

Bike Sharing Demand Prediction based on Knowledge Sharing across Modes: A Graph-based Deep Learning Approach

Yuebing Liang¹, Guan Huang² and Zhan Zhao³

Abstract—Bike sharing is an increasingly popular part of urban transportation systems. Accurate demand prediction is the key to support timely re-balancing and ensure service efficiency. Most existing models of bike-sharing demand prediction are solely based on its own historical demand variation, essentially regarding bike sharing as a closed system and neglecting the interaction between different transport modes. This is particularly important because bike sharing is often used to complement travel through other modes (e.g., public transit). Despite some recent efforts, there is no existing method capable of leveraging spatiotemporal information from multiple modes with heterogeneous spatial units. To address this research gap, this study proposes a graph-based deep learning approach for bike sharing demand prediction (B-MRGNN) with multimodal historical data as input. The spatial dependencies across modes are encoded with multiple intra- and inter-modal graphs. A multi-relational graph neural network (MRGNN) is introduced to capture correlations between spatial units across modes, such as bike sharing stations, subway stations, or ride-hailing zones. Extensive experiments are conducted using real-world bike sharing, subway and ride-hailing data from New York City, and the results demonstrate the superior performance of our proposed approach compared to existing methods.

Index Terms—bike sharing, demand prediction, inter-modal relationships, graph neural networks, deep learning

I. INTRODUCTION

In the past decade, bike sharing has emerged as a sustainable, convenient, and generally affordable travel mode, and become an integral component of urban transportation systems in cities around the world. The trend is likely to continue under the ongoing COVID-19 pandemic, as biking is usually seen as a safer and healthier travel option. Because of its positive effects on the environment, public health and traffic congestion, bike sharing systems (BSS) should be promoted to play a bigger role in urban mobility. Currently, efficient operations of BSS rely on the dynamic rebalancing of bikes between locations to better match the ever-changing demand patterns [1]. Therefore, the accurate short-term demand prediction at high spatial resolution is crucial because it is the basis to ensure the availability, efficiency, and user experience of bike sharing service.

Recent years have seen growing interests in short-term demand prediction for intelligent BSS, with a particular focus on deep learning methods, because of their demonstrated

effectiveness in extracting the complex and nonlinear knowledge hidden in large-scale mobility data [2], [3]. Despite the success of these methods, they regard bike sharing as a closed system and neglect the potential rich information of the interaction between BSS and other transportation modes. This is especially important to consider for bike sharing, because it is mainly used for short-distance trips or the first-mile/last-mile portion of longer trips. In practice, BSS are often designed as feeders to public transport systems or support multimodal transport connections [4]. Extensive prior studies have explored the relationship between BSS usage with public transit, taxi and ride-hailing, and unrevealed significant complementary or competitive relationships subject to trip purpose, trip length, availability, etc [5], [6]. As a result, the demand for bike sharing will inevitably be influenced by other transportation modes, which should be considered in demand prediction. Incorporating demand information across modes can also help mitigate the data sparsity problem commonly seen in BSS, since bike sharing is rarely one of the primary travel modes in cities.

To incorporate inter-modal demand information for bike sharing demand prediction, recent studies have proposed to relate each BSS station to its adjacent subway stations or bus stops and use their accumulated usage as an additional attribute of BSS stations [4], [7]. Another approach is to divide the study area into uniform grids and aggregate multimodal data to the same grid system [8], [9]. However, these methods are not always straightforward because of the heterogeneous spatial units of different transportation modes: some are station-based (e.g., subway, station-based BSS), while others are stationless (e.g., ride-hailing, dockless BSS) [10]. Instead of arbitrarily aligning multimodal networks, we need a more flexible approach that can learn the inter-modal relationships across heterogeneous spatial units directly from data. Also, the learning should be done with high spatial resolution so that it is useful for BSS operations. For example, dynamic rebalancing for station-based BSS would require good demand prediction at the station level.

In this paper, we propose a graph-based deep learning approach for bike sharing demand prediction (B-MRGNN). To integrate multimodal data with diverse spatial units, we encode spatial dependencies across different modes with multiple intra- and inter-modal graphs. To extract cross-mode relationships, we introduce an improved version of the multi-relational spatiotemporal graph neural network from our prior work [10]. The proposed model is demonstrated using the Citi Bike data from New York City (NYC), with the subway and ride-hailing data as additional inter-modal

¹Yuebing Liang is with Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China yuebingliang@connect.hku.hk

²Guan Huang is with Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China guanhuang@connect.hku.hk

³Zhan Zhao is with Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China zhanzhao@hku.hk

demand input. Because of data constraint, we will focus on station-based BSS in this study, but the methodology should be generalizable to stationless (or dockless) BSS as well. The specific contributions of this research are as follows:

- We propose a deep learning model for bike sharing demand prediction based on inter-modal relationships learnt from historical demand data.
- We introduce a multi-relational graph neural network (MRGNN) to model spatial correlations between heterogeneous spatial units across different modes.
- Extensive experiments are conducted based on real-world datasets from NYC, and the results demonstrate the superior performance of our proposed model compared to existing methods.

II. LITERATURE REVIEW

Traditional methods for bike sharing demand prediction mainly focus on finding the relationships between bike demand and its historical demand and exogenous factors (e.g. weather, land use) through regression models, such as auto-regressive integrated moving average (ARIMA) [11] and linear regression (LR) [12]. Later research adopts other machine learning models, including random forest [13] and gradient boosting machines [14]. In recent years, extensive studies have presented the power of deep learning models to capture the nonlinear and complex relationships for demand prediction tasks. Xu et al. [2] introduced a recurrent-based model considering the effect of exogenous factors. Jin et al. [3] applied temporal convolution networks (TCNs) to capture temporal dependencies for bike sharing demand prediction. To incorporate spatial information, researchers employed convolutional neural networks (CNNs) along with recurrent layers to capture complex spatiotemporal influences, using artificial grids for demand aggregation [15], [16]. Graph neural networks (GNNs) exempt the requirement of artificial segmentation and can capture relationships at the station level, which is more useful for BSS operations. Yu et al. [17] introduced a convolution-based model with Graph convolutional networks (GCNs) to model spatial dependencies and TCNs to model temporal dependencies. A multi-graph learning approach was proposed in [18] for ride-hailing demand prediction encoding multiple types of spatial dependencies. Wu et al. [19] introduced an adaptive adjacency matrix to learn spatial dependencies hidden in data. However, these methods are mode-specific and do not consider inter-modal relationships.

Only a few studies have considered the influence of other modes for bike sharing demand prediction. Zhang et al. [4] incorporated the historical demand of public transit and used LSTM for bike sharing demand prediction. Cho et al. [7] enhanced the accuracy of bike sharing demand during peak hours with public transit usage information using a graph learning approach. These methods assign the accumulated flow of adjacent public transit stations to each BSS station, and do not model the spatiotemporal dependencies among adjacent stations/zones from different modes directly. Recent research has also investigated the co-prediction of bike

sharing demand and other transport modes. In [8] and [9], the demands for taxis and bike sharing are aggregated to a grid system to enable shareable feature learning, before co-predicted using a convolutional recurrent network. Despite existing relevant works, a model that can directly leverage spatiotemporal knowledge across modes with heterogeneous spatial units is still needed.

III. PROBLEM STATEMENT

In this section, we introduce some notations in this research and then formulate our problem.

Definition 1 (Demand Sequence): Consider a transport mode m with N_m nodes (i.e. stations/service zones). For each node $i = 1, 2, \dots, N_m$, its inflow and outflow demand at time step t is denoted as $x_{m,i}^t \in \mathbb{R}^2$. Next, we represent the demand of all the nodes from mode m at time step t as $X_m^t = \{x_{m,0}^t, x_{m,1}^t, \dots, x_{m,N_m}^t\}$, $X_m^t \in \mathbb{R}^{N_m \times 2}$. Further, we use $X_m^{t-T:t} = \{X_m^{t-T}, \dots, x_m^{t-1}, x_m^t\}$ to denote the demand sequence of mode m over time steps T .

Problem (Bike Sharing Demand Prediction): This research aims to predict the station-level bike sharing demand given historical demand of BSS as well as other modes. Formally, given the historical demand of bike sharing denoted as $X_b^{t-T:t}$ and auxiliary modes, i.e., subway and ride-hailing in our case, denoted as $X_s^{t-T:t}$ and $X_h^{t-T:t}$, the goal is to predict bike sharing demand X_b^{t+1} at the next time step :

$$X_b^{t+1} = F(X_b^{t-T:t}, X_s^{t-T:t}, X_h^{t-T:t}), \quad (1)$$

where $F(*)$ is the prediction function to be learned by our proposed model. This formulation can be easily adapted to other demand prediction problems with multimodal historical demand as input.

IV. METHODOLOGY

This section presents an improved version of the multi-relational spatiotemporal graph neural network framework [10] for station-level bike sharing demand prediction (B-MRGNN) by taking advantage of historical demand sequences of subway and ride-hailing systems. Figure 1 displays the overall architecture of our proposed model. It is composed of L multi-relational spatiotemporal blocks (ST-MR blocks) for multimodal representation learning, each comprising TCNs to model temporal patterns and multi-relational graph neural networks (MRGNNs) to model the spatial influence of adjacent subway stations and ride-hailing zones on BSS stations. Based on the learned representation from ST-MR blocks, a prediction layer is used to generate bike sharing demand prediction. We further introduce a prediction-based regularization term that incorporates the demand prediction error for subway and ride-hailing during model training. Details of each module are introduced below.

A. Multi-relational graph neural network

In this subsection, we provide more details on the MRGNN that is used to capture the interactions between heterogeneous spatial units across modes. As shown in

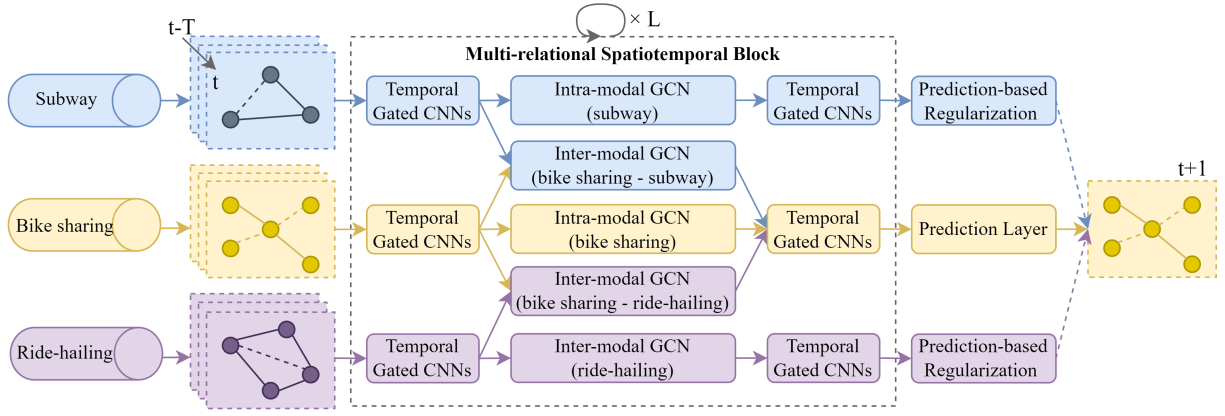


Fig. 1. The architecture of B-MRGNN

Figure 2, MRGNN is composed of two major parts: *multi-relational graph construction* to encode cross-mode spatial dependencies and *multi-relational graph convolutions* to capture correlations between nodes through message passing.

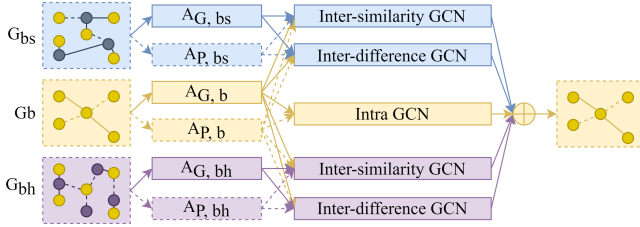


Fig. 2. The framework of MRGNN

1) *Multi-relational graph construction*: As subway and BSS (in our case) are station-based and ride-hailing is stationless, it is difficult to model spatial dependencies on a single homogeneous graph. To encode spatial dependencies within and across modes, we define multiple intra- and inter-modal graphs. Specifically, **intra-modal graphs** are used to capture spatial correlations among stations/zones of the same mode. Taking bike sharing as an example, its intra-modal graph is defined as $G_b = (V_b, A_b)$, where V_b is a set of BSS stations, and $A_b \in \mathbb{R}^{N_b \times N_b}$ is an adjacency matrix representing the spatial dependencies between BSS stations. Similarly, an intra-modal graph is defined for subway and ride-hailing, denoted as G_s and G_r . **Inter-modal graphs** are defined to capture the pairwise correlations among stations/zones between the target mode (i.e., bike sharing) and each of the auxiliary modes (i.e., subway and ride-hailing). For example, the inter-modal graph between bike sharing and subway is represented as $G_{bs} = (V_b, V_s, A_{bs})$, where V_b and V_s denote BSS and subway stations and $A_{bs} \in \mathbb{R}^{N_b \times N_s}$ is a weighted matrix indicating the cross-mode dependencies between adjacent BSS and subway stations. Similarly, an inter-modal graph is defined between bike sharing and ride-hailing denoted as G_{bh} .

Prior studies have demonstrated that strong correlations may occur between locations that are either geographically close or semantically similar (i.e., demand patterns) [18]. To

encode both relationships, we define two adjacency matrices for each graph: one for **geographical proximity** denoted as A_G and, the other for **semantic similarity** denoted as A_P . The former is computed as a function of distance, while the latter based on demand patterns. The specific definitions of A_G and A_P follow our prior work [10].

Figure 3 illustrates the constructed multi-relational graph of bike sharing, subway and ride-hailing. A total of $(3+2) \times 2 = 10$ relations is defined to encode spatial dependencies between nodes from different modes, including 3 intra-modal and 2 inter-modal graphs, each with 2 adjacency matrices. This formulation can be easily adapted to encode spatial dependencies across other modes with diverse network structures spatial.

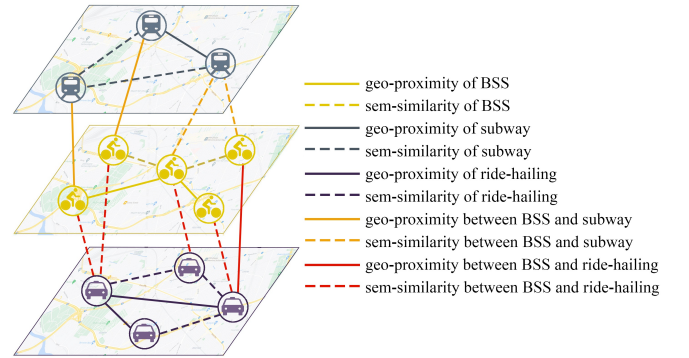


Fig. 3. Modeling multimodal spatial dependencies for BSS

2) *Multi-relational graph convolutions*: Graph convolutions have been an effective way to aggregate information of connected nodes. However, most existing GCNs cannot be applied to the multi-relational graph defined above for two main reasons: first, they cannot process inter-modal graphs with heterogeneous nodes and a non-square adjacency matrix; second, they fail to consider the disparity between different modes, which might result in negative transfer. To tackle these issues, we introduce an inter-modal graph convolution network to aggregate information of connected subway stations and ride-hailing zones for each BSS station, considering both cross-mode similarity and difference. Given the

inter-modal graphs $G_{bs} = (V_b, V_s, A_{bs})$ and $G_{bh} = (V_b, V_h, A_{bh})$, as well as the input representations of subway and ride-hailing, denoted as $H_s \in \mathbb{R}^{N_s \times c_{in}}$, $H_h \in \mathbb{R}^{N_h \times c_{in}}$ respectively, the **inter-modal similarity** is modeled as:

$$Z_{bs}^{(c)} = \text{ReLU}(\tilde{A}_b(\tilde{A}_{bs}H_s)W_{bs}^{(c)} + l_{bs}^{(c)}), \quad (2)$$

$$Z_{bh}^{(c)} = \text{ReLU}(\tilde{A}_b(\tilde{A}_{bh}H_h)W_{bh}^{(c)} + l_{bh}^{(c)}), \quad (3)$$

where $Z_{bs}^{(c)} \in \mathbb{R}^{N_b \times c_{out}}$, $Z_{bh}^{(c)} \in \mathbb{R}^{N_b \times c_{out}}$ are the aggregated features of connected subway stations and ride-hailing zones on BSS stations respectively, $W_{bs}^{(c)}, W_{bh}^{(c)} \in \mathbb{R}^{c_{in} \times c_{out}}$ and $l_{bs}^{(c)}, l_{bh}^{(c)} \in \mathbb{R}^{c_{out}}$ are the learned model parameters, c_{in} and c_{out} represents the input and output vector dimension of each node (i.e. station/zone). $\tilde{A} = \frac{A}{\text{rowsum}(A)}$ denotes the normalized adjacency matrix constructed from A .

Our preliminary analysis shows that there is a notable disparity between the demand patterns of different modes: generally, the demand of subway and ride-hailing are more concentrated during rush hours, while bike sharing usage is more active during off-peak period. Simply modeling cross-mode similarity can easily lead to negative transfer. Inspired by [20], we model **inter-modal difference** between bike sharing and the two auxiliary modes as:

$$Z_{bs}^{(d)} = \text{ReLU}(\tilde{A}_b(\tilde{A}_{bs}H_s - H_b)W_{bs}^{(d)} + l_{bs}^{(d)}), \quad (4)$$

$$Z_{bh}^{(d)} = \text{ReLU}(\tilde{A}_b(\tilde{A}_{bh}H_h - H_b)W_{bh}^{(d)} + l_{bh}^{(d)}), \quad (5)$$

where $|\tilde{A}_{bs}H_s - H_b|$, $|\tilde{A}_{bh}H_h - H_b|$ represent the information gap between BSS stations and the aggregated information of connected subway stations and ride-hailing zones respectively, $W_{bs}^{(d)}, W_{bh}^{(d)}, l_{bs}^{(d)}, l_{bh}^{(d)}$ are the learned model parameters.

In addition, we apply a standard GCN layer to model pairwise correlations between connected BSS stations. Given the intra-modal graph G_b and the input representations of bike $H_b \in \mathbb{R}^{N_b \times c_{in}}$, the correlations between BSS stations are modeled as:

$$Z_b = \text{ReLU}(\tilde{A}_b H_b W_b^c + l_b^c), \quad (6)$$

where W_b^c, l_b^c are model parameters. In practice, the inter-modal similarity and difference graph convolution layers as well as the intra-modal graph convolution layer can be modeled in parallel using batch matrix multiplication operations. Through the intra- and inter-modal graph convolutions, each BSS station receives multiple feature vectors from geographically adjacent or semantically similar subway stations, ride-hailing zones and other BSS stations. The learned feature vectors from heterogeneous neighborhood nodes are then aggregated using an adding function.

B. Multi-relational spatiotemporal block

In this subsection, we describe the design of ST-MR blocks, which are stacked to capture correlations between nodes of different modes from both spatial and temporal domains. To capture temporal dependencies, we employ a temporal gated convolution network (TCN) proposed by [17], due to its high training efficiency and good prediction performance. Briefly, given the historical demand series of a

node, TCN models the relationship between each time step and its neighborhoods using a 1-D causal convolution. As introduced above, different modes usually exhibit notably different temporal patterns. Therefore, we apply a separate TCN layer for each mode to capture mode-specific temporal information. The learned features of different modes from TCN layers are then fed into inter-modal graph convolution layers to jointly model spatial and temporal dependencies of BSS stations on connected subway stations and ride-hailing zones. We also apply an intra-modal graph convolution layer on each mode to capture spatiotemporal correlations within modes. Following [10], we use an additional TCN layer for each mode after MRGNN layers. The input of the second TCN layer is the sum of the output of the first TCN layer and the MRGNN layer, as a residual connection function to speed up model training. To stabilize model parameters, we employ a layer normalization function at the end of each ST-MR block. After the L stacked ST-MR blocks, each BSS station gets a learned representation vector that summarizes the spatiotemporal information from multimodal historical demand. Based on the learned representations, we generate future demand predictions using a fully-connected feed-forward network.

C. Prediction-based Regularization

To train our model more efficiently, we introduce a prediction-based regularization term. Specifically, in addition to bike sharing demand prediction, we generate future predictions of subway and ride-hailing demand with additional feed-forward networks, and use the prediction error of subway and ride-hailing demand as a regularization term in the loss function. The intuition behind it is that by jointly optimizing the demand prediction of auxiliary modes, we can guide the model to extract useful spatiotemporal information from them, which can in turn benefit bike sharing demand prediction. In this way, the loss function is defined as:

$$L(\theta) = \|\hat{X}_b^{t+1} - X_b^{t+1}\| + \epsilon_r \left(\sum_{m \in \{s, h\}} \|\hat{X}_m^{t+1} - X_m^{t+1}\| \right) \quad (7)$$

where $\hat{X}_m^{t+1}, X_m^{t+1}$ are the predicted and true demand values of mode m at time step $t+1$, ϵ_r is a pre-determined weight for the prediction-based regularization term.

V. EXPERIMENTS

A. Data Description

Our proposed model is validated on real-world public datasets from NYC. In this study, we use Manhattan as the research area and collect travel demand data of bike sharing, subway and ride-hailing from 2018-03-01 to 2018-08-31. Specifically, we use the following three datasets: (1) **NYC Citi Bike**: the data consists of the pick-up and drop-off time and station of each Citi Bike trip records. During our study period, there are around 36 thousand trip records per day. We filter out the stations with an average of fewer than three orders per hour and keep 246 BSS stations for demand prediction. (2) **NYC Subway**: the data contains the

entries and exists counts of each turnstile in subway stations every four hours, with around 2.4 million entry/exist counts every day on average. We filter out subway stations with no demand for long time periods, which results in 107 subway stations. (3) **NYC Ride-hailing**: we use the for-hire vehicle (FHV) data from NYC Taxi & Limousine Commission (TLC), which is provided by ride-hailing companies such as Uber and Lyft. It contains the pick-up and drop-off time and zone of each trip records during the study period. On average there are 234 thousand trips per day during our study period. The zones are pre-determined by TLC and there are 63 TLC zones in Manhattan.

B. Experiment settings

The demands of all three modes are aligned into 4-hour intervals and min-max normalization is used for each mode to mitigate the effect of demand variance. Data from the first 60% time steps are used for model training, the following 20% for validation, and the last 20% for model testing. We set the historical time step $T = 6$, the number of training epochs $E = 500$, the learning rate of 0.002, the batch size of 32 and the dropout ratio of 0.3. To prevent overfitting, we use early stopping on the validation set and a L2 regularization on the loss function with a weight decay of $1e-5$. The prediction-based regularization weight ϵ_r is tuned from 0 to 0.3, and we find that our model achieves the best prediction performance for bike sharing demand when $\epsilon_r = 0.2$, verifying the effectiveness of jointly optimizing multimodal demand prediction for model training.

The following models are used as baselines for benchmarking: (1) **HA**: a statistical method that makes predictions based on the historical average; (2) **LR**: a regression model that captures the linear relationship between historical and future demand patterns; (3) **XGBoost**: a gradient boosting machine model to uncover nonlinear relationships; (4) **LSTM**: a recurrent-based model to capture long- and short-term temporal dependencies in demand sequences; (5) **STGCN** [17]: a convolutional framework using GCNs for spatial correlations and TCNs for temporal dependencies; (6) **MGCN** [18]: a multi-graph convolution network that captures multiple types of spatial correlations with different GCNs; (7) **Graph WaveNet** [19]: a graph learning approach using node embedding to learn an adaptive adjacency matrix. For fair comparison, we use the same experiment settings for all models, and evaluate them using three widely used metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2). We run 10 independent experiments for each model and report the average values on the test set.

VI. RESULTS

A. Comparison of Model Performance

The performance of different models are summarized in Table I. Compared with the baseline models, our proposed model achieves significantly superior performance for all evaluation metrics. This is likely because our model can

directly leverage spatiotemporal information of related subway stations and ride-hailing zones to enhance the prediction of bike sharing demand. In our model, the spatial effects across modes are encoded with inter-modal graphs. Figure 4 presents the performance variance of 10 independent runs of our proposed model and several selected baselines. We can find that our proposed model has the best performance regarding both RMSE and MAE in all experiments with relatively high stability.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT MODELS

Models	RMSE	MAE	R^2
HA	26.832	19.322	0.167
LR	16.750	10.431	0.676
XGBoost	15.139	9.262	0.734
LSTM	16.148	10.255	0.695
STGCN	13.820	9.097	0.789
MGCN	14.436	9.359	0.760
Graph WaveNet	12.689	8.092	0.814
B-MRGNN	11.598	7.270	0.845

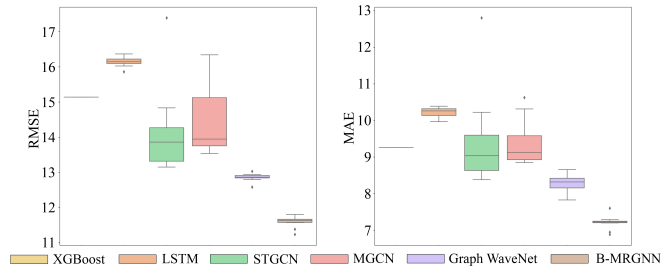


Fig. 4. Comparison of model stability

Among baseline models, the performance of HA is relatively poor, indicating the low temporal regularity of bike sharing demand patterns. This is reasonable as people use bikes for different purposes (e.g. leisure) that are less regular (compared to commuting). LR performs worse than other machine learning models, suggesting the necessity of capturing nonlinear relationships in demand series. XGBoost performs notably better than LSTM, showing the advantage of gradient boosting machines in some cases. The poor performance of LSTM can be potentially explained by its inability to capture spatial dependencies between BSS stations. Although MGCN encodes multiple types of spatial dependencies, it performs worse than STGCN in our experiments, which might be due to its low model stability as illustrated in Figure 4. Graph WaveNet achieves the best performance among baseline models given its ability to learn spatial dependencies hidden in demand data with an adaptive adjacency matrix. Compared with Grave WaveNet, our proposed model can further reduce the prediction error, with RMSE and MAE improvement of 8.6% and 10.2% respectively.

B. The effect of input modes

To further investigate the effect of incorporating inter-modal relationships on bike sharing demand prediction, we compare the performance of our model using different input mode combinations: bike sharing alone, bike sharing and subway, bike sharing and ride-hailing, and all three modes. The results are shown in Table II. Note that they share the same model structure, just with different input modes. We find that the inter-modal relations from geographically nearby or semantically similar subway stations and ride-hailing zones can indeed help with the prediction of bike sharing demand: without any inter-modal relations, the prediction error regarding RMSE and MAE would increase by 8.0% and 8.5% respectively. Incorporating either subway or ride-hailing demand patterns can already significantly improve the prediction performance of bike sharing demand, with the RMSE reduced by 6.2% and 7.2% respectively. It is also worth noting that with only bike sharing demand as input, our proposed model can still perform better than Graph WaveNet, suggesting that our spatiotemporal framework can better extract intra-modal relationships.

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT MODE COMBINATIONS

Models	RMSE	MAE	R^2
Bike sharing	12.528	7.889	0.821
Bike sharing + subway	11.754	7.391	0.843
Bike sharing + ride-hailing	11.630	7.275	0.845
Bike sharing + subway + ride-hailing	11.598	7.270	0.845

VII. CONCLUSIONS

This paper proposes a bike sharing demand prediction approach that allows for information sharing across modes. Specifically, we adapt a multi-relational spatiotemporal graph neural network approach to capture the complex spatiotemporal correlations among heterogeneous spatial units from different modes. The spatial dependencies across modes are encoded with multiple intra- and inter-modal graphs, and an inter-modal graph convolution layer is introduced to capture both correlations and difference between nodes from different modes. To integrate dependencies from spatial and temporal domains, we further incorporate MRGNN layers with temporal convolution networks in stacked ST-MR blocks. Empirical validation is performed on real-world bike sharing, subway and ride-hailing datasets from NYC. The results show that our proposed model achieves the best performance compared to existing methods, suggesting that the knowledge of subway and ride-hailing demand can indeed benefit the demand prediction of bike sharing. In future works, this research can be further improved by examining how the effects of cross-mode information vary over space and time.

REFERENCES

[1] J. Liu, L. Sun, W. Chen, and H. Xiong, "Rebalancing bike sharing systems: A multi-source data smart optimization," in *Proceedings*

of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1005–1014.

[2] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets," *Transportation research part C: emerging technologies*, vol. 95, pp. 47–60, 2018.

[3] K. Jin, W. Wang, S. Li, P. Liu, and H. Sun, "Dockless shared-bike demand prediction with temporal convolutional networks," in *CICTP 2020*, 2020, pp. 2851–2863.

[4] C. Zhang, L. Zhang, Y. Liu, and X. Yang, "Short-term prediction of bike-sharing usage considering public transport: A lstm approach," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 1564–1571.

[5] Y. Song and Y. Huang, "Investigating complementary and competitive relationships between bikeshare service and public transit: a spatial-temporal framework," *Transportation research record*, vol. 2674, no. 1, pp. 260–271, 2020.

[6] T. Gu, I. Kim, and G. Currie, "Measuring immediate impacts of a new mass transit system on an existing bike-share system in china," *Transportation research part A: policy and practice*, vol. 124, pp. 20–39, 2019.

[7] J.-H. Cho, S. W. Ham, and D.-K. Kim, "Enhancing the accuracy of peak hourly demand in bike-sharing systems using a graph convolutional network with public transit usage data," *Transportation Research Record*, vol. 2675, no. 10, pp. 554–565, 2021.

[8] J. Ye, L. Sun, B. Du, Y. Fu, X. Tong, and H. Xiong, "Co-prediction of multiple transportation demands based on deep spatio-temporal neural network," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 305–313.

[9] Y. Wang, Z. Zhan, Y. Mi, A. Sobhani, and H. Zhou, "Nonlinear effects of factors on dockless bike-sharing usage considering grid-based spatiotemporal heterogeneity," *Transportation Research Part D: Transport and Environment*, vol. 104, p. 103194, 2022.

[10] Y. Liang, G. Huang, and Z. Zhao, "Joint demand prediction for multimodal systems: A multi-task multi-relational spatiotemporal graph neural network approach," *arXiv preprint arXiv:2112.08078*, 2021.

[11] J. W. Yoon, F. Pinelli, and F. Calabrese, "Cityride: a predictive bike sharing journey advisor," in *2012 IEEE 13th International Conference on Mobile Data Management*. IEEE, 2012, pp. 306–311.

[12] C. Rudloff and B. Lackner, "Modeling demand for bikesharing systems: neighboring stations as source for demand and reason for structural breaks," *Transportation Research Record*, vol. 2430, no. 1, pp. 1–11, 2014.

[13] S. Guidon, D. J. Reck, and K. Axhausen, "Expanding a (n)(electric) bicycle-sharing system to a new city: Prediction of demand with spatial regression and random forests," *Journal of Transport Geography*, vol. 84, p. 102692, 2020.

[14] S. Feng, H. Chen, C. Du, J. Li, and N. Jing, "A hierarchical demand prediction method with station clustering for bike sharing system," in *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*. IEEE, 2018, pp. 829–836.

[15] X. Zhou, Y. Shen, Y. Zhu, and L. Huang, "Predicting multi-step citywide passenger demands using attention-based neural networks," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 2018, pp. 736–744.

[16] X. Li, Y. Xu, Q. Chen, L. Wang, X. Zhang, and W. Shi, "Short-term forecast of bicycle usage in bike sharing systems: a spatial-temporal memory network," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[17] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.

[18] X. Geng, Y. Li, L. Wang, L. Zhang, Q. Yang, J. Ye, and Y. Liu, "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 3656–3663.

[19] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," *arXiv preprint arXiv:1906.00121*, 2019.

[20] Q. Zhou, J. Gu, X. Lu, F. Zhuang, Y. Zhao, Q. Wang, and X. Zhang, "Modeling heterogeneous relations across multiple modes for potential crowd flow prediction," *arXiv preprint arXiv:2101.06954*, 2021.