_T^A , and the extracted information is dynamically fused through a novel gated function.

4.1.1. Temporal Attention and Spatial Attention

In recent years, attention mechanisms have been widely used in various tasks and have achieved excellent performance. The goal of the attention mechanism is to select valuable information for the current task from all inputs. As shown in Figure 1b, historical traffic conditions can also have an impact on future road conditions, and it presents a nonlinear relationship. As shown in Figure 4, in the STA module, the proposed method employs temporal attention and spatial attention to extract nonlinear temporal dependencies and complex spatial correlations, respectively. Instead of the idea of extracting spatial–temporal information in stages, spatial–temporal attention simultaneously extracts road information to improve accuracy.



Figure 4. Spatial-temporal attention module.

In the temporal dimension, the HGM uses a novel attention mechanism, which can dynamically capture long-range temporal information. The proposed method uses a multi-head attention mechanism to stabilize the training process of the proposed method. Multi-head attention can focus on information in different subspaces, which is beneficial for extracting rich spatial-temporal information.

As shown in Figure 5, the *k*th attention score of node between *t* and t_j time steps can be defined as:

$$\omega_{t,t_{j}}^{k} = \frac{\langle h_{v_{i},t}^{l-1} || STE_{v_{i},t_{j}} h_{v_{i},t_{j}}^{l-1} || STE_{v_{i},t_{j}} \rangle}{\sqrt{d_{k}}}$$
$$\beta_{t,t_{j}}^{k} = \frac{exp(\omega_{t,t_{j}}^{k})}{\sum_{t_{j}=t_{1}}^{t_{h}} exp(\omega_{t,t_{j}}^{k})}$$

where || represents the vector concatenation operation, $h_{v_i,t}^{l-1}$ is the hidden layer state of the node v_i at time $t, < \cdot, \cdot >$ denotes the inner product operator, and $d_k = \frac{D}{K}$ means the output dimension of the *k*th attention head. Therefore, the dynamic information extracted by the node v_i at the *l*th layer through temporal attention at time *t* can be defined as:

$$ht_{v_i,t}^{l} = ||_{k=1}^{K} \{ \sum_{t_j=t_1}^{T_p} \beta_{t,t_j}^k \cdot h_{v_i,t_j}^{l-1} \}$$

It can be seen from Figure 1 that the influence relationship between nodes is changing dynamically at different time steps. Similar to temporal attention, the HGM also takes multi-head attention to extract spatial information. Spatial attention assigns different weights at different times according to the correlation between nodes. The proposed method comprehensively considers the dynamic spatial and temporal correlation, which is beneficial to improving the prediction performance.



Figure 5. Temporal attention.

4.1.2. Gated Function

To fuse the results from different branches, the proposed method designs a novel gated function as shown in Figure 6. Instead of assigning fixed weights, the gated function aims to calculate a dynamic weight based on the result from branches A and B, which can be defined as:

$$C = z \times A + (1 - z) \times B$$
$$z = \sigma(\epsilon_1 A + \epsilon_2 B + b)$$

where *A*, *B* indicate different components $, \epsilon_1, \epsilon_2 \in \mathbb{R}^{N \times N}$ and $b \in \mathbb{R}^D$ represent the trainable parameters, and $\sigma(\cdot)$ is the sigmoid activation function. The gated function effectively combines the results from different branches. In this paper, the proposed method mainly uses the gated function to fuse the results from the spatial attention and temporal attention and to fuse the results from the STA module branch and DGCN branch.



Figure 6. Gated function.

4.2. DGCN Block

As shown in Figure 7, the HGM uses a graph convolutional network to mine potential topological information. The HGM dynamically updates the adjacency matrix during model training to construct topological information of the traffic network. Meanwhile, the DGCN adopts the 1D convolution to extract nonlinear temporal dependencies. The previous GNN-based methods mainly construct the topology information of the traffic network through distance or similarity between nodes. The adjacency matrix is considered prior knowledge and is fixed during the model training process. These methods are subjective and incomplete and may cause a huge impact on the predictive performance. Veličković et al. [43] combined attention with a graph to reflect the dynamic information of the network, but it still relies on prior knowledge to a certain extent. However, the constructed distance-based adjacency matrix cannot fully represent the topological information of the traffic network.



Figure 7. Dynamic graph convolution network.

The HGM proposes a novel way of generating network topology information. Specifically, as shown in Figure 8, the proposed method randomly initializes two node embeddings $E_1, E_2 \in \mathbb{R}^{N \times D_d}$ with learnable parameters, which can be updated during the model training process and can be defined as:

$$A_{adv} = softmax(RELU(E_1E_2^T))$$

The proposed method multiplies E_1 and E_2 to calculate the adaptive adjacency matrix, removes the weak correlation of the adaptive adjacency matrix by the ReLU activation function, and normalizes the adjacency matrix through the softmax function. From a spatial-based perspective, the graph convolutional neural network updates the signal of the current node by aggregating and transforming the information of adjacent nodes. After obtaining the dynamic adjacency matrix, the graph convolutional neural network in the proposed method can be defined as:

$$h^l = A_{adn} h^{l-1} W$$

where A_{adp} represents the proposed dynamic adjacency matrix, h^{l-1} and h^{l} denote the input and output of the hidden layer, and W indicates trainable weights.



Figure 8. Dynamic adjacency matrix.

In the temporal dimension, instead of building a complex neural network, a 1D convolutional neural network is used to extract nonlinear temporal dependencies after aggregating spatial signals through a dynamic adjacency matrix. Specifically, the proposed method sets the size of the 1D convolution kernel to size = 3, and the sliding step size to s = 1 to extract temporal features, and then the size of the time dimension is converted into Q through linear projection operations to obtain the future road conditions.

As mentioned above, we detail the core principles of the STA module and the DGCN. As shown in Figure 3, in order to extract deep features, the proposed method adopts a residual network to stack *L* layers of STA module and DGCN. The extraction of static and dynamic information is achieved through different branches. In addition, the HGM dynamically calculates weights according to branch results to improve the predictive performance of the model.

5. Experiments

5.1. Settings

The proposed method evaluates the prediction performance of the HGM on two public traffic datasets, METR-LA and PEMS-BAY. The proposed method uses mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) to evaluate the performance of the HGM [1,2]. MAPE is the average value of the absolute value of the relative percentage error, which can be used to evaluate the prediction results. The calculation formula can be defined as follows:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y} - y}{y} \right|$$

where \hat{y} represents the predicte value, y is the real data, and n denotes the total number of samples. *MAE* is the average absolute error, and the calculation method can be expressed:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y} - y}{y} \right|$$

RMSE is used to measure the deviation between ground truth and predict value, which can be defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|\hat{y} - y|^2}$$

5.2. Parameter Settings

The HGM predicts the road conditions at Q future time steps based on the historical traffic conditions at P = 12 time steps. The proposed method uses the Adam optimizer to update network parameters. The HGM sets the learning rate, the number of iterations, and output dimension as lr = 0.001, epoch = 150, and D = 64. The proposed method stacks L = 3 layers of the STA module and DGCN to extract spatial–temporal information from different components, respectively. Furthermore, to avoid unnecessary overhead, the proposed method uses an early stopping strategy during the model training process. In other words, if the validation loss does not decrease within 15 epochs, we consider the proposed method to have converged. In addition, all experiments run on the computer with Nvidia_V100 GPU, 16 G RAM, and Intel I9700 CPU.

5.3. Compared Methods

In order to further verify the effectiveness of the proposed method, we compare the predicted results of the proposed method with the following compared methods:

- HA: HA uses the average of historical data to predict future road conditions.
- ARIMA [11]: ARIMA employs autoregressive and moving average methods for traffic prediction.
- SVR [44]: SVR uses support vector machine to extract spatial-temporal features of the traffic network.
- FC-LSTM [45]: FC-LSTM uses LSTM to analyze spatial-temporal dependencies.
- DCRNN [26]: DCRNN extracts spatial-temporal correlations with diffusion convolution and recurrent neural network.
- STGCN [31]: STGCN combines graph convolution and 1D convolution operations to extract spatial-temporal dependencies.
- GraphWaveNet [29]: GraphWaveNet proposes a hybrid model of graph convolution and dilated convolution to dynamically extract features.
- ASTGCN [30]: ASTGCN extracts spatial and temporal correlations through a convolutional neural network and attention mechanism.
- MTGNN [46]: MTGNN is a GNN-based and CNN-based model which employs adaptive graph, mix-hop propagation layers, and dilated inception layers to capture spatial-temporal correlations.
- GMAN [38]: GMAN captures spatial-temporal information using a spatial-temporal attention mechanism.
- MRA-DGCN [27]: MRA-DGCN captures complex dependencies among nodes using a dynamic adjacency matrix.

6. Experimental Results

6.1. Predictive Performance

Table 1 shows the performance comparison of the proposed method and compared methods on the METR-LA and PeMS datasets. On the PeMS-BAY and METR-LA datasets, the proposed method achieves the best prediction results at different time steps. In order to mine the potential correlation between nodes, the HGM uses a dynamic adjacency matrix to represent the spatial information of the traffic network and uses a convolutional neural network to extract nonlinear temporal dependencies, which is beneficial for achieving better predictive performance. For example, on the METR-LA dataset, the error of the proposed method drops to 6.60%, 8.19%, and 9.65% at 3, 6, and 12 future time steps on MAPE, respectively, which shows that the proposed method achieves the best predictive

Data	Mathad	15 min				30 min		60 min			
Data	Method	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	
	HA	2.88	5.59	6.80%	2.88	5.59	6.80%	2.88	5.59	6.80%	
	ARIMA	1.62	3.30	3.50%	2.33	4.76	5.40%	3.38	6.50	8.30%	
	SVR	1.85	3.59	3.82%	2.48	5.18	5.50%	3.28	7.08	8.00%	
	FC-LSTM	2.05	4.19	4.80%	2.20	4.55	5.20%	2.37	4.96	5.70%	
	DCRNN	1.38	2.95	2.90%	1.74	3.97	3.90%	2.07	4.74	4.90%	
DoMC	STGCN	1.36	2.96	2.90%	1.81	4.27	4.17%	2.49	5.69	5.79%	
rems	Graph WaveNet	1.30	2.74	2.73%	1.63	3.70	3.67%	1.95	4.52	4.63%	
	ASTGCN	1.53	3.13	3.22%	2.01	4.27	4.48%	2.61	5.42	6.00%	
	MTGCN	1.32	2.79	2.77%	1.65	3.74	3.69%	1.94	4.49	4.53%	
	GMAN	1.34	2.82	2.81%	1.62	3.72	3.63%	1.86	4.32	4.31%	
	MRA-DGCN	1.28	2.75	2.68%	1.59	3.62	3.60%	1.87	4.33	4.42%	
	HGM (Ours)	1.25	2.66	2.64%	1.54	3.57	3.47%	1.85	4.31	4.29%	
Data	Method	15 min				30 min		60 min			
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	
	HA	4.16	7.80	13.00%	4.16	7.80	13.00%	4.16	7.80	13.00%	
	ARIMA	3.99	8.21	9.60%	5.15	10.45	12.70%	6.90	13.23	17.40%	
	SVR	3.99	8.45	9.30%	5.05	10.87	12.10%	6.72	13.76	16.70%	
	FC-LSTM	3.44	6.30	9.60%	3.77	7.23	10.90%	4.37	8.69	13.20%	
	DCRNN	2.77	5.38	7.30%	3.15	6.45	8.80%	3.60	7.60	10.50%	
METD I A	STGCN	2.88	5.74	7.62%	3.47	7.24	9.57%	4.59	9.40	12.70%	
ME1K-LA	Graph WaveNet	2.69	5.15	6.90%	3.07	6.22	8.37%	3.53	7.37	10.01%	
	ASTGCN	4.86	9.27	9.21%	5.43	10.61	10.13%	6.51	12.52	11.64%	
	MTGCN	2.69	5.18	6.86%	3.05	6.17	8.19%	3.49	7.23	9.87%	
	GMAN	2.80	5.55	7.41%	3.12	6.49	8.73%	3.44	7.35	10.07%	
	MRA-DGCN	2.62	5.14	6.68%	3.01	6.13	8.05%	3.38	7.22	9.98%	
	HGM (Ours)	2.58	5.06	6.60%	3.04	6.10	8.19%	3.35	7.03	9.65%	

the following results can be drawn. **Table 1.** Performance comparison of different methods on METR-LA and PeMS datasets.

performance for both short-range and long-range. From the experimental results in Table 1,

First, since traditional time series prediction methods such as HA and ARIMA focus on historical records and ignore the spatial characteristics of the traffic network, these methods cannot compete with the proposed method. Compared with HA and ARIMA, SVR further considers the spatial correlation and achieves better performance. Due to the complex spatial–temporal correlation of road conditions and high-dimensional feature information, the performance of traditional time series prediction methods is limited. The proposed method adopts a deep learning-based approach to build models, which is beneficial for extracting complex spatial–temporal information.

Second, deep learning-based methods (such as FC-LSTM, STGCN, DCRNN, MTGCN, ASTGCN, etc.) alleviate the shortcomings of traditional methods. These methods utilize deep neural networks to automatically mine the complex spatial–temporal features of the traffic network for better predictive performance. However, LSTM considers temporal features while ignoring spatial features, which leads to poor predictive performance. Other methods represent spatial information through a static adjacency matrix, which cannot effectively model the global correlation. The proposed method devises a dynamic adjacency matrix to represent the topology of the transportation network. The HGM dynamically updates the correlation between nodes during the model training process, which is conducive to mining potential correlations, and has achieved the best performance in MAE, RMSE, and MAPE. For example, on the PeMS-BAY dataset, the proposed method achieves a reduction of 0.61% and 1.71% on MAPE over 12 future time steps compared to DCRNN and ASTGCN, respectively.

Finally, the proposed method constructs the topology information from different visions, and the STA module and DGCN are designed to perform information extraction

on different graph structures. Compared to Graph WaveNet and GMAN, the proposed method simultaneously considers shallow and potential relationships between nodes through a dynamic and static adjacency matrix. The importance of adjacent nodes can be highlighted through different network topology information and the hidden relationship between nodes can be captured in the proposed method. On the METR-LA dataset, the HGM reduces the RMSE metric by 0.36 and 0.42 over 12 future time steps compared to Graph WaveNet and GMAN, respectively.

6.2. Ablation Studies

Some experiments are performed to analyze the impact of different modules in the proposed method on the predictive performance. In the ablation analysis, we mainly explore the impact of STA module and DGCN. In addition, in order to analyze the impact of the gated function of the information fusion of different components, we also performed related experiments.

6.2.1. Impact of STA Module

The proposed method uses the STA module which mainly includes spatial attention and temporal attention to extract complex spatial-temporal features. Some experiments are conducted to evaluate the impact of the STA module on the proposed method. We analyze the impact of different variants on the predictive performance based on the proposed method. The HGM-V1 uses 1D convolutions instead of temporal attention to extract temporal information. The HGM-V2 indicates that the spatial-temporal attention module consists of static graph convolution and 1D convolution operations. HGM-V3 performs graph convolution operations on static adjacency matrices. The HGM-V4 represents the extraction of spatial-temporal information of traffic networks in stages. In other words, the features extracted by spatial attention are used as the input information of temporal attention. The HGM-V5 indicates that the model only considers dynamically generated topology. Table 2 records the predictive performance of different methods on the two datasets. According to the experimental results, it can be concluded that the proposed method achieves the best prediction performance.

Table 2. Prediction results of different variants of STA module. The HGM-V1 stands for using 1D convolution instead of temporal attention. The HGM-V2 means that the spatial-temporal attention module is composed of static graph convolution and 1D convolution operation. The HGM-V3 performs a static graph convolution operation on static graph. The HGM-V4 means extracting spatial-temporal information in stages. The HGM-V5 means using dynamic graphs instead of static graphs.

Data	Method	MAE	15 min RMSE	MAPE	MAE	30 min RMSE	MAPE	MAE	60 min RMSE	MAPE
	HGM-V1	1.33	2.76	2.74%	1.66	3.71	3.68%	1.95	4.47	4.57%
	HGM-V2	1.50	3.15	3.10%	2.02	4.54	4.52%	2.66	5.93	6.59%
DoMC	HGM-V3	1.32	2.78	2.75%	1.70	3.95	3.80%	1.93	4.45	4.58%
Pems	HGM-V4	1.32	2.79	2.73%	1.64	3.71	3.64%	1.97	4.52	4.60%
	HGM-V5	1.30	2.74	2.67%	1.62	3.69	3.60%	1.90	4.41	4.48%
	HGM	1.25	2.66	2.64%	1.54	3.57	3.47%	1.85	4.31	4.29%
Data	Method		15 min			30 min			60 min	
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
	HGM-V1	2.66	5.11	6.86%	3.03	6.15	8.35%	3.45	7.20	10.14%
METR-LA	HGM-V2	3.19	6.09	8.41%	3.83	7.60	10.74%	4.69	9.20	14.55%
	HGM-V3	2.65	5.08	6.77%	3.04	6.15	8.42%	3.50	7.34	10.36%
	HGM-V4	2.70	5.25	6.97%	3.12	6.42	8.64%	3.55	7.47	10.21%
	HGM-V5	2.67	5.17	6.84%	3.10	6.30	8.57%	3.57	7.49	10.54%
	HGM	2.58	5.06	6.60%	3.04	6.10	8.19%	3.35	7.03	9.65%

Compared with the HGM-V1, the proposed method uses temporal attention to extract temporal information. On the PeMS-BAY and METR-LA datasets, MAPE decreases by 0.28% and 0.49% over the next 12 time steps, respectively. The HGM-V2 combines 1D convolution and static graph convolution operations to extract nonlinear temporal dependencies and complex spatial correlations. Compared with the HGM-V2, the proposed method shows a significant drop in both long-range and short-range predictive performance. On the PeMS-BAY and METR-LA datasets, the proposed method outperforms the HGM-V2 by 0.81 and 1.34 in MAE over the next 12 time steps. The HGM-V3 performs graph convolution operations in a static adjacency matrix to aggregate topological information of the traffic network, which cannot represent the potential topological information and has limited predictive performance. In the HGM-V4 model, we try to extract spatial-temporal information in stages. According to the experimental results in Table 2, the HGM-V4 did not achieve excellent prediction performance. On the METR-LA dataset, MAE, RMSE, and MAPE increase by 0.2, 0.44, and 0.56% over the next 12 time steps, respectively, compared to our proposed method. The HGM-V5 ignores the spatial information contained in the static adjacency matrix, but we all know that the nodes directly adjacent to the current node have a great influence on the current one. From the experimental results in Table 2, we can see that our proposed method achieves the best prediction performance. By combining dynamic and static adjacency matrices, the HGM model can not only strengthen the relationship between nodes but also mine the underlying topology information.

6.2.2. Impact of DGCN

In order to verify the importance of DGCN, some experiments were conducted to analyze the impact of different variants of DGCN on predictive performance. The HGM-V6 uses temporal attention to model temporal information in DGCN and the HGM-V7 removes the entire DGCN branch. The experimental results are recorded in Table 3. Compared with the HGM, the HGM-V6 uses temporal attention to extract information. Since the proposed method continuously updates the dynamic adjacency matrix during the model training process, the dynamic adjacency matrix cannot effectively represent the topological information of the network at the initial stage, so temporal attention does not perform well on the dynamic adjacency matrix. Furthermore, if the proposed method removes the DGCN branch, the predictive performance can also drop. On the METR-LA dataset, MAE, RMSE, and MAPE decreased by 0.22, 0.46, and 0.89% over the next 12 time steps, respectively.

Table 3. Prediction results of different variants of DGCN. The HGM-V6 uses temporal attention to model temporal information in DGCN and the HGM-V7 removes the DGCN module.

Data	Method	MAE	15 min RMSE	MAPE	MAE	30 min RMSE	MAPE	MAE	60 min RMSE	MAPE
PeMS	HGM-V6	1.30	2.73	2.68%	1.62	3.68	3.62%	1.93	4.49	4.60%
	HGM-V7	1.30	2.74	2.67%	1.62	3.69	3.60%	1.90	4.41	4.48%
	HGM	1.25	2.66	2.64%	1.54	3.57	3.47%	1.85	4.31	4.29%
Data	Method	MAE	15 min RMSE	MAPE	MAE	30 min RMSE	MAPE	MAE	60 min RMSE	MAPE
METR-LA	HGM-V6	2.68	5.18	6.87%	3.06	6.21	8.42%	3.53	7.44	10.43%
	HGM-V7	2.67	5.17	6.84%	3.10	6.30	8.57%	3.57	7.49	10.54%
	HGM	2.58	5.06	6.60%	3.04	6.10	8.19%	3.35	7.03	9.65%

6.2.3. Impact of Gated Function

In order to verify the effectiveness of the proposed method to dynamically fuse information from different branches, we conducted experiments on the PeMS-BAY and METR-LA datasets. Figure 9 records the experimental results under different weights. From the experimental results, it can be seen that assigning different weights to different branches affects the prediction performance, which verifies the effectiveness of the proposed gated function. It is not easy to obtain the optimal performance of the model by artificially specifying fixed weights. Therefore, in the HGM model, the method calculates the corresponding weights according to the branch results to realize the dynamic fusion of different components in the experiment.



Figure 9. Effect of different branch weights on prediction performance. λ represents the weight ratio. For example, $\lambda = 9$ indicates that the proposed method sets the weights of STA branch and DGCN branch to 0.9 and 0.1, respectively.

6.3. Time Cost

Some experiments are conducted to explore the training time and inference time of the proposed method and compared methods. The experiments run on the computer with Nvidia_V100 GPU, 16 G RAM, Intel I9700 CPU. Table 4 records the experimental results of different methods in the next 60 min on the METR-LA dataset. According to the experimental results in Table 4, traditional statistics-based methods and machine learning-based methods require less training time. Since FC-LSTM, DCRNN, STGCN, and Graph WaveNet use a single topology to predict road conditions, the training time is less than the proposed method. The proposed method uses a graph neural network, convolutional neural network, and attention mechanism to capture complex spatial-temporal information of the traffic network, and designs a learnable adjacency matrix to represent the potential topology of the network. Although the proposed method requires more training time than most of the compared methods, it achieves the best prediction performance due to the superiority of the proposed method.

Table 4. Time cost (s) of the proposed method and compared methods.

Dataset	Methods	HA	ARIMA	SVR	FC-LSTM	DCRNN	STGCN	Graph WaveNet	ASTGCN	MTGCN	GMAN	MRA-DGCN	HGM
METR-LA	Training time	0.00	64.21	84.37	463.49	4598.65	6021.55	5967.32	6218.35	7463.74	6458.56	8165.25	7489.54
	Inference	0.11	0.74	0.42	1.21	0.98	1.31	0.98	1.56	1.01	1.32	1.79	1.01

7. Conclusions

Traffic prediction is of great significance for travel arrangement and resource planning. In this paper, we propose a graph-based approach to extract complex spatial-temporal information about traffic networks, which achieves excellent predictive performance. The proposed method constructed diverse topology information of traffic network and used different branches to extract complex spatial-temporal dependencies. A dynamic adjacency matrix is used to represent the underlying topology of the traffic network to provide richer information. Furthermore, the gated function improves the fusion results from different branches. In the future, we will try to design a more lightweight network to capture the spatial-temporal information of the traffic network to further reduce the number of network parameters.

Author Contributions: Conceptualization, R.C. and H.Y.; methodology, R.C. and H.Y.; validation, R.C.; formal analysis, H.Y.; investigation, R.C.; resources, H.Y.; writing—original draft preparation, R.C.; writing—review and editing, R.C. and H.Y.; visualization, R.C.; supervision, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: In this paper we conduct extensive experiments using two open-source datasets. We can get datasets from https://gitcode.net/mirrors/liyaguang/DCRNN (accessed on 9 May 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Dai, G.; Ma, C.; Xu, X. Short-term traffic flow prediction method for urban road sections based on space–time analysis and GRU. *IEEE Access* **2019**, *7*, 143025–143035. [CrossRef]
- 2. Li, F.; Feng, J.; Yan, H.; Jin, G.; Yang, F.; Sun, F.; Jin, D.; Li, Y. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution. *ACM Trans. Knowl. Discov. Data* **2023**, *17*, 1–21.
- 3. Medina-Salgado, B.; Sanchez-DelaCruz, E.; Pozos-Parra, P.; Sierra, J. E. Urban traffic flow prediction techniques: A review. *Sustain. Comput. Inform. Syst.* 2022, 35, 100739. [CrossRef]
- 4. Ye, X.; Fang, S.; Sun, F.; Zhang, C.; Xiang, S. Meta graph transformer: A novel framework for spatial-temporal traffic prediction. *Neurocomputing* **2022**, *491*, 544–563. [CrossRef]
- Kashyap, A.A.; Raviraj, S.; Devarakonda, A.; Nayak K, S.R.; KV, S.; Bhat, S.J. Traffic flow prediction models—A review of deep learning techniques. *Cogent Eng.* 2022, 9, 2010510. [CrossRef]
- 6. Bui, K.H.N.; Cho, J.; Yi, H. Spatial-temporal graph neural network for traffic forecasting: An overview and open research issues. *Appl. Intell.* **2022**, *52*, 2763–2774.
- 7. Lohrasbinasab, I.; Shahraki, A.; Taherkordi, A.; Delia Jurcut, A. From statistical-to machine learning-based network traffic prediction. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e4394. [CrossRef]
- 8. Zhou, X.; Zhang, Y.; Li, Z.; Wang, X.; Zhao, J.; Zhang, Z. Large-scale cellular traffic prediction based on graph convolutional networks with transfer learning. *Neural Comput. Appl.* **2022**, *34*, 5549–5559. [CrossRef]
- 9. Jin, J.; Rong, D.; Zhang, T.; Ji, Q.; Guo, H.; Lv, Y.; Ma, X.; Wang, F. A GAN-based short-term link traffic prediction approach for urban road networks under a parallel learning framework. *IEEE Trans. Intell. Transp. Syst.* **2022**, 23, 16185–16196. [CrossRef]
- 10. Jiang, W. Internet traffic prediction with deep neural networks. Internet Technol. Lett. 2022, 5, e314. [CrossRef]
- 11. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
- Ghimire, S.; Deo, R.C.; Wang, H.; Al-Musaylh, M.S.; Casillas-Pérez, D.; Salcedo-Sanz, S. Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: A review and new modeling results. *Energies* 2022, 15, 1061.
- 13. Fan, W.; Ma, Y.; Li, Q.; He, Y.; Zhao, E.; Tang, J.; Yin, D. Graph neural networks for social recommendation. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 417–426.
- 14. Jiang, W.; Luo, J. Graph neural network for traffic forecasting: A survey. Expert Syst. Appl. 2022, 207, 117921.
- Zeng, Z.; Huang, Y.; Wu, T.; Deng, H.; Xu, J.; Zheng, B. Graph-based Weakly Supervised Framework for Semantic Relevance Learning in E-commerce. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, 17–21 October 2022; pp. 3634–3643.
- 16. Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph neural networks: A review of methods and applications. *AI Open* **2020**, *1*, 57–81.
- Pan, B.; Demiryurek, U.; Shahabi, C. Utilizing real-world transportation data for accurate traffic prediction. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 595–604.
- Boukerche, A.; Wang, J. Machine learning-based traffic prediction models for intelligent transportation systems. *Comput. Netw.* 2020, 181, 107530. [CrossRef]

- Zhang, W.; Zhu, F.; Lv, Y.; Tan, C.; Liu, W.; Zhang, X.; Wang, F.Y. AdapGL: An adaptive graph learning algorithm for traffic prediction based on spatiotemporal neural networks. *Transp. Res. Part C Emerg. Technol.* 2022, 139, 103659. [CrossRef]
- Zhou, J.; Han, T.; Xiao, F.; Gui, G.; Adebisi, B.; Gacanin, H.; Sari, H. Multiscale network traffic prediction method based on deep echo-state network for internet of things. *IEEE Internet Things J.* 2022, *9*, 21862–21874. [CrossRef]
- Yao, H.; Tang, X.; Wei, H.; Zheng, G.; Li, Z. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 5668–5675.
- Yu, H.; Wu, Z.; Wang, S.; Wang, Y.; Ma, X. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors* 2017, 17, 1501. [CrossRef]
- 23. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3848–3858.
- Zhang, X.; Huang, C.; Xu, Y.; Xia, L.; Dai, P.; Bo, L.; Zhang, J.; Zheng, Y. Traffic flow forecasting with spatial-temporal graph diffusion network. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 15008–15015.
- Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Multi-stage attention spatial-temporal graph networks for traffic prediction. *Neurocomputing* 2021, 428, 42–53. [CrossRef]
- 26. Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv* 2017, arXiv:1707.01926.
- Yao, H.; Chen, R.; Xie, Z.; Yang, J.; Hu, M.; Guo, J. MRA-DGCN: Multi-Range Attention-Based Dynamic Graph Convolutional Network for Traffic Prediction. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1613–1621.
- Lv, M.; Hong, Z.; Chen, L.; Chen, T.; Zhu, T.; Ji, S. Temporal multi-graph convolutional network for traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* 2020, 22, 3337–3348. [CrossRef]
- 29. Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Zhang, C. Graph wavenet for deep spatial-temporal graph modeling. *arXiv* 2019, arXiv:1906.00121.
- Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 922–929.
- Yu, B.; Yin, H.; Zhu, Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv* 2017, arXiv:1709.04875.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 2017, 30. [CrossRef]
- Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* 2018, 13, 55–75. [CrossRef]
- 34. Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.; Le, Q.V.; Salakhutdinov, R. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv* 2019, arXiv:1901.02860.
- Hossain MD, Z.; Sohel, F.; Shiratuddin, M.F.; Laga, H. A comprehensive survey of deep learning for image captioning. ACM Comput. Surv. (CsUR) 2019, 51, 1–36. [CrossRef]
- Deng, L.; Hinton, G.; Kingsbury, B. New types of deep neural network learning for speech recognition and related applications: An overview. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 8599–8603.
- Chen, W.; Chen, L.; Xie, Y.; Cao, W.; Gao, Y.; Feng, X. Multi-range attentive bicomponent graph convolutional network for traffic forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 3529–3536.
- Zheng, C.; Fan, X.; Wang, C.; Qi, J. Gman: A graph multi-attention network for traffic prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 1234–1241.
- Shuman, D.I.; Narang, S.K.; Frossard, P.; Ortega, A.; Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* 2013, 30, 83–98. [CrossRef]
- 40. Micheli, A. Neural network for graphs: A contextual constructive approach. IEEE Trans. Neural Netw. 2009, 20, 498–511. [CrossRef]
- 41. Oloulade, B.M.; Gao, J.; Chen, J.; Lyu, T.; Al-Sabri, R. Graph neural architecture search: A survey. *Tsinghua Sci. Technol.* **2021**, 27, 692–708. [CrossRef]
- 42. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. arXiv 2016, arXiv:1609.02907.
- Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* 2017, arXiv:1710.10903.
 Wu, C.-H.; Ho, J.-M.; Lee, D.T. Travel-time prediction with support vector regression. *IEEE Trans. Intell. Transp. Syst.* 2004, 5, 276–281. [CrossRef]

- 45. Graves, A.; Graves, A. Long short-term memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45.
- Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; Zhang, C. Connecting the dots: Multivariate time series forecasting with graph neural networks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual, 6–10 July 2020; pp. 753–763.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.