Embedding Transportation Knowledge Graphs for Enhancing Traffic Prediction Models

Traffic Prediction with Transportation Knowledge Graph

Md Mobasshir Rashid*

University of Central Florida, mdmobasshir.rashid@ucf.edu

Samiul Hasan

University of Central Florida, samiul.hasan@ucf.edu

To develop better traffic prediction models, it is crucial to consider the spatiotemporal characteristics of traffic network and the effects of many external factors such as weather, land use patterns etc. A knowledge graph stores relationships among these external factors by turning data into machine-understandable information. It can accurately represent the correlations between heterogeneous external data sources and transportation network. However, integrating knowledge graphs and traffic networks is challenging due to the inherent heterogeneity of information present in external factors. To address this issue, this study introduces a traffic prediction model fused with a knowledge representation technique. A knowledge representation learning model called 'TransE' has been used to produce knowledge graph embeddings which are added with traffic features as inputs to several machine learning-based traffic prediction models. The models are trained over traffic data collected in Florida's Seminole county from January to December 2019. Without knowledge graph embeddings, Random Forest model outperforms other prediction models with R^2 value of 0.78 (RMSE=148.18). However, when knowledge graph embeddings are used as additional features, the R^2 value of a Random Forest model increased to 0.85 (RMSE=120.67). Experimental results show that embedding knowledge graphs can significantly improve the accuracy of traffic prediction models.

CCS CONCEPTS •Applied computing~Physical sciences and engineering~Engineering

Additional Keywords and Phrases: Traffic Prediction, Transportation Knowledge Graph, Transportation Semantics Representation

ACM Reference Format:

* Corresponding Author

IWCTS'24, October 29-November 1 2024, Atlanta, GA, USA

ACM 979-8-4007-1151-0/24/10…
https://doi.org/10.1145/3681772.3698214

1 INTRODUCTION

Cities are getting more congested due to increasing urban development and vehicle demand on roads. Improving the efficiency of urban transportation by alleviating traffic congestions is an important problem. Intelligent transportation systems (ITS) technologies are utilizing artificial intelligence (AI) methods and big data coming from road sensors to enhance traffic prediction systems. Accurate prediction is beneficial for proactive decision making of urban traffic management and optimal traffic flows. Recent traffic prediction models use advanced computational techniques to integrate data from heterogeneous sources such as detector data, incident data, weather data, congestion reports etc. Accurate traffic prediction can benefit different stakeholders such as road users, policy makers, and traffic management agencies. It also helps to decrease travel time, increase road capacity, and enhance traffic efficiency [4].

Traffic states can be influenced by several external factors such as weather, incidents, holidays, special events, and land use etc. [7]. For example, point of interest (POI) can affect the surrounding traffic in a road network. A shopping mall or school can attract more vehicles in a certain period of a day. Weather condition also influences traffic flow propagation. Additionally, traffic congestion occurs due to a traffic incident such as a crash or a disabled vehicle. The influence of these factors on a network-level traffic prediction model is twofold as these factors can vary in temporal or spatial level. Furthermore, the traffic condition in two roads can vary significantly due to flow variations. Also, traffic flow in one road can be influenced by surrounding roads. As a result, the interdependency between these heterogeneous factors combinedly affect the traffic state of a road network [17].

It is challenging to incorporate the effect of heterogeneous sources in network-wide traffic prediction problem. Previous studies have used these external factors as additional input variables and ignore the interdependencies between traffic state and these external factors. Instead of using these external features as additional input features, a knowledge graph (KG) concept can be utilized to improve the traffic prediction model. KG can provide the relationship such as connectivity between different roads, association of roads and detectors, or association between roads and Point of Interest (POI) etc. One of the major advantages of KG concept is that it can handle data from multiple sources and provides output in different vector embeddings which can be added to the traffic prediction model. It can also explain the interaction between different components of a traffic network, which is not possible when different external factors are considered only as features in the model. Additionally, traditional data-driven models generate results that do not consider the semantic relationship among different segments of transportation networks and lack interpretation of the generated results. KG can address this issue as it can compute complex semantical relationships [13]. It can also provide information such as traffic situation analysis (which roads are congested in a certain period, traffic flow in surrounding roads), impact of traffic incidents on a location and adjacent neighboring roads etc.

Traffic prediction models can be divided into two branches: simulation-based and data-driven approaches. Simulation models rely on several rules to predict traffic, but they are difficult to calibrate due to extensive data requirements and have limits in capturing real-time traffic dynamics [16]. Data-driven approaches overcome these limitations as they learn from the data by extracting traffic-related features and can enhance predictive accuracy [9]. Several studies have used external factors as additional features by concatenating with traffic features [2]. Researchers have included weather, holiday, POI, traffic incidents, event data, temperature data as additional features to their model [6]. Other researchers utilized knowledge representation learning (KRL) methods to generate vector embeddings that can capture semantics of different entities and their associated relationships. Zhu et al. [18] proposed a prediction model which integrates external factors such as POI and weather data to predict traffic flow. Another study used adaptive fusion mechanism to predict future traffic flow via KRL [17]. Although these studies have showed the potential of knowledge graph, they didn't use several entities of a transportation network. For example, these studies didn't consider the influence of trip production or attraction values at traffic analysis zone (TAZ) level, and land use features at TAZ level in the knowledge representation. These external features can influence the traffic state of a network. For example, residential areas have high trip production and office areas have high trip attraction, which influences adjacent road network. Also, previous researchers didn't include detector-specific information such as road to detector information or detector to detector information to their KG. By utilizing additional information of detectors' relationship with their surrounding detectors and adjacent roads, the prediction model can benefit by learning how congestion can propagate at a localized level. This additional information can help model to understand the impact of sudden fluctuations of traffic flow caused by incidents or weather condition effectively. To address this issue, we have integrated several heterogeneous information such as TAZ- specific trip production/attraction values and land use features, road network data, connectivity between different spatial components of a traffic network in our knowledge graph. We also illustrate how to use KG to gain more insights of prediction results during traffic congestion via situation analysis as the KG can help us identify upstream and downstream locations of congestion area. It can also give insights on how congestion is impacting surrounding road networks of the affected location. As a result, we have shown the effectiveness of KG to improve the accuracy of traffic prediction models and how to use the knowledge graph to optimize large-scale network level traffic management system. This study contributes to literature in the following ways:

- Integration of different heterogeneous information from traffic detectors, incidents, weather reports, land use, travel demand, and road networks to develop a knowledge graph and generate vector embeddings to represent the semantic relationships among different components. Previous studies didn't consider TAZ-level features and spatial connections of detectors while constructing transportation knowledge graphs.
- We evaluate the developed approach by using real-world data from a transportation network of varying capacity such as highways, arterials, and signalized intersections. Previous studies did not consider such complex transportation networks to generate KG embeddings. The results show that incorporating such complex network information in the KG enhances model's performance.

2 DATA DESCRIPTION

2.1 Detector Data

We selected Florida's Seminole County as our study area and collected traffic detector data from January 1, 2019, to December 31, 2019. The study area contains different types of roads including Interstate highway (I-4), expressway (SR-417) and arterials (SR-17). We used Regional Integrated Transportation Information System (RITIS) platform for gathering traffic detector data for I-4 and surrounding arterials [19]. Additionally, the area contains several signalized streets that contain Automated Traffic Signal Performance Measures (ATSPM) detectors. After extracting traffic detector data, we selected total 262 detectors from available 498 detectors in the selected region (183 detectors from ATSPM data and 79 detectors from RITIS data). We filtered out 236 detectors by data processing as they contained missing values, noises and outliers. The traffic detector data were processed for 15-minute intervals. The study area and selected detectors are illustrated in Figure 1.

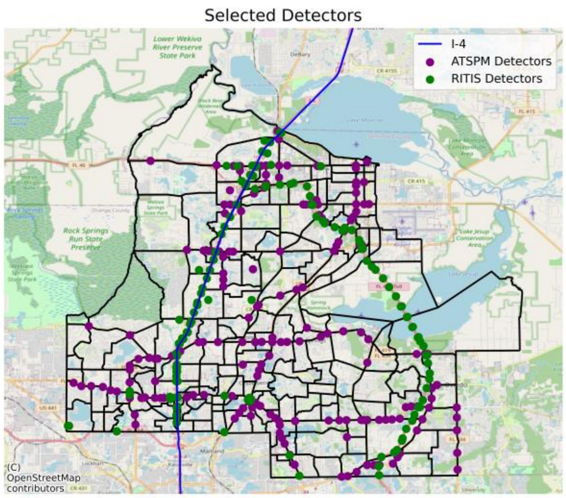

Figure 1: Selected detectors of the study area

2.2 Traffic Incident Data

We used traffic incident data downloaded from RITIS as additional features in our model. We considered only traffic crashes in this study and extracted 1912 crashes during the study period (January 2019 - December 2019). We used the start and end time of each crash and preprocess them to match with traffic detector data at 15-minute intervals.

2.3 Weather Data

We used weather data and extracted five types of weather patterns during the study period [20]. The categories are cloudy, fair, foggy, light rain and heavy rain conditions. The weather data is also processed to match with each 15-minute observation of the traffic detector data.

2.4 Knowledge Graph Data

To generate the knowledge graph of the selected transportation network, we need to create several relationships. First, we extracted TAZ information of the study area from the Central Florida Regional Planning Model and selected 208 TAZs in Seminole county. Then we extracted five land use features for each TAZ: area of residential, industrial, institutional, recreational, and retail area. We also collected total number of trips generated from a TAZ and number of trips attracted by a TAZ in 2019. We collected the road network information in the area. We used road name and associated geometry to generate relationships between roads and other roads/detectors. Finally, we generated the transportation knowledge graph of our study area which represents the following relationships: *TAZ to TAZ adjacency, TAZ to associated land use features, TAZ to trip production/Attraction, TAZ to road (*which TAZ contains a specific road section*), road to road (*which roads intersect with each other*), road to detector (*which detector falls into which road*) and detector to detector (*if two detectors are neighbor or adjacent to each other*)*. By using all of these relationships, we generated a knowledge graph as shown in Figure 2. To generate the knowledge graph embeddings, we need to create triple to explain relationship between two entities in this way: *(head, relation, tail).* For example, we used following triples: *(TAZ1, contains, Road 1), (Road 1, intersects, Road 2), (Road 3, has, Detector 6)* etc. In total, we have 548 entities, 1247 unique relationships, and 7343 unique triples obtained from the knowledge graph.

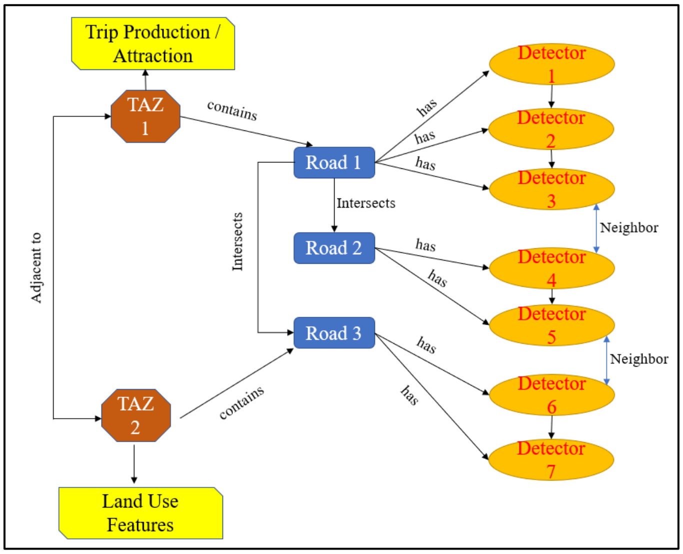

Figure 2: Schematics of the generated knowledge graph

3 METHODOLOGY

3.1 Knowledge Graph (KG) Representation

KG is represented by nodes or entities, the associated relationships/edges between two entities, and each node or edge can contain respective attributes [15]. We used 'Neo4j Desktop' to generate the KG triples. This is a graph database management software that can be used to store nodes and edges of a graph along with their respective attributes. The generated graph of our study area by Neo4j desktop is shown in Figure 3. Different colors are used to represent various entities such as TAZs, roads, trip production/attraction, different land use features and detectors of the developed KG.

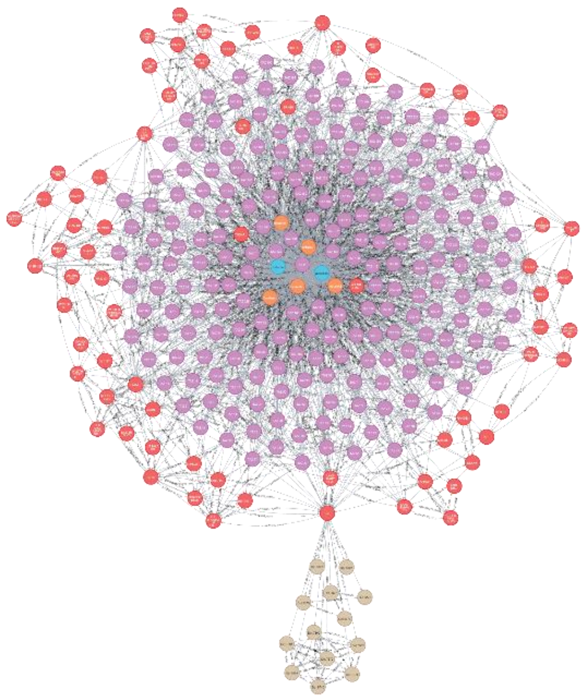

Figure 3: Generated Knowledge Graph in Neo4j Desktop

The study of extracting and using useful information from large-scale knowledge graphs for downstream prediction tasks is known as Knowledge Representation Learning (KRL) [8]. It can effectively represent underlying knowledge of the graph and efficiently maps implicit semantic information in graph structures to low-dimensional Euclidean space, revealing previously hidden relationships between different entities [12]. The main objective of these representation models is to generate vector embeddings for each node or entity up to a certain dimension. In this study, we generated 16-dimensional embeddings from the KG for each detector.

We used TransE [3], RotatE [10], DistMult [14], and ComplEx [11] models for generating the knowledge graph embedding. These models utilize the head (h), relationship (r), and tail (t) information of each 7343 triples generated from the knowledge graph. A brief description of models' formulation is provided below:

3.1.1 TransE Model.

It represents each relation consists of (h, r, t) as a translation in vector space as shown below.

$$h + r \approx t$$

3.1.2 RotatE Model.

It treats each relation (h, r, t) as movements in a vector embedding space. Unlike TransE, it treats relations as rotations in a complex space, rather than translations in a real space. Here, $\circ$ is elementwise (Hadamard) product.

$$h \circ r \approx t$$

3.1.3 DistMult Model.

It generates vector embeddings in real space by considering entity and relation as a single real value vector and creates a generalized dot product of these three vectors (h, r, t). Here, h and r are multiplied first, and the product (a scalar vector represents a transformed head) is multiplied with the t vector.

$$< h, r, t >$$

3.1.4 ComplEx Model.

It is an extension of the DistMult that uses complex vector space for embedding entity and relations. It generates two sets of embeddings (real and complex with same dimension), but the real part of the complex vector is used to generate embeddings as below.

$$Real(< h, r, t >)$$

3.2 Traffic Prediction Model

We used following models in this study: linear regression, adaboost regression, decision tree (DT) regression, extreme gradient boosting (XGboost) regression, random forest (RF) regression and artificial neural network (ANN) regression models to predict traffic flow at 15-minute intervals. First, we checked each model's performance without the knowledge graph. Then, we generate knowledge graph embeddings for each traffic detector via KRL. Then we merged them together with detector, weather and incident data via fusion block as shown in Figure 4. The model was trained on first 9 months (Jan.-Sept.) of data and tested on last 3 months' (Oct.-Dec.) data. We also maintained the temporal sequence of each detector in both training and testing data. The modeling framework is shown in Figure 4.

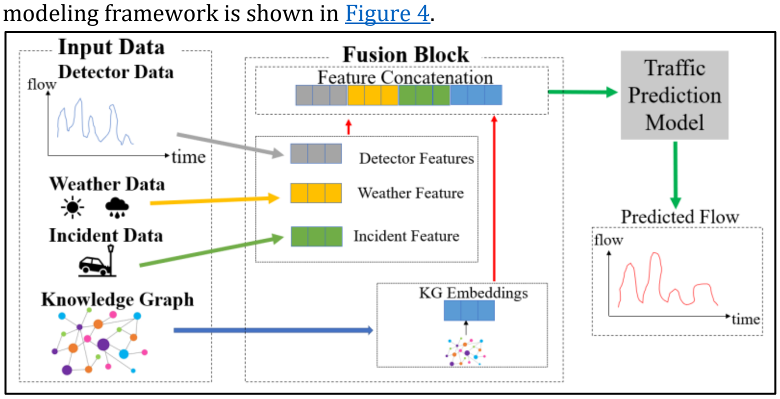

Figure 4: Modeling framework for traffic flow prediction model

3.3 Detector Data Processing

The raw traffic detector data may contain noises and outliers. To fix this issue, we conducted a data cleaning process. We calculated the percentage of missing information for each detector and discarded those detectors having more than 30% missing data. Then we checked for capacity for each detector over 15-minute intervals. For RITIS, we selected 2200 vehicles as the maximum capacity of freeways [1]. For other signalized intersections, we fixed 525 vehicles as maximum capacity at 15-minute intervals [5]. Then we checked for outliers by using isolation forest, local outlier factor and interquartile range algorithms. Then, we applied multivariate iterative imputation methods to generate the final traffic flow data for the prediction model. The detector data processing steps conducted in this study are shown in Figure 5.

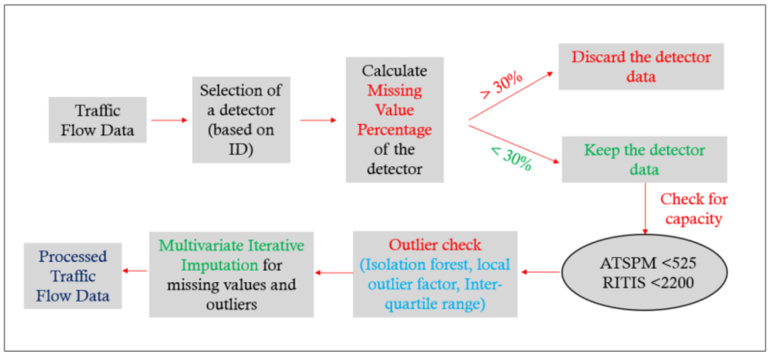

Figure 5: Detector data processing steps

4 RESULTS

4.1 Knowledge Graph Embedding Evaluation

After generating knowledge graph embeddings from different knowledge representation models, we compared their performances based on two evaluation metrics. The first metric is called ***hits @ n***, which denotes how well a model ranks true (h, r, t) triples among top n triples. For example, if a model can put a certain triple at rank 2, and another model puts the same triple at rank 5; then knowledge embeddings generated by the first model should be selected.

Additionally, we used another metric named ***Mean Reciprocal Rank (MRR)*** which is the average reciprocal rank of all triples generated by the model. If a model achieves high MRR value, then it is performing better than other models. The equation of how to calculate MRR for all triples (Q) predicted by a model is shown below:

$$MRR = \frac{1}{Q}\sum_{i=1}^{Q}\frac{1}{rank_i}$$

We compared model performance of TransE, RotatE, DistMult and ComplEx models for different hits @ n metrics. When we used hits@1 ranking, the TransE model performed better than other models. Later, we used hits@5 and hits@10 ranking, the performance of all models improved as we allow them to rank with more hits. In all scenarios, the TransE model outperformed other models and achieved 66% accuracy when hits@10 is used. The MRR value of the TransE model was higher than other models (around 0.36). As a result, we selected embeddings generated by the TransE model to further use in our traffic prediction model. Our findings also match with previous studies where translation-based knowledge representation technique was used only for traffic prediction [17,18]. We used other KRL techniques to check whether they can better capture the knowledge regarding spatial connectivity of a transportation network, but translational-based KRL technique (TransE) was found more suitable for the traffic prediction task. Figure 6 illustrates the comparison among different knowledge graph representation models.

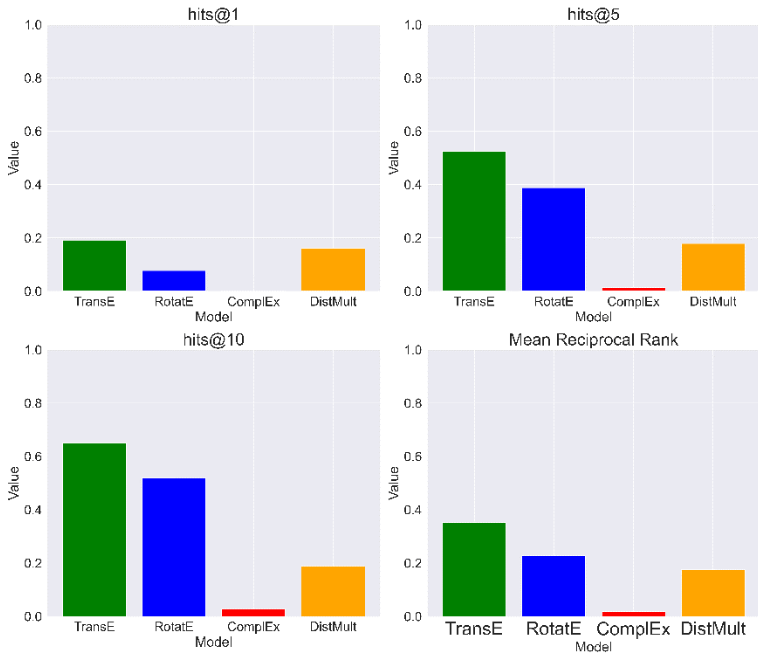

Figure 6: Comparison over different Knowledge Graph embedding models

4.2 Input Features

We selected several features related to TAZ level associated with each detector. For instance, we used the frequency of a TAZ (number of occurrences in the dataset). If a TAZ has more frequency, it means that more traffic detectors are situated in that TAZ. So, the model will give more priority to that TAZ compared to another TAZ which has lower frequency. We also used five types of land use features and associated trip production/attraction values for each TAZ. We used road name where detectors are situated by replacing with each road's frequency in the dataset. We also used several temporal features such as hour, day of the week, and weekend of the observation. Additionally, we used incident information (whether crashes occur or not) by assigning to the closest detector of the incident location. We used five types of weather conditions data by using their respective frequency. After the knowledge graph embedding was generated, we also used it as additional features. Finally, we used our target variable which is traffic flow (i.e. volume) at 15-minute intervals. Figure 7 illustrates the feature importance values generated by the Random Forest regression model. It can be seen that the embeddings generated by knowledge graph achieved high significance values. Top 10 significant features include 6 features that are obtained from the knowledge graph. The features used in this study are provided in Table. 1.

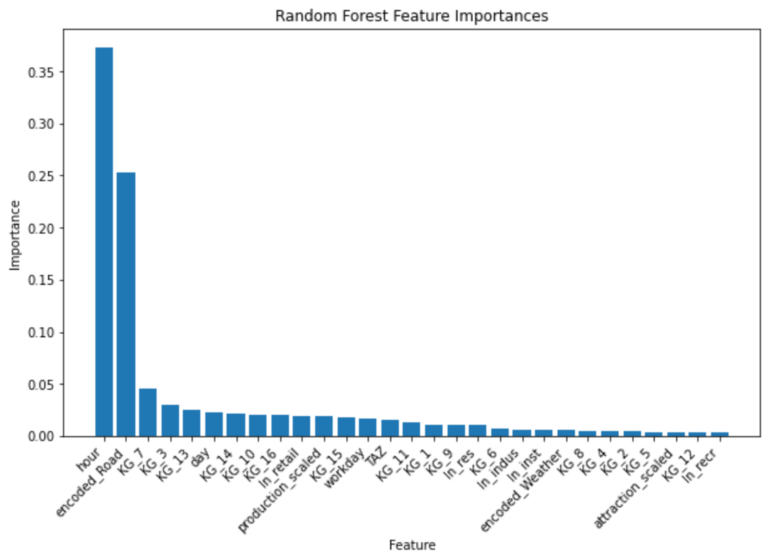

Figure 7: Feature importance from Random Forest Regression

Table 1: Features used in the traffic prediction model

Feature Type	Description	Feature Type	Description
TAZ Specific Features	TAZ ID	Temporal Features	Hour
	Residential Area		Day of the week
	Industrial Area		Weekend or not
	Institutional Area	Incident	Binary (0, 1)
	Recreational Area	Weather	Categorical (five weather categories)
	Retail/Office Area	Road Information	Road Name
	Trip Production	Knowledge Graph Embeddings	Generated from TransE model
	Trip Attraction	Target Variable	Flow at 15-minute Interval

4.3 Traffic Prediction Model Results

We selected data from January 1, 2019 to September 30, 2019 as the training data, and we maintained temporal sequence of each detector for consistency. Then, we checked the model performances on remaining 3 months data by using several evaluation metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and R^2 value. Equations to calculate RMSE and MAE values are provided below respectively. Here, $F_{actual,i}$ denotes the actual traffic flow at timestep i, and $F_{predicted,i}$ is the predicted traffic flow at timestep i.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(F_{actual,i} - F_{predicted,i}\right)^2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|F_{actual,i} - F_{predicted,i}\right|$$

The comparison of model performances is provided in Table 2. Without Knowledge Graph embeddings, the Random Forest regression model outperformed other regression models with a highest R^2 of 0.78 and RMSE of

148.18. When we incorporated the knowledge graph embeddings to the regression models, the Random Forest model's RMSE value further reduced to 120.67 and R^2 increased to 0.85. Apart from that, all models performed significantly better when the KG embeddings were included. The results indicate that KG captures semantic relationships among different spatial factors of the transportation network and increases prediction capabilities of all machine learning algorithms.

Table 2: Comparison of model performances (min flow 5, max flow 1862, mean flow 352)

Models	Without KG			With KG		
	RMSE	MAE	R^2	RMSE	MAE	R^2
Linear Regression	272.37	216.28	0.27	264.70	210.16	0.31
Adaboost	263.46	207.30	0.31	258.73	204.89	0.34
Decision Tree	209.71	149.77	0.56	203.02	147.19	0.59
XGBoost	180.30	126.04	0.68	160.59	116.50	0.74
ANN	159.91	102.92	0.75	125.63	76.24	0.84
Random Forest	148.18	85.82	0.78	120.67	66.29	0.85

4.4 Practical Application of the Knowledge Graph

We can use knowledge graph to understand the impact of congestion in a network. For example, we can identify both upstream and downstream detectors of a specific detector where congestion occurs by running a query in the graph database in Neo4j and analyze the congestion situation in that road segment. After upstream and downstream detectors are detected via automated queries, we can use predicted traffic flows of those detectors to analyze future traffic condition of a specific road segment. Figure 8 (left) shows one upstream and one downstream detectors of a selected location obtained from the knowledge graph and the corresponding traffic volume predictions.

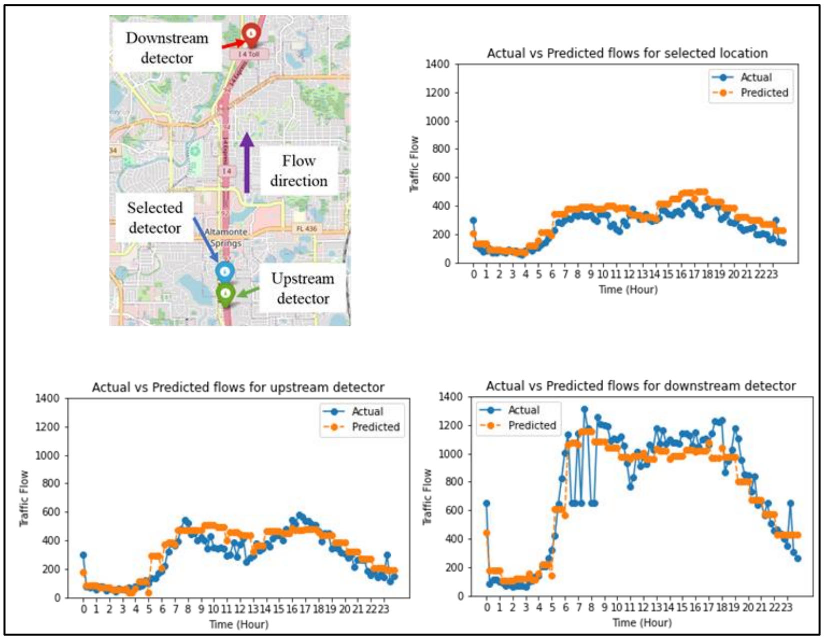

Figure 8: Use of knowledge graph to identify one upstream and one downstream detectors of a selected detector in the same road section along with traffic predictions

Similarly, knowledge graph can help analyze congestion patterns of a specific road and its adjacent road sections. Figure 9 shows two neighboring detectors that are situated in different roads obtained from automated knowledge graph queries and the corresponding traffic predictions. Here, KG enables us, in an automated way, to identify these neighboring detectors along with how congestion can propagate from one road to another road. Traffic managers can query the knowledge graph database to find alternative routes of a congestion location and then use predicted traffic flows of detectors placed in alternate routes to facilitate optimal traffic management strategies.

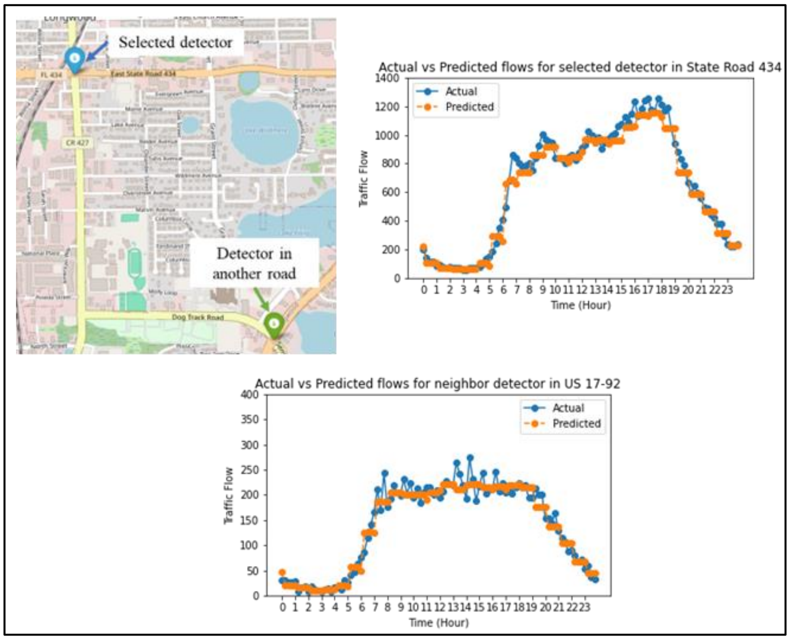

Figure 9: Actual vs. predicted traffic for two neighboring detectors in different roads

5 CONCLUSIONS

In this study, we adopted a new concept of incorporating a knowledge graph to predict network-wide traffic volume. We combined heterogeneous data sources such as traffic detector, incident, weather, land use, and road network data to increase model's performance by incorporating semantical relationships of different spatial factors of the transportation network into the prediction model. In the network, we used detectors placed in freeways, adjacent arterials, and signalized intersections and flow capacity of these detectors varied significantly. The model successfully understands the flow propagation at different roads with promising accuracy. A Random Forest model performed well with an R^2 of 0.85 when knowledge graph embeddings were fused with input features. We also conducted situation analysis to demonstrate the utility of knowledge graph to analyze congestion pattern in the network. The proposed methodology has potential towards developing a generalizable traffic prediction model that can be applied over any transportation network.

The study also has several limitations. More vehicles may travel through Seminole county due to tourist activity during summer, but our test dataset doesn't have that seasonal variation. Also, we used categorical weather information. More enriched weather information such as temperature and precipitation can be used to understand the impact of weather on network-level traffic state. Furthermore, we used machine-learning algorithms for traffic prediction, but more sophisticated models such as long short-term memory (LSTM) model or graph neural network (GNN) can be used to check whether better predictive capabilities can be achieved.

ACKNOWLEDGMENTS

The authors are supported by NSF EAGER Grant no: 2122135 and Florida Department of Transportation. However, the authors are solely responsible for the facts and accuracy of the information presented in the paper.

REFERENCES

<bib id="bib1"><number>[1]</number> Highway Capacity Manual 1994. Traffic Level of Service Calculation Methods Appendix B. Retrieved September 20, 2023 from https://ccag.ca.gov/wp-content/uploads/2014/07/cmp_2005_Appendix_B.pdf </bib>
<bib id="bib2"><number>[2]</number> Ahmad Ali, Yanmin Zhu, and Muhammad Zakarya. 2022. Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction. *Neural Networks* 145: 233–247. https://doi.org/10.1016/j.neunet.2021.10.021</bib>
<bib id="bib3"><number>[3]</number> Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, 2787–2795.</bib>
<bib id="bib4"><number>[4]</number> Cen Chen, Kenli Li, Sin G. Teo, Xiaofeng Zou, Keqin Li, and Zeng Zeng. 2020. Citywide Traffic Flow Prediction Based on Multiple Gated Spatio-temporal Convolutional Neural Networks. *ACM Transactions on Knowledge Discovery from Data* 14, 4. https://doi.org/10.1145/3385414</bib>
<bib id="bib5"><number>[5]</number> Federal Highway Administration (FHWA). 2004. Signalized Intersections: Informational Guide. Retrieved November 04, 2023 from https://www.fhwa.dot.gov/publications/research/safety/04091/</bib>
<bib id="bib6"><number>[6]</number> Siyu Hao, Der Horng Lee, and De Zhao. 2019. Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system. *Transportation Research Part C: Emerging Technologies* 107, July: 287–300. https://doi.org/10.1016/j.trc.2019.08.005</bib>
<bib id="bib7"><number>[7]</number> Ibai Lana, Javier Del Ser, Manuel Velez, and Eleni I. Vlahogianni. 2018. Road Traffic Forecasting: Recent Advances and New Challenges. *IEEE Intelligent Transportation Systems Magazine* 10, 2: 93–109. https://doi.org/10.1109/MITS.2018.2806634</bib>
<bib id="bib8"><number>[8]</number> Xinyu Li, Mengtao Lyu, Zuoxu Wang, Chun Hsien Chen, and Pai Zheng. 2021. Exploiting knowledge graphs in industrial products and services: A survey of key aspects, challenges, and future perspectives. *Computers in Industry* 129: 103449. https://doi.org/10.1016/j.compind.2021.103449</bib>
<bib id="bib9"><number>[9]</number> Nidhi Nidhi and D. K. Lobiyal. 2022. Traffic flow prediction using support vector regression. *International Journal of Information Technology (Singapore)* 14, 2: 619–626. https://doi.org/10.1007/s41870-021-00852-2</bib>
<bib id="bib10"><number>[10]</number> Zhiqing Sun, Zhi Hong Deng, Jian Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. *7th International Conference on Learning Representations, ICLR 2019*: 1–18.</bib>
<bib id="bib11"><number>[11]</number> Theo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Ciaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. *33rd International Conference on Machine Learning, ICML 2016* 5: 3021–3032.</bib>
<bib id="bib12"><number>[12]</number> Shilpa Verma, Rajesh Bhatia, Sandeep Harit, and Sanjay Batish. 2023. Scholarly knowledge graphs through structuring scholarly communication: a review. *Complex and Intelligent Systems* 9, 1: 1059–1095. https://doi.org/10.1007/s40747-022-00806-6</bib>
<bib id="bib13"><number>[13]</number> Bingcong Xue and Lei Zou. 2023. Knowledge Graph Quality Management: A Comprehensive Survey. *IEEE Transactions on Knowledge and Data Engineering* 35, 5: 4969–4988. https://doi.org/10.1109/TKDE.2022.3150080</bib>
<bib id="bib14"><number>[14]</number> Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*: 1–12.</bib>
<bib id="bib15"><number>[15]</number> Zi Ye, Yogan Jaya Kumar, Goh Ong Sing, Fengyan Song, and Junsong Wang. 2022. A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs. *IEEE Access* 10, May: 75729–75741. https://doi.org/10.1109/ACCESS.2022.3191784</bib>
<bib id="bib16"><number>[16]</number> Xueyan Yin, Genze Wu, Jinze Wei, Yanming Shen, Heng Qi, and Baocai Yin. 2022. Deep Learning on Traffic Prediction: Methods, Analysis, and Future Directions. *IEEE Transactions on Intelligent Transportation Systems* 23, 6: 4927–4943. https://doi.org/10.1109/TITS.2021.3054840</bib>
<bib id="bib17"><number>[17]</number> Yi Zhou, Yihan Liu, Nianwen Ning, Li Wang, Zixing Zhang, Xiaozhi Gao, and Ning Lu. 2023. Integrating knowledge representation into traffic prediction: a spatial–temporal graph neural network with adaptive fusion features. *Complex and Intelligent Systems* 10, 2: 2883–2900. https://doi.org/10.1007/s40747-023-01299-7</bib>
<bib id="bib18"><number>[18]</number> Jiawei Zhu, Xing Han, Hanhan Deng, Chao Tao, Ling Zhao, Pu Wang, Tao Lin, and Haifeng Li. 2022. KST-GCN: A Knowledge-Driven Spatial-Temporal Graph Convolutional Network for Traffic Forecasting. *IEEE Transactions on Intelligent Transportation Systems* 23, 9: 15055–15065. https://doi.org/10.1109/TITS.2021.3136287</bib>
<bib id="bib19"><number>[19]</number>] Regional Integrated Transportation Information System (RITIS) 2022. Retrieved September 4, 2022 from https://ritis.org/intro </bib>
<bib id="bib20"><number>[20]</number> Weather Underground Historical Data 2023. Retrieved September 24, 2023 from https://www.wunderground.com/history </bib>