A Graph Deep Learning Model for Station Ridership Prediction in Expanding Metro Networks

A Graph Deep Learning Model for Station Ridership Prediction in Expanding Metro Networks

Fangyi Ding
The University of Hong Kong
Hong Kong SAR, China
fyding@connect.hku.hk

Yuebing Liang
The Singapore-MIT Alliance for Research and Technology (SMART)
Singapore, Singapore
ybliang@mit.edu

Yamin Wang
The University of Hong Kong
Hong Kong SAR, China
u3603952@connect.hku.hk

Yan Tang
Tongji University
Shanghai, China
2010763@tongji.edu.cn

Yang Zhou
The University of Hong Kong
Hong Kong SAR, China
zhouyhk@hku.hk

Zhan Zhao*
The University of Hong Kong
Hong Kong SAR, China
zhanzhao@hku.hk

Abstract

Due to their reliability, efficiency, and environmental friendliness, metro systems have become a crucial solution to transportation challenges associated with urbanization. Many countries have constructed or expanded their metro networks over the past decades. During the planning stage, accurately predicting station ridership post-expansion, particularly for new stations, is essential to enhance the effectiveness of infrastructure investments. However, station-level metro ridership prediction under expansion scenarios (MRP-E) has not been thoroughly explored, as most advanced models currently focus on short-term predictions. MRP-E presents significant challenges due to the absence of historical data for newly built stations and the dynamic, complex spatiotemporal relationships between stations during expansion phases. In this study, we propose a Metro-specific Multi-Graph Attention Network model (Metro-MGAT) to address these issues. Our model leverages multi-sourced urban context data and network topology information to generate station features. Multi-relation graphs are constructed to capture the spatial correlations between stations, and an attention mechanism is employed to facilitate graph encoding. The model has been evaluated through realistic experiments using multi-year metro ridership data from Shanghai, China. The results validate the superior performance of our approach compared to existing methods, particularly in predicting ridership at new stations.

CCS Concepts

• **Applied computing** → **Transportation**; *Forecasting*.

*Corresponding author

UrbanAI'24, October 29–November 01, 2024, Atlanta, GA

ACM ISBN 979-8-4007-1156-5/24/10
https://doi.org/10.1145/3681780.3697247

Keywords

Ridership Prediction, Metro Expansion, Graph Attention Network, Transport Planning, Urban Development

ACM Reference Format:
Fangyi Ding, Yuebing Liang, Yamin Wang, Yan Tang, Yang Zhou, and Zhan Zhao. 2024. A Graph Deep Learning Model for Station Ridership Prediction in Expanding Metro Networks. In *2nd ACM SIGSPATIAL International Workshop on Advances in Urban-AI (UrbanAI'24), October 29-November 1 2024, Atlanta, GA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3681780.3697247

1 Introduction

Recent decades have witnessed accelerating urbanization, marked by a significant migration from rural to urban areas, especially in developing countries. While urbanization has propelled substantial global advancements, it has also introduced a series of transportation challenges such as traffic congestion and air pollution. Metro systems have emerged as a key solution to these issues, offering a reliable, efficient, and environmentally friendly mode of urban transportation to support burgeoning populations and economic activities. In light of these benefits, metro systems have seen a significant surge in popularity and have flourished in numerous countries globally. During the planning stage of metro expansion, accurately predicting station ridership (e.g., inflow and outflow) after expansion, particularly for new stations, is crucial for enhancing the effectiveness of infrastructure investments and the sustainability of the urban metro system. In this study, we define this task as the Station-Level Metro Ridership Prediction under Expansion Scenario (MRP-E) problem.

Metro network expansion generally includes planning new metro lines or extending existing ones. In the expansion scenario, stations can be divided into three types, as illustrated in Fig. 1. Newly built stations are those not present before the expansion but added to the metro system afterward. Updated stations are those that existed before the expansion but have new lines passing through or extending after the expansion. Existing stations are those that were part of the network before the expansion and have no new lines passing through or extending after the expansion. Among these three types, new stations have no historical ridership data before expansion. Although existing and updated stations have historical ridership

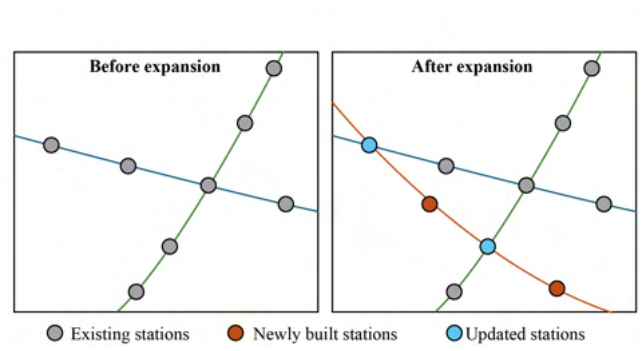

Figure 1: Schematic diagram of station types in the network expansion scenario.

data, their ridership dynamics are affected differently during the expansion (this will be elaborated in Sec. 5.1).

The MRP-E task can be viewed as a crucial precursor to metro network expansion design. Effective expansion design requires accurate demand predictions for new stations during the planning year. While previous studies have made significant contributions to metro network expansion design using optimization models and reinforcement learning algorithms [22, 27, 28, 33], they often overlook the demand prediction step. Many studies either use the current year's all-mode travel demand as a proxy for metro demand in the planning year or rely on unvalidated estimated demands. This approach can compromise the effectiveness of design models, as metro demand is not equivalent to all-mode travel demand, and actual metro demand in the planning year may differ significantly from current ridership figures.

Traditional methods for estimating metro ridership, whether through travel demand modeling [19, 21] or regression models [2, 12], either involve substantial costs and produce approximate estimates, or fail to capture more complex patterns in the data. While significant efforts have been made to apply deep learning methods for metro ridership prediction [4, 29], these methods predominantly target short-term prediction scenarios (e.g., 15 minutes) and assume an unchanged network structure. The potential of artificial intelligence (AI) methods to address the MRP-E problem merits in-depth exploration.

In summary, the MRP-E task is challenging for several reasons. First, under the expansion scenario, the metro network is dynamically evolving, and the spatial interaction between stations can change accordingly. Therefore, deep learning models designed for short-term ridership prediction may not be applicable to the MRP-E task. Secondly, the demand at a given station varies but remains correlated over different times. Previous research primarily focuses on capturing these temporal relationships to forecast future demand. However, these models depend on the availability of historical demand data. In the case of newly built stations within the MRP-E task, where historical data is absent, potential demand must be inferred from surrounding urban contexts and other relevant factors. Third, metro expansion is not a frequent event; new stations and updated stations are much fewer than existing stations, especially in mature metro systems. This limits our ability to obtain sufficient new samples for model training, leading to high performance in predicting existing stations but poor performance in predicting new and updated stations.

In this research, we propose a Metro-specific Multi-Graph Attention Network model (Metro-MGAT) to address the MRP-E problem. Utilizing various urban context data and multi-year metro network data, we extract station-based spatiotemporal features. Then, to capture the complex spatial dependencies, we construct multiple metro-specific correlation graphs. The spatiotemporal features and spatial dependency features are fed into a prediction network to generate predicted station demand. The main contributions of this paper are summarized as follows:

- We develop a Metro-specific Multi-Graph Attention Network model (Metro-MGAT) to address the metro station ridership prediction challenge under expansion scenarios. To the best of our knowledge, this work presents the first attempt to predict expanded metro station demand using deep learning algorithms.
- We design multiple graphs specific to metro systems to encode the complex spatial dependencies among stations. Specifically, a geographical distance graph, a functional similarity graph, and a travel impedance graph are incorporated into the model, and a graph attention mechanism is adopted to facilitate the learning of spatial interaction features.
- To address the class imbalance issue (i.e., new stations are much less numerous than existing stations) in the MRP-E task, we design an age-weighted loss function. This gives "younger" stations, typically those newly built, higher priority, enhancing the model's performance for new stations.

2 Related Works

2.1 Traditional Metro Ridership Prediction Methods

The traditional approach to metro ridership prediction has predominantly utilized four-step methods [11, 19] or activity-based models [1, 21], which are part of the broader travel demand modeling framework. In this framework, metro ridership is viewed as the demand for a specific mode of transportation. While these methods are behaviorally consistent, they require extensive data and a comprehensive calibration of explanatory variables.

As alternatives, several effective models have been proposed to enhance the detail and accuracy of ridership predictions. Time series models, extensively applied in forecasting transportation demand, predict ridership based on regression analyses of past values [7, 34]. Direct ridership models, such as Ordinary Least Squares (OLS) regression and Geographically Weighted Regression (GWR), directly link ridership to accessible factors like local demographics, economic indices, and geographic information [2, 3, 12]. While these statistical methods provide good interpretability, they often struggle to capture the nonlinear dynamics of traffic data and may falter under complex conditions and large datasets.

Additionally, machine learning techniques have been explored for ridership prediction. For example, [26] employed the XGBoost model to predict travel demand in expanding metros. However, existing models have not been deployed on real-world expansion datasets, which may fail to capture the complex spatial interactions

between stations during expansion. Consequently, there is a pressing need to develop more accurate models that can realistically represent the planning and deployment of real-world metro system expansions over time.

2.2 Deep Learning Methods in Ridership Prediction

To address the limitations of traditional forecasting methods, some researchers have turned to neural networks (NN) and hybrid models to enhance demand forecasting [20, 24]. Neural networks excel at handling complex nonlinear problems without requiring prior knowledge of the relationships between input and output variables. With the rapid advancements in AI, an increasing number of deep learning algorithms are being utilized to predict metro ridership [17, 29]. Prominent models include Long Short-Term Memory (LSTM) [9] and Gated Recurrent Unit (GRU) [8, 30]. Furthermore, spatial dependency has been integrated into models, employing Convolutional Neural Networks (CNNs) and their hybrid counterparts to explore spatial correlations and achieve network-scale passenger predictions [18, 32].

However, as these CNN-based models often represent metro ridership as grid-based data, there are concerns that this approach may not provide satisfactory accuracy due to the neglect of realistic network topology [4], especially since metro stations are sparsely distributed across urban areas, potentially leading to inadequate capture of representative spatiotemporal patterns [4, 15, 29]. Recently, graph-based models such as the Graph Convolutional Network (GCN) have gained popularity for their ability to consider the topological information of the network [4, 31].

Despite these advancements, most AI applications have focused on short-term prediction scenarios. Over long planning horizons, the evolving nature of metro networks due to expansion can significantly influence ridership distribution, while short-term deep learning prediction models often assume an unchanged network structure. Additionally, although these studies demonstrate the effectiveness of deep learning models for metro demand prediction, they generally rely on sequential/temporal dependencies, which are not applicable for metro expansion scenarios where new stations lack historical data.

Although some recent studies have addressed demand prediction problems in long-term expansion scenarios [14, 16], they predominantly focus on bike-sharing and electric vehicles. Metro expansion differs from these modes in several key ways: 1) the spatial interaction between stations in metro systems is more complex due to the distribution of metro lines; 2) metro expansions are rarer, thus exacerbating the new/existing station imbalance problem; and 3) in metro expansion, there are new/existing stations as well as updated stations, each showing distinct dynamics. To the best of our knowledge, very limited efforts have been devoted to exploring deep learning for demand forecasting in metro expansion scenarios. Given the growing interest in applying deep learning-based methods to future urban and transportation planning, there is a compelling need to further investigate these advanced methods for planning metro systems of the future.

3 Problem Formulation

3.1 Preliminaries

3.1.1 Metro Station Demand. For practical purposes, metro station demand under expansion scenarios does not always necessitate fine-grained forecasting such as hourly or daily levels as is typical in short-term prediction models. Instead, the goal is to capture the total passenger flow at relatively macroscopic time scales for newly planned and updated stations to ensure a reasonable allocation of resources. In this study, to capture the seasonal dynamics, metro station demand is observed at the monthly level. For station i, we aim to predict its demand $d_{i,t}$ at month t as the average daily inflow and outflow. This can be computed as $d_{i,t}^{in/out} = \sum_{m=1}^{M} d_{i,t,m}^{in/out}/M$, where $d_{i,t,m}^{in/out}$ represents the daily inflow and outflow of station i on day m of month t, and M is the total number of days in month t.

3.1.2 Multi-relation Graph. We model the urban metro network as a weighted undirected relational graph. Due to metro expansion, the structure of the metro network evolves dynamically over time, which can be denoted as $G = \{V, E, W, X, D, t\}$, where V refers to the metro stations at month t, E denotes the relationships between stations at month t, W represents the weights of the edges at time t, X indicates the features of stations at month t and D signifies the demand at stations for month t. In this study, we define multiple graphs according to different relationships between stations, which will be introduced in Sec. 4.2.

3.2 Metro Expansion Demand Prediction Problem

Suppose month p marks the transition from the training to the testing period. Given the metro network structure and station features from the earlier training period $\{1, 2, ...p-1\}$, this study aims to develop a mapping function F that correlates station demand with station features, denoted as $D_{i,t}^{in/out} = F(X_{i,t}^{in/out})$. The training period encompasses several metro expansions, enabling this function to capture dynamic mapping relationships under expansion scenarios.

Subsequently, using the anticipated metro network structure and current station features, we apply the mapping function to future expansion scenarios in the testing period $\{p, p+1, ..., P\}$ to generate demand predictions $\hat{D}_{i,t}^{in/out} = F(X_{i,t}^{in/out})$. It is important to note that stations will have their features based solely on the urban environment, network structure, and temporal information, rather than historical demand.

3.3 Framework

Fig. 2 presents an overview of the proposed framework for predicting demand in metro expansion scenarios, which comprises three main components. First, we utilize multi-sourced urban information to extract station-based urban context features. Additionally, temporal factors that influence station demand dynamics are encoded to capture the station-level temporal information. Detailed features will be presented in Sec. 4.1. Second, to capture the spatial correlations between stations, we construct multiple relational graphs to encode various spatial dependency relationships, which

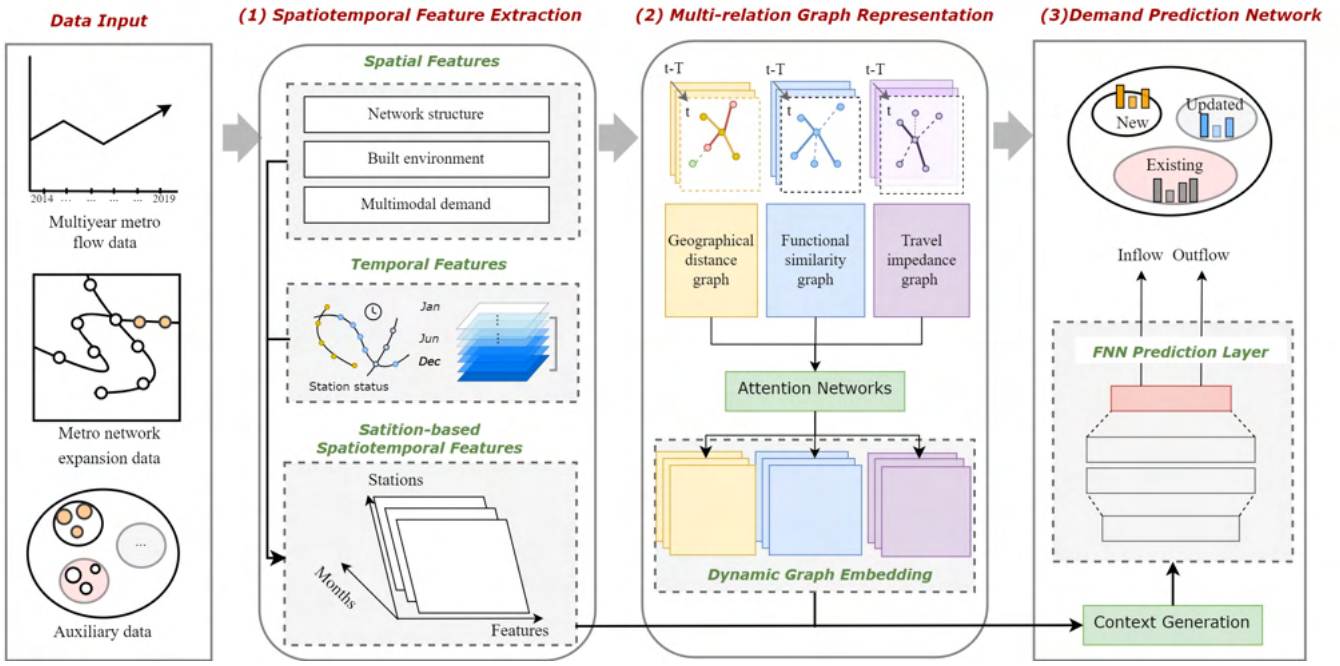

Figure 2: Overview of the proposed framework for expanded demand prediction.

will be detailed in Sec. 4.2. Third, utilizing the spatiotemporal encodings and spatial interaction features, we integrate these features and input them into the prediction network, generating the anticipated demand for newly-built, existing, and updated stations. This component is outlined in Sec. 4.3.

4 Methodology

4.1 Spatiotemporal Feature Extraction

4.1.1 Spatial Features. We utilize diverse urban and metro network information to extract station-based spatial features. These primarily include features related to the metro network structure, the built environment, and multimodal demand. A detailed list of these features is provided in Tab. 1.

Network structure feature are extracted from the metro network's topology, denoted as $x^s_{i,t} \in \mathbb{R}^{12}$. This 12-dimensional vector includes features such as station centrality, the number of connected stations, and the number of connected metro lines. Due to the dynamically evolving topology, we extract these structure features for every observation month t.

Built environment features have been found to be highly correlated with metro flows in previous studies. In this study, we extract a 21-dimensional built environment feature vector $x^b_i \in \mathbb{R}^{21}$. These features, calculated within a 500-meter buffer around stations, primarily include population density [6], house value [1], and the density of 14 categories of Points of Interest (POIs) [2]. Due to data availability, the built environment data is static and was primarily recorded in 2017.

[1]Extracted from *Lianjia*, one of China's largest real estate intermediary companies. https://m.lianjia.com/
[2]Obtained from the Gaode API: https://lbs.amap.com/

Table 1: Summary of features and categories.

Features
Spatial features (dim=37)
1) Network structure (dim=12)
station centralities, metro station connections, metro line connections, interchange station, terminal station…
2) Built environment (dim=21)
population density, house value, POI density of 14 categories, road density, intersection density, landuse diversity…
3) Multimodal demand (dim=4)
taxi pick-ups, taxi drop-offs, bus stop density, bus line density
Temporal features (dim=14)
1) Station status (dim=2)
station age, updated status
2) Month (dim=d_m)
observation month embedding

Multimodal demand can potentially reflect metro demand. For instance, buses and taxis can either compete with or complement metro systems, especially over long distances. Hence, these modes may exhibit complex relationships with metro usage. In this study, we incorporate bus and taxi demand into the model. Taxi demand is calculated as the average daily number of taxi pickups and drop-offs within a 500-meter radius of metro stations. Bus demand is represented by the density of bus stops and bus lines within the same radius. We combine these as the multimodal demand feature vector $x^m_i \in \mathbb{R}^4$, with taxi data specifically collected from Shanghai between April 1st and April 30th, 2015.

Based on the above features, the spatial feature vector for station i at month t can be represented as:

$$x_{i,t}^{P} = [x_{i,t}^{s} \parallel x_{i}^{b} \parallel x_{i}^{m}], \tag{1}$$

where $\parallel$ denotes the concatenation function.

4.1.2 Temporal Features. We include two types of features to encode temporal information: station status and month. The former includes the station's age $x_{i,t}^{a} \in \mathbb{R}$ (the number of months since it first operated) and update status $x_{i,t}^{u} \in \mathbb{R}$ (whether it has had new lines pass through within the past year). The latter feature records the month of observation, which is transformed into a month embedding vector $X_{i,t}^{m} \in \mathbb{R}^{d_m}$, where d_m is the vector dimension. Each month is represented as a unique vector to capture seasonal fluctuations. The temporal features of station i at month t can be represented as:

$$x_{i,t}^{T} = [x_{i,t}^{a} \parallel x_{i,t}^{u} \parallel x_{i,t}^{m}] \tag{2}$$

4.2 Multi-relation Graph Representation

4.2.1 Multiple Graph Construction. The demand at one station can be influenced by other stations within the metro system. Rather than solely relying on physical topologies or geographic proximity, we construct multiple graphs specifically tailored for metro systems to capture the complex spatial dependencies among stations.

Geographical Distance Graph: Generally, stations that are geographically closer tend to be more correlated. We define a distance graph to model this spatial correlation. In this graph, the geographical proximity between station i and j at month t is computed as:

$$a_{ij,t}^{g} = \exp(-(\frac{dist_{ij,t}}{\sigma_g})^2), \tag{3}$$

where $a_{ij,t}^{g}$ is the geographical adjacency weight between station i and j in month t, $dist_{ij,t}$ denotes their geographical distance, and σ_g is set as the standard deviations of $dist_{ij,t}$.

Functional similarity graph: Stations with similar distributions of land use functions may exhibit closer demand patterns. For example, if two stations are located within the same functional zone, the people arriving and departing from these stations are likely to have similar activity patterns, leading to more similar flow distributions. In this study, we use the built environment features surrounding each station to encode their land use function. The function similarity between a pair of stations $\{i, j\}$ is is computed as:

$$a_{ij,t}^{b} = \exp(-(\frac{Euc(x_i^b, x_j^b)}{\sigma_b})^2), \tag{4}$$

where $a_{ij,t}^{b}$ is the functional adjacency weight between station i and j in month t, $Euc(x_i^b, x_j^b)$ is the Euclidean distance function, x_i^b and x_j^b are built environment vectors described in Sec. 4.1, and σ_b is set as the standard deviations of $Euc_{x_i^b, x_j^b}$.

Travel impedance graph: For transportation systems such as bikesharing or taxis, travel cost is highly correlated with the distance between two locations. However, in metro systems, transfer times play a critical role in people's travel decision-making. Thus, station pairs with lower travel impedance may exhibit higher spatial interaction. The travel impedance proximity between a station pair $\{i, j\}$ is computed as:

$$a_{ij,t}^{c} = \exp(-(\frac{Imp_{ij,t}}{\sigma_c})^2), \tag{5}$$

$$Imp_{ij,t} = \min_{r}\{l_{ij,t}^{r} + \beta n_{ij,t}^{r}\}, \tag{6}$$

where $a_{ij,t}^{c}$ is the impedance-based adjacency weight between station i and j in month t. $Imp_{ij,t}$ is the impedance function, which factors in network distance and transfer times. $l_{ij,t}^{r}$ is the number of link segments in route r between station pair $\{i, j\}$ at month t, $n_{ij,t}^{r}$ denotes the number of transfers in route r between station pair $\{i, j\}$ at month t, β is transfer penalty parameter, and σ_c is set as the standard deviations of $Imp_{ij,t}$.

Based on the above three types of adjacency matrices, we select the top-K nearest stations for station i as its neighborhood stations for each category and construct three graphs for each month, denoted as $G_{i,t}^{g}$, $G_{i,t}^{b}$, and $G_{i,t}^{c}$, respectively. Note that for the same station i, its associated graphs may change in different months due to the addition of new stations.

4.2.2 Graph Attention Network Encoding. Based on the constructed graphs $G_{i,t}^{g}$, $G_{i,t}^{b}$, and $G_{i,t}^{c}$, we can extract the spatial features of neighboring stations for each station and fuse them through a graph attention mechanism to generate spatial interaction features $h_{i,t}^{g}$, $h_{i,t}^{b}$, and $h_{i,t}^{c}$. Given the input spatial feature vector of station i $x_{i,t}^{P}$, the spatial interaction feature is computed as the weighted sum of its neighbors' spatial feature vectors $x_{j,t}^{P}$. Taking the geographical proximity graph $G_{i,t}^{g}$ as an example, the geographical interaction feature vector $h_{i,t}^{g}$ can be represented as:

$$h_{i,t}^{g} = \textstyle\sum_{j \in N_{i,t}^{g}} \alpha_{ij,t} W x_{j,t}^{P}, \tag{7}$$

$$\alpha_{ij,t} = \frac{exp(Attn(Wx_{i,t}^{P} \parallel Wx_{j,t}^{P}))}{\sum_{k \in N_{i,t}^{g}} exp(Attn(Wx_{i,t}^{P} \parallel Wx_{k,t}^{P}))}, \tag{8}$$

where $N_{i,t}^{g}$ denotes the geographical neighborhood station set of station i at month t, W is a shared parameter matrix used to perform a linear transformation on all spatial feature vectors. $\alpha_{ij,t}$ is the normalized attention weight between station i and its neighboring station j at month t, and $Attn(\cdot)$ is a two-layer feed-forward network used to generate pairwise attention scores.

4.3 Demand Prediction Network

4.3.1 FNN Prediction Layer. In the prediction network, the extracted spatiotemporal features along with the learned spatial interaction features are fed into the network, generating the inflow and outflow predictions. In this study, we employ a feed-forward network (FNN) as the prediction layer. Specifically, for a target station i at month t, the prediction network performs the following computations:

$$
\begin{aligned}
z_1 &= ReLU(W_1([x^P_{i,t} \parallel x^T_{i,t} \parallel h^g_{i,t} \parallel h^b_{i,t} \parallel h^c_{i,t}])),\\
z_2 &= ReLU(W_2 z_1),\\
z_3 &= sigmoid(W_3 z_2),\\
\hat{d}_{i,t} &= W_4 z_3,
\end{aligned} \tag{9}
$$

where $x^P_{i,t}$, $x^T_{i,t}$, $h^g_{i,t}$, $h^b_{i,t}$, and $h^c_{i,t}$ represent the spatial features, temporal features, and three types of spatial interaction features, respectively. W_1, W_2, W_3, and W_4 are learnable parameter matrices. $\hat{d}_{i,t} \in \mathbb{R}^2$ is the predicted inflow and outflow demand of station i at month t.

4.3.2 Age-weighted Loss Function. Typically, the loss function is designed as the sum of squared errors between the predicted and real-world demand. However, in this model, we prioritize the prediction of newly built stations. Specifically, we design a weighted loss function that adjusts the observation weight based on the station's opening age. The loss function is calculated as follows:

$$
\begin{aligned}
L_\theta &= \sum_{t=1}^{T} \sum_{i=1}^{N_t} \gamma_{i,t} (\hat{d}_{i,t} - d_{i,t})^2,\\
\gamma_{i,t} &= \frac{1}{x^a_{i,t}+1},
\end{aligned} \tag{10}
$$

where T is the number of months in the training period, N_t is the number of stations at month t, and $\gamma_{i,t}$ is the sample weight, which is inversely proportional to the station age $x^a_{i,t}$.

5 Experiments

5.1 Data Analysis

In this study, we utilize Shanghai as the case city. The Shanghai metro ridership dataset spans daily passenger inflow and outflow for each station from January 2014 to December 2019. Due to a technical issue, data from April 2014 to October 2014 are missing. Fig. 3 (b) illustrates the number of stations and the average daily ridership for each month from December 2014 to December 2019. It is evident that the Shanghai metro network underwent several expansions during this period, resulting in a general increase in system-wide average ridership. Fig. 3 (a) depicts the network topology in December of 2014, 2017, and 2019. Purple stations represent existing stations, red stations denote newly built stations since the previous stage, and blue stations indicate updated stations that were passed by new lines or received line extensions compared to the previous stage.

The three types of stations exhibit distinct flow dynamics due to the expansion. For example, consider the expansion in December 2015, as illustrated in Fig. 3 (c), where we track the ridership dynamics from January 2015 (one year before expansion) to December 2016 (one year after expansion) for the three types of stations. The flows of 20 sampled existing stations, all new stations, and all updated stations in this expansion are illustrated in this subfigure. The flow at newly built stations showed a significant increasing trend since their opening in December 2015. In contrast, the flow at existing stations remained relatively stable, with only seasonal fluctuations evident. Updated stations, compared to existing ones, displayed more pronounced fluctuations; particularly in December 2015, when the updated stations were first served by new lines or had extensions, some updated stations experienced a noticeable initial surge in ridership followed by a sudden drop. On the other hand, contemporaneous existing stations did not exhibit an increasing trend but rather a decline leading up to the Spring Festival holidays. This suggests that the expansion impacts different types of stations differently: new stations experience a significant and sustained increase in traffic, updated stations show a relatively minor and short-term growth trend, while existing stations are largely unaffected.

5.2 Experiment Settings

In this study, each station-month pair is treated as an individual observation. To capture the dynamics of metro expansion, we utilize data from January 2014 to June 2017 for training and validation, and data from July 2017 to December 2019 for testing. During the training and validation period, 303 stations were operational, while the testing period saw the addition of 31 new stations and updates to 10 stations (i.e., these stations had new lines introduced or line extensions during the testing period).

As discussed in Sec. 5.1, the flow fluctuations exhibit differences between new and updated stations under the influence of expansion. For newly built stations in the testing period, all observations from their months of operation are treated as new observations. In contrast, for updated stations in the testing period, due to their relatively minor fluctuations resulting from the expansion, only observations within one year of the update are considered new observations. All other station-month observations in the testing period are classified as existing observations. Consequently, there are 10,347 observations in the training and validation period, and 9,883 observations in the testing period, which includes 8,960 existing observations and 923 new observations.

5.3 Baseline Models

We evaluate our Metro-MGAT model against five baseline models, which include three non-deep learning models and two advanced deep learning models. The details of these models are as follows:

Linear Regression [25] is a commonly used regression model that assumes linear relationships between metro station demand and influencing factors.

Ridge Regression [10] addresses the issue of multicollinearity in linear regression by incorporating L2 regularization. The regularization strength is controlled by the parameter α, which we set at 0.01.

XGBoost [5] is a tree-based ensemble learning algorithm. It utilizes multiple regression trees in a boosting framework to improve prediction accuracy.

Feed Forward Network (FNN) [23] typically consists of an input layer, several hidden layers, and an output layer. Unlike graph-based models, FNNs do not inherently capture spatial relationships. To ensure a fair comparison, we use the prediction layer from our model as the benchmark for the FNN.

Multi-graph Convolutional Network (MGCN) [13] is designed specifically for graph data. It aggregates features from neighboring stations using predefined weights rather than an attention mechanism. For a fair comparison in this baseline model, we employ the same multiple graphs as those constructed in our Metro-MGAT model.

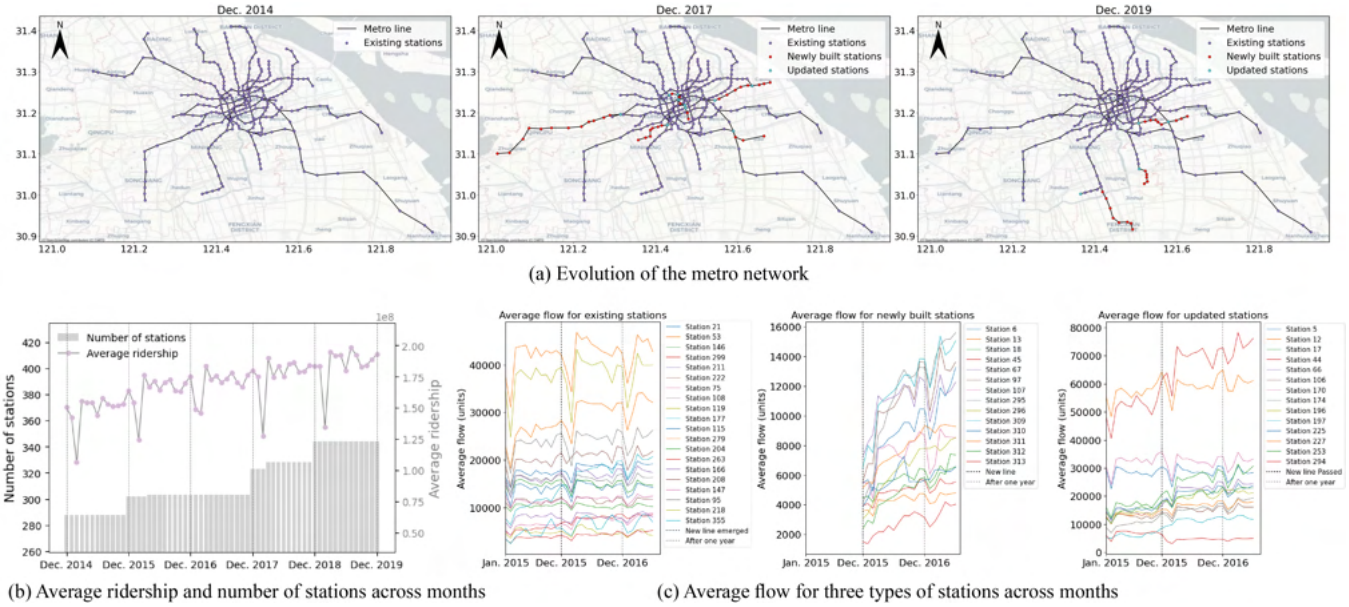

Figure 3: Spatiotemporal analysis of Shanghai metro data.

Spatially-dependent Multi-graph Attention Network (Spatial-MGAT) [14] is a graph neural network approach for predicting the station-level demand in bike-sharing system under expansion scenarios. This model constructs graphs based on geographical proximity and built environment similarity, and uses attention mechanisms to learn the correlation weights between bike stations.

The model performance is evaluated using four metrics: Root Mean Square Error (*RMSE*), Mean Absolute Error (*MAE*), Mean Absolute Percentage Error (*MAPE*), and the Coefficient of Determination (R^2), which are defined as follows:

$$
\begin{aligned}
RMSE &= \sqrt{\frac{1}{2\times\sum_{t=1}^{T} N_t}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{e=0}^{1}(d_{i,t}^{e}-\hat{d}_{i,t}^{e})^2},\\
MAE &= \frac{1}{2\times\sum_{t=1}^{T} N_t}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{e=0}^{1}\left|d_{i,t}^{e}-\hat{d}_{i,t}^{e}\right|,\\
MAPE &= \frac{100\%}{2\times\sum_{t=1}^{T} N_t}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{e=0}^{1}\frac{\left|d_{i,t}^{e}-\hat{d}_{i,t}^{e}\right|}{d_{i,t}^{e}},\\
R^2 &= 1-\frac{\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{e=0}^{1}(d_{i,t}^{e}-\hat{d}_{i,t}^{e})^2}{\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{e=0}^{1}(d_{i,t}^{e}-\bar{d})^2},
\end{aligned}
\tag{11}
$$

where T is the number of months in the test period, N_t is the number of stations in month m, $d_{i,t}^{e}$ and $\hat{d}_{i,t}^{e}$ are the true and predict demand of inflow/outflow (e=1 denotes inflow, and $e=0$ denotes outflow). $\bar{d}$ is the average demand.

6 Results

6.1 Performance Analysis

In this section, we compare the performance of the Metro-MGAT model with the aforementioned baseline models. For each model, we conduct experiments five times and report the average performance, as listed in Tab. 2. For new observations, our model outperforms all other models across all four metrics. Compared to the second-best model, FNN, our model reduces RMSE, MAE, and MAPE by 7.2%, 6.8% and 8.3%, respectively, while improving R^2 by 4.5%. This improvement may be attributed to our model's enhanced ability to leverage spatial knowledge compared to classical deep neural networks.

Linear regression and ridge regression exhibit similar performance, with XGBoost performing slightly better. Unexpectedly, MGCN performs poorly regarding R^2 in predicting new observations, which may be due to the predefined adjacency weights introducing biased spatial dependencies. In contrast, the graph attention mechanism in Metro-MGAT effectively captures the complex spatial relationships between stations.

For existing observations, XGBoost achieves the best performance across all four metrics, while our model generally ranks third. This is not surprising, as our model prioritizes new observations by assigning higher weights to newly built stations, which can lead to relatively lower performance in predicting existing demand. Prediction models for existing stations have been extensively explored in previous work and have achieved very high accuracy, making them less of a focus in this study.

6.2 Ablation Analysis

In this subsection, we conduct an ablation analysis to verify the effectiveness of key components in our model. Specifically, we create variant models by removing different components and compare their performance with the full Metro-MGAT model. The variant models include: ***-SS:*** Removing station status features, i.e., station age $x_{i,t}^{a}$ and update status $x_{i,t}^{u}$. ***-MD:*** Removing multimodal demand features x_i^m. ***-WL:*** Using an unweighted loss function, i.e., applying uniform weights to all samples. ***-IG:*** Removing the impedance graph. The performance of these variants is displayed in Tab. 3.

Table 2: Comparison of model performance.

	New observations				Existing observations			
Models	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
Linear regression	8.471e3	6.226e3	1.381	0.680	8.338e3	6.145e3	0.535	0.768
Ridge regression	8.480e3	6.211e3	1.375	0.678	8.335e3	6.141e3	0.536	0.768
XGBoost	7.712e3	3.833e3	0.536	0.698	***3.789e3***	***2.119e3***	***0.165***	***0.952***
FNN	3.584e3	2.618e3	0.460	0.709	5.710e3	3.857e3	0.305	0.902
MGCN	7.557e3	5.503e3	1.310	0.354	5.689e3	3.894e3	0.348	0.902
Spatial-MGAT	3.829e3	2.894e3	0.473	0.694	5.024e3	3.378e3	0.283	0.923
Metro-MGAT (our model)	***3.325e3***	***2.441e3***	***0.422***	***0.741***	5.068e3	3.449e3	0.323	0.919

Table 3: Performance of variant models.

	New observations				Existing observations			
Models	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
Metro-MGAT	***3.325e3***	***2.441e3***	***0.422***	***0.741***	5.068e3	3.449e3	0.323	0.919
-SS	3.810e3	2.820e3	0.469	0.631	5.692e3	3.456e3	0.266	0.909
-MD	3.736e3	2.812e3	0.474	0.696	5.390e3	3.696e3	0.338	0.908
-WL	3.565e3	2.691e3	0.438	0.734	***4.643e3***	***3.014e3***	***0.270***	***0.932***
-IG	3.432e3	2.520e3	0.451	0.713	5.657e3	3.735e3	0.347	0.902

The key components of our Metro-MGAT model are shown to be effective in predicting demand for new observations, as indicated by the lower performance across all four metrics when these components are removed. For existing observations, all components, except for the weighted loss function, also contribute significantly to improving model performance, which is expected given that the weighted loss function is specifically designed to prioritize new observations.

7 Conclusion and Discussion

In this study, we propose a Metro-specific Multi-Graph Attention Network (Metro-MGAT) model to address the challenge of station-level metro ridership prediction under expansion scenarios (MRP-E). MRP-E is a complex task due to the absence of historical demand data for newly built stations and the intricate relationship between urban context and station ridership. Additionally, the dynamically evolving network structure and the scarcity of new or updated stations further complicate the problem, rendering recent deep learning models designed for short-term prediction unsuitable. To tackle the MRP-E task, our Metro-MGAT model incorporates three key components. First, we extract station-based spatiotemporal features for each month from various urban data sources and network topology. Next, to capture the spatial dependencies between stations, we construct multiple specialized graphs for metro systems, including geographical distance graphs, functional similarity graphs, and travel impedance graphs. An attention mechanism is employed to effectively capture complex spatial correlations. All features are then fused and fed into a feed-forward network to generate potential metro demand predictions. To address the issue of observation imbalance, we design an age-weighted loss function. To validate our model and capture the complex impacts of network expansion, we conduct experiments on a real-world dataset from the Shanghai metro system, covering the period from 2014 to 2019. The results demonstrate the superiority of our proposed model in predicting demand for new observations and highlight the effectiveness of the model's components.

Additionally, we acknowledge the limitations of our work concerning the adequacy of the experiments and datasets. In the future, we plan to conduct further experiments to differentiate the performance of updated stations from that of newly built stations, and we will utilize additional datasets to validate our model's performance. Furthermore, we aim to enhance our model to develop a demand-driven framework for metro network design.

Acknowledgment

This research is supported by the National Natural Science Foundation of China (42201502) and Seed Fund for Basic Research at The University of Hong Kong (109001117).

References

[1] Muriel Beser and Staffan Algers. 2002. SAMPERS—The new Swedish national travel demand forecasting tool. In *National transport models: Recent developments and prospects*. Springer, 101–118.

[2] Osvaldo Daniel Cardozo, Juan Carlos García-Palomares, and Javier Gutiérrez. 2012. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography* 34 (2012), 548–558.

[3] Robert Cervero. 2006. Alternative approaches to modeling the travel-demand impacts of smart growth. *Journal of the American Planning Association* 72, 3 (2006), 285–295.

[4] Pengfei Chen, Xuandi Fu, and Xue Wang. 2021. A graph convolutional stacked bidirectional unidirectional-LSTM neural network for metro ridership prediction. *IEEE Transactions on Intelligent Transportation Systems* 23, 7 (2021), 6950–6962.

[5] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[6] CIESIN. 2018. Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4JW8BX5. Accessed: day month year.

[7] Chuan Ding, Jinxiao Duan, Yanru Zhang, Xinkai Wu, and Guizhen Yu. 2017. Using an ARIMA-GARCH modeling approach to improve subway short-term ridership forecasting accounting for dynamic volatility. *IEEE Transactions on Intelligent Transportation Systems* 19, 4 (2017), 1054–1064.

[8] Rui Fu, Zuo Zhang, and Li Li. 2016. Using LSTM and GRU neural network methods for traffic flow prediction. In *2016 31st Youth academic annual conference of Chinese association of automation (YAC)*. IEEE, 324–328.

[9] Jianyuan Guo, Zhen Xie, Yong Qin, Limin Jia, and Yaguan Wang. 2019. Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM. *IEEE Access* 7 (2019), 42946–42955.

[10] Arthur E Hoerl and Robert W Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 1 (1970), 55–67.

[11] Goran Jovicic and Christian Overgaard Hansen. 2003. A passenger travel demand model for Copenhagen. *Transportation Research Part A: Policy and Practice* 37, 4 (2003), 333–349.

[12] Myung-Jin Jun, Keechoo Choi, Ji-Eun Jeong, Ki-Hyun Kwon, and Hee-Jae Kim. 2015. Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. *Journal of transport geography* 48 (2015), 30–40.

[13] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[14] Yuebing Liang, Fangyi Ding, Guan Huang, and Zhan Zhao. 2023. Deep trip generation with graph neural networks for bike sharing system expansion. *Transportation Research Part C: Emerging Technologies* 154 (2023), 104241. https://doi.org/10.1016/j.trc.2023.104241

[15] Yang Liu, Zhiyuan Liu, and Ruo Jia. 2019. DeepPF: A deep learning based architecture for metro passenger flow prediction. *Transportation Research Part C: Emerging Technologies* 101 (2019), 18–34.

[16] Man Luo, Hongkai Wen, Yi Luo, Bowen Du, Konstantin Klemmer, and Hongming Zhu. 2019. Dynamic demand prediction for expanding electric vehicle sharing systems: A graph sequence learning approach. *arXiv preprint arXiv:1903.04051* (2019).

[17] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. 2014. Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems* 16, 2 (2014), 865–873.

[18] Xiaolei Ma, Jiyu Zhang, Bowen Du, Chuan Ding, and Leilei Sun. 2018. Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction. *IEEE Transactions on Intelligent Transportation Systems* 20, 6 (2018), 2278–2288.

[19] Michael G McNally. 2007. The four-step model. In *Handbook of transport modelling*. Vol. 1. Emerald Group Publishing Limited, 35–53.

[20] Brian L Smith and Michael J Demetsky. 1997. Traffic flow forecasting: comparison of modeling approaches. *Journal of transportation engineering* 123, 4 (1997), 261–266.

[21] Laron Smith, Richard Beckman, and Keith Baggerly. 1995. *TRANSIMS: Transportation analysis and simulation system.* Technical Report. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

[22] Hongyuan Su, Yu Zheng, Jingtao Ding, Depeng Jin, and Yong Li. 2024. MetroGNN: Metro Network Expansion with Reinforcement Learning. In *Companion Proceedings of the ACM on Web Conference 2024*. 650–653.

[23] Daniel Svozil, Vladimir Kvasnicka, and Jiri Pospichal. 1997. Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems* 39, 1 (1997), 43–62.

[24] JWC Van Lint, SP Hoogendoorn, and Henk J van Zuylen. 2005. Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C: Emerging Technologies* 13, 5-6 (2005), 347–369.

[25] Gerald Walters and Robert Cervero. 2003. Forecasting transit demand in a fast growing corridor: The direct-ridership model approach. *Fehrs and Peers Associates* (2003).

[26] Kaipeng Wang, Pu Wang, Zhiren Huang, Ximan Ling, Fan Zhang, and Anthony Chen. 2022. A Two-Step model for predicting travel demand in expanding subways. *IEEE Transactions on Intelligent Transportation Systems* 23, 10 (2022), 19534–19543.

[27] Lebing Wang, Jian Gang Jin, Gleb Sibul, and Yi Wei. 2023. Designing metro network expansion: Deterministic and robust optimization models. *Networks and Spatial Economics* 23, 1 (2023), 317–347.

[28] Yu Wei, Minjia Mao, Xi Zhao, Jianhua Zou, and Ping An. 2020. City metro network expansion with reinforcement learning. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 2646–2656.

[29] Jie Zeng and Jinjun Tang. 2023. Combining knowledge graph into metro passenger flow prediction: A split-attention relational graph convolutional network. *Expert Systems With Applications* 213 (2023), 118790.

[30] Da Zhang and Mansur R Kabuka. 2018. Combining weather condition data to predict traffic flow: a GRU-based deep learning approach. *IET Intelligent Transport Systems* 12, 7 (2018), 578–585.

[31] Jinlei Zhang, Feng Chen, Zhiyong Cui, Yinan Guo, and Yadi Zhu. 2020. Deep learning architecture for short-term passenger flow forecasting in urban rail transit. *IEEE Transactions on Intelligent Transportation Systems* 22, 11 (2020), 7004–7014.

[32] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 31.

[33] Liqing Zhang, Leong Hou U, Shaoquan Ni, Dingjun Chen, Zhenning Li, Wenxian Wang, and Weizhi Xian. [n. d.]. City Metro Network Expansion Based on Multi-Objective Reinforcement Learning. ([n. d.]). Available at SSRN: https://ssrn.com/abstract=4837977 or http://dx.doi.org/10.2139/ssrn.4837977.

[34] Yanru Zhang, Yunlong Zhang, and Ali Haghani. 2014. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transportation Research Part C: Emerging Technologies* 43 (2014), 65–78.