Journal of Information Systems Engineering and Management
2025, 10(10s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Exploring the Efficacy of Graph Neural Networks in Predicting Traffic Flow in India

Paras Patil[1], Vishesh Mittal[2], Dhruvisha Mondhe[3], Akshita Upadhyay[4], Dr. Nupur Giri[5]
[1]Dept. of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai, India 2019paras.patil@ves.ac.in
[2]Dept. of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai, India 2019vishesh.mittal@ves.ac.in
[3]Dept. of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai, India 2019dhruvisha.mondhe@ves.ac.in
[4]Dept. of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai, India 2019akshita.upadhyay@ves.ac.in
[5]Dept. of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai, India nupur.giri@ves.ac.in

ARTICLE INFO

Received: 02 Nov 2024

Revised: 26 Dec 2024

Accepted: 04 Jan 2025

ABSTRACT

Accurate and real-time traffic forecasting is essen- tial for managing vehicle movement, reducing congestion, and generating optimal routes. However, existing traffic prediction systems rely on datasets specific to certain regions, making them unreliable for Indian traffic and road networks due to the lack of suitable training data. To address this issue, in this paper, we propose a Dataset for Indian Vehicular Traffic Analysis (DIVA) capturing the traffic features of the Indian Province. By training Graph Neural Networks on this dataset, we achieved comparable results to the state-of-the-art datasets (PEMSD7, PEMSD8), demonstrating the efficacy of the DIVA dataset in accurate flow prediction. Using this dataset, appreciable accuracy was achieved for traffic speed prediction using GNNs with STGCN achieving the lowest RMSE values. Moreover, we propose a novel route navigation approach that utilizes predicted future traffic speeds at nodes, represented as a future graph thus, provide a path that maximizes the traffic speed hence, minimizing the travel time.

Keywords: Traffic flow prediction, spatial-temporal graph data, graph neural networks, intelligent transportation systems

I. INTRODUCTION

The rapid growth of urban populations due to urbanization has led to increased pressure on urban traffic management. It is estimated that the cost of delay in urban traffic is approximately $6.6 billion per year, while the additional fuel consumption due to delays costs about $14.7 billion per year [1]. To address this issue, Intelligent Transportation Systems (ITS) have become an essential component of smart cities, with traffic prediction being a crucial element of the ITS. Traffic prediction involves the forecasting of traffic flow parameters, such as speed, density, and volume, to manage vehicle movement, reduce congestion, and generate optimal routes. Thus, accurate and real-time traffic forecasting plays an important role in the intelligent traffic system and is of great significance for urban traffic planning, traffic manage- ment, and traffic control.

Currently, a number of advanced systems for predicting traffic data such as road volume, average speed, travel times, etc, have been developed based on well-known datasets such as PEMSD7, PEMSD8, and TaxiBJ [2]. However, these systems are based on datasets collected from specific regions, making them unreliable when applied to Indian traffic and road networks due to the lack of available training data. While existing traffic flow prediction systems exist, the effectiveness of the models in specific regions depends on the training data. There exist systems that make use of machine learning, genetic and soft computing, and deep learning algorithms along with Image Processing to recognize traffic flow [3], or introduce line graph transformation into the construction of road traffic topology along with using LSTM for temporal feature extraction [4]. Other systems work on creating a hybrid traffic flow prediction system based on LSTM and Sparse Auto Encoder (SAE) [5]. By training the model on spatiotemporal data collected for the Indian province, it is possible to achieve the required foresight and enable early action, such as altering routes by users with minimal intervention from officials.

In this paper, we aim to tackle the above issues by collecting traffic data for roads and highways in Mumbai - Navi Mumbai Area in India and creating a Dataset for Indian Vehicular Traffic Analysis (DIVA). We found that this dataset contains sufficient spatiotemporal information to capture the traffic settings in India. By training Graph Neural Networks on this data, we observed that the results are comparable to current state-of-the-art datasets (PEMSD7, PEMSD8) which proves the utility of the DIVA dataset in accurate flow prediction. Then, we propose a novel approach for route navigation that utilizes the predicted future traffic speed at the nodes under consideration in the form of a future graph. The edge weights of the future graph are taken as the average of the nodes between which the edge is formed and finally, the Bellman-Ford algorithm is applied to this future graph to get the path between the given set of nodes that maximizes overall speed.

II. RELATED WORK

Many datasets [6]–[10] have been proposed for Traffic Flow Prediction. In a nutshell, the datasets have been created by collecting sensor data placed at several nodes across a pre-determined road network. These and many other datasets have been utilized to train several models for Traffic data prediction. In the early years of traffic forecasting, various statistical and machine-learning techniques such as Auto- Regressive Integrated Moving Average (ARIMA), Historical Average (HA), and Support Vector Regression (SVR) [11]–[13] were widely used. These methods relied on analyzing historical data and making predictions based on patterns and trends observed in the data. However, in recent years, the advent of Graph Neural Networks (GNN) has revolutionized the field of traffic modeling. STGCN [14] combines graph convolutional networks with temporal convolutional networks to model both the spatial and temporal aspects of traffic data by considering the connectivity of the road network and the correlation between different locations. On the other hand, DCRNN [8] leverages a diffusion process to model the traffic networks in DCRNN that capture the spatial relations by simulating the spread of information through the network. In addition to STGCN and DCRNN, there have been other notable works [15]–[20] in the field of traffic forecasting. These studies have achieved impressive performance on state- of-the-art datasets by employing innovative techniques and leveraging advancements in deep learning and graph-based models. However, these systems are difficult to scale due to the sensors required at all the nodes to capture traffic data.

Although existing methods have made achievements in their own lines of research, we observe the following issues:

- The lack of specific datasets for training models on Indian traffic and road networks has been a significant challenge in optimizing transportation systems within India. This is primarily due to the unique characteristics and complexities of Indian traffic and road networks that differ significantly from those in other countries. Indian traffic patterns are known to be highly com- plex and dynamic. Congestion levels, peak hours, and traffic flow exhibit unique patterns that are influenced by factors such as cultural practices, festivals, events, and heterogeneity of road conditions. Models trained on foreign datasets lack the necessary understanding of these patterns and are therefore unable to predict or optimize traffic conditions in Indian settings accurately.
- Current navigation systems allow users outside of con- gestion zones to benefit from real-time information and adjust their routes accordingly. However, for users who are already within congested areas, their options to avoid traffic are often limited or non-existent. They are essentially trapped in the congestion and contribute to its buildup. This is particularly problematic in densely populated urban areas where traffic congestion is a com- mon occurrence. Therefore, there is a need to develop a system that minimizes the possibility of traffic issues by providing such optimal routes to users that congestion is avoided.

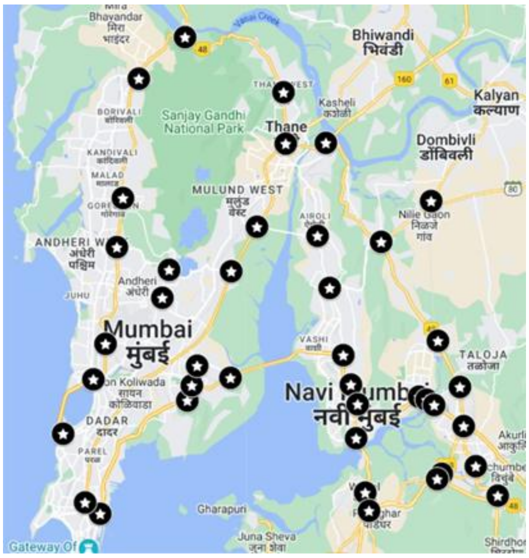

Fig. 1. Selected 40 points for data collection.

III. OUR CONTRIBUTION

- We construct the DIVA dataset that encapsulates the vehicular traffic data across different nodes across Navi Mumbai - Mumbai Area in India. Our experiments demonstrate that it effectively represents traffic in the region.
- We extensively test the DIVA dataset by training multi- ple Graph Neural Networks (GNNs) and comparing their performance with other datasets. The results indicate that the models achieve similar accuracy across all datasets, confirming the suitability of the DIVA dataset.
- We propose a system that utilizes spatiotemporal pre- dictions from the GNN model to determine an optimal path between a starting node and an end node. The algorithm considers future time intervals' speeds to minimize congestion and travel time by selecting a path with the least speed loss.

TABLE I

RESULTS AT 15, 30, AND 45 MINS TIME LAG.

Models	Dataset	15 mins			30 mins			45 mins		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
STGCN	DIVA	0.614	1.404	1.355	1.326	1.991	2.778	3.334	1.524	2.594
	PeMSD7	2.26	4.07	5.24	3.09	5.77	7.39	3.79	7.03	9.12
	PeMSD8	1.19	2.62	2.34	1.59	3.61	3.24	1.92	4.21	3.91
DCRNN	DIVA	0.365	5.231	0.012	0.552	5.426	0.023	0.694	5.603	0.038
	PeMSD7	2.22	4.25	5.16	3.04	6.02	7.46	3.64	7.24	9
	PeMSD8	1.17	2.59	2.32	1.49	3.56	3.21	1.71	4.13	3.83
A3TGCN	DIVA	0.445	5.312	0.011	0.644	5.689	0.072	0.824	5.923	0.125
	PeMSD7	2.58	4.52	5.77	3.21	5.41	7.22	3.34	5.11	4.72
	PeMSD8	1.27	2.35	2.31	1.67	2.89	2.76	2.03	3.92	3.18
SSTGNN	DIVA	2.04	3.53	1.77	2.67	4.8	4.6	3.17	5.79	6.24
	PeMSD7	1.03	2.08	1.86	1.39	2.8	2.67	1.62	3.28	2.67
	PeMSD8	0.95	1.76	2.01	0.96	1.69	2.03	0.91	1.69	1.94

IV. DIVA DATASET

We introduce the DIVA (Dataset for Indian Vehicular Traffic Analysis) dataset, focused on the road network near the Mumbai-Navi Mumbai area in India. It represents graphs of the road network having 40 nodes at several timestamps. Similar to PeMSD7 and PeMSD8, the dataset consists of the Adjacency matrix, representing the Euclidean distance between two nodes in the road network and traffic data, representing the current speed at several timestamps on 40 sensors. It is an ensemble of different sources of vehicle traffic speeds at specific roads as collected by TomTom which can be attained using the Traffic API and then, formatted to be utilized for the prediction tasks. The speed data is collected at 15-minute intervals and calculated as the root mean squared of the collected speeds at all edges for each node. While the real-time traffic data is constantly getting added to the dataset, till the time of this paper, it consists of 3150 data samples per node.

V. OPTIMAL PATH NAVIGATION

The optimal path can be obtained using the trained GNN model to infer optimal paths between any pair of nodes in the graph. Provide the starting and destination nodes as input to the model, along with the forecasted node speeds at N future timestamps (N > no. of nodes) [21]. At each timestamp, the future travel time (FTT) is calculated by dividing the weight of the edge (distance between nodes) by the average of the speeds at the nodes of each edge. Finally, the Bellman-Ford Algorithm is applied with the weights as the Future Travel Time initialized at starting timestamp. These weights are then, updated with the FTT at the next timestamp in accordance with the consideration of farther nodes for calculating distance. Finally, the algorithm outputs the shortest path from the starting node to every other node. This path can thus, be utilized to get the required shortest path from starting to the ending node.

VI. EXPERIMENTATION

In order to demonstrate the effectiveness and robustness of the proposed DIVA dataset, different GNNs are trained on DIVA and other state-of-the-art datasets. Finally, the results calculated on each dataset at different time lags are compared.

A. Datasets

Our experiments used two well-known traffic datasets from Caltrans Performance Measurement System: PeMSD7 and PeMSD8 [15], along with the proposed DIVA dataset.

- PeMSD7: This dataset includes the data in California's District 7 that measures traffic speed using 228 sensors for the period of May to June 2012 (only on weekdays), with a time interval of 5 minutes.
- PEMSD8: This dataset includes data from 170 detectors on 8 highways in San Bernardino, California, collected every five minutes from July through August of 2016.
- DIVA: This proposed dataset includes traffic speed data from 40 nodes on the Highway Network in the Mumbai

  - Navi Mumbai Area, collected every fifteen minutes from February through April of 2023.

B. Models under Consideration

- AT3GCN [22]: The Attention Temporal Graph Convolu- tional Network (A3T-GCN) model utilizes both gated re- current units (GRUs) and a graph convolutional network (GCN) to understand short-term patterns in time series data and capture spatial relationships within the road network topology. By incorporating an attention mech- anism, the model can dynamically assign importance to different time points and effectively integrate global temporal information, leading to enhanced accuracy in traffic prediction.
- STGCN [14]: Unlike traditional approaches using reg- ular convolutional and recurrent units, Spatio-Temporal Graph Convolutional Network (STGCN) formulates the problem using graphs and constructs a model with com- plete convolutional structures. This design choice leads to faster training speed and requires fewer parameters.

C.

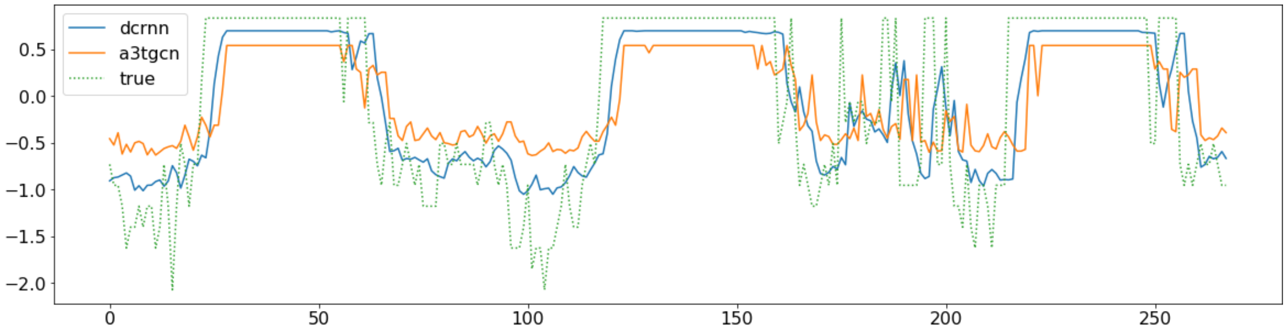

Fig. 2. True Vs. Predicted Speeds at 15 mins time lag.

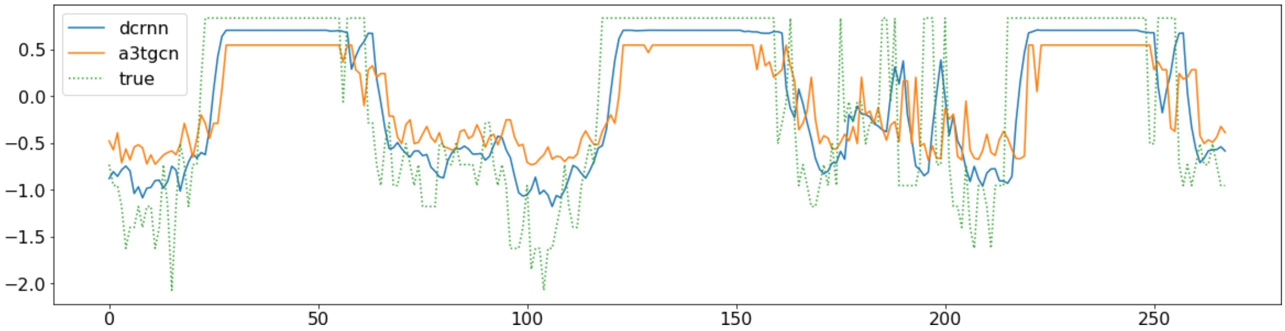

Fig. 3. True Vs. Predicted Speeds at 30 mins time lag.

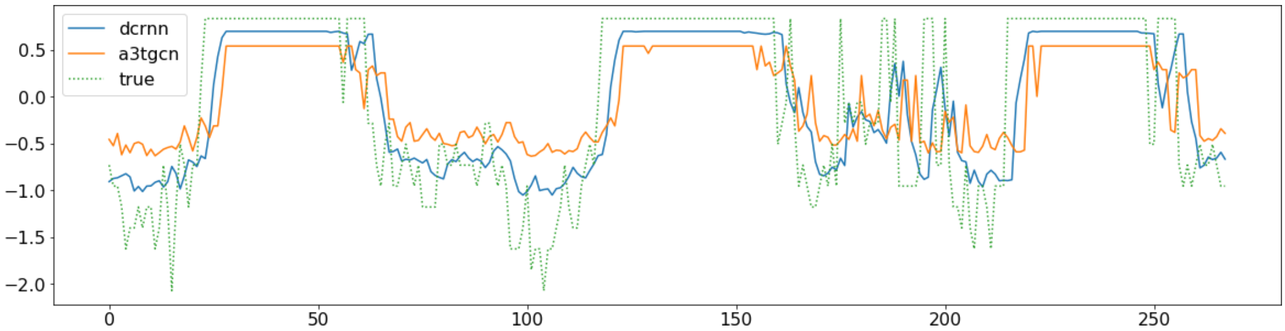

Fig. 4. True Vs. Predicted Speeds at 45 mins time lag.

- DCRNN [8]: Diffusion Convolutional Recurrent Neural Network (DCRNN) effectively incorporates both spatial and temporal dependencies in the traffic flow. It cap- tures spatial dependencies by employing bidirectional random walks on the graph and temporal dependencies through an encoder-decoder architecture with scheduled sampling.
- SSTGNN [15]: The Simplified Spatio-temporal Traffic forecasting Graph Neural Network (SST-GNN) efficiently incorporates spatial dependency by aggregat- ing different neighborhood representations individually instead of using multiple layers. It captures temporal dependency through a weighted spatiotemporal aggregation mechanism. To capture periodic traffic patterns, the model uses a unique position encoding scheme with historical and current data in two separate models.

D. *Evaluation Metrics*

To assess the performance of models on datasets, we employed three widely used metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These metrics offer valu- able insights into model performance and enable comparisons between different models or evaluations of prediction accu- racy. To ensure a dependable and robust evaluation process, we computed the average values across five experimental rounds.

E. Implementation Details

Z-score normalization is applied to data in our experiments which is a normalization technique that involves subtracting the mean and dividing by the standard deviation, resulting in a standardized representation of the data. De-normalization was performed on the predicted values to convert them back to their original scale. For training and testing the models, we divided the dataset into two subsets. The training set consisted of 80% of the data, while the remaining 20% was used for testing. This split allowed us to train the models on a majority of the data and assess their performance on unseen examples. SSTGNN and STGCN were trained and implemented using Pytorch while Pytorch Geometric Temporal was used for DCRNN and A3TGCN. We employed the Adam optimizer with a learning rate of 0.01. The training process was run for 500 epochs for each model.

F. Results and Analysis

Table I presents a comparison between different models and the state-of-the-art PeMS datasets at time lags of 15, 30, and 45 minutes. The models were fed with the previous four traffic speed values across the output label graph, capturing the speeds from the current to the prior 60 minutes. The graph models demonstrate favorable performance in terms of RMSE and MAPE on the DIVA dataset, comparable to PeMSD7 and PeMSD8. Figures 2, 3, and 4 visualize the test outcomes of DCRNN (blue), A3TGCN (orange), and the actual values (dotted green) for an individual sensor at time lags of 15, 30, and 45 minutes, respectively. Time lags refer to the interval between input sequences and output values. These figures indicate that the models effectively capture the dynamic traffic patterns across road networks, with DCRNN delivering the most optimal results. To enhance accuracy, it is advisable to explore better models or expand the graph's size, both vertically (increasing the number of nodes) and horizontally (increasing the number of data samples per node).

VII. CONCLUSION

In this paper, we propose a novel dataset for traffic pre- diction, encapsulating feature space within traffic data in the Indian Province. Experiments show that our models trained on the DIVA dataset provide similar results as other state- of-the-art real-world datasets, indicating its great potential for exploring spatiotemporal traffic data in India. We also propose a potential system to get an optimal path across a given set of nodes using the future predictions from the GNN models. In the future, we will improve the dataset by aggregating data from more sources and increasing the number of nodes thus, enlarging the area under consideration. Additionally, it is possible to further optimize the algorithm for path navigation. Moreover, this dataset can be applied to understand traffic patterns, predictions, Intelligent Transport Systems, etc.

VIII. ACKNOWLEDGMENTS

We would like to express our sincere gratitude to iHub- Data Mobility, our mentor, and our college faculty for their invaluable support of this research. We would also like to acknowledge the guidance and insights shared especially by the iHub-Data Mobility team, which helped to shape the direction and outcome of our work. We believe that this collaboration with iHub Data Mobility marks a significant milestone in our research journey, and it will serve as a foundation for further exploration and discovery in the field. We hope that this research will bring about meaningful and positive impacts for the community and industry, and contribute towards the continued growth and advancement of data mobility.

REFERENCES

[1] I. TCI, "Calcutta, 2016," *Operational Efficiency of Freight Transporta- tion by Road in India, Joint Study Report, Transport Corporation of India (TCI) and Indian Institute of Management Calcutta,*, 2016.

[2] W. Jiang and J. Luo, "Graph neural network for traffic forecasting: A survey," *Expert Systems with Applications*, vol. 207, p. 117921, nov 2022.

[3] G. Meena, D. Sharma, and M. Mahrishi, "Traffic prediction for intelligent transportation system using machine learning," in *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, 2020, pp. 145–148.

[4] Y. Wang, Y. Guo, Z. Wei, Y. Huang, and X. Liu, "Traffic flow prediction based on deep neural networks," in *2019 International Conference on Data Mining Workshops (ICDMW)*, 2019, pp. 210–215.

[5] C. Chen, Z. Liu, S. Wan, J. Luan, and Q. Pei, "Traffic flow prediction based on deep learning in internet of

vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3776–3789, 2021.

[6] L. Ge, S. Li, Y. Wang, F. Chang, and W. Kunyan, "Global spatial- temporal graph convolutional network for urban traffic speed predic- tion," *Applied Sciences*, vol. 10, p. 1509, 02 2020.

[7] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 922–929, Jul. 2019. [Online]. Available: https://doi.org/10.1609/aaai.v33i01.3301922

[8] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017. [Online]. Available: https://arxiv.org/abs/1707.01926

[9] X. Kong, W. Xing, X. Wei, P. Bao, J. Zhang, and W. Lu, "Stgat: Spatial-temporal graph attention networks for traffic flow forecasting," *IEEE Access*, vol. PP, pp. 1–1, 07 2020.

[10] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017. [Online]. Available: https://doi.org/10.1609/aaai.v31i1.10735

[11] C. Xu, Z. Li, and W. Wang, "Short-term traffic flow prediction using a methodology based on auto-regressive integrated moving average and genetic programming," *Transport*, vol. 31, no. 3, pp. 343–358, Sep. 2016. [Online]. Available: https://doi.org/10.3846/16484142.2016.1212734

[12] L. Jing and G. Wei, "A summary of traffic flow forecasting methods," *Journal of Highway and Transportation Research and Development*, vol. Vol. 21, pp. 82–85, 03 2004.

[13] J. Ahn, E. Ko, and E. Y. Kim, "Highway traffic flow prediction using support vector regression and bayesian classifier," in *2016 International Conference on Big Data and Smart Computing (BigComp)*, 2016, pp. 239–244.

[14] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, jul 2018. [Online]. Available: https://doi.org/10.24963+Fijcai.2018+F505

[15] A. Roy, K. K. Roy, A. A. Ali, M. A. Amin, and A. K. M. M. Rahman, "SST-GNN: Simplified spatio-temporal traffic forecasting model using graph neural network," in *Advances in Knowledge Discovery and Data Mining*. Springer International Publishing, 2021, pp. 90–102.

[16] C. Park, C. Lee, H. Bahng, Y. Tae, S. Jin, K. Kim, S. Ko, and J. Choo, "ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed," in *Pro- ceedings of the 29th ACM International Conference on Information and Knowledge Management*. ACM, oct 2020.

[17] C. Zheng, X. Fan, C. Wang, and J. Qi, "Gman: A graph multi-attention network for traffic prediction," 2019.

[18] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," 2019.

[19] G. Khodabandelou, W. Kheriji, and F. H. Selem, "Link traffic speed forecasting using convolutional attention-based gated recurrent unit," *Applied Intelligence*, vol. 51, no. 4, pp. 2331–2352, Nov. 2020. [Online]. Available: https://doi.org/10.1007/s10489-020-02020-8

[20] T. Mallick, P. Balaprakash, E. Rask, and J. Macfarlane, "Graph- partitioning-based diffusion convolutional recurrent neural network for large-scale traffic forecasting," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 9, pp. 473–488, jul 2020.

[21] X. Hu, T. Binaykiya, E. Frank, and O. Cirit, "Deepreta: An eta post- processing system at scale," 2022.

[22] J. Zhu, Y. Song, L. Zhao, and H. Li, "A3t-gcn: Attention temporal graph convolutional network for traffic forecasting," 2020. [Online]. Available: https://arxiv.org/abs/2006.11583