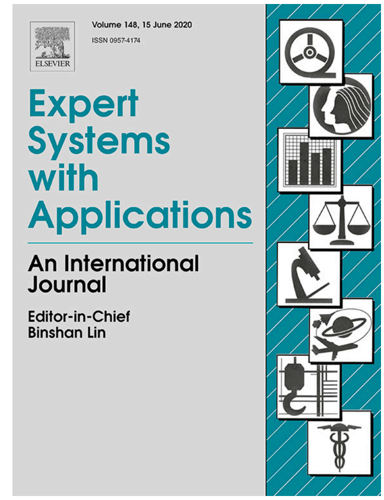
Journal Pre-proofs



Combining knowledge graph into metro passenger flow prediction: a split-attention relational graph convolutional network

Jie Zeng, Jinjun Tang

PII:	S0957-4174(22)01808-5
DOI:	https://doi.org/10.1016/j.eswa.2022.118790
Reference:	ESWA 118790
To appear in:	*Expert Systems with Applications*
Received Date:	30 March 2022
Revised Date:	25 August 2022
Accepted Date:	4 September 2022

Please cite this article as: Zeng, J., Tang, J., Combining knowledge graph into metro passenger flow prediction: a split-attention relational graph convolutional network, *Expert Systems with Applications* (2022), doi: https://doi.org/10.1016/j.eswa.2022.118790

Combining knowledge graph into metro passenger flow prediction: a split-attention relational graph convolutional network

Jie Zeng
Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering
Central South University, Changsha, 410075, China
Email: zj991130@csu.edu.cn

Jinjun Tang*
Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering
Central South University, Changsha, 410075, China
Email: jinjuntang@csu.edu.cn

*Corresponding Author

Abstract:

With the rapid development of intelligent operation and management in metro systems, accurate network-scale passenger flow prediction has become an essential component in real-time metro management. Although numerous novel methods have been applied in this field, critical barriers still exist in integrating travel behaviors and comprehensive spatiotemporal dependencies into prediction. This study constructs the metro system as a knowledge graph and proposes a split-attention relational graph convolutional network (SARGCN) to address these challenges. Breaking the limitations of physical metro networks, we develop a metro topological graph construction method based on the historical origin-destination (OD) matrix to involve travel behaviors. Then, we design a metro knowledge graph construction method to incorporate land-use features. To adapt prior knowledge of metro systems, we subsequently propose the SARGCN model for network-scale metro passenger flow prediction. This model integrates the relational graph convolutional network (R-GCN), split-attention mechanism, and long short-term memory (LSTM) to explore the spatiotemporal correlations and dependence between passenger inflow and outflow. According to the model validation conducted on the metro systems in Shenzhen and Hangzhou, China, the SARGCN model outperforms the advanced baselines. Furthermore, quantitative experiments also reveal the effectiveness of its component and the constructed metro knowledge graph.

Keywords: passenger flow prediction, urban metro system, knowledge graph, graph neural network, deep learning

1. Introduction

Among all modes of urban public transportation, the metro system has attracted attention from transportation planners and managers due to its advantages of high speed, large capacity, and punctuality. Although these advantages help the metro system attract more passengers, the imbalance between travel demands and services has become increasingly severe. Many metro stations, especially the critical nodes in the metro network, always face the intense challenge of congestion. These issues negatively affect on the travel experience of passengers and reduce the attractiveness of metro systems.

In recent years, since many compelling scenarios of the internet of things (IoT) have been applied in metro operations, transportation administration can quickly obtain real-time states of the metro system. Based on real-time information from advanced data collection and processing technologies, numerous emerging applications have been conducted. To relieve pressure in metro operations, the administration not only needs to grasp the current conditions but also to forecast their future variations in advance. Therefore, accurate passenger flow prediction in metro systems has become an essential task in recent years.

According to the prediction horizons, metro passenger flow prediction can be classified into three major categories (Ma et al., 2019): long-term, medium-term, and short-term prediction. Long-term and medium-term passenger flow predictions are vital in metro planning and development. For these purposes, researchers mainly employ the four-step travel demand forecasting model (Agrawal et al., 2018; McNally, 2007) and geographically weighted regression (GWR) (Daniel et al., 2012) to predict future (e.g., monthly, annual, etc.) demand. Generally, the abovementioned two prediction tasks focus on metro policy and planning, but they cannot meet the needs of real-time applications. Since real-time information is beneficial to avoiding congestion and balancing transportation resources (Zhang et al., 2020b), the importance of short-term passenger flow prediction is increasingly highlighted.

Previous studies on short-term passenger flow prediction can be further divided into station-level and network-scale predictions. The former aims to forecast the future passenger flow of a specific station. Since the temporal dependence of passenger flow is the basis of this task, the involved prediction methods include statistical methods, machine learning methods, and recurrent neural network (RNN) based deep learning methods. In the opposite, the latter focuses on predicting future passenger flows of each station simultaneously. In addition to the temporal dependence, exploring the spatial correlation is also essential to achieving accurate predictions. Therefore, several novel deep learning models with powerful spatial correlation extraction capability, e.g., convolutional neural network (CNN) and graph neural network (GNN), have widely attracted the attention of researchers.

Since the abovementioned station-level prediction models rarely consider the spatial correlation in the metro network, these models always suffer from limited prediction performance. Furthermore, this category of prediction models needs to conduct training on each station separately, so the training and storage costs approximately linearly increase with the metro network scale. Overall, the station-level prediction methods are unsuitable for predicting future passenger flows of all the stations in the metro network. Therefore, in recent years, researchers have preferred network-scale prediction models. As there are typical topological structures in metro systems, GNNs are widely applied in this field and have consistently achieved state-of-the-art performance.

Although numerous novel methods have been applied in this field, several critical issues are still unaddressed:

(1) In GNN-based models, researchers must construct a reasonable graph in advance, which plays a vital role in prediction performance. In numerous studies, the graphs are directly established according to the physical adjacency relationship (i.e., the topology of the metro network) (Han et al., 2019; Ye et al., 2020; Zhang et al., 2020b). However, these graph construction methods always overlook metro travel behaviors. Thus, how to fuse travel behaviors and features into graph construction principles still needs further exploration.

(2) Travel demands and behaviors are significantly associated with land-use characteristics (Jun et al., 2015), but only a few studies (He et al., 2020; Lin et al., 2020) consider this vital factor in metro passenger flow prediction. These studies attempt to address metro passenger flow prediction using statistical methods or shallow machine learning methods, so there is a lack of exploration of the spatiotemporal dependencies in network-scale passenger flow. GNNs are always regarded as powerful tools in this field (Han et al., 2019; Liu et al., 2020; Ye et al., 2020; Zhang et al., 2020b), but land-use features are rarely introduced into GNNs for metro passenger flow prediction. Thus, effectively integrating the land-use features and GNN model to improve prediction accuracy is still a challenge.

(3) Spatiotemporal correlations of passenger flow at different stations need further mining to improve prediction performance. There are two passenger flows at each metro station: inflow (i.e., the number of passengers in the origin station) and outflow (i.e., the number of passengers in the destination station). Since the outflow of each station consists of the inflow of the remaining stations in the metro network, the inflow and outflow are correlated in the spatial and temporal dimensions. However, few current studies explore this relationship in passenger flow prediction. Additionally, passenger flows in different regions also follow specific travel patterns. For instance, commuting will cause the metro passenger flow around industrial and residential regions to show the opposite

trend. These spatiotemporal regularities are also vital for metro passenger flow analysis and prediction.

To fill these gaps, we propose a deep learning framework for short-term metro passenger flow prediction, named split-attention relational graph convolutional network (SARGCN). To adapt metro travel behaviors and operation principles, we extract the historical OD matrix from the smart card data as the similarity measure and employ the complex network construction method (Cupertino et al., 2013) to establish a topological graph. Since land-use features significantly impact metro travel demands and behaviors, we introduce the point of interest (POI) data to transform the constructed topological graph into a metro knowledge graph. Then, a spatiotemporal learning framework, i.e., the SARGCN, is proposed for network-scale metro passenger flow learning and prediction based on the established knowledge graph. In summary, the major contributions of this study are concluded as follows:

(1) We develop a metro topological graph construction method based on the historical OD matrix and complex network construction algorithm. Compared with the physical metro network, this data-driven graph construction method is adaptive to metro travel patterns, so the spatial correlation on the graph is enhanced.

(2) Based on the land-use features around metro stations, a metro knowledge graph construction method is designed and applied to the constructed metro topological graph. In this way, each station is assigned a specific semantic type to provide essential prior knowledge for the deep learning model.

(3) We propose the SARGCN model for network-scale metro passenger flow prediction. In this model, the R-GCN, split-attention mechanism, and LSTM are effectively incorporated to learn the spatiotemporal correlations and dependencies between inflow and outflow on the constructed metro knowledge graph.

(4) Validated on the metro systems in Shenzhen and Hangzhou, China, the proposed SARGCN model expresses a superior performance than the advanced baselines in terms of accuracy and efficiency. Additionally, the ablation experiment results also demonstrate the effectiveness of each component.

The organization of this paper is summarized as follows. Section 2 discusses the existing studies in the field of short-term metro passenger flow prediction. We briefly describe the metro passenger flow and land-use data involved in this study in Section 3. Section 4 introduces the detailed methodology of the proposed metro knowledge construction method and SARGCN model. Section 5 shows the experimental results and discussions. Finally, we conclude this study and summarize the research directions for future works in Section 6.

2. Literature review

2.1 Station-level prediction methods

Due to the limitation of computing capability, statistical methods won the favor of researchers in the early stage of metro passenger flow prediction. Statistical methods always regard the previous ridership of each station as sequence data and employ time-series analysis models to make predictions. Among all the time-series analysis models, the autoregressive integrated moving average (ARIMA) model (Chen et al., 2020; Wen et al., 2022) and its variants are the most famous methods in metro passenger flow prediction. Meanwhile, other statistical methods, such as the

Kalman filter (Sun et al., 2014) and generalized autoregressive conditional heteroskedasticity (GRACH) (Ding et al., 2018), are also widely applied in this task. However, these statistical methods always encounter limitations in exploring the nonlinear characteristics of traffic data (Zhao et al., 2020a) and have high computational complexity (Zhang et al., 2019). Additionally, these methods may encounter challenges when facing complex conditions and big data (Zhou et al., 2020).

To fill these gaps, numerous machine learning models have been developed for metro passenger flow prediction, including artificial neural network (ANN) (Li et al., 2019; Li et al., 2017b; Wei & Chen, 2012; Zhao et al., 2011), support vector machine (SVM) (Sun et al., 2015; Tang et al., 2019a), decision trees (Ding et al., 2016; Zhao et al., 2020b), and Bayesian networks (Lin et al., 2017; Roos et al., 2017). Although these machine learning methods can usually achieve higher prediction accuracies than traditional statistical methods, their prediction performances are still unsatisfactory for real-time applications of metro systems. Meanwhile, machine learning methods always face significant challenges to capturing the temporal dynamics in passenger flow. Facing a dramatic increasement in the scale of metro data and metro management demands, researchers widely apply deep learning methods in this field and have demonstrated their superiority to traditional methods. Since metro passenger flow is highly temporally dependent, recurrent neural network (RNN) and its famous variants, i.e., long short-term memory (LSTM) (Tang et al., 2019b) and gated recurrent unit (GRU) (Zhang & Kabuka, 2018) are widely employed to mine its time-varying dynamics. Meanwhile, to improve prediction performance under anomalous large passenger flow, (Zheng et al., 2020) employed the complex network theory to collective behavior modeling, and then a hybrid model was subsequently proposed to capture the time-varying characteristics of passenger flow.

2.2 Network-scale prediction methods

To solve the limitation of traditional station-level prediction methods, many researchers have paid attention to network-scale passenger flow prediction models. Hao et al. (2019) proposed a sequence to sequence (Seq2Seq) model based on LSTM and the attention mechanism for network-scale passenger prediction. Additionally, this model further introduced external features (e.g., weather, special events, etc.) into the prediction framework. Ma et al. (2019) transformed metro ridership into grid-based data and then combined CNN with bidirectional LSTM to construct a parallel architecture for prediction. Ning et al. (2018) designed a residual unit and introduced external factors into metro passenger flow prediction. However, since metro stations are sparsely distributed in the urban areas, Liu et al. (2019) demonstrated that metro networks are unsuitable for transforming into grid-based data. Hence, they manually designed high-level features to represent the spatial correlation to achieve accurate prediction.

To further promote passenger flow prediction accuracy, researchers also turned to GNN-based methods to involve topological information in prediction models. Zhang et al. (2020b) integrated the ResNet (He et al., 2016), GCN (Kipf & Welling, 2016), and attention-based LSTM to construct the ResLSTM model for metro passenger flow prediction. Ye et al. (2020) proposed a Multi-STGCnet model, which employed the LSTM and GCN to extract the temporal and spatial dependencies of metro passenger flow, respectively. Relying on the physical metro network, Wang et al. (2021) constructed metro hypergraphs to involve OD passenger flow and proposed a dynamic spatiotemporal hypergraph neural network (DSTHGCN) for prediction. Ou et al. (2020) integrated diffusion graph convolutional networks with a novel temporal convolutional model (i.e., TrellisNet (Bai et al., 2019b)) to explore the spatiotemporal dependencies of metro passenger flow.

Actually, there is a typical *tunnel effect* from the perspective of metro operation regularities and passenger travel behaviors. That is, passengers are more inclined to take the metro on long-distance travel instead of going to the nearby stations. Therefore, it is challenging for physical metro networks to capture this travel behavior. Liu et al. (2020) proposed a physical-virtual collaboration graph network (PVCGN) model, which integrates a physical graph, a similarity graph, and a correlation graph for metro ridership and online OD ridership prediction. However, the similarity graph and correlation graph are based on the k-nearest neighbors (i.e., connecting each node to its k most similar nodes) and the ε-radius principles (i.e., connecting each node to nodes within distance threshold ε). Thus, the constructed network only focuses on the similarity at the node level while ignoring the optimality at the network level. Meanwhile, land-use features are also a typical factor to represent the travel behavior of metro passenger flow, but they are rarely considered in network-scale passenger flow prediction. Overall, a brief summary of this study and current novel models in this field is displayed in Table 1.

Table 1 Comparison of network-scale prediction models in metro passenger prediction

	Land-use features	Travel behavior	Correlation of in & out flow
ResLSTM	×	×	√
DSTHGCN	×	√	×
PVCGN	×	√	√
SARGCN	√	√	√

3. Data description

3.1 Metro data

This study applies the smart card data collected in Shenzhen and Hangzhou, China, to validate the proposed model. The smart card records include smart card ID, collection machine ID, state (i.e., enter or exit), collection time, metro line, and metro station. Passenger travel features, such as the travel time matrix and OD matrix, can be extracted from this data source. We count the number of passengers entering and leaving each station from the original smart card records, named inflow and outflow. Furthermore, we also extract the historical OD matrix from the smart card records to measure the spatial dependence between different metro stations.

The detailed descriptions of these two datasets are introduced as follows:

(1) **Shenzhen city**. The Shenzhen metro network includes 166 metro stations, and the collection duration of smart card records ranges from May 1st to May 31st in 2019. According to the actual operation time of the metro system, only the records between 6:00 am and 11:00 pm are used in this study. For this dataset, we aggregate the inflow and outflow of each station every 10 minutes. Thus, the inflow and outflow of each station can be regarded as a time-series with 102 records per day.

(2) **Hangzhou city**. This dataset is released by (Liu et al., 2020) via an accessible link[1]. In this dataset, there are 80 metro stations in total. All the data is collected in January 2019, and the time interval of passenger flow is set as 15 minutes.

[1]https://github.com/HCPLab-SYSU/PVCGN

3.2 Land-use data

Lin et al. (2020) demonstrated that land-use features vitally affected metro passenger flows. POI data refer to specific points with different functional attributions in urban areas and are widely used in travel pattern analysis (Krause & Zhang, 2019). Therefore, we utilize the POI data to reflect the land-use characteristics around each metro station for passenger flow analysis and prediction. The POI data involved in this study are collected by the application programming interface (API) of Baidu Map[2]. The original POI data have 19 categories and 140 subcategories. We merge similar categories according to the definition of land-use attributes and finally obtain five categories. The detailed descriptions of these five merged categories are displayed in Table 2.

Table 2 Classification and description of POI data

Category	Contents
residential area	residential area, dormitory, etc.
leisure and entertainment	restaurants, cinema, shopping center, etc.
education institution	colleges, high schools, kindergartens, etc.
corporate company	company, factory, etc.
transportation hub	airports, railway stations, bus stations, etc.

4. Methodology

4.1 Problem formulation

Assume $\mathbf{I}_t \in \mathbb{R}^{\mathrm{N}\times\mathrm{M}}$ and $\mathbf{O}_t \in \mathbb{R}^{\mathrm{N}\times\mathrm{M}}$ denote the feature matrices of inflow and outflow at time interval t, respectively, where N represents the number of stations and M denotes the number of previous time steps. The metro passenger flow prediction task in this study can be summarized as follows: given the previous passenger flow ($\mathbf{I}_t$ and $\mathbf{O}_t$) and a knowledge graph ($\mathbb{G}$), aim to learn a mapping function $\mathbb{F}(\cdot)$ to predict inflow ($\boldsymbol{i}_{t+p} \in \mathbb{R}^{\mathrm{N}}$) and outflow ($\boldsymbol{o}_{t+p} \in \mathbb{R}^{\mathrm{N}}$) at the p-th step afterward at each station.

$$(\boldsymbol{i}_{t+p},\boldsymbol{o}_{t+p}) = \mathbb{F}(\mathbf{I}_t,\mathbf{O}_t;\mathbb{G}) \tag{1}$$

4.2 Metro knowledge graph construction

4.2.1 Metro topological graph construction

The *tunnel effect* mentioned above is an unignored characteristic in metro travel behavior. According to the smart card data used in this study, the travel distance distribution in the Shenzhen metro system is displayed in Figure 1. Here, we employ the Floyd-Warshall algorithm and metro topological network to calculate the distance among metro stations. From this figure, an intuitive finding is that the majority of passengers select the metro as middle-distance and long-distance transportation, and few passengers take adjacent stations as their destinations. Specifically, the average travel distance is 7.60 stations, and only 6% of passengers stop at adjacent stations. This finding indicates that the passenger flow interactions between the adjacent metro stations are not

[2]https://lbsyun.baidu.com/index.php?title=webapi/guide/webservice-placeapi

strong enough.

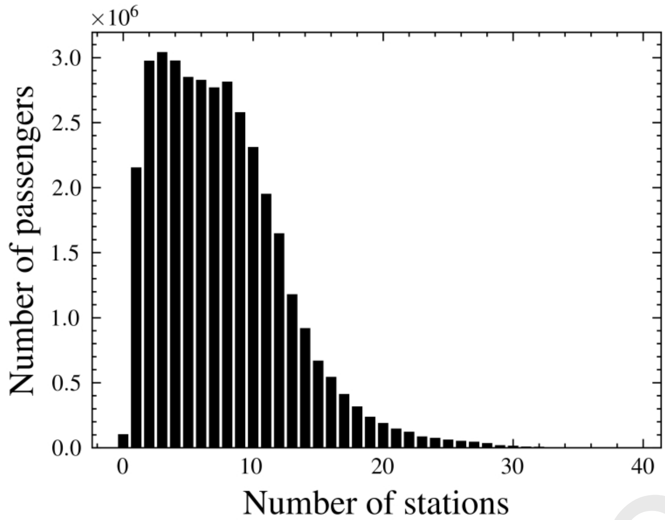
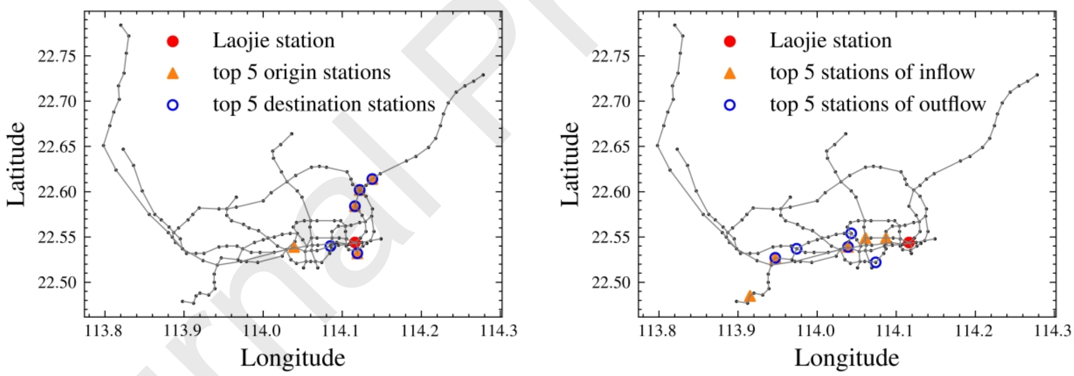


Figure 1 Travel distance distribution in the Shenzhen metro network

Furthermore, we illustrate the spatial distributions of the correlated stations for Laojie Station in Figure 2. Laojie Station is an important node in the Shenzhen metro network, which is the transfer station for Lines 1 and 3. Here, these four figures explore the positional relationship between Laojie Station and its correlated stations from the traffic perspective (i.e., origin and destination stations) and statistical perspective (i.e., Pearson correlation coefficient). Similarly, we can find that all the correlated stations are far apart from Laojie Station, instead of the neighbor stations.



(a) top 5 origin and destination stations (b) top 5 correlated stations

Figure 2 Spatial distributions of correlated stations with Laojie Station

Following these findings, since the topological network overlooks these travel patterns, it is unsuitable for passenger flow prediction. Thus, it is essential to construct a reasonable graph that is adaptive to the travel behaviors of the metro system. This study employs the OD relationship among all the stations as the similarity measure for graph construction. Then, the complex network construction algorithm proposed by (Cupertino et al., 2013) is applied to build a directed graph. Unlike the simple k-nearest neighbors and the ε-radius approaches (Liu et al., 2020), this data-driven construction method can connect correlated stations at the node level and consider optimality at the network level. Cupertino et al. (2013) utilized a distance measure for graph construction, so this method aimed to connect nodes with short distances. However, in this study, since we employ the OD relationship as a similar measure, we prefer the OD passenger flow on the edges in the constructed graph to be as large as possible. Thus, we replace the $\min(\cdot)$ and $\max(\cdot)$

operations in this method with the corresponding $\max(\cdot)$ and $\min(\cdot)$ operations, respectively. Finally, we summarize this metro graph construction algorithm in Algorithm 1.

Algorithm 1. Metro topological graph construction method.

Input: number of nodes, N; similarity matrix, $\mathbf{W}_{\mathrm{OD}} \in \mathbb{R}^{\mathrm{N}\times N}$;
hyperparameters, K and λ; node set $\mathbf{V} = \{v_1,v_2,...,v_N\}$.
Output: adjacency matrix, $\mathbf{W}_g \in \mathbb{R}^{\mathrm{N}\times N}$.
Process:
1.$\mathbf{W}_g \leftarrow \mathrm{zeros(N,N)}$
2.$\mathbf{\Omega} \leftarrow \{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,...,\boldsymbol{\omega}_{\mathrm{N}}\}$ $(\boldsymbol{\omega}_i = \{v_i\})$
3.$\mathbf{W}_\Omega \leftarrow \mathbf{W}_{\mathrm{OD}}$
4.**while** len($\mathbf{\Omega}$) > 1 **do**
5. $[\boldsymbol{\omega}_m,\boldsymbol{\omega}_n] \leftarrow \mathrm{argmax}(\mathbf{W}_\Omega)$
6. $d_\varepsilon \leftarrow \lambda \cdot \min(d_m,d_n)$
7. $[\boldsymbol{v}_s,\boldsymbol{v}_e] \leftarrow \mathbf{select}(\boldsymbol{\omega}_m,\boldsymbol{\omega}_n,\mathrm{K})$
8. **for** k=1, …, K **do**
9. **if** $\mathbf{W}[\boldsymbol{v}_s^k, \boldsymbol{v}_e^k] > d_\varepsilon$ **do**
10. $\mathbf{W}_g[\boldsymbol{v}_s^k, \boldsymbol{v}_e^k] \leftarrow 1$
11. **end**
12. **end**
13. $\boldsymbol{\omega}_m \leftarrow \mathrm{concat}(\boldsymbol{\omega}_m,\boldsymbol{\omega}_n)$ and delete $\boldsymbol{\omega}_n$
14. update $\mathbf{W}_\Omega$ among the current groups
15.**end**
16.$\mathbf{W}_g[i,i] \leftarrow 1,\ \forall i \in [1, \mathrm{N}]$

In this algorithm, the $\mathbf{select}(\cdot)$ operation in Step 7 aims to select the most similar K node pairs from $\boldsymbol{\omega}_m$ and $\boldsymbol{\omega}_n$. Meanwhile, d_m and d_n represent the average similarity within node groups $\boldsymbol{\omega}_m$ and $\boldsymbol{\omega}_n$, respectively. In Step 13, $\mathbf{concat}(\cdot)$ denotes the concatenation operation, which aims to join node groups $\boldsymbol{\omega}_m$ and $\boldsymbol{\omega}_m$ into a larger group. In Step 14, $\mathbf{W}_\Omega[i,\ j]$ is updated by the similarity of the most similar node pair between node groups $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_j$. Furthermore, since the previous passenger flow of each station significantly affects its own future states, we apply a self-loop connection (i.e., Step 16) to each node to retain its previous influence.

4.2.2 Knowledge graph construction

According to the definition in (Hogan et al., 2020), a knowledge graph is a network that consists of entities with semantic types and relations between these entities. Since knowledge graphs can truthfully and powerfully reflect the dependencies between entities in the real world, it has been widely used in search engines, social networks, question answering, etc. In this study, we employ the POI data around metro stations to represent their semantic types to obtain the metro knowledge graph.

Determining the semantic types of nodes and relationships between them is the key step in knowledge graph construction. In the previous study (Tang et al., 2020), POI data are generally classified into several categories and employed to assign a label to each station, according to the POI category with the maximum number around it. This method depends on the number of POI categories around each metro station, but overlooks the differences in the total number of each category. If there is a significant gap in the number of each category, it would be unfair to determine the station type only based on the quantity. To obtain a more reasonable classification, we utilize the distribution frequency of each category to determine the station type, which is summarized in Equations 2 and 3. In these equations, c_i^j and p_i^j denote the number and distribution frequency of POI category j around station i, respectively. $\mathcal{R}_i$ represents the determined semantic type of station i. In Equation 3, the type of each station is assigned as the POI category with the highest distribution frequency.

$$p_i^j = \frac{c_i^j}{\sum_{k=1}^{\mathrm{N}} c_k^j} \tag{2}$$

$$\mathcal{R}_i = \underset{j}{\operatorname{argmax}}\, p_i^j \tag{3}$$

Using the directed graph $\mathcal{G}$ constructed above, we can establish a metro knowledge graph by assigning the semantic types to corresponding stations. Therefore, the established knowledge graph can be denoted as $\mathbb{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$. Here, $\mathcal{V}$, $\mathcal{E}$, and $\mathcal{R}$ represent the node set, edge set, and node types, respectively.

4.3 Framework of the SARGCN

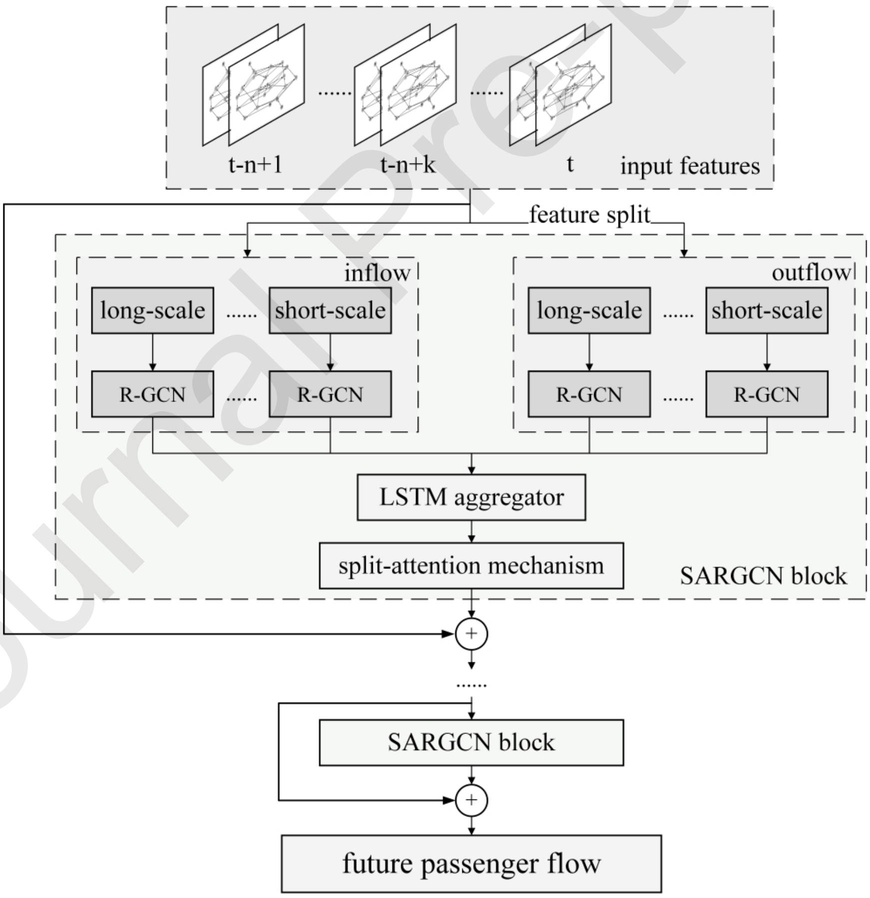


Figure 3 The framework of the proposed SARGCN model

Figure 3 illustrates the framework of the proposed prediction method. Based on the constructed knowledge graph, we integrate the R-GCN, LSTM, and split-attention mechanism to construct the SARGCN block. In each block, the R-GCN layer is applied to extract the spatial correlation on the established knowledge graph. Meanwhile, we employ the split-attention mechanism and LSTM to capture the temporal dynamics of passenger flow and explore the dependence between inflow and outflow. The split-attention mechanism allows the input features to be divided into several groups. Then, the unique characteristics of each group are extracted and aggregated to incorporate global contextual information. After that, several SARGCN blocks are stacked together to improve the capability of hidden feature extraction.

4.4 Spatial correlation modeling

R-GCN (Schlichtkrull et al., 2018) is an effective variant of GCN, and it develops a powerful capability to learn realistic knowledge bases. Thus, we adopt it to model the spatial dependence on the constructed metro knowledge graph. Supposing the input feature of R-GCN is $\mathbf{H} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_N\}$, where $\boldsymbol{h}_i$ denotes feature of node i, the calculation details of R-GCN are summarized in Equation 4 and Figure 4.

$$\boldsymbol{h}_i' = \sigma(\sum_{r\in\mathcal{R}}\sum_{j\in\mathbb{N}_i^r}\frac{1}{c_{ir}}\mathbf{W}_r\boldsymbol{h}_j + \mathbf{W}_0\boldsymbol{h}_i) \tag{4}$$

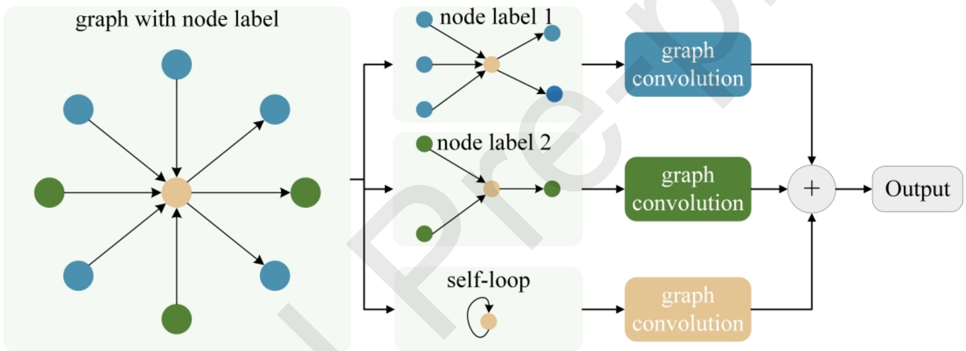


Figure 4 The calculation framework of R-GCN

Here, $\mathcal{R}$ stands for the number of node types which is set as 5 according to Table 1, and $\mathbb{N}_i^r$ denotes the neighbors of node i with type r. c_{ir} is a problem-specific normalization constant. To highlight the importance of each node itself compared with its neighbors, $\mathbf{W}_0$ is employed to represent the particular connection type of the self-loops.

Moreover, a regularization, named basis decomposition, is applied in Equation 4 to reduce the parameters by weight sharing. The regularization of $\mathbf{W}_r$ is a linear combination and described in Equation 5, where $\mathbf{V}_b$ and α_{rb} represent the learnable basis transformations and coefficients, respectively.

$$\mathbf{W}_r = \sum_{b=1}^{B}\alpha_{rb}\mathbf{V}_b \tag{5}$$

4.5 Temporal dependence modeling

In addition to the spatial correlation mentioned above, the metro passenger flow data still have two dependencies: (i) temporal dynamics; (ii) the dependence between inflow and outflow. Both of these dependencies are critical factors in improving prediction accuracy. Motivated by the breakthrough of the ResNeSt model (Zhang et al., 2020a) in computer vision, we integrate its core component, i.e., the split-attention mechanism, with R-GCN and LSTM model to model these two vital dependencies. Specifically, the original split-attention mechanism employs the group convolution operation (Krizhevsky et al., 2017; Xie et al., 2017) to extract the feature-map attention and utilizes a weighted combination operation to mine global contextual information. Following this opinion, using the split operation, we design a graph-based group convolution operation on the previous passenger flow. Furthermore, since evident temporal dependence exists in passenger flow data, we employ LSTM to address this time-series characteristic.

4.5.1 Feature split operation

Considering the dependence between inflow and outflow, we divide the previous passenger flow into S split groups (e.g., inflow and outflow). In traffic prediction, researchers have demonstrated that traffic data at different time steps show different influences on future states (Yang et al., 2019). Therefore, in this study, we further classify each split group into C cardinal groups (e.g., long-scale, middle-scale, and short-scale) according to the temporal dimension. Supposing the temporal dimension of $\mathbf{I}_t$ and $\mathbf{O}_t$ is 6, the developed feature split operation under $S = 2$ and $C = 3$ is illustrated in Figure 5. Since each subgroup uniquely affects future passenger flow, using different models to capture the characteristics of each subgroup will help achieve accurate prediction results.

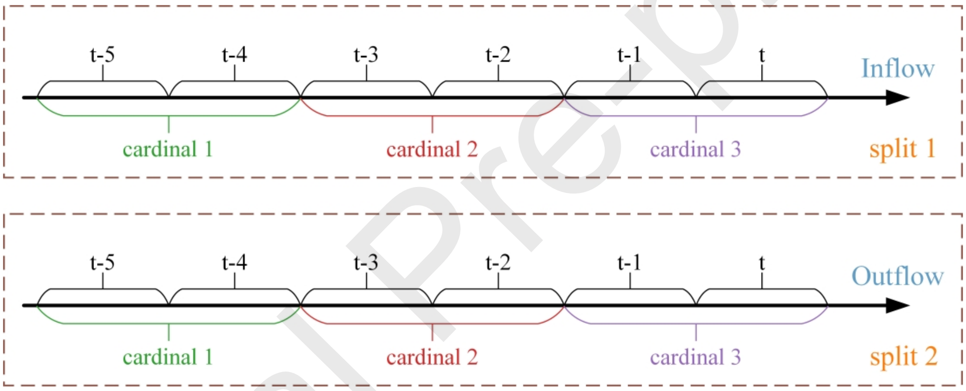


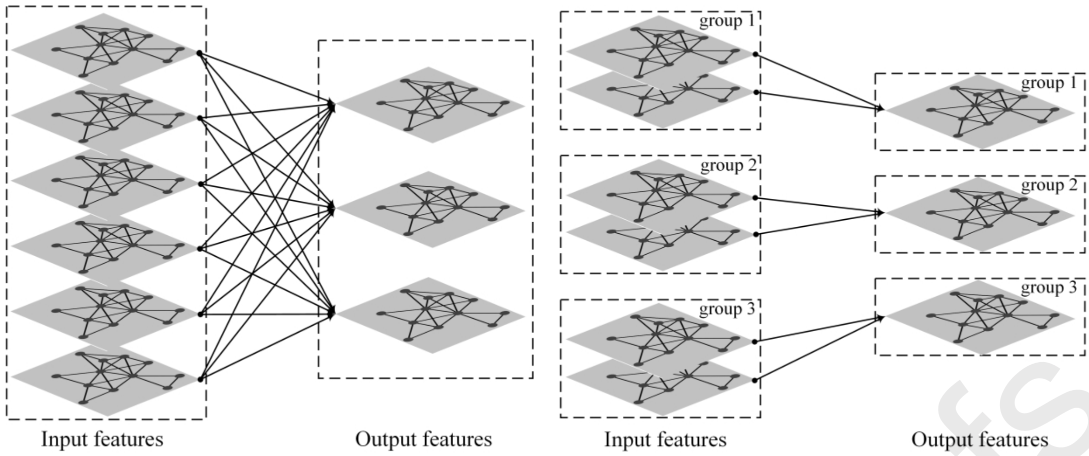
Figure 5 An example of the feature split operation under $S = 2,\ C = 3$

4.5.2 Group graph convolution operation

The group convolution plays a vital role in the split-attention mechanism. The principle of this operation can be summarized as: output features can only receive information from input features in the same group. Hence, we adopt this opinion and propose a group convolution operation. The calculation process of the involved group R-GCN operation is summarized in Equations 6 and 7, where $\boldsymbol{h}_i^l$ denotes the input features of node i in group l, and $\boldsymbol{h}_i^{l'}$ represents the corresponding output features. Meanwhile, the difference between the graph convolution operation and group graph convolution is displayed in Figure 6.

$$\boldsymbol{h}_i^{l'} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathbb{N}_i^r} \frac{1}{c_{ir}} \mathbf{W}_r \boldsymbol{h}_j^l + \mathbf{W}_0 \boldsymbol{h}_i^l\right) \tag{6}$$

$$\boldsymbol{h}_i' = \mathrm{concat}(\boldsymbol{h}_i^{1'}, \boldsymbol{h}_i^{2'}, \ldots, \boldsymbol{h}_i^{L'}) \tag{7}$$



(a) graph convolution operation (b) group graph convolution operation

Figure 6 The structure of graph convolution operation and group graph convolution operation

4.5.3 LSTM layer

As shown in Figure 5, we group the input features according to the temporal dimension. Thus, temporal dependence exists among these groups. Many studies have demonstrated the strong ability of LSTM to handle time-series data (Ma et al., 2015), so we employ it to model the temporal dynamics among the output features of R-GCNs. The core composition of the LSTM unit includes an input gate i_t, a forget gate f_t, an output gate o_t, and a memory cell c_t. Assuming x_t denotes the input vector, the calculation process of the LSTM unit is described below.

$$\boldsymbol{i}_t = \sigma(\boldsymbol{x}_t\mathbf{W}_{xi} + \boldsymbol{h}_{t-1}\mathbf{W}_{hi} + \boldsymbol{b}_i) \tag{8}$$

$$\boldsymbol{f}_t = \sigma(\boldsymbol{x}_t\mathbf{W}_{xf} + \boldsymbol{h}_{t-1}\mathbf{W}_{hf} + \boldsymbol{b}_f) \tag{9}$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{x}_t\mathbf{W}_{xo} + \boldsymbol{h}_{t-1}\mathbf{W}_{ho} + \boldsymbol{b}_o) \tag{10}$$

$$\tilde{\boldsymbol{c}}_t = \tanh(\boldsymbol{x}_t\mathbf{W}_{xc} + \boldsymbol{h}_{t-1}\mathbf{W}_{hc} + \boldsymbol{b}_c) \tag{11}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tilde{\boldsymbol{c}}_t \tag{12}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot \tanh(\boldsymbol{c}_t) \tag{13}$$

Here, $\mathbf{W}$ and $\boldsymbol{b}$ represent the weight matrices and bias, respectively. In addition, $\odot$ is the elementwise product operation, and $\sigma(\cdot)$ represents the sigmoid activation function.

4.5.4 SARGCN block

Figure 3 indicates that the proposed SARGCN model consists of several stacked SARGCN blocks. Relying on the description of R-GCN, feature split operation, and LSTM, we introduce the SARGCN block in detail in this subsection (shown in Figure 7). Compared with the naive split-attention mechanism in ResNeSt, we replace the convolution operation with R-GCN and utilize LSTM to explore the temporal dynamics among groups.

As described in Section 4.1, $\mathbf{I}_t$ and $\mathbf{O}_t$ denote the input features of the SARGCN model. Taking the first SARGCN block as an example, its output features $\hat{\mathbf{V}}$ can be computed by the following equations. Here, Equation 14 denotes the feature split process, and Equation 15 represents the R-GCN operation. Equations 17-21 present the split-attention mechanism applied to the output features of R-GCNs. Finally, a residual structure (described in Equation 22) is employed to enhance the stability and improve the convergence speed in the training process.

In Equation 19, $\xi^c(\cdot)$ represents two stacked dense layers with ReLU as the activation function. And $\mathbf{W}_F$ shown in Equation 22 is a learnable weight matrix that aims to transfer the dimension of input features $\mathbf{F}$ to be equal to that of $\mathbf{V}$.

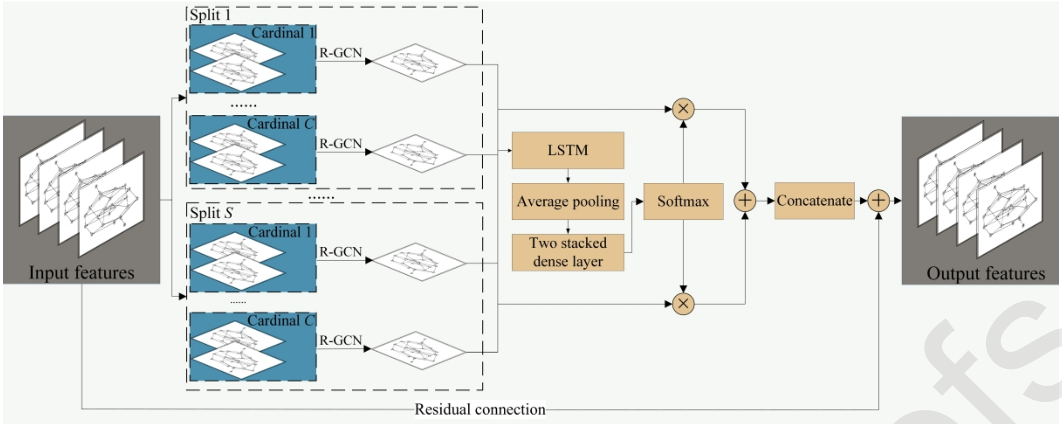


Figure 7 The detailed description of the SARGCN block

According to Equations 14-22, the proposed SARGCN block has a solid capability to explore the spatiotemporal dependencies by integrating R-GCN, split-attention mechanism, and LSTM. Specifically, R-GCN can capture the spatial correlation on the established knowledge graph and explore the interactions between stations with different semantic types. The split-attention mechanism can assign traffic significance to the deep learning model and effectively mine the temporal dynamics and dependencies between inflow and outflow. Moreover, since passenger flow data have typical time-series characteristics, LSTM is employed to enhance the capability of SARGCN to handle this temporal correlation.

$$\mathbf{F} = \mathrm{split}(\mathbf{I}_t,\mathbf{O}_t) = \{\mathbf{F}_1^1,\ldots,\mathbf{F}_1^C;\ldots;\mathbf{F}_S^1,\ldots,\mathbf{F}_S^C\} \quad (14)$$

$$\mathbf{U}_s^c = \mathrm{RGCN}_s^c(\mathbf{F}_s^c) \quad (15)$$

$$\tilde{\mathbf{U}} = \mathrm{LSTM}(\mathbf{U}) = \{\tilde{\mathbf{U}}_1^1,\ldots,\tilde{\mathbf{U}}_1^C;\ldots;\tilde{\mathbf{U}}_S^1,\ldots,\tilde{\mathbf{U}}_S^C\} \quad (16)$$

$$\tilde{\mathbf{U}}_s = \sum_{c=1}^{C} \tilde{\mathbf{U}}_s^c \quad (17)$$

$$\mathbf{S}_s = \mathrm{average\_pooling}(\tilde{U}_s) = \frac{1}{N}\sum_{n=1}^{N} \tilde{\mathbf{U}}_s(n) \quad (18)$$

$$a_s^c = \frac{\exp(\xi^c(\mathbf{S}_s))}{\sum_{j=1}^{C}\exp(\xi^j(\mathbf{S}_s))} \quad (19)$$

$$\mathbf{V}_s = \sum_{c=1}^{C} (a_s^c \cdot \tilde{\mathbf{U}}_s^c) \quad (20)$$

$$\mathbf{V} = \mathrm{concat}(\mathbf{V}_1,\mathbf{V}_2,\ldots,\mathbf{V}_S) \quad (21)$$

$$\hat{\mathbf{V}} = \mathbf{W}_V\mathbf{V} + \mathbf{W}_F\mathbf{F} \quad (22)$$

The proposed SARGCN model is a modular design by stacking SARGCN blocks. This modular design idea makes the network structure relatively compact and convenient for building complex and deep models. In this way, the model structure can be modified easily by changing the number of groups (including split groups and cardinal groups) and the output dimension of R-GCN. Compared with the naive R-GCN model, the split-attention mechanism can also help SARGCN reduce model parameters and construct lightweight models.

5. Experiment

5.1 Evaluation metrics

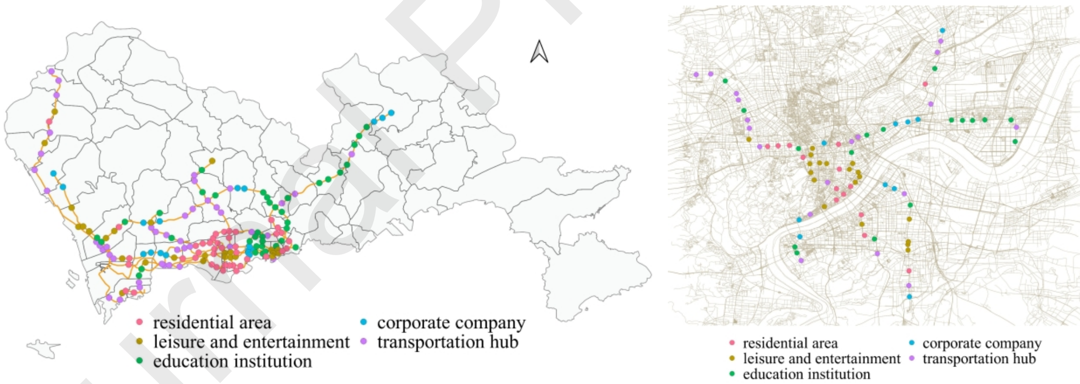
In this study, three metrics are employed to evaluate prediction performance, including the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The definitions of these three metrics are introduced in the following equations. Here, N denotes the number of metro stations, and n presents the number of testing samples. Meanwhile, y can represent both the ground-truth inflow and outflow, and $\hat{y}$ denotes the corresponding predicted values.

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{nN}}\sum_{i=1}^{\mathrm{N}}\sum_{j=1}^{\mathrm{n}}\left(y_j^i - \hat{y}_j^i\right)^2} \tag{23}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{nN}}\sum_{i=1}^{\mathrm{N}}\sum_{j=1}^{\mathrm{n}}\left|y_j^i - \hat{y}_j^i\right| \tag{24}$$

$$\mathrm{MAPE} = \frac{1}{\mathrm{nN}}\sum_{i=1}^{\mathrm{N}}\sum_{j=1}^{\mathrm{n}}\left|\frac{y_j^i - \hat{y}_j^i}{y_j^i}\right| \times 100\% \tag{25}$$

However, many studies have demonstrated that MAPE always faces significant challenges when encountering zero or close-to-zeros ground truth (Kim & Kim, 2016). When the metro system begins to operate in the early morning, no passengers exit at many stations, thus leading to zero data in outflows. Therefore, we employ MAPE@10 (Zhang et al., 2019) to address this problem. Specifically, we calculate MAPE on metro stations with the top 10% largest passenger flow.



(a) Shenzhen city (b) Hangzhou city

Figure 8 The spatial distribution of metro stations with different semantic types

5.2 Experimental setting

5.2.1 Network construction method

By using the construction method described in Section 4.2, we can transfer the employed Shenzhen and Hangzhou metro systems into knowledge graphs, which are shown in Figure 8. According to Algorithm 1, the network construction method has two critical parameters, named K and λ, which significantly impact network density. In this study, we set the value of λ as 0.1 and assign the value of K to these two datasets as 7 and 13, respectively. Finally, the Shenzhen dataset contains 166 nodes and 832 edges (including 166 self-loop edges). And in the Hangzhou dataset, we obtain 80 nodes and 715 edges (including 80 self-loop edges).

5.2.2 SARGCN model

The implementation details of the proposed SARGCN model on the Shenzhen and Hangzhou metro systems are described as follows.

(1) **Shenzhen metro system.** As mentioned in Section 3.1, we aggregate both inflow and outflow into 10 minutes and finally obtain 3,162 records for each station. These passenger flow records are divided into a training set, a validation set, and a testing set according to a splitting rate of 70%: 10%: 20%. We employ the previous 12 time-steps inflow and outflow to predict network-scale ridership at the next 1-step, 4-step, 7-step, and 10-step, respectively. Two stacked SARGCN blocks are utilized to compose the SARGCN model. The hidden dimension of the proposed SARGCN model is set to be the same as the number of stations, i.e., $\mathrm{N} = 166$. Considering the actual traffic significance, we set the number of groups in SARGCN as $S = 2$ and $C = 3$. That is, we first divide passenger flow into two split groups (i.e., inflow and outflow), and each split group is further divided into three cardinal groups (i.e., short-scale flow, middle-scale flow, and long-scale flow). Meanwhile, we apply a grid search strategy on {16,32,48,64,96} to search for the optimal values of the hidden dimension of ξ. Finally, the prediction accuracy reaches the peak at 48, so we employ it as the optimal hyperparameter.

(2) **Hangzhou metro system**. Since this dataset is obtained from the open-source data (Liu et al., 2020), we follow all the settings of the original dataset. That is, we utilize the inflow and outflow of the previous 4 intervals to simultaneously predict the next 4 steps. Furthermore, a two-layer SARGCN model is employed to conduct passenger flow prediction in the Hangzhou metro network, and the hidden dimension is set as 224. After the grid search strategy, the hidden unit number of ξ is set as 96. Since the horizontal of historical passenger flows is 4, we set the number of groups as $S = 2$ and $C = 2$.

Before inputting to the deep learning model, we first normalize these two datasets. The batch size of the mini-batch training strategy is set to 40 for Shenzhen and 32 (the same as in (Liu et al., 2020)) for Hangzhou, and the Adam optimizer (Kingma & Ba, 2014) with a learning rate of 0.001 is applied for model training. All the parameters are initialized by Xavier initialization (Glorot & Bengio, 2010). The number of training epochs is set to 300, and an early-stop strategy is adopted on the validation set to avoid overfitting.

5.3 Performance comparison

5.3.1 Shenzhen metro system

This subsection first introduces and employs 11 widely-used traffic state prediction models to be compared with the proposed SARGCN model on the Shenzhen metro system, including traditional statistical models, machine learning models, shallow deep learning models, and novel graph neural networks. The brief descriptions of the selected baselines are summarized below.

(1) **HA**. The historical average (HA) method utilizes the average passenger flows of each period to represent its future values. For instance, the future passenger flow at 7:00 am-7:10 am on the testing set is calculated by the average passenger flows during 7:00 am-7:10 am on the training set. We utilize the prediction performance of 1-step to denote that of multistep.

(2) **MLP**. The multilayer perceptron (MLP) is a basic machine learning model, and neither

temporal correlation nor spatial dependence is involved in this method. For each metro station, we employ a 3-layer MLP, including an input layer, a hidden layer with ReLU as the activation function, and an output layer to predict short-term passenger flow.

(3) **LSTM** (Ma et al., 2015). Long short-term memory (LSTM) can effectively capture the temporal dynamics of metro passenger flow, but the spatial dependence is also overlooked.

(4) **GRU**. Gated recurrent unit (GRU) is a variant of LSTM. Similarly, only the temporal correlation is used in this model.

(5) **GCN**. Considering the prediction performance, instead of the original GCN framework proposed in (Kipf & Welling, 2016), we employ the R-GCN model without node types and edge relationships (i.e., without a knowledge graph) as GCN to forecast short-term passenger flows.

(6) **GAT** (Velicković et al., 2017). The graph attention network (GAT) can explore spatial dependence by the self-attention and multi-head attention mechanisms, and it has advantages in directed graphs and inductive learning tasks. However, only spatial dependence is involved in the naive GAT model.

(7) **STGCN** (Yu et al., 2018). In this model, the gated CNN is employed to capture the temporal dynamics and combined with GCN to formulate the spatiotemporal graph convolutional network (STGCN).

(8) **Graph-WaveNet** (Wu et al., 2019). In Graph-WaveNet, a novel adaptive correlation matrix and stacked temporal convolutional layers are employed to handle spatial dependence and temporal dynamics, respectively.

(9) **T-GCN** (Zhao et al., 2020a). This method utilizes GRU to capture the time-varying characteristics of traffic data and applies GCN to explore the spatial correlation.

(10) **TGC-LSTM** (Cui et al., 2020). In this method, a traffic graph convolution operation based on GCN is proposed and stacked with LSTM.

(11) **PVCGN** (Liu et al., 2020). This model constructs three topological graphs (including a physical graph, a similarity graph, and a correlation graph) to explore the comprehensive spatial correlations in the metro system. Then, a physical-virtual collaboration graph network (PVCGN) is proposed to predict network-scale passenger flow.

All the experiments are conducted on a Windows 10 workstation (CPU: Intel Core (TM) i9-9900K @ 3.6GHz, RAM: 32GB random-access memory, GPU: NVIDIA GTX 2080Ti with 11GB memory) with Python 3.6.10. We implement the proposed SARGCN model with an open-source graph learning framework, i.e., deep graph library (DGL) (Wang et al., 2019), and utilize MXNet (Chen et al., 2015) as the backend.

Table 3 to Table 5 show the quantitative comparisons between SARGCN and baselines in the Shenzhen metro system. The prediction performances shown in Table 3 consist of the whole ridership, so we also separate the prediction metrics of inflow and outflow into Table 4 and Table 5, respectively. Here, the performances shown in Table 4 and Table 5 correspond to the metrics in Table 3. Furthermore, we illustrate the ground-truth passenger flow and the predicted values of three typical stations in Figure 9.

Table 3 Quantitative comparison of the whole ridership in Shenzhen metro system

	1-step			4-step			7-step			10-step		
	RMSE	MAE	MAPE@10	RMSE	MAE	MAPE@10	RMSE	MAE	MAPE@10	RMSE	MAE	MAPE@10
HA	47.32	22.62	30.07%	47.32	22.62	30.07%	47.32	22.62	30.07%	47.32	22.62	30.07%
MLP	23.52	14.03	16.70%	34.20	19.30	25.74%	47.48	25.19	33.41%	59.87	30.61	38.78%
LSTM	23.73	13.82	18.11%	31.22	17.16	25.98%	40.80	20.78	35.07%	48.47	23.76	36.73%
GRU	23.93	14.07	19.24%	33.84	18.19	32.37%	45.45	22.46	47.16%	50.09	24.98	39.98%
GCN	23.35	13.80	15.57%	30.49	17.63	18.10%	42.63	24.36	26.37%	50.22	26.99	25.66%
GAT	20.24	12.41	21.38%	26.27	15.52	26.51%	33.03	18.85	31.62%	39.81	21.99	36.44%
STGCN	25.25	14.21	15.29%	34.98	18.80	19.51%	50.73	24.97	28.68%	67.79	32.07	41.65%
Graph-WaveNet	20. 63	12.11	14.16%	24.47	13.29	15.42%	30.46	14.64	17.13%	32.09	15.49	18.84%
T-GCN	23.41	14.09	16.05%	28.11	16.32	19.40%	30.28	17.24	20.10%	32.18	18.75	23.16%
TGC-LSTM	22.59	14.22	18.56%	24.55	15.00	19.47%	28.33	16.32	23.08%	29.86	16.39	20.80%
PVCGN	22.63	12.76	14.27%	23.77	13.39	15.37%	24.46	13.64	16.39%	25.36	**13.99**	17.39%
SARGCN	**18.14**	**11.24**	**13.49%**	**20.85**	**12.52**	**15.10%**	**23.35**	**13.34**	**15.46%**	**24.51**	14.22	**17.34%**

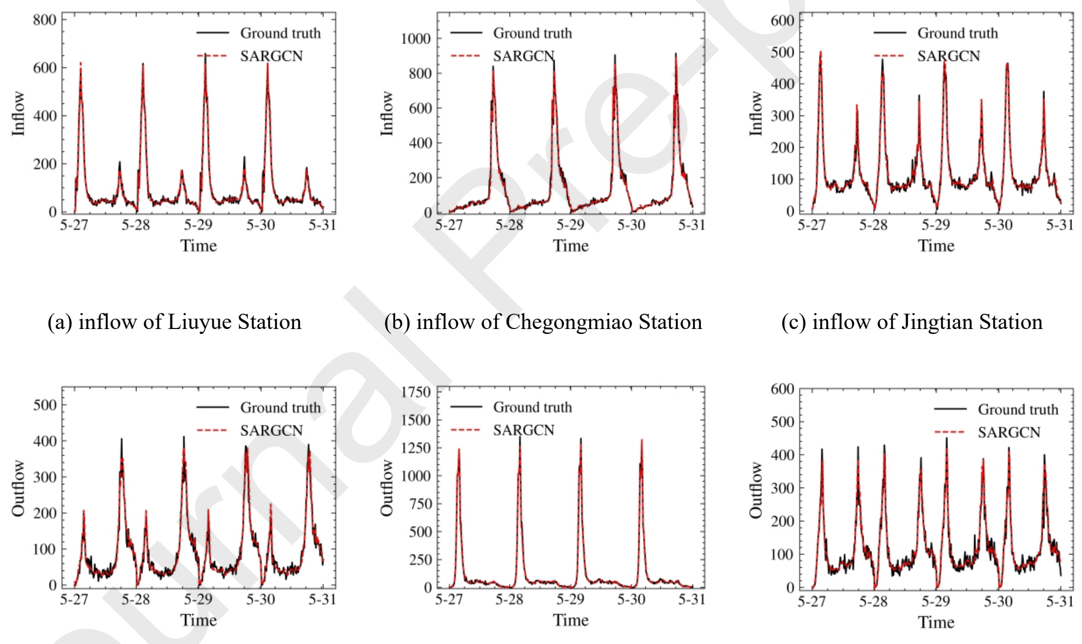
Table 4 Quantitative comparison of inflow in Shenzhen metro system

	1-step			4-step			7-step			10-step		
	RMSE	MAE	MAPE@10	RMSE	MAE	MAPE@10	RMSE	MAE	MAPE@10	RMSE	MAE	MAPE@10
HA	46.32	22.77	29.86%	46.32	22.77	29.86%	46.32	22.77	29.86%	46.32	22.77	29.86%
MLP	21.50	13.17	16.65%	34.47	19.56	28.54%	48.26	25.74	35.12%	60.17	30.99	39.72%
LSTM	21.36	12.82	18.73%	30.59	16.75	28.43%	40.77	20.61	33.83%	48.64	23.65	35.04%
GRU	21.48	13.01	19.77%	32.85	17.69	33.48%	44.51	22.10	43.77%	49.71	24.72	36.68%
GCN	21.37	12.79	15.05%	31.25	17.52	18.82%	44.37	24.69	27.73%	52.74	27.52	27.18%
GAT	18.55	11.49	14.03%	26.47	15.22	18.06%	34.77	19.28	20.65%	41.90	22.47	23.55%
STGCN	21.39	12.90	14.74%	33.35	17.64	19.01%	49.62	24.06	30.32%	69.17	31.94	45.74%
Graph-WaveNet	18.65	10.98	13.60%	22.99	12.48	**14.90%**	27.70	13.87	17.16%	28.76	14.77	19.36%
T-GCN	19.97	12.63	16.05%	26.70	15.58	19.40%	27.52	15.92	20.10%	31.44	18.29	23.16%
TGC-LSTM	21.52	13.79	18.98%	24.24	14.57	20.62%	28.71	15.91	23.94%	29.64	15.79	21.31%
PVCGN	20.99	11.74	13.75%	23.72	12.86	15.31%	24.82	13.21	16.39%	26.02	**13.60**	**17.14%**
SARGCN	**16.46**	**10.40**	**13.27%**	**20.58**	**12.17**	15.98%	**23.49**	**12.79**	**15.15%**	**24.35**	13.84	17.33%

These tables show that in terms of prediction accuracy on both the 1-step and multistep, SARGCN always expresses superior performances to the baselines. For instance, compared with the most accurate baseline in Table 3, SARGCN improves the RMSE by 10.38%, 12.28%, 4.54%, and 3.35% on these four steps, respectively. These results demonstrate that the proposed SARGCN model has a powerful capability to explore the spatiotemporal dependencies on the metro system. A comparison of the prediction results and ground truth (shown in Figure 9) shows that the time-varying patterns of passenger flow vary from station to station. And the proposed SARGCN model is effective in capturing these temporal dynamics in different time-varying patterns. Furthermore, compared with the baselines without spatial correlations (i.e., HA, MLP, LSTM, and GRU), almost all the GNN-based models perform higher accuracies in the 1-step prediction task and stabilities in multistep prediction tasks. This phenomenon further proves the importance of spatial correlations in metro passenger flow prediction. In summary, these comparisons can indicate the superior

Table 5 Quantitative comparison of outflow in Shenzhen metro system

	1-step			4-step			7-step			10-step		
	RMSE	MAE	MAPE@10	RMSE	MAE	MAPE@10	RMSE	MAE	MAPE@10	RMSE	MAE	MAPE@10
HA	48.30	22.48	30.29%	48.30	22.48	30.29%	48.30	22.48	30.29%	48.30	22.48	30.29%
MLP	25.35	14.88	16.75%	33.92	19.05	22.88%	46.69	24.64	31.66%	59.56	30.23	37.82%
LSTM	25.87	14.82	17.48%	31.82	17.56	23.47%	40.83	20.94	36.34%	48.31	23.87	38.45%
GRU	26.13	15.11	18.68%	34.79	18.67	31.23%	46.37	22.82	50.63%	50.47	25.23	43.37%
GCN	25.17	14.81	16.10%	29.72	17.75	17.36%	40.83	24.03	24.98%	47.58	26.45	24.10%
GAT	21.79	13.31	15.01%	26.07	15.82	16.82%	31.21	18.42	19.16%	37.62	21.51	20.26%
STGCN	26.78	15.51	15.86%	36.53	19.94	20.03%	51.81	25.88	27.00%	66.38	32.19	37.47%
Graph-WaveNet	22.42	13.23	14.74%	25.84	14.09	15.96%	32.96	15.39	17.11%	35.08	16.21	18.31%
T-GCN	26.38	15.53	16.32%	29.43	17.05	17.96%	32.79	18.55	20.37%	32.90	19.21	20.76%
TGC-LSTM	23.61	14.65	18.13%	24.85	15.43	18.29%	27.95	16.73	22.21%	30.08	16.99	20.28%
PVCGN	24.15	13.77	14.81%	23.83	13.92	15.44%	24.03	14.06	16.38%	24.69	**14.37**	17.65%
SARGCN	**19.66**	**12.07**	**13.72%**	**21.10**	**12.88**	**14.19%**	**23.22**	**13.89**	**15.78%**	**24.67**	14.59	**17.34%**



(a) inflow of Liuyue Station (b) inflow of Chegongmiao Station (c) inflow of Jingtian Station

(d) outflow of Liuyue Station (e) outflow of Chegongmiao Station (f) outflow of Jingtian Station

Figure 9 Comparison of prediction results and ground truth based on SARGCN model

performance of SARGCN on the Shenzhen metro dataset.

5.3.2 Hangzhou metro system

To evaluate the prediction performance of SARGCN model on the Hangzhou metro system, we directly introduce the experimental results in the previous study (Liu et al., 2020). The prediction performance in the Hangzhou metro system is summarized in Table 6. In the baselines, there are three traditional time series models, three general deep learning models, and six recently-proposed graph networks (i.e., ASTGCN (Guo et al., 2019), STG2Seq (Bai et al., 2019a), DCRNN (Li et al., 2017a), GCRNN, Graph-WaveNet (Wu et al., 2019), PVCGN (Liu et al., 2020)). From this table, we can find that the proposed SARGCN model can achieve the lowest RMSE and MAE at 15-min among all the state-of-the-art methods. Although the MAE of PVCGN is lower than SARGCN with the prediction horizontal increases, our SARGCN model always shows superior performance on RMSE. Meanwhile, compared with the MAE of SARGCN, PVCGN reduces 0.55% (30-min), 0.86% (45-min), and 1.42% (60-min). However, compared with SARGCN, PVCGN increases the RMSE by 3.99%, 3.91%, and 2.45% for 30-min, 45-min, and 60-min, respectively. According to these experimental results, the improvement of SARGCN in RMSE is higher than the decrease in MAE. Hence, these comparisons indicate that the proposed SARGCN model can achieve higher prediction accuracy. However, since our SARGCN model lacks the Seq2Seq structure (Sutskever et al., 2014), its prediction performance will decrease to a certain extent in long-scale prediction.

Table 6 Quantitative comparison of the whole ridership in Hangzhou metro system

	15 min			30 min			45 min			60 min		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
HA	64.19	36.37	19.14%	64.10	36.37	19.31%	63.92	36.23	19.57%	63.72	35.99	20.01%
RF	53.52	32.19	18.34%	64.54	38.00	21.46%	80.06	45.78	26.51%	94.29	52.95	37.12%
GBDT	51.50	30.88	17.60%	61.94	36.48	20.49%	76.70	44.12	25.75%	91.21	51.l0	38.10%
MLP	46.55	26.57	16.26%	47.96	27.44	17.10%	50.66	28.79	19.01%	54.62	30.52	22.56%
LSTM	45.30	25.76	14.91%	45.52	26.01	15.10%	46.30	26.38	15.40%	47.53	26.76	16.34%
GRU	45.10	25.69	15.13%	45.26	25.93	15.35%	46.13	26.36	15.79%	47.69	26.98	17.20%
ASTGCN	46.19	27.34	15.05%	46.16	27.74	15.56%	46.79	28.20	16.48%	49.70	28.85	17.75%
STG2Seq	39.52	23.80	17.09%	40.72	24.72	19.51%	43.36	25.98	23.59%	46.05	26.50	27.93%
DCRNN	40.39	23.76	14.00%	42.57	25.22	14.99%	46.26	26.97	16.19%	49.35	28.47	18.16%
GCRNN	40.24	23.84	14.08%	41.95	25.14	14.86%	45.53	26.82	16.05%	50.28	28.75	17.89%
Graph-WaveNet	40.78	24.07	14.27%	42.80	25.48	15.23%	45.84	27.15	17.34%	49.89	29.14	19.37%
PVCGN	37.76	22.68	**13.70%**	39.34	**23.33**	**13.81%**	40.95	**24.22**	**14.45%**	42.61	**24.93**	**15.49%**
SARGCN	**36.22**	**22.48**	13.94%	**37.83**	23.46	14.99%	**39.41**	24.43	16.25%	**41.59**	25.29	17.60%

In addition to the prediction performance, we further compare the computational costs between SARGCN and PVCGN on the Hangzhou dataset in Table 7. These comparisons are conducted on the same workstation and batch size to ensure fairness. This table indicates that SARGCN can reduce computational costs significantly. In particular, the number of parameters in SARGCN is only 3.51% of those in PVCGN. This phenomenon demonstrates that the proposed SARGCN model is much lighter than the state-of-the-art baseline. Furthermore, concerning the computational efficiency, SARGCN is just 49.1% of PVCGN in the average training time of each epoch. Hence, we can conclude that SARGCN can reduce training costs in terms of both the number of parameters and training efficiency. Therefore, it is more friendly and suitable for real-world applications in metro management.

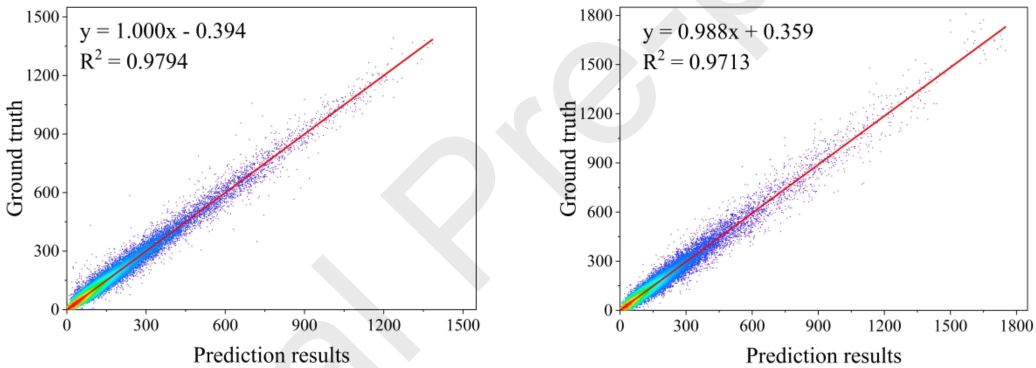
Table 7 Computational efficiency comparisons on the Hangzhou metro dataset

	PVCGN	SARGCN
Parameter amount	3.76×10^7	1.32×10^6
GPU occupation	7655 MiB	4859 MiB
Average training time	22.88 s/epoch	11.23 s/epoch

5.4 Error distribution analysis

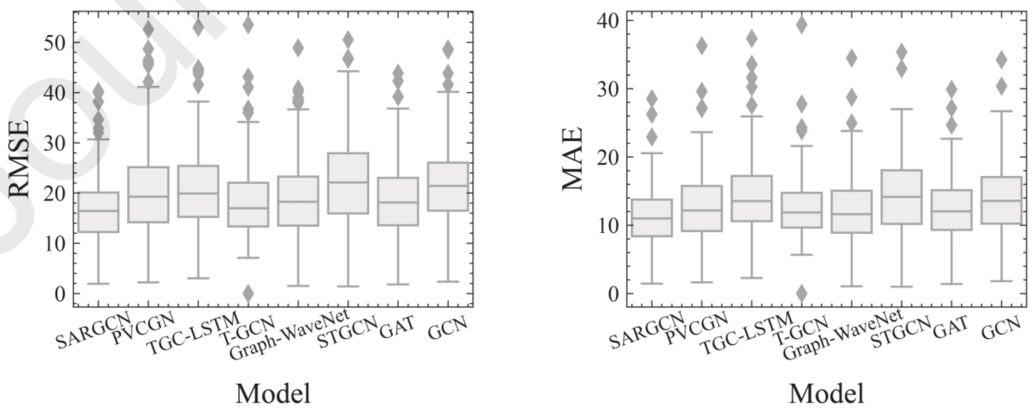
The evaluation metrics shown in the above tables measure the prediction performance in terms of average errors. To further validate the prediction accuracy, it is also essential to explore the error distributions of each data point and station. Figure 10 expresses the actual distribution of ground truth and prediction results of both inflow and outflow. As shown in these figures, the slopes of the fitted lines are close to 1, thus indicating that the predicted values can effectively match the ground truth. Although there are still errors, these data points are closely fitted and evenly distributed on both sides of the fitted line. Meanwhile, the heatmaps show that the metro passenger flow is concentrated at a lower level, and the quantity of data points decreases significantly with increasing passenger flow. Overall, regardless of whether the passenger flow is large or small, SARGCN can achieve reliable and accurate predictions.

In addition, we further explore the error distributions among all the metro stations. Different travel patterns of each station may lead to different prediction performances. Thus, we employ RMSE and MAE of each station to illustrate boxplots in Figure 11(a) and Figure 11(b), respectively. These two figures indicate that there are fewer abnormal data in SARGCN, and its error indicators of boxplot are generally lower than other models. This phenomenon reveals that SARGCN can always perform more accurately on each metro station and reduce outliers.



(a) inflow (b) outflow

Figure 10 Scatters and fitting of prediction results and ground truth of SARGCN



(a) RMSE (b) MAE

Figure 11 Boxplot of RMSE and MAE for all the GNN-based models on the Shenzhen metro system

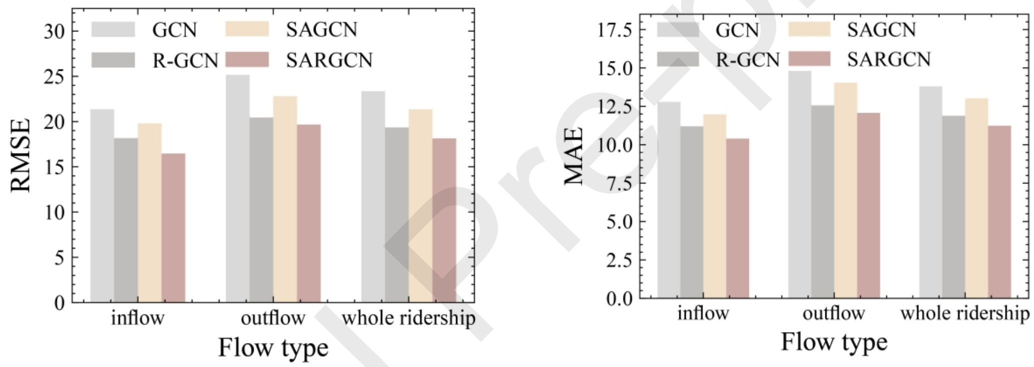
5.5 Ablation experiments

According to the model structure, R-GCN (or the knowledge graph) and split-attention mechanism are the vital components of the SARGCN model. In this subsection, we further conduct an ablation experiment to evaluate the importance of these two components.

The prediction performances of GCN, R-GCN, SAGCN (i.e., split-attention graph convolutional network), and SARGCN model in the Shenzhen metro system are illustrated in Figure 12, and differences among these models are displayed in Table 8. From the prediction results shown in the following figures, several conclusions can be summarized.

Table 8 Differences of models in ablation experiment

Category	Knowledge graph	Split-attention
GCN	×	×
R-GCN	√	×
SAGCN	×	√
SARGCN	√	√



(a) RMSE (b) MAE

Figure 12 Ablation experiment results of SARGCN on the Shenzhen metro system

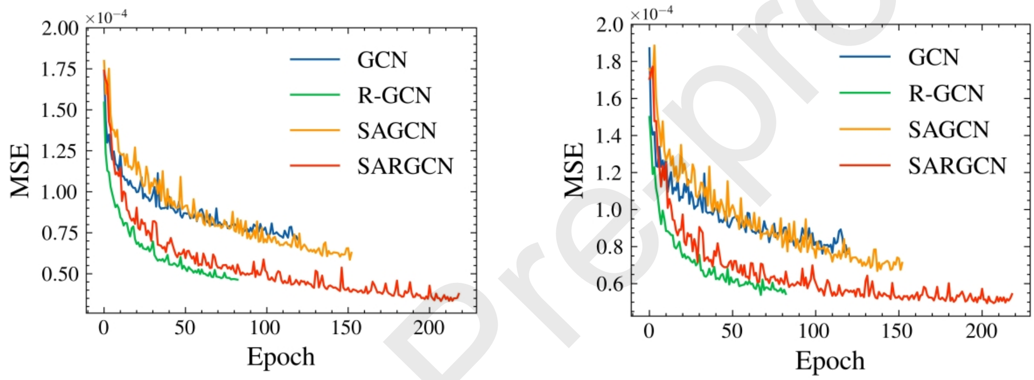
(1) It is evident that R-GCN and SARGCN significantly outperform GCN and SAGCN, respectively. For the whole ridership, the RMSE of R-GCN decreases by 17.15% compared with GCN on the 1-step prediction task, and SARGCN achieves a 15.10% improvement than SAGCN. Therefore, we can conclude that the knowledge graph plays a vital role in metro passenger flow prediction and dramatically improves prediction performance.

(2) From the prediction comparison of R-GCN and SARGCN on the whole ridership, we can find that RMSE and MAE decrease by 6.23% and 5.44%, respectively. And from the comparison between GCN and SAGCN, the improvement reaches 8.51% and 5.67%. Meanwhile, comparing SARGCN and R-GCN models on the inflow and outflow, the evaluation metrics express that SARGCN can achieve improvement on both inflow and outflow. This is because the split-attention mechanism can explore both the temporal dynamics and the dependencies between inflow and outflow. Many studies note that the group convolution operation can reduce the number of parameters. Although we replace all the convolution operations in the split-attention mechanism with R-GCN operation, SARGCN can also achieve parameters reducing by 11.29% (from 829,006 in R-GCN to 735,372 in SARGCN), thereby leading to a lightweight model.

(3) Since the prediction accuracy of SARGCN further outperforms its components, the results of the ablation experiment also indicate that the combination of knowledge graph and split-attention

mechanism is helpful in improving prediction performance.

Furthermore, Figure 13 illustrates the training loss and validation loss during training on the Shenzhen metro system. Because we apply the early-stop strategy on these deep learning models to avoid overfitting, there is a difference in the epochs when they stop training. Since GCN and SAGCN overlook the information on the established knowledge graph, their MSEs are always much higher than that of R-GCN and SARGCN. For the models with the split-attention mechanism (i.e., SAGCN and SARGCN), although their convergence speeds are slower than other models, they can avoid precocious convergence effectively. Hence, they can search for more appropriate parameters to obtain higher accuracies than GCN and R-GCN. According to the principles of SAGCN and SARGCN, the main reason for precocious convergence is that there are several parallel models in the split-attention mechanism. The prediction performance of SARGCN and SAGCN not only relies on the prediction results of each parallel model but also depends on the aggregation of these parallel models. Therefore, searching for the appropriate parameters for each model and the aggregators will increase the training time, but this increased training time cost can improve prediction performance.



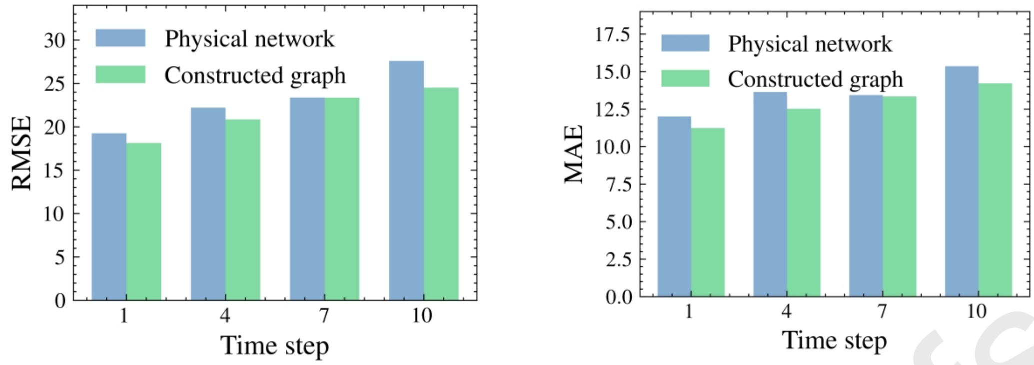
(a) training loss (b) validation loss

Figure 13 Performance comparison of SARGCN and its corresponding models at different training epochs

5.6 Graph construction method analysis

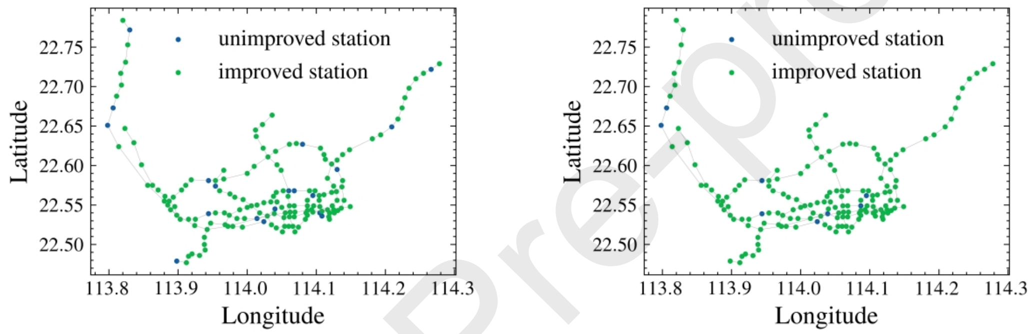
To validate the effectiveness of the graph construction method, we compare the prediction performance of the physical metro network and constructed directed graph of the Shenzhen metro system. The comparison results are displayed in Figure 14. Here, both of these graphs are transformed to knowledge graphs. To distinguish them, we name these two graphs as the physical network and constructed graph, respectively. The results show that the SARGCN based on the constructed graph consistently outperforms that on the physical network. Concerning the constructed graph, the RMSE and MAE of 1-step prediction are reduced by 5.80% and 6.42%, respectively. Moreover, Figure 15(a) and Figure 15(b) show that most metro stations can achieve more accurate prediction performance on the constructed graph. Specifically, the percentages of stations with a reduced RMSE and MAE are 88.6% and 95.2%, respectively.

Besides, we also explore the distributions of the reduced prediction errors in Figure 15(c) and Figure 15(d). According to these figures, the reduced prediction errors similarly follow the normal distribution, and the average improvement is 1.07 for RMSE and 0.77 for MAE. These improvements suggest that the constructed graph has a more powerful capability to capture the spatial dependence of the metro system, and combining it with the proposed SARGCN model is an appropriate direction to improve metro passenger flow prediction performance.

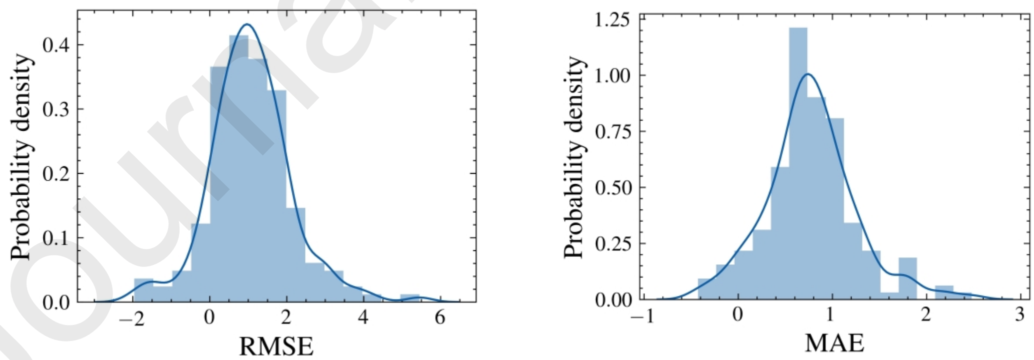


(a) RMSE (b) MAE

Figure 14 Performance comparison of SARGCN on the constructed graph and physical network in the Shenzhen metro system



(a) spatial distributions of metro stations with reduced RMSE (b) spatial distributions of metro stations with reduced MAE



(c) histogram distribution of reduced RMSE (d) histogram distribution of reduced MAE

Figure 15 Distribution of reduced prediction errors by the metro graph construction method in the Shenzhen metro system

Moreover, we further analyze the OD passenger flow distributions on these two graphs, and the results are illustrated in Figure 16. In this figure, the OD passenger flows on the self-loop edges are not involved. It is evident that OD passenger flows on the connected edges of the physical network are significantly lower than those of the constructed graph. In other words, there are more frequent passenger flow interactions between the adjacent stations on the constructed graph, while the physical network always ignores these features. Therefore, since the constructed graph can explore this dependence from travel behaviors, it can obtain higher prediction accuracy.

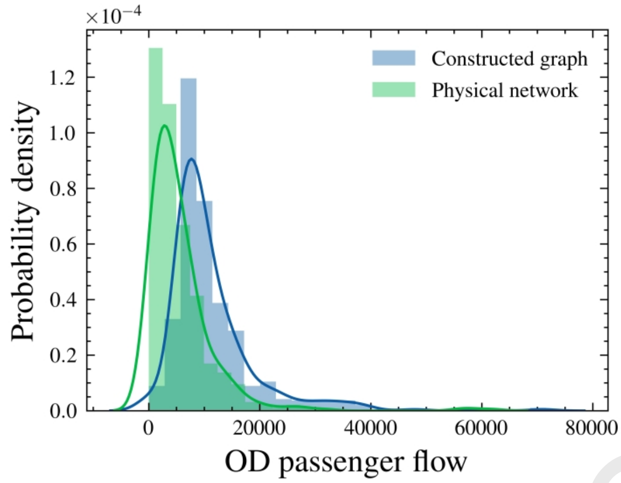
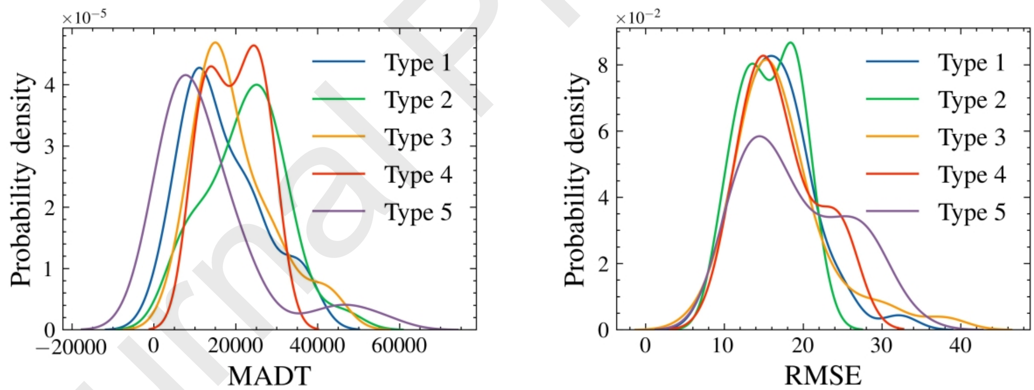


Figure 16 OD passenger flow distribution of the connected edges on the constructed graph and physical network in the Shenzhen metro system

5.7 Knowledge graph analysis

The prediction performance comparisons in Section 5.5 demonstrate that the established knowledge graph positively impacts prediction accuracies. In this subsection, we further explore the properties of the knowledge graph.



(a) probability density of MADT (b) probability density of RMSE

Figure 17 Traffic features distributions between different types of stations in the Shenzhen metro system

In the established knowledge graph, the proportion of metro stations with different semantic types is 24.7%: 15.7%: 22.9%: 10.8%: 25.9%. Among all the types, stations with type 1 (i.e., resident area) account for the largest proportion. Moreover, Figure 17(a) and Figure 17(b) illustrate the probability density distribution of monthly average day traffic (MADT) and RMSE under different semantic types. According to the MADT distribution, each type has its unique characteristics, and these differences can be effectively distinguished by the knowledge graph construction method. Meanwhile, an interesting finding is obtained from the probability density distribution of MADT and RMSE. For instance, the MADT distribution of type 5 is always lower than that of type 2. However, as the RMSE distribution comparison shows, the probability density of type 5 is much higher than type 2 when RMSE is higher than 20. The reason may be that type 5 includes many transfer nodes with other modes of transportation (e.g., buses, railways, and airplanes). Hence, passenger flows at these stations mainly originate from other transportation modes. The imbalanced schedules of other transportation modes easily produce randomness and instabilities in metro passenger flows. Therefore, it brings significant challenges for accurate prediction, leading to higher prediction errors. This finding reveals that the knowledge graph can capture and distinguish the travel patterns of each station type to promote spatial correlation mining.

6. Conclusion

This study proposes a deep learning framework named split-attention relational graph convolutional network (SARGCN) to address the network-scale metro passenger flow prediction. Unlike previous studies, which directly apply the physical metro network for GNNs, we develop a metro knowledge graph construction method to adapt the travel behavior. Specifically, the historical OD matrix is extracted and employed as the similarity measure to construct the metro topological graph. Then, we utilize the land-use features to represent the semantic types of each station, aiming to establish a knowledge graph based on the constructed directed graph. To further explore the spatiotemporal dependencies on the established knowledge graph, we propose the SARGCN model, by integrating the R-GCN, split-attention mechanism, and LSTM. Validated on the Shenzhen and Hangzhou metro system, SARGCN expresses superiority compared to widely-used baselines and state-of-the-art methods.

However, this study still has several limitations. For instance, various payment methods, such as smart cards and mobile payments, have been developed for metro systems in recent years. However, due to the barriers in research data collection, this study only extracts the smart card data for passenger flow analysis and prediction. Since passenger travel behavior may differ by payment method, fully considering these differences might enhance spatiotemporal correlation analysis. Moreover, many studies demonstrate that external information, such as weather and special events (Xue et al., 2022), also impacts the metro passenger flow, which is not involved in this study.

According to the limitations and challenges of this study, the following suggestions might be interesting directions for future work.

(1) Metro is a critical component in the urban transit system, and its passengers always source from other transportation modes, e.g., buses, bike-sharing, etc. So, introducing the real-time passenger distributions of other transportation modes may improve prediction performance.

(2) In addition to predicting the inflow and outflow, OD passenger flow prediction (Dai et al., 2018; Hussain et al., 2021; Zhang et al., 2021) is also a hot topic in this field. From the essence of these two passenger flows, the former denotes the number of passengers entering and exiting the metro system, and the latter reflects the passenger flow direction and evolution process within this system. Future works can further explore the dependence between these two flows and integrate these two tasks to improve prediction performance.

(3) An increasing number of researchers have paid attention to inductive learning tasks in traffic prediction (Wu et al., 2020). Graph neural networks with inductive learning ability can be applied to different topology networks and achieve acceptable performance. Since metros are rapidly constructed and developed, a robust model with strong generalization to different topological graphs is needed. Thus, inductive learning has excellent potential in metro passenger flow prediction.

Appendix

The important abbreviations used in this study is summarized as follows.

Table 9 The important abbreviations used in this paper

Abbreviations	Description
GCN	Graph convolutional network.
GNN	Graph neural network.
LSTM	Long short-term memory network.
MADT	Monthly average day traffic of passenger flow.
MAE	Mean absolute error, mathematically expressed by Equation 24.
MAPE@10	Mean absolute percentage error (MAPE) on metro stations with the top 10% largest passenger flow.
OD matrix	Origin-destination (OD) matrix.
POI	Point of interest data, which represents the land-use characteristics in the urban area.
R-GCN	Relational graph convolutional network.
RMSE	Root mean square error, mathematically expressed by Equation 23.
SAGCN	Split-attention graph convolutional network, which use the GCN layer to replace the R-GCN layer in SARGCN.
SARGCN	Split-attention relational graph convolutional network proposed by this study.

Acknowledgments

This research is funded by the National Natural Science Foundation of China (No. 52172310), Humanities and Social Sciences Foundation of the Ministry of Education (No.21YJCZH147), Innovation-Driven Project of Central South University (No.2020CX041).

Reference

Agrawal, A., Udmale, S. S., & Sambhe, V. K. (2018). Extended Four-Step Travel Demand Forecasting Model for Urban Planning. *Lecture Notes in Networks and Systems*, *10*(January), 191–198. https://doi.org/10.1007/978-981-10-3920-1_19

Bai, L., Yao, L., Kanhere, S. S., Wang, X., & Sheng, Q. Z. (2019a). StG2seq: Spatial-temporal graph to sequence model for multi-step passenger demand forecasting. *IJCAI International Joint Conference on Artificial Intelligence*, *2019-Augus*, 1981–1987. https://doi.org/10.24963/ijcai.2019/274

Bai, S., Zico Kolter, J., & Koltun, V. (2019b). Trellis networks for sequence modeling. *7th International Conference on Learning Representations, ICLR 2019*, *103*, 1–18.

Chen, E., Ye, Z., Wang, C., & Xu, M. (2020). Subway Passenger Flow Prediction for Special Events Using Smart Card Data. *IEEE Transactions on Intelligent Transportation Systems*, *21*(3), 1109–1120. https://doi.org/10.1109/TITS.2019.2902405

Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., & Zhang, Z. (2015). MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems. *ArXiv*, 1–6. http://arxiv.org/abs/1512.01274

Cui, Z., Henrickson, K., Ke, R., & Wang, Y. (2020). Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, *21*(11), 4883–4894.

Cupertino, T. H., Huertas, J., & Zhao, L. (2013). Data clustering using controlled consensus in complex networks. *Neurocomputing*, *118*, 132–140. https://doi.org/10.1016/j.neucom.2013.02.026

Dai, X., Sun, L., & Xu, Y. (2018). Short-Term Origin-Destination Based Metro Flow Prediction with Probabilistic Model Selection Approach. *Journal of Advanced Transportation*, *2018*, 1–15. https://doi.org/10.1155/2018/5942763

Daniel, O., García-palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, *34*, 548–558. https://doi.org/10.1016/j.apgeog.2012.01.005

Ding, C., Duan, J., Zhang, Y., Wu, X., & Yu, G. (2018). 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), 1054–1064. https://doi.org/10.1109/TITS.2017.2711046

Ding, C., Wang, D., Ma, X., & Li, H. (2016). Predicting Short-Term Subway Ridership and Prioritizing Its Influential Factors Using Gradient Boosting Decision Trees. *Sustainability*, *8*(11), 1100. https://doi.org/10.3390/su8111100

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research*, *9*, 249–256.

Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 922–929. https://doi.org/10.1609/aaai.v33i01.3301922

Han, Y., Wang, S., Ren, Y., Wang, C., Gao, P., & Chen, G. (2019). Predicting Station-Level Short-Term Passenger Flow in a Citywide Metro Network Using Spatiotemporal Graph Convolutional Neural Networks. *ISPRS International Journal of Geo-Information*, *8*(6), 243. https://doi.org/10.3390/ijgi8060243

Hao, S., Lee, D.-H., & Zhao, D. (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*, 287–300. https://doi.org/10.1016/j.trc.2019.08.005

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

He, Y., Zhao, Y., & Tsui, K. L. (2020). An adapted geographically weighted LASSO (Ada-GWL) model for predicting subway ridership. *Transportation*, 1–32. https://doi.org/10.1007/s11116-020-10091-2

Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., de Melo, G., Gutierrez, C., Gayo, J. E. L., Kirrane, S., Neumaier, S., Polleres, A., Navigli, R., Ngomo, A.-C. N., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2020). Knowledge Graphs. *ArXiv*. http://arxiv.org/abs/2003.02320

Hussain, E., Bhaskar, A., & Chung, E. (2021). Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. *Transportation Research Part C: Emerging Technologies*, *125*, 103044. https://doi.org/10.1016/j.trc.2021.103044

Jun, M., Choi, K., Jeong, J., Kwon, K., & Kim, H. (2015). Land use characteristics of subway catchment areas and their in fl uence on subway ridership in Seoul. *Journal of Transport Geography*, *48*, 30–40. https://doi.org/10.1016/j.jtrangeo.2015.08.002

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *ArXiv Preprint ArXiv:1609.02907*.

Krause, C. M., & Zhang, L. (2019). Short-term travel behavior prediction with GPS, land use, and point of interest data. *Transportation Research Part B: Methodological*, *123*, 349–361. https://doi.org/10.1016/j.trb.2018.06.012

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Li, H., Wang, Y., Xu, X., Qin, L., & Zhang, H. (2019). Short-term passenger flow prediction under passenger flow control using a dynamic radial basis function network. *Applied Soft Computing*, *83*, 105620. https://doi.org/10.1016/j.asoc.2019.105620

Li, Yaguang, Yu, R., Shahabi, C., & Liu, Y. (2017a). Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. *ArXiv*. http://arxiv.org/abs/1707.01926

Li, Yang, Wang, X., Sun, S., Ma, X., & Lu, G. (2017b). Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transportation Research Part C: Emerging Technologies*, *77*, 306–328. https://doi.org/10.1016/j.trc.2017.02.005

Lin, C., Wang, K., Wu, D., & Gong, B. (2020). Passenger Flow Prediction Based on Land Use around Metro Stations: A Case Study. *Sustainability*, *12*(17), 6844. https://doi.org/10.3390/su12176844

Lin, E., Park, J. D., & Züfle, A. (2017). Real-Time Bayesian Micro-Analysis for Metro Traffic Prediction. *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*, 1–4. https://doi.org/10.1145/3152178.3152190

Liu, L., Chen, J., Wu, H., Zhen, J., Li, G., & Lin, L. (2020). Physical-Virtual Collaboration Modeling for Intra- and Inter-Station Metro Ridership Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 1–15. https://doi.org/10.1109/TITS.2020.3036057

Liu, Y., Liu, Z., & Jia, R. (2019). DeepPF: A deep learning based architecture for metro passenger flow prediction. *Transportation Research Part C: Emerging Technologies*, *101*, 18–34. https://doi.org/10.1016/j.trc.2019.01.027

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, *54*, 187–197.

Ma, X., Zhang, J., Du, B., Ding, C., & Sun, L. (2019). Parallel Architecture of Convolutional Bi-Directional LSTM Neural Networks for Network-Wide Metro Ridership Prediction. *IEEE Transactions on Intelligent Transportation Systems*, *20*(6), 2278–2288. https://doi.org/10.1109/TITS.2018.2867042

McNally, M. G. (2007). *The Four-Step Model* (pp. 35–53). https://doi.org/10.1108/9780857245670-003

Ning, Y., Huang, Y., Li, J., Liu, Q., Yang, D., Zheng, W., & Liu, H. (2018). ST-DRN: Deep Residual Networks for Spatio-Temporal Metro Stations Crowd Flows Forecast. *2018 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN.2018.8489234

Ou, J., Sun, J., Zhu, Y., Jin, H., Liu, Y., Zhang, F., Huang, J., & Wang, X. (2020). STP-TrellisNets: Spatial-Temporal Parallel TrellisNets for Metro Station Passenger Flow Prediction. *International Conference on Information and Knowledge Management, Proceedings*, 1185–1194. https://doi.org/10.1145/3340531.3411874

Roos, J., Bonnevay, S., & Gavin, G. (2017). Dynamic Bayesian networks with Gaussian mixture models for short-term passenger flow forecasting. *Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2017*, *2018-Janua*, 1–8. https://doi.org/10.1109/ISKE.2017.8258756

Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., & Welling, M. (2018). Modeling Relational Data with Graph Convolutional Networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10843 LNCS*(1), 593–607. https://doi.org/10.1007/978-3-319-93417-4_38

Sun, Yujuan, Zhang, G., & Yin, H. (2014). Passenger Flow Prediction of Subway Transfer Stations Based on Nonparametric Regression Model. *Discrete Dynamics in Nature and Society*, *2014*, 1–8. https://doi.org/10.1155/2014/397154

Sun, Yuxing, Leng, B., & Guan, W. (2015). A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing*, *166*, 109–121. https://doi.org/10.1016/j.neucom.2015.03.085

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, *4*(January), 3104–3112.

Tang, J., Wang, X., Zong, F., & Hu, Z. (2020). Uncovering Spatio-temporal Travel Patterns Using a Tensor-based Model from Metro Smart Card Data in Shenzhen, China. *Sustainability*, *12*(4), 1475. https://doi.org/10.3390/su12041475

Tang, L., Zhao, Y., Cabrera, J., Ma, J., & Tsui, K. L. (2019a). Forecasting Short-Term Passenger Flow: An Empirical Study on Shenzhen Metro. *IEEE Transactions on Intelligent Transportation Systems*, *20*(10), 3613–3622. https://doi.org/10.1109/TITS.2018.2879497

Tang, Q., Yang, M., & Yang, Y. (2019b). ST-LSTM: A Deep Learning Approach Combined Spatio-Temporal Features for Short-Term Forecast in Rail Transit. *Journal of Advanced Transportation*, *2019*. https://doi.org/10.1155/2019/8392592

Velicković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2017). Graph attention networks. *ArXiv*, 1–12.

Wang, J., Zhang, Y., Wei, Y., Hu, Y., Piao, X., & Yin, B. (2021). Metro Passenger Flow Prediction via Dynamic Hypergraph Convolution Networks. *IEEE Transactions on Intelligent Transportation Systems*, 1–13. https://doi.org/10.1109/TITS.2021.3072743

Wang, M., Zheng, D., Ye, Z., Gan, Q., Li, M., Song, X., Zhou, J., Ma, C., Yu, L., Gai, Y., Xiao, T., He, T., Karypis, G., Li, J., & Zhang, Z. (2019). Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *ArXiv*, 1–18. http://arxiv.org/abs/1909.01315

Wei, Y., & Chen, M.-C. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies*, *21*(1), 148–162. https://doi.org/10.1016/j.trc.2011.06.009

Wen, K., Zhao, G., He, B., Ma, J., & Zhang, H. (2022). A decomposition-based forecasting method with transfer learning for railway short-term passenger flow in holidays. *Expert Systems with Applications*, *189*(September 2021), 116102. https://doi.org/10.1016/j.eswa.2021.116102

Wu, Y., Zhuang, D., Labbe, A., & Sun, L. (2020). Inductive Graph Neural Networks for Spatiotemporal

Kriging. *ArXiv*, 1–18. http://arxiv.org/abs/2006.07527

Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for Deep Spatial-Temporal Graph Modeling. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 1907–1913. https://doi.org/10.24963/ijcai.2019/264

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 5987–5995. https://doi.org/10.1109/CVPR.2017.634

Xue, G., Liu, S., Ren, L., Ma, Y., & Gong, D. (2022). Forecasting the subway passenger flow under event occurrences with multivariate disturbances. *Expert Systems with Applications*, *188*(October 2021), 116057. https://doi.org/10.1016/j.eswa.2021.116057

Yang, B., Sun, S., Li, J., Lin, X., & Tian, Y. (2019). Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing*, *332*, 320–327. https://doi.org/10.1016/j.neucom.2018.12.016

Ye, J., Zhao, J., Ye, K., & Xu, C. (2020). Multi-STGCnet: A Graph Convolution Based Spatial-Temporal Framework for Subway Passenger Flow Forecasting. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN48605.2020.9207049

Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *IJCAI International Joint Conference on Artificial Intelligence*, *2018-July*, 3634–3640. https://doi.org/10.24963/ijcai.2018/505

Zhang, D., & Kabuka, M. R. (2018). Combining weather condition data to predict traffic flow: a GRU-based deep learning approach. *IET Intelligent Transport Systems*, *12*(7), 578–585. https://doi.org/10.1049/iet-its.2017.0313

Zhang, H., Wu, C., Zhang, Z. Z., Zhu, Y., Lin, H., Zhang, Z. Z., Sun, Y., He, T., Jonas, M., Manmatha, R., Li, M., & Alexander, S. (2020a). ResNeSt: Split-Attention Networks. *ArXiv*.

Zhang, J., Che, H., Chen, F., Ma, W., & He, Z. (2021). Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method. *Transportation Research Part C: Emerging Technologies*, *124*, 102928. https://doi.org/10.1016/j.trc.2020.102928

Zhang, J., Chen, F., Cui, Z., Guo, Y., & Zhu, Y. (2020b). Deep Learning Architecture for Short-Term Passenger Flow Forecasting in Urban Rail Transit. *IEEE Transactions on Intelligent Transportation Systems*, 1–11. https://doi.org/10.1109/TITS.2020.3000761

Zhang, Y., Cheng, T., & Ren, Y. (2019). A graph deep learning method for short-term traffic forecasting on large road networks. *Computer-Aided Civil and Infrastructure Engineering*, *34*(10), 877–896. https://doi.org/10.1111/mice.12450

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2020a). T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, *21*(9), 3848–3858. https://doi.org/10.1109/TITS.2019.2935152

Zhao, S.-Z., Ni, T.-H., Wang, Y., & Gao, X.-T. (2011). A new approach to the prediction of passenger flow in a transit system. *Computers & Mathematics with Applications*, *61*(8), 1968–1974. https://doi.org/10.1016/j.camwa.2010.08.023

Zhao, Y., Ren, L., Ma, Z., & Jiang, X. (2020b). Novel Three-Stage Framework for Prioritizing and Selecting Feature Variables for Short-Term Metro Passenger Flow Prediction. *Transportation Research Record*, *2674*(8), 192–205. https://doi.org/10.1177/0361198120926504

Zheng, Z., Ling, X., Wang, P., Xiao, J., & Zhang, F. (2020). Hybrid model for predicting anomalous large passenger flow in urban metros. *IET Intelligent Transport Systems*, *14*(14), 1987–1996.

https://doi.org/10.1049/iet-its.2020.0054

Zhou, F., Yang, Q., Zhang, K., Trajcevski, G., Zhong, T., & Khokhar, A. (2020). Reinforced Spatiotemporal Attentive Graph Neural Networks for Traffic Forecasting. *IEEE Internet of Things Journal*, *7*(7), 6414–6428. https://doi.org/10.1109/JIOT.2020.2974494

Highlight

- This paper models the metro system as knowledge graph for passenger flow prediction.
- It combines traffic patterns and land-use features for knowledge graph construction.
- It proposes a SARGCN model for spatiotemporal prediction on metro knowledge graphs.
- It uses an attention mechanism to learn the correlation between inflow and outflow.
- Validated on two metro datasets, it outperforms numerous advanced baselines.

Jie Zeng: Conceptualization, Methodology, Software, Writing-Original draft preparation. **Jinjun Tang**: Conceptualization, Writing- Reviewing and Editing, Visualization.