Proceedings of the 3rd International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2025.DL27990

Graph Neural Networks for Urban Traffic Flow Forecasting: A Comprehensive Review and Future Perspectives

Yuchen Zhu

College of Liberal Arts, University of Minnesota, Minneapolis, USA
zhu01042@umn.edu

Abstract. Traffic flow prediction is an essential part of intelligent transportation systems (ITS), facilitating dynamic traffic control, congestion alleviation, and route planning. The past few years have seen the booming of Graph Neural Networks (GNNs), providing a strong tool to capture the sophisticated spatial compositions and dynamic temporal patterns inherent in urban road networks. In this paper, I provide a systematic and comprehensive review of GNN-based solutions for short-term traffic prediction. I first briefly review the basic concepts and categorization of GNNs, then elaborate on representative models such as spatial and spectral convolutional networks, temporal graph structures, and hybrid models with attention mechanisms and recurrent units. Besides, I recap typical traffic datasets and evaluation metrics, compare the performance of different models from multiple aspects, and highlight key technical challenges such as spatiotemporal heterogeneity, scalability, and interpretability. Finally, I suggest future research lines to improve the accuracy, efficiency, and robustness of GNN-based traffic prediction models for real-world ITS applications.

Keywords: Traffic flow forecasting, Graph neural networks, Spatiotemporal modeling, Deep learning, Intelligent transportation systems

1. Introduction

With the accelerating pace of urbanization and motorization, cities worldwide are facing increasing traffic congestion, environmental pollution, and pressure on transportation infrastructure. Effective traffic flow prediction has thus become a vital component of intelligent transportation systems (ITS), enabling dynamic traffic signal control, congestion mitigation, and real-time route optimization. In particular, short-term traffic forecasting—predicting traffic conditions in the next 5 to 30 minutes—has received growing attention due to its practicality in real-time urban management.

Previous traffic forecasting methods, ranging from statistical approaches like ARIMA and Kalman filters, to conventional machine learning approaches like support vector regression (SVR) and random forests, have achieved some success in terms of modeling the trends of past traffic. However, these models typically assume linearity or stationarity and cannot grasp the nonlinear, dynamic, and spatially dependent characteristics of current urban traffic networks. Furthermore, they are not scalable when applied to large-scale, high-dimensional traffic data.

The advent of deep learning has brought unprecedented improvement in traffic prediction. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have been

DOI: 10.54254/2753-8818/2025.DL27990

shown to have strong capability in modeling temporal dependencies in sequential traffic data. Similarly, convolutional neural networks (CNNs) have also been applied to model spatial patterns by considering the traffic maps as images. While these models are good at modeling spatial or temporal features independently, they inherently make Euclidean assumptions and are not able to well model non-grid, irregular traffic networks. This spatial inflexibility limits their application in real-world urban settings.

To bypass these constraints, researchers have resorted to Graph Neural Networks (GNNs), a family of deep learning models natively tailored for graph-structured data. In traffic forecasting, road networks can be naturally cast as graphs, with nodes representing road segments or sensors and edges defining physical or functional relationships. GNNs are able to model the intricate spatial dependencies between road elements in a flexible manner and incorporate temporal dynamics via recurrent units, temporal convolutions, or attention mechanisms. Consequently, GNN-based models have achieved state-of-the-art performance on short-term traffic forecasting tasks.

This article offers an in-depth survey of GNN-based approaches to short-term traffic flow prediction. We start with an introduction to the basic concepts and taxonomies of GNNs. Then we delve into representative model structures, such as static spatial GCNs, dynamic graph-based approaches, and hybrid models with temporal modeling modules. We also recap typical datasets and evaluation metrics, compare the performance of existing models in terms of accuracy and efficiency, and summarize key technical challenges including spatiotemporal heterogeneity, model scalability, and interpretability. Finally, we suggest future research directions and indicate how GNN-based traffic prediction applies to real-world scenarios like urban areas in Shandong Province, China.

2. Foundations and taxonomies of GNNs

2.1. Page setup graph structure and traffic representation

Urban traffic networks have complicated topologies with strong spatial correlations, which are difficult to model with conventional time-series or grid-based approaches. Therefore, a number of recent works have moved in the direction of formulating traffic systems as graphs, where nodes represent road intersections, sensors, or spatial areas, and edges signify the physical or functional connections among them.

This graph-based representation offers a natural means of encoding road networks' non-Euclidean structure. For instance, in the popular DCRNN model, Li et al. create a directed graph with loop detectors as nodes and road connectivity as edges with temporal diffusion processes [1]. Likewise, Yu et al. suggest a graph convolutional network based on spatial adjacency matrices built upon physical distances between sensors [2]. These works demonstrate that the choice of adjacency definition—be it distance, connectivity, or correlation-based—significantly impacts the performance of the model.

Moreover, node attributes in traffic graphs commonly include features such as average speed, traffic volume, and occupancy rate, while edge features may carry information about travel distance, congestion level, or turn restrictions. As noted by Zhang et al.,such multimodal input allows richer representations of local traffic conditions and their mutual influence [3]. To further enhance adaptability, Bai et al. introduce an adaptive graph learning module that learns dynamic adjacency matrices during training, enabling models to discover latent spatial dependencies [4].

An illustrative example of traffic network graph modeling is shown in Figure 1, where nodes and edges are annotated with representative features such as vehicle count, average speed, and road segment length.

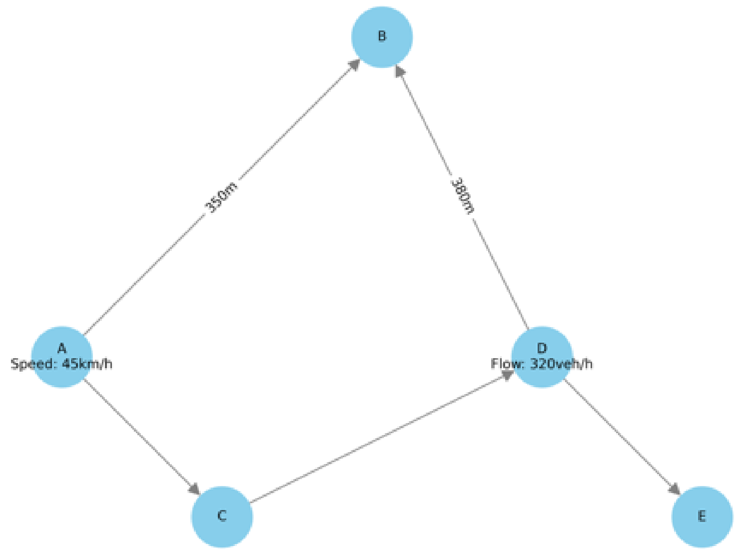

Figure 1. Traffic network graph modeling

Several studies have also explored hybrid or hierarchical graph representations. For instance, Song et al. propose multi-scale graphs to capture regional and sub-regional dependencies, and Wu et al. employ spatio-temporal bipartite graphs to represent interactions across both space and time [5] [6]. These variations reflect a growing consensus in the field: that flexible, semantically meaningful graph construction is a prerequisite for effective GNN-based traffic forecasting.

In summary, graph structures offer a powerful abstraction for encoding spatial relationships in traffic systems. Their effectiveness hinges on the choice of topology, the selection of node/edge features, and the ability to adapt to changing traffic dynamics over time.

2.2. Basic GNN concepts and operations

Graph Neural Networks (GNNs) provide a natural framework for learning on graph-structured data, which is particularly valuable in traffic flow prediction due to the irregular nature of road networks. Unlike standard neural networks, GNNs are designed to operate on nodes and their neighborhoods, enabling them to capture both local interactions and broader topological patterns.

The core idea behind most GNNs lies in message passing. At each layer, nodes receive and aggregate information from their neighbors to update their own state. This process is typically formulated as:

$$h_v^{(l)} = \sigma\left(W^{(l)} \bullet AGG\left(\left\{h_u^{(l-1)} : u \epsilon N(v) \cup \{v\}\right\}\right)\right) \tag{1}$$

Here, $h_v^{(l)}$ is the hidden representation of node v at layer l, $N(v)$ denotes its neighbors, AGG is an aggregation function such as mean or sum, $W^{(l)}$ is a learnable weight matrix, and σ is a non-linear activation function (e.g., ReLU).

Over the past few years, several variants of GNNs have emerged, each proposing different ways of aggregating neighbor information:

1.GCN (Graph Convolutional Network), introduced by Kipf and Welling, defines a normalized graph convolution operation in the spectral domain. It assumes homophily—that connected nodes

DOI: 10.54254/2753-8818/2025.DL27990

are likely to have similar features [7].

2.GraphSAGE was proposed to improve scalability. It samples a fixed number of neighbors and uses functions like mean, LSTM, or pooling to aggregate information, supporting inductive learning [8].

3.GAT (Graph Attention Network)takes a different approach by learning attention coefficients, allowing the model to weigh each neighbor differently during aggregation. This is particularly useful in graphs with heterogeneous connectivity [9].

In traffic forecasting, these architectures are often adapted into spatio-temporal models. A typical design pattern involves stacking a GNN layer (to extract spatial dependencies) with a temporal component such as GRU, LSTM, or 1D convolution (to model temporal dynamics). For instance, the STGCN model applies temporal convolutions before and after GCN blocks, while AGCRN learns both the adjacency matrix and temporal dependencies jointly [10].

One interesting aspect of recent work is the move toward adaptive and dynamic graphs. Rather than relying solely on fixed road topology, some models learn graph structure as part of the training process. This allows the network to uncover hidden patterns such as functional relationships between roads that are not physically adjacent but experience correlated traffic conditions.

An overview of message passing in GNNs is illustrated in Figure 2. The diagram shows how a central node collects information from its neighbors and updates its representation through aggregation and transformation.

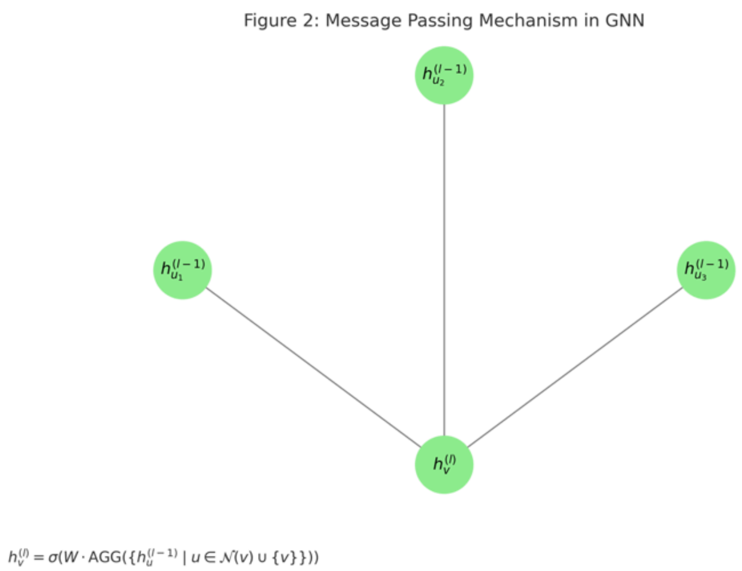

Figure 2. Message passing in GNNs

2.3. Spatial-Temporal Graph Neural Networks (STGNNs)

Even though traditional Graph Neural Networks (GNNs) solely focus on learning static spatial relationships, traffic data necessarily have both spatial and temporal correlations. Spatial-Temporal Graph Neural Networks (STGNNs) are specially designed in an attempt to model these two types of dependencies and therefore are particularly well-adopted for traffic flow prediction tasks.

In STGNNs, the spatial domain is frequently represented as a graph where nodes are sensors, intersections, or road sections and edges are the connectivity or proximity. The temporal domain, on the other hand, represents sequential patterns such as traffic speed variation with time. By

DOI: 10.54254/2753-8818/2025.DL27990

integration of both, STGNNs become powerful enough to model the dynamics of traffic systems better.

One of the universal design schemes for STGNNs is to employ spatial modules (such as Graph Convolutional Networks or Graph Attention Networks) to capture structure information, and temporal modules (such as Recurrent Neural Networks, Temporal Convolutional Networks, or self-attention mechanism) to capture sequential relations [1][6]. Some models adopt a spatio-temporal block structure, where a block processes spatial and temporal signals in parallel, enabling deeper interaction between the two directions.

For example, the DCRNN (Diffusion Convolutional Recurrent Neural Network) introduced by Li et al. employs diffusion convolution to represent directed spatial relationships, while the temporal relationships are represented using Gated Recurrent Units (GRUs) [1]. Similarly, STGCN (Spatio-Temporal Graph Convolutional Network) employs gated temporal convolutions in conjunction with graph convolutions to encode dependencies in a computationally efficient way [5]. Subsequent architectures, such as ASTGCN and GMAN, use attention mechanisms to dynamically modulate the influence of the neighbor and the prior time step, improving adaptability in highly dynamic traffic scenarios [6][11].

Mathematically, a typical STGNN layer can be formulated as:

$$H^{(l+1)} = \sigma\left(Temp\left(Spatial\left(H^{(l)}, A\right)\right)\right) \tag{2}$$

where $H^{(l)}$ is the input feature matrix at layer l, A is the adjacency matrix, $Spatial(\cdot)$ denotes a graph-based operation, $Temp(\cdot)$ represents a temporal modeling function, and $\sigma(\cdot)$ is a non-linear activation function.

STGNNs' architecture addresses some of the most important traffic flow forecast challenges, including:

1. Dynamic spatial dependencies – Traffic interactions change over time due to congestion, accidents, and events.

2. Multi-scale temporal patterns – Short-run fluctuations and long-term trends must both be captured.

3. Computational efficiency – Real-time prediction models need to be balanced between accuracy and latency.

In brief, STGNNs are a considerable advancement of GNNs for traffic applications, with one architecture that is capable of modeling spatial and temporal dependencies within a single framework.In brief, STGNNs are a considerable advancement of GNNs for traffic applications, with one architecture that is capable of modeling spatial and temporal dependencies within a single framework.

3. Datasets review

3.1. Well-known traffic flow datasets

Publicly available datasets are widely used in traffic flow prediction for model training and performance evaluation. Datasets are distributed over different geographical regions and traffic flow directions, with a variety of temporal granularities and features to support both short-term prediction and long-term trend analysis.

(1) METR-LA

DOI: 10.54254/2753-8818/2025.DL27990

METR-LA dataset consists of traffic speed measurements from 207 loop detectors along Los Angeles highways, USA, during the period from March 2012 to June 2012, with a 5-minute sampling interval [1].

(2) PEMS-BAY

PEMS-BAY dataset is derived from California's Performance Measurement System (PeMS) with data of 325 Bay Area sensor stations, recorded from January 2017 to May 2017, with a 5-minute sample interval [1].

(3) PEMS-D

The PEMSD4, PEMSD7, and PEMSD8 datasets are different highway networks in California with 170 to 307 nodes, a 5-minute sampling interval, and vast amount of data, for which they are appropriate in the assessment of generalization performance of models [12].

(4) BJ-Taxi

The BJ-Taxi dataset is the GPS traces of taxis in Beijing with one month in 2013 as the shared sample period and a 1-minute sampling rate, with contents covering latitude, longitude, speed, and heading data [3].

(5) NYC-Bike

The NYC-Bike data is accessed from New York City's Citi Bike system, with time-stamped trip records at the station level, as well as station coordinates, most often used to study how shared mobility coalesces with metropolitan traffic flow [13].

3.2. Comparison of dataset characteristics

To better understand the characteristics of different datasets, Table 1 compares the above datasets in terms of node count, time span, sampling interval, and data type.

Table 1. Data statistics

Dataset	Nodes	Time Span	Interval	Data Type
METR-LA	207	2012-03 ~ 2012-06	5 min	Speed
PEMS-BAY	325	2017-01 ~ 2017-05	5 min	Speed
PEMSD4	307	Several months (2018)	5 min	Speed
PEMSD7	228	Several months (2018)	5 min	Speed
PEMSD8	170	Several months (2016)	5 min	Speed
BJ-Taxi	34,000+	1 month (2013)	1 min	GPS, Speed
NYC-Bike	~800 stations	Several years	Trip-based	Trips

3.3. Applicability and limitation of datasets

The applicability of a dataset for traffic forecasting activity depends heavily on its collection mode, area covered, and time span.

Second, METR-LA and PEMS-BAY have good spatial and temporal resolution and are well-suited for short-term traffic speed prediction applications (e.g., 15–60 min forecast horizons).The PEMSD series, with mixed coverage and node counts, is better for generalizability performance evaluation across networks of different sizes [5].

By contrast, BJ-Taxi and NYC-Bike have more diverse attributes, with weather, holiday, and social occurrences affecting their traffic patterns, thus being more appropriate for travel-demand-

DOI: 10.54254/2753-8818/2025.DL27990

based prediction [3]. Nevertheless, their temporal sampling tends to be periodic, and nodes are defined as vehicles or stations, not fixed sensors, which results in less stable spatial topologies [13].

Moreover, most public datasets suffer from missing values due to sensor faults or communication latencies [14]. Data imputation might need to be done by using interpolation, matrix factorization, or deep generative models to avoid performance degradation.

3.4. Case study: shandong province traffic data

Shandong Province has recently invested heavily in intelligent transportation systems, deploying extensive fixed sensor networks, video detection, and car tracking based on GPS in cities such as Jinan and Qingdao. Such data is employed in real-time traffic control and offers high-quality spatio-temporal data for GNN/STGNN models.

For example, Jinan's main road network encompasses over 200 sensor nodes with sampling frequencies as low as one minute for highways, arterial roads, and part of secondary roads. In addition, taxi and ride-hailing platform traffic data integration in Qingdao enables high-frequency observation of city-level travel behavior and robust data support for models to identify flash congestion or holiday traffic surges.

It should be pointed out that compared with worldwide used overseas datasets, Shandong traffic data also have room for improvement in open sharing and standardized format, which would have possible impacts on model reproducibility and cross-regional transferability [15]. Openning some of the data under privacy protection is a direction for future studies to push forward cross-regional generalization studies.

4. Methodologies

4.1. Traditional traffic forecasting methods

Before deep learning, traffic flow forecasting relied heavily on statistical and classical machine learning methods that assumed stationarity and periodicity presumptively in traffic flow sequences.
(1)Time Series Models

Conventional time series approaches consist of the Autoregressive Integrated Moving Average (ARIMA) and its seasonal counterpart (SARIMA). While ARIMA can grasp linear dependence between past flow and present states, it is poor at representing nonlinear trends or complex spatio-temporal correlations [16].

(2)Regression and feature engineering based approach

Regression methods, such as Multiple Linear Regression (MLR) and Support Vector Regression (SVR), are prone to incorporate external features such as weather, holidays, and road types [17]. They are, however, reliant on extensive manual feature engineering, making it hard to automatically discover complex patterns.

(3)Kalman filter based approach

Kalman Filter (KF) and its extensions, i.e., Extended KF (EKF) and Unscented KF (UKF), are well-liked for real-time traffic state estimation, especially for noisy sensor data [18]. Their performance is, however, limited for multi-node, long-horizon prediction scenarios, with relatively high computational costs.

DOI: 10.54254/2753-8818/2025.DL27990

4.2. Deep learning-based traffic forecasting methods

When deep learning became popular, researchers began employing neural networks to learn nonlinear spatio-temporal patterns among traffic flows. Unlike traditional methods, deep models can learn features by themselves, perceive subtle patterns, and generalize well to large data.

(1)Recurrent Neural Networks (RNN) and extensions

Recurrent Neural Networks (RNNs) have the ability to model long-term relationships in time series data but suffer from gradient vanishing in long-horizon forecasting. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) reduce this issue by adding gating mechanisms [19]. LSTM/GRU models for traffic forecasting capture periodic trends and sudden changes in traffic flow, and are typically applied for single-node or small networks [20].

(2)CNN vs. TCN

Convolutional Neural Networks (CNNs) employ their local receptive fields to learn spatial and temporal features by performing 1D or 2D convolutions [3]. Temporal Convolutional Networks (TCNs) employ causal and dilated convolutions to model long-term dependencies and enjoy strengths in parallelization and training efficiency over RNN-based models.

(3)Attention mechanism-based methods

Attention mechanisms dynamically provide weights of importance to different time steps or nodes, enhancing the model's ability to identify critical spatio-temporal patterns. In traffic forecasting, Transformer models have been used to replace recurrent architecture entirely, enabling better modeling of long-term dependency [20].

4.3. Deep learning-based traffic forecasting methods graph neural network-based methods

Graph Neural Networks (GNNs) have the ability to learn spatial relationships between nodes with graph-structured data [7]. The nodes in the graph can be intersections or sensors, and road links are connected by edges.

(1)Spectral-based GCNs

Spectral methods require convolution operations via eigen-decomposition of the graph Laplacian matrix (e.g., ChebNet, GCN), offering a mathematically rigorous spatial filtering process but taking significant computational effort on large dynamic graphs [8].

(2)Spatial-based GCNs

Spatial methods aggregate near node attributes directly (e.g., GraphSAGE, GAT), and improve more to traffic flow dynamics in graph models. Graph Attention Networks (GATs) assign different weights to neighbors for more accurate predictions [2].

4.4. Spatio-temporal graph neural network models

Spatio-Temporal Graph Neural Networks (STGNNs) integrate GNNs with temporal units for modeling (RNNs, TCNs, Attention) to jointly learn spatial and temporal relations [10].

•DCRNN(Diffusion Convolutional Recurrent Neural Network):DCRNN replaces traditional GCN convolution with diffusion convolution and involves GRUs to represent time, showing superior performance on METR-LA and PEMS-BAY datasets.

•STGCN(Spatio-Temporal Graph Convolutional Network):STGCN alternatively stacks graph convolution layers and temporal convolution (TCN) to improve parallel computation efficiency [4].

•ASTGCN(Attention-based STGCN):ASTGCN introduces spatial and temporal attention mechanisms into STGCN and dynamically adjusts the weights of different nodes and time steps.

DOI: 10.54254/2753-8818/2025.DL27990

Recently, adaptive graph learning STGNNs such as AGCRN (Adaptive Graph Convolutional Recurrent Network) have emerged, which learn non-stationary traffic patterns by dynamically adapting the adjacency matrix during training.

5. Challenges and future research directions

5.1. Data sparsity and quality

Good quality continuous and complete traffic data are necessary to build good prediction models. In real traffic sensing systems, however, traffic data is often contaminated by sensor failures, communication delay, and environmental noise and hence contains missing or incorrect values.[13] Moreover, in sparse traffic areas or recently opened roads, data sparsity is particularly crucial, which significantly influences model generalization [11].

To counteract this, the future research could explore data imputation and generative models (e.g., Variational Autoencoders, VAEs; Generative Adversarial Networks, GANs), and cross-region transfer learning techniques to transfer learning from high-data regions to data-poverty regions [6].

5.2. Dynamic topology and non-stationarity

The road network's topology is not static, and traffic dynamics are dynamically altered because of construction, accidents, weather, or large events, The majority of STGNN models present now, however, have recourse to a static adjacency matrix and are not very reflective of the effects of changes in topology [10].

The subsequent research can be focused on dynamic graph modeling and adaptive graph learning, whereby the adjacency matrix is dynamically adjusted in real time while predicting to fit non-stationary traffic situations.

5.3. Model credibility and interpretability

As more advanced deep traffic forecasting models are being developed, model interpretability has become a critical constraint in applying models to real-world environments. Transport authorities often require model predictions to be explained clearly to promote belief in the output [19].

Including Explainable AI (XAI) techniques, such as attention weight visualisation, causal inference, and feature contribution analysis (e.g., SHAP, LIME),can provide additional transparency and verifiability to the decision-making process.

5.4. Model credibility and interpretability5.4 computational efficiency and scalability

Traffic forecasting tasks typically face massive-scale networks and high-frequency sampled data, leading to costly computation at training time and inference time [2].Real-time prediction is vital in Intelligent Transportation Systems (ITS), requiring stringent demands on computational efficiency [3].

Subsequent work can be done on model compression (i.e., pruning, quantization),distributed computation, and edge deployment of computing to reduce latency, save resources, and be flexible in resource-constrained environments.

DOI: 10.54254/2753-8818/2025.DL27990

5.5. Multi-source heterogeneous data fusion

The inclusion of weather, social media, event reports, and shared mobility data outside traditional sensor data such as speed, volume, and density can improve the quality of prediction.Spatial resolution, temporal alignment, and formats of data between sources, however, pose obstacles to direct fusion.

Future work can investigate unified graph-based multimodal learning methods, projecting different data sources into the same representation space to enhance the general predictive performance of STGNN models.

6. Conclusion

In this review, the evolution of Graph Neural Networks (GNNs) and their use for short-term urban traffic flow prediction has been comprehensively explored. Through the analysis of some leading architectures such as GCN, GAT, DCRNN, STGCN, and their complemented extension in terms of attention or adaptivity, the study demonstrates how GNNs are able to effectively model the inherent spatiotemporal correlations that define actual traffic networks. Compared to standard statistical approaches and earlier deep learning approaches, GNN-based approaches achieve significant advances in predictive power, scalability, and ability to learn from irregular, non-Euclidean road topologies.

Aside from model advancements, this survey also focuses heavily on data availability and quality. Benchmark datasets used by most researchers such as METR-LA, PEMS-BAY, and BJ-Taxi have facilitated developments but are incomplete in handling missing values, skewed features, and local bias. The case study of the Shandong Province's ITS programs illustrates real-world practicable challenges and opportunities of utilizing GNNs in novel applications.

For the future, some of these areas of research must be examined more intensely. They are connecting sparsity in data and transfer learning with generative models, enhancing flexibility with dynamic graph learning, improving explainability with explainable AI approaches, and achieving real-time scalability using model compression and edge deployment. Moreover, integrating multi-source heterogeneous data, i.e., weather, mobility services, and social context, may empower more robust and contextualized forecasting models.

In short, GNN-based traffic forecasting is where innovation in machine learning converges with urban mobility management. Despite the challenge, ongoing innovations are gradually nudging the practice towards increasingly accurate, efficient, and transparent machine learning models. The advances will not just make intelligent transport systems (ITS) more efficient but also help to ease congestion, reduce emissions, and facilitate more sustainable urban planning.

References

[1] Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. International Conference on Learning Representations (ICLR).

[2] Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. IJCAI, 3634–3640.

[3] Zhang, J., Zheng, Y., & Qi, D. (2020). Deep spatio-temporal residual networks for citywide crowd flows prediction. IEEE TKDE, 32(1), 98–111.

[4] Bai, L., Yao, L., Li, C., Wang, X., & Wang, C. (2020). Adaptive graph convolutional recurrent network for traffic forecasting. NeurIPS, 33, 17804–17815.

[5] Song, C., Lin, Y., Guo, S., & Wan, H. (2020). Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. AAAI Conference on Artificial Intelligence, 34(1), 914–

DOI: 10.54254/2753-8818/2025.DL27990

921.
[6] Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2020). Connecting the dots: Multivariate time series forecasting with graph neural networks. Proceedings of the 26th ACM SIGKDD, 753–763.
[7] Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. ICLR.
[8] Hamilton, W. L., Ying, Z., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. NeurIPS.
[9] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph Attention Networks. ICLR.
[10] Guo, K., Hu, Y., Qiu, X., Wang, Y., & Gao, J. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. AAAI Conference on Artificial Intelligence, 33(1), 922–929.
[11] Zheng, C., Fan, X., Wang, C., & Qi, J. (2020). GMAN: A graph multi-attention network for traffic prediction. AAAI Conference on Artificial Intelligence, 34(1), 1234–1241.
[12] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
[13] Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., … & Li, Z. (2018). Deep multi-view spatial-temporal network for taxi demand prediction. AAAI Conference on Artificial Intelligence, 32(1).
[14] Chen, L., Li, Y., Yu, R., & Shahabi, C. (2021). Multi-range attentive bicomponent graph convolutional network for traffic forecasting. AAAI Conference on Artificial Intelligence, 35(1), 104–112.
[15] Sun, L., & Axhausen, K. W. (2016). Understanding urban mobility patterns with a probabilistic tensor factorization framework. Transportation Research Part B: Methodological, 91, 511–524.
[16] Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. Journal of Transportation Engineering, 129(6), 664–672.
[17] Wu, C. H., Ho, J. M., & Lee, D. T. (2004). Travel-time prediction with support vector regression. IEEE Transactions on Intelligent Transportation Systems, 5(4), 276–281.
[18] Okutani, I., & Stephanedes, Y. J. (1984). Dynamic prediction of traffic volume through Kalman filtering theory. Transportation Research Part B: Methodological, 18(1), 1–11.
[19] Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies, 54, 187–197.
[20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 30.