Empowering Spatial Knowledge Graph for Mobile Traffic Prediction

Jiahui Gong[1], Yu Liu[1], Tong Li[1], Haoye Chai[1], Xing Wang[2], Junlan Feng[2], Lin Zhu[2], Chao Deng[2], Depeng Jin[1], Yong Li[1]
[1] Department of Electronic Engineering, Tsinghua University, Beijing, China
[2] China Mobile Research Institute, Beijing, China
tongli@mail.tsinghua.edu.cn,liyong07@tsinghua.edu.cn

ABSTRACT

Accurately predicting base station traffic volumes and understanding mobile traffic patterns is essential for smart city development, enabling efficient resource allocation and ensuring high-quality communication services. However, existing works have limitations in capturing spatial information, though the surrounding environment plays a critical role in mobile traffic prediction. In this paper, we utilize a spatial knowledge graph to represent spatial information and add important urban components to augment it making it a more effective tool for capturing environmental information. we further propose a multi-relational knowledge graph convolutional network model for mobile traffic prediction, which consists of three parts. The environmental context modelling captures spatial information from the augmented spatial knowledge graph using tucker decomposition and relational graph convolutional network. The semantic relationship modelling extracts semantic relationships between base stations and employs transformer and causal convolution to capture temporal features. The inter-attentional fusion modelling utilizes the self-attention mechanism to further capture base station relationships and predict future traffic volumes. Extensive experiments demonstrate that our proposed model significantly outperforms the state-of-the-art models by over 10% in mobile traffic prediction.

CCS CONCEPTS

• **Information systems** → **Spatial-temporal systems**; • **Networks** → *Mobile networks*; • **Computing methodologies** → *Machine learning algorithms.*

KEYWORDS

knowledge graph, mobile traffic prediction, graph neural networks

1 INTRODUCTION

The rise of smart cities has brought about a surge demand for Internet of Things (IoT) devices, which play a crucial role in various aspects of smart cities. However, the rapid proliferation of IoT has also posed significant challenges to mobile networks, putting a strain on network resources and leading to issues such as congestion and latency [37]. To address these issues, it is crucial to accurately predict mobile traffic and implement refined allocation and proactive scheduling of network resources [25]. Accurate prediction of network traffic is the key to rational allocation of energy and resources and ensures high-quality service and communication [24]. By predicting traffic patterns in advance, resources can be allocated efficiently, and proactive scheduling of network resources becomes possible, which ensures that network performance meets the demands of smart cities.

Traditional methods models treat mobile traffic prediction as a general time series prediction problem, some machine-learning models such as support vector regression (SVR) [4], and auto-regressive integrated moving average (ARIMA) [33], have been applied. Additionally, recurrent neural networks (RNNs) [39], long short-term memory (LSTM) [12], and gated recurrent units (GRU) [2] have been employed to improve prediction accuracy. Recently, several models have employed graph convolution networks (GCN) to capture the proximity relationships between base stations [3, 38, 41]. Despite progress in these areas, two fundamental limitations remain in existing models.

- **Existing models have limitations in capturing semantic relationships between base stations.** Most existing models [3, 7, 38, 41] only consider the proximity relationship between base stations, neglecting other relevant factors such as the relatedness to traffic patterns and the functional region in which the base station is located.
- **Existing research often overlooks the critical role of the surrounding environment in mobile traffic prediction.** The surrounding environment plays a critical role in mobile traffic prediction, yet it is often overlooked in existing research. For example, during rush hour in a densely populated area, mobile traffic may increase due to the high volume of commuters. By incorporating environmental factors into mobile traffic prediction models, models can gain a more comprehensive understanding of traffic patterns and improve the performance of prediction.

To mitigate such limitations in mobile traffic prediction, we first introduce the spatial knowledge graph for accurate mobile traffic prediction. A spatial knowledge graph is a type of knowledge graph that represents spatial information and the relationships between spatial entities. Specifically, we first construct the basic spatial knowledge graph consisting of four semantic relationships between base stations which could capture both spatial and temporal features. Based on this, we employ the relational graph convolution network (RGCN) [22] and transformer-based traffic encoder [28] to model the dynamic traffic series and generate the domain embeddings. Furthermore, to capture the complex interaction between base stations and the surrounding environment, we augment the basic spatial KG with regions, business areas, points of interest (POIs) and categories included, which could generate the environmental embeddings. Finally, the self-attention mechanism is applied to fuse domain embeddings and environmental embedding together for future traffic prediction. By leveraging the spatial KG and RGCN, our proposed model effectively captures the comprehensive features of base stations over time, providing more accurate predictions of mobile traffic patterns.

Our major contributions are summarized as follows:

- We are the first to propose the knowledge-driven paradigm for mobile traffic prediction and conduct a systematic study of knowledge discovery from multi-source urban data via KG construction, which identifies key elements and complex relationships in the city as entities and relationships, respectively. By leveraging the spatial KG, we can incorporate multiple types of spatial data and capture a comprehensive representation of the spatial context surrounding base stations.
- We propose a novel multi-relational knowledge graph convolutional network model for improving mobile traffic prediction. Specifically, by leveraging the spatial KG and RGCN, the model could capture various relationships between base stations and environmental features. And model could capture dynamic temporal features through the transformer. The model then employs the inter-attention mechanism to fuse these features and predict future traffic.
- We conduct extensive experiments on two real-world datasets and the proposed model outperforms state-of-the-art approaches by more than 10% on precision, which demonstrates its accuracy and effectiveness. We carried out an analysis to identify the semantic relationships and entities that could improve the prediction of base station traffic. Besides, our model has been integrated into the Jiutian platform as a crucial component of the AI-powered modules for communication modelling.

2 RELATED WORKS

2.1 Mobile Traffic Prediction

Mobile traffic prediction is a widely recognized time series prediction task, and significant efforts have been made.

Initially, models utilized sufficient historical data to learn the statistical rules of data automatically. For example, Hong et al. [14] used SVR, and Shu et al. [23] used seasonal auto-regression integrated moving average (SARIMA) for mobile data traffic prediction. However, SARIMA's reliance on past traffic volume averages makes it unable to capture fast changes or model non-linear relationships, and it requires significant effort to set suitable hyperparameters for good performance. Li et al. [17] proposed a software-defined cellular radio access network architecture, and Xu et al. [36] proposed a Gaussian Process (GP) method. However, these models focus on single base stations for predicting short-term mobile traffic and cannot be applied to large-scale mobile traffic prediction.

The recent advancements in deep learning have led to the development of various neural network models for mobile traffic prediction. Fu et al. [8] apply the LSTM and GRU to model time series data, which solely consider temporal data and disregard geographical elements, despite the fact that each node in the base station network can be influenced by other nodes. To address this limitation, convolution neural networks (CNN) and graph neural networks (GNN) [21] are applied to incorporate both temporal and geographical data. Zhang et al. [40] apply CNN and LSTM to encode spatial-temporal features, but the effectiveness of CNN is constrained in non-Euclidean base station networks.

Fang et al. [6] employ Graph Convolutional Networks (GCN) to describe the geographic dependency, where the graph's edge indicates the spatial relationship between base stations. Feng et al. [7] present an end-to-end model to acquire the spatially dependent and long-term mobile traffic, which employs LSTM for modelling complex temporal changes and correlation selection mechanism for modelling spatial relationships and encoding the external data. Wang et al. [30] propose a hybrid deep learning model, which consists of LSTM units and a unique autoencoder-based deep model for spatial-temporal modelling and prediction. Xu et al. [32] present a unique breakdown of in-cell and inter-cell data flow, and apply GNN to interact with the base stations around for mobile traffic prediction.

However, current models only utilize the distance relationship to capture spatial features, which can be imprecise due to the diverse and complex urban functional areas. Furthermore, existing studies tend to overlook the potential influence of environmental factors that could enhance the accuracy of mobile traffic prediction.

2.2 Traffic Prediction in Other Applications

There are other types of traffic series predictions that share a similar mathematical expression with mobile traffic prediction, such as road traffic prediction and network traffic prediction.

Road traffic prediction is to predict the change of future flow through the road topology map and the historical traffic flow recorded by sensors. In the road traffic prediction, Yu et al. [38] propose the STGCN model, which integrates GCN and gated CNN to capture both the temporal dependency of dynamic mobile data traffic and the topological structure of the graph. Guo et al. [9] propose a new attention-based spatial-temporal graph convolutional network (ASTGCN) model, which contains three independent components to model three temporal properties of mobile data traffic, and the three patterns are weighted fused to be the final output. Zhao et al. [41] introduce the T-GCN model, which combines GCN and GRU to capture both the spatial dependence of topology similarity and the temporal dependence of dynamic mobile data traffic change of node attribution. Wu et al. [34] propose GraphWaveNet to model the spatial-temporal dependency, which develops a novel and learnable adaptive dependency matrix through node embedding and a stacked dilated convolution is applied to expand the receptive field. Zheng et al. [42] propose a graph multi-attention network that adapts an encoder-decoder architecture, where both the encoder and the decoder consist of multiple spatial-temporal attention blocks to model the impact of the spatial-temporal factors. Diao et al. [3] propose a DGCNN model to track dynamic spatial dependencies by a dynamic Laplacian matrix estimator which could capture the stable global long-term temporal-spatial traffic relation and the local traffic fluctuations.

While the network traffic prediction is to predict the futural network traffic through the network topology and the historical traffic columns recorded by routers. In the network traffic prediction, Davide et al. [1] apply DCRNN to predict the traffic and the events of congestion, which utilizes a graph-based machine learning method to learn a representation of each node considering both its properties and the structure of the network. Laisen et al. [20] propose a reinforcement learning-based mechanism to model the traffic prediction problem as a Markov decision process, which consists of a residual-based dictionary learning algorithm to find the features of temporal factors. He et al. [11] propose a meta-learning scheme which consists of a set of predictors, each optimized to predict a particular kind of traffic, and of a master policy that is trained for choosing the best-fit predictor dynamically based on these performances.

Though road traffic prediction and network traffic prediction share the same mathematical formulation, these models could not be directly applied to mobile traffic prediction due to differences in the nature of the data and the underlying dynamics.

3 PRELIMINARY

3.1 Problem Definition

Given the historical mobile traffic series of each base station, the mobile traffic prediction problem is to develop a model to predict the traffic of all base stations in future time steps.

Mathematically, we denote a set of base stations as $B = \{b_1, b_2, ..., b_N\}$, and N is the number of base station. And the spatial knowledge graph as G. Let $x_i^t \in \mathbb{R}$ represent the traffic value of base station b_i at time step t, and $s_i^t = \left[x_i^{t-T+1}, x_i^{t-T+2}, ..., x_i^t\right] \top \in \mathbb{R}^T$ represents the past T time steps of base station b_i at time step t. Then, we denote the $S^t = \left[s_1^t, s_2^t, ..., s_N^t\right] \in \mathbb{R}^{N\times T}$ as historical traffic matrix of all base stations at time step t.

Mobile traffic prediction aims to forecast future mobile traffic based on past traffic data, which involves training a mapping function (f) using historical traffic series (S^t) and a spatial knowledge graph (G) to predict the traffic value for the next time step. The mapping function can be expressed as:

$$x_1^{t+1}, x_2^{t+1}, ..., x_N^{t+1} = f(G, S^t). \tag{1}$$

3.2 Spatial Knowledge Graph

A knowledge graph is a type of database that represents knowledge in a structured format using nodes, edges, and attributes. It is designed to capture the relationships between different entities and concepts in a particular domain or field [13]. While the spatial knowledge graph contains key components of urban areas to capture comprehensive spatial information. Formally, the spatial knowledge graph is made up of facts. The fact set includes triplets on factual knowledge as $\mathbf{F} = \{(h, r, t)|h, t \in \varepsilon, r \in R\}$, where h is the head entity, t is tail entity and r is the relation between h and t [5]. ε and $\mathbf{R}$ are sets of entities and relations. Then we define the spatial knowledge graph as $G = (\varepsilon, \mathbf{R}, \mathbf{F})$.

3.3 TuckER Decomposition

To obtain the representation of each entity in the augmented spatial knowledge graph, we utilize the widely adopted TuckER model. This model is highly effective at capturing the environmental information present within the graph and generating low-dimensional vectors, or embeddings, for all entities. To accomplish this, we feed the augmented spatial knowledge graph into the TuckER model. The model then learns and understands the complex relationships between the entities and the environmental features present in the graph and generates embeddings that accurately capture the most important features of each entity and its surrounding environment. For each triplet $(h, r, t) \in \mathbf{F}$, the model measures the plausibility as follows,

$$\phi(h, r, t) = \mathbf{W} \times_1 \mathbf{e}_h \times_2 \mathbf{r} \times_3 \mathbf{e}_t, \tag{2}$$

where $\mathbf{W} \in \mathbb{R}^{d\times d\times d}$ is the core tensor in Tucker decomposition [27], and the $\mathbf{e}_h, \mathbf{r}, \mathbf{e}_t \in \mathbb{R}^d$ are embeddings of entities and relations. Respectively, d and $\times_i$ represent the embedding dimension and the tensor product along the $i - th$ mode. Based on the observed triplets in the augmented spatial knowledge graph, we calculate the plausibility scores via the above scoring function and develop the cross-entropy loss functions for parameter learning such that valid triplets obtain higher scores than invalid ones.

3.4 Relational Graph Convolutional Network

To make the embeddings more suited for mobile traffic prediction, we apply the Relational Graph Convolutional Network (RGCN) [22] which could get the learnable entity embeddings. RGCN is an extension of GCN, which could only operate on homogeneous graphs. In contrast, RGCN can handle heterogeneous graphs with multiple types of nodes and edges, each with its own set of features. The RGCN model is able to handle heterogeneous graphs by utilizing a different weight matrix for each edge type present in the graph. This allows the model to learn different representations for nodes and edges based on their respective types. By performing convolutional operations on the graph, the RGCN model is able to propagate information across the nodes and edges, resulting in the generation of node embeddings, which can be formed as,

$$\mathbf{h}_i^{l+1} = \sigma(\mathbf{W}_0^{(l)}\mathbf{h}_i^{(l)} + \sum_{r\in R}\sum_{j\in N_i^r}\frac{1}{c_{i,r}}\mathbf{W}_r^{(l)}\mathbf{h}_j^{(l)}), \tag{3}$$

where N_i^r denotes the set of neighbourhoods of node i under relation $r \in R$, $\mathbf{W}_r^{(l)}$ denoted the learnable parameters under relation r, and $\mathbf{h}_j^{(l)}$ represents the embedding of node i at layer l. $c_{i,r}$ is a normalization constant that can be learned or chosen in advance, and l denotes the $l - th$ RGCN layer. The ability of RGCN to handle heterogeneous graphs makes it particularly useful for modelling knowledge graphs. By applying RGCN to mobile traffic prediction, we are able to generate embeddings that are better suited for capturing the relationships between the various entities in the graph.

4 METHODS

4.1 Framework Overview

Our model, shown in Figure 1, predicts future mobile traffic for all base stations using the spatial knowledge graph G and historical mobile traffic series S^t as inputs, with three main components: semantic relationship modelling, environmental context modelling, and inter-attentional fusion modelling. The semantic relationship modelling component uses a historical mobile traffic matrix and basic spatial knowledge graph to capture temporal features, producing a domain embedding output for each base station to capture relationships and traffic patterns. The environmental context modelling component enhances the knowledge graph and employs representation learning to capture the spatial structure and functional similarity, producing an environmental embedding for each base

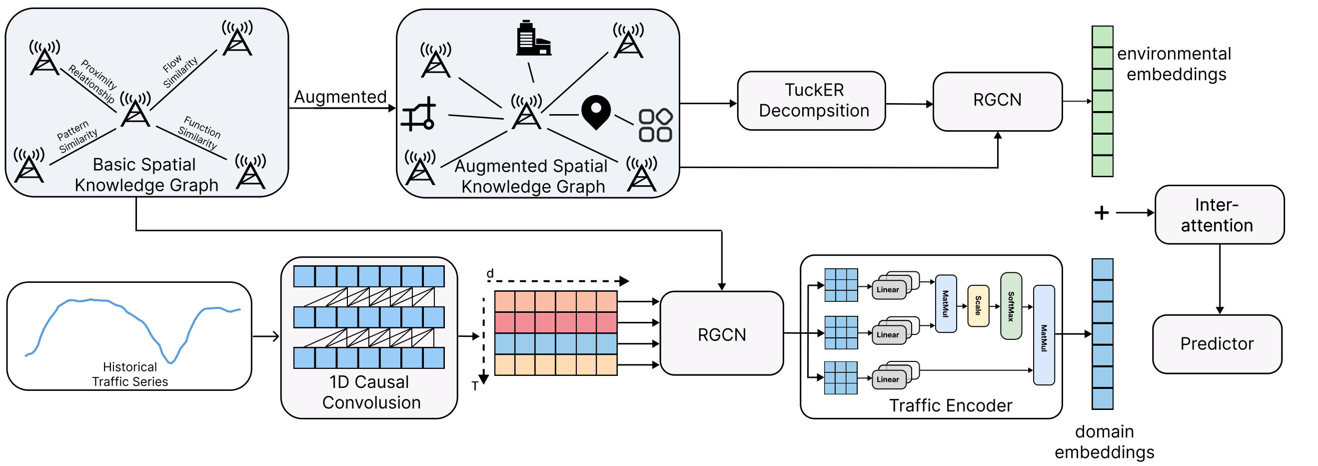


Figure 1: The Framework Overview of our model.

station to improve accuracy and robustness. The inter-attentional fusion model combines domain and environmental embeddings, providing comprehensive predictions that incorporate temporal features and environmental information for greater accuracy.

4.2 Spatial Knowledge Graph Construction

4.2.1 Basic Spatial Knowledge Graph. To better understand the mobile traffic prediction task, we discover four relationships between base stations and form the basic spatial knowledge graph to capture their spatial and temporal features. These four relationships model the relationships between base stations from different temporal and spatial perspectives.

- **Proximity Relationship.** Two base stations have a proximity relationship when their physical locations are within a certain distance of each other, formulated as,

$$a_{i,j} = \begin{cases} exp(-\frac{dis_{i,j}}{\sigma^2}), & exp(-\frac{dis_{i,j}}{\sigma^2}) \geq \epsilon \\ 0, & exp(-\frac{dis_{i,j}}{\sigma^2}) < \epsilon \end{cases} \tag{4}$$

  where $dis_{i,j}$ denotes the distance between base station b_i and base station b_j, and σ and ϵ are thresholds to control the distribution and sparsity of the matrix of proximity relationship, where we set σ and ϵ are 37 and 0.5. Close base stations likely transfer traffic due to user behaviour of connecting to nearby stations for weak signals. This affects overall traffic patterns and mobile network usage.
- **Function Similarity.** POIs are functional units in a city where people engage in social and productive activities. Similar POIs generate similar traffic patterns. To determine this, we calculate POI numbers in each category near each base station and the cosine similarity of each base station to create a similarity matrix. The formulation can be formed as,

$$b_{i,j} = cos(v_i^{POI}, v_j^{POI}), \tag{5}$$

  where $b_{i,j}$ denotes the cosine similarity of POI distribution between base station b_i and base station b_j, and the v_i^{POI} represents the POI distribution vector of base station b_i. We select the 20 most similar base stations for each station to establish functional similarity relationships.
- **Pattern Similarity.** After thorough analysis, we found that each base station has a unique and consistent traffic pattern that reflects its typical usage by mobile users. However, there is some variability in real traffic flow, which fluctuates around the pattern on a weekly scale. Location can also affect a base station's traffic pattern. To group similar base stations based on normalized patterns, we use hierarchical clustering [35], an unsupervised machine-learning technique. We consider base stations in the same cluster and within a certain distance to have a pattern similarity relationship, expressed as:

$$c_{i,j} = \begin{cases} 1 \cdot a_{i,j}\ , & v_i^p = v_j^p \\ 0, & otherwise. \end{cases} \tag{6}$$

  where v_i^p denotes the results of clustering the pattern series of base station b_i.
- **Flow Similarity.** Base stations may have pattern similarity relationships, but their traffic flow can still vary significantly in terms of absolute values and speed of changes. To examine temporal characteristics, we introduce the concept of Similar Series relationships. Using dynamic time warping methods [15], we calculate traffic series similarity between base stations to generate a similarity matrix D, expressed as:

$$d_{i,j} = DTW(v_i^{flow}, v_j^{flow}), \tag{7}$$

  where $d_{i,j}$denotes the traffic series similarity between base station b_i and base station b_j, and v_i^{flow} represents the traffic flow of base station b_i. We select the 20 most similar base stations for each station to establish flow similarity relationships.

4.2.2 Augmented Spatial Knowledge Graph. To enhance the accuracy of mobile traffic prediction by capturing environmental information, we have expanded the basic spatial knowledge graph to include additional elements and relationships beyond just those between base stations. By establishing connections between base stations and other important components of urban areas, such as

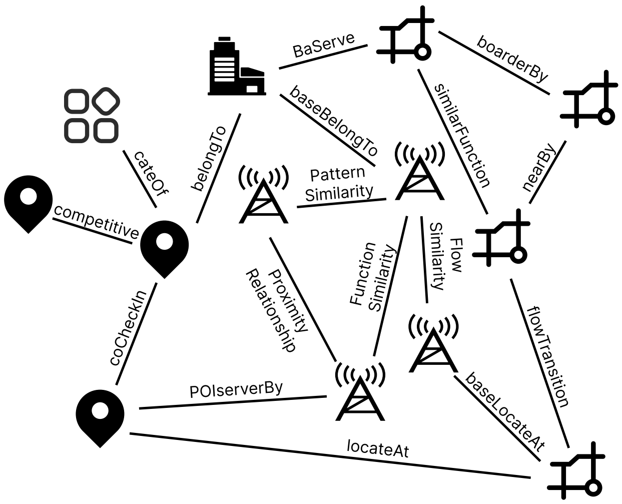


Figure 2: The schema of augmented spatial knowledge graph.

regions, business areas, POIs, and categories, we have enriched the graph and made it more informative. We have also added various relationships, including *cateOf*,*coCheckIn*,*competitive*, *broderBy*, *nearBy*, *flowTransition*, *similarFunction*,*locateAt*, *belongTo*, *POIservedBy*, *baseLocateAt*, *BaServe*, and *baseBelongTo* [19, 31], which are crucial for capturing the complex spatial relationships that influence mobile traffic behaviour. More details are provided in Appendix A.

The augmented spatial knowledge graph, as shown in Figure 2, provides a more comprehensive and informative representation of the urban environment and its impact on mobile traffic. The expanded graph includes relationships between base stations and various urban components, enabling the graph to capture the interdependencies between them and making it a more powerful tool for mobile traffic prediction.

4.3 Semantic Relationship Modelling

We opt to use the basic spatial knowledge graph to model the semantic relationship between base stations. By utilizing both the historical mobile traffic series and the basic spatial knowledge graph as inputs, our model can effectively capture the temporal features of all base stations, ultimately resulting in domain embeddings for each base station.

To capture the local shapes of the time series accurately, we employ one-dimensional causal convolution. Unlike regular convolution, causal convolution transforms the traffic of the base stations while ensuring that the model has no access to future information at the current position. This enables the model to focus on the local context and shapes of the time series rather than just point-wise values. The one-dimensional causal convolution takes the historical traffic series S^t as its input and can be expressed as,

$$x_i^t = \sigma(\sum_{j=0}^{k-1} \mathbf{W} x_i^{t-j} + b), \tag{8}$$

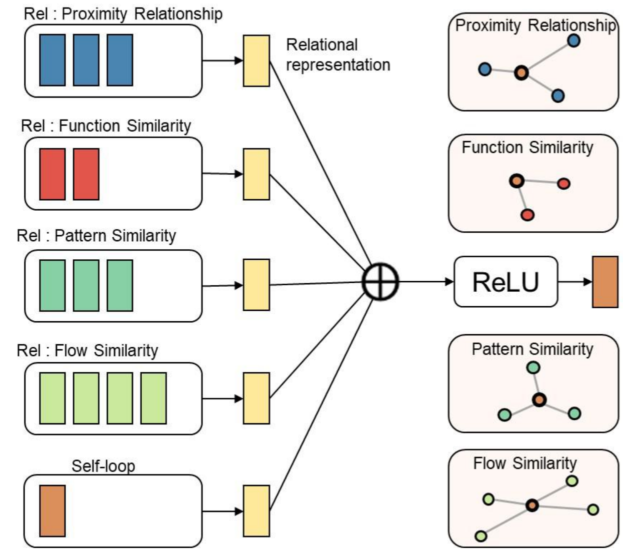


Figure 3: Nodes updating process of RGCN. Features from neighbour nodes with specific relations are gathered and transformed into relational representations. After being gathered, the relational representation is sent via an activation function.

where x_i^t denotes the traffic value of base station b_i at time step t, and k denotes the kernel size. The one-dimensional causal convolution can capture the local shapes of time series and outputs the convolution feature matrix $\boldsymbol{E}^{conv} \in \mathbb{R}^{N \times T \times d}$, where N denotes the number of base stations, T denotes the length of historical traffic series and d denotes the embedding dimension. In this way, enable the attention mechanism in the transformer behind take concentrates on the time series shapes instead of point-wise values [18].

We next employ the RGCN to propagate the local shapes of each base station's time series. The RGCN takes the convolution feature matrix E^{conv} as input in time order which can be expressed as,

$$\mathbf{E}_i^{RGCN} = RGCN(\mathbf{E}_i^{conv}) \qquad i \in \{1, 2, ..., T\}, \tag{9}$$

where $\mathbf{E}_i$ denotes the $i-th$ step in the T time steps, $RGCN$ denotes the RGCN layers, which can be expressed as Eq. 3. Then, we collocate the output of RGCN to $\mathbf{E}^{RGCN} \in \mathbb{R}^{N \times r \times d}$.

We next employ the intra-attention mechanism, which focuses on the feature vector of each base station, to model the time series. This mechanism is composed of a time position encoding model and transformer encoders. To leverage the sequence order [28], we first apply time position embedding. We then utilize a transformer to model the time series. To capture the interdependence of both long and short-term temporal features, the transformer utilizes a multi-head self-attention mechanism, with each head focusing on different aspects of temporal features, which can be expressed as,

$$\mathbf{E}^{do} = Attention(\mathbf{E}^{RGCN}) = softmax(\frac{\mathbf{Q}_h \mathbf{K}_h^T}{\sqrt{d_k}})\mathbf{V}_h, \tag{10}$$

$$\mathbf{Q}_h = \mathbf{E}^{RGCN}\mathbf{W}_h^Q, \quad \mathbf{K}_h = \mathbf{E}^{RGCN}\mathbf{W}_h^K, \quad \mathbf{V}_h = \mathbf{E}^{RGCN}\mathbf{W}_h^V, \tag{11}$$

where $\mathbf{W}_h^Q, \mathbf{W}_h^K \in \mathbb{R}^{d\times d_k}, \mathbf{W}_h^V \in \mathbb{R}^{d\times d_v}$, are learnable parameters and $h \in \{1, 2, ..., H\}$ denotes the $h-th$ head. The intra-attention mechanism takes the E^{RGCN} as the input and outputs the domain embeddings $E^{do} \in \mathbb{R}^{N\times d}$.

4.4 Environmental Context Modeling

The environmental context modelling takes the augmented spatial knowledge graph G^a as input, captures the environment information of base stations, and outputs the environmental embeddings of base stations.

To capture environmental information, we first use the TuckER model to pre-train and learn the complex relationships between entities and environmental features, resulting in learned embeddings of all entities in the augmented spatial knowledge graph $E^{pre} \in \mathbb{R}^{N_a\times d}$, where N_a represents the number of entities in the augmented spatial knowledge graph. We then utilize the RGCN to improve the pre-trained embeddings for mobile traffic prediction, which is not learnable. The RGCN takes $\mathbf{E}^{pre}$ as its input and outputs environmental embeddings $\mathbf{E}^{en} \in \mathbb{R}^{N\times d}$, where we only retain the embeddings of the base stations.

To fuse the environmental embeddings E^{en} (output of environmental context modelling) and domain embeddings E^{do} (output of semantic relationship modelling), we utilize the Inter-attentional fusion module. This module first employs an inter-attention mechanism to capture the relationships between base stations, which can be expressed as:

$$\mathbf{E}^{att} = Attention(\mathbf{E}^{en}||\mathbf{E}^{do}), \tag{12}$$

where || denotes the concatenate operation, and the attentional mechanism is expressed as Eq.10. Next, we use a Multilayer Perception (MLP) [26] to be the output layer, which can be formed as,

$$M = f(\mathbf{E}^{att}) = W_2(\sigma(\mathbf{W}_1\mathbf{E}^{att} + b_1)) + b_2, \tag{13}$$

where $\mathbf{W}, b$ are the trainable weight matrix and the bias matrix. And the output of the MLP is the predicted future mobile traffic.

Algorithm 1 outlines the multi-relational knowledge graph convolutional network model training procedure. The model takes the historical traffic matrix S^t and the spatial knowledge graph as input and outputs the mobile traffic for the next time step. The model is trained in batches to minimize the difference between predicted values $\left[\hat{x}_1^{t+1}, \hat{x}_2^{t+1}, ..., \hat{x}_N^{t+1}\right]$ and their corresponding true values $\left[x_1^{t+1}, x_2^{t+1}, ..., x_N^{t+1}\right]$. We use the Mean Squared Error (MSE) loss function to optimize the model's parameters, which can be formed as,

$$\mathcal{L} = \sum_{i=1}^{N} \left\|\hat{x}_i^{t+1} - x_i^{t+1}\right\|_2. \tag{14}$$

5 EXPERIMENT AND RESULT

5.1 Experimental Settings

5.1.1 Datasets.

- **Shanghai Dataset.** The Shanghai Datasets consist of anonymous mobile traffic data collected by China Mobile in Shanghai during August 2014. The dataset comprises 4505 base stations and over 150,000 users, with each entry containing the anonymous device ID, start and end time of data collection, anonymous base station ID, and amount of data used in the connection. We contributed 1.96 billion tuples of entries to 4505 base stations in Shanghai every 30 minutes, according to the tracing logs. Moreover, the augmented spatial knowledge graph for Shanghai includes five types of entities, such as regions, business areas, POIs, categories, and base stations, with the number of relationships totalling 17.
- **Nanjing Dataset.** The Nanjing Datasets consist of anonymous mobile traffic data collected by China Mobile from Nanjing between February 2nd and March 31st, 2021. The dataset comprises 8000 base stations, which is larger than the Shanghai Datasets. We contributed to 8000 base stations in Nanjing every

Algorithm 1: Multi-relational Knowledge Graph Convolutional Network Model

Input: Basic spatial knowledge Graph G, Augmented spatial knowledge graph G^a, historical mobile traffic matrix $S^t \in \mathbb{R}^{N\times T}$

Result: Mobile traffic prediction for the next time step $\left[\hat{x}_1^{t+1}, \hat{x}_2^{t+1}, ..., \hat{x}_N^{t+1}\right]$

```
E^conv ← CausalConv1d(S^t) // E^conv ∈ R^{N×T×d}
for i = 1 to T do
    E_i^RGCN ← RGCN(E_i^conv, G) // according to Equ 9
end
E^pos ← PositionEmbedding(E^RGCN) // E^pos ∈ R^{N×T×d}
E^tem ← Attention(E^pos) // according to Equ 10
E^do ← Mean(E^tem) // E^do ∈ R^{N×d}
E^pre ← TuckER(G^a) // E^pre ∈ R^{N_a×d}
E^en ← RGCN(E^pre, G^a) // E^en ∈ R^{N×d}
E^fu ← Concat(E^en, E^do) // E^fu ∈ R^{N×2d}
E^att ← Attention(E^fu) // E^att ∈ R^{N×2d}
[x̂_1^{t+1}, x̂_2^{t+1}, ..., x̂_N^{t+1}] ← MLP(E^att)
```

Table 1: Statistics of the datasets used in our experiments.

Dataset	Shanghai	Nanjing
Collection Duration	Aug. 1st - 31st, 2014	Feb. 2nd - Mar. 31st, 2021
Time Interval	30 minutes	
Covered Users	≥150,000	≥450,000
Covered BSs	4505	8000
Covered Area	6340 km^2	6587 km^2
Flow Records	8.65×10^8	8.18×10^8
Regions	2579	1022
Business Areas	280	228
POIs	85018	51264
Categories	14	14
relationships	17	14

30 minutes. Furthermore, the augmented spatial knowledge graph for Nanjing includes five types of entities, such as regions, business areas, POIs, categories, and base stations, with the number of relationships totalling 13.

Table 1 shows the statistics of the Shanghai dataset and Nanjing dataset. The large-scale and fine-grained datasets can ensure the validity of mobile traffic reality and the model test.

5.1.2 Metrics. To address the issue of the large absolute value of mobile traffic and focus on its magnitude, we apply log-normalization to the mobile traffic data. In evaluating the performance of mobile traffic prediction, we carefully consider three metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).

5.1.3 Baselines. We elaborately select the following seven representatives to be compared with our proposed algorithms, which cover both representative classical traffic prediction models [4, 33] and state-of-the-art deep learning models [7, 10, 29, 38, 41].

- **SVR [4].** SVR is an extension of SVM for regression tasks that predicts continuous output based on input features.
- **ARIMA [33].** ARIMA is a statistical model for analyzing and predicting time series data that uses three parameters to represent auto-regressive terms, moving average terms, and differences or orders.
- **GAT [29].** GAT is a dynamic graph neural network that learns node weights, capturing varying connection importance. It handles changing graphs or contextual variables. We use mask graph attention with GAT.
- **GraphSAGE [10].** GraphSAGE is a graph representation learning method that enhances the scalability and performance of GNN, which maximizes the ratio of sampling the current neighbour node to sampling the entire graph.
- **DeepTP [7].** DeepTP is an end-to-end deep learning model that predicts spatial-temporally dependent cellular traffic over a lengthy period. It handles complex and dynamic traffic patterns influenced by spatial and temporal factors, using a sequential module and a broad feature extractor.
- **STGCN [38].** STGCN combines GCN and gated CNN architectures to capture spatial-temporal patterns in graph-structured data. It uses GCN to mine the graph's topology and gated CNN to explore dynamic mobile traffic features.
- **T-GCN [41].** T-GCN combines GCN and GRU to model time series and capture the dynamic mobile traffic change of node attribution.
- **GMAN[42].** GMAN utilizes an encoder-decoder architecture with spatial-temporal attention blocks to capture the impact of spatial-temporal variables on traffic conditions, where the input traffic characteristics are encoded by the encoder and the decoder predicts the output time step sequence.

5.1.4 Parameter Settings. Our model utilizes the Adam optimizer [16] with a learning rate of 0.0005 and applies MSE loss for training. When using the TuckER model for pre-training, we set the embedding vector dimension d to 32 and the epoch to 50 to balance accuracy and efficiency. We set the length of the historical traffic series T at 12 for higher accuracy and faster speed. The kernel size of the one-dimensional causal convolution k is 3. We set the embedding vector dimension for environmental context modelling and semantic relationship modelling between base stations to 32. The number of RGCN layers is 1, and we set the transformer encoder layers as 2 with $H = 8$ heads. In the inter-attention mechanism, we set the dimension of the key and value as 32 and 16, respectively. The datasets are divided into three parts: training, validation, and testing, with a ratio of 0.7:0.15:0.15.

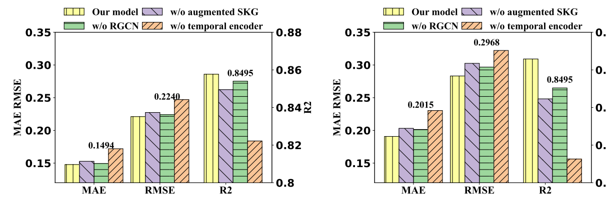
Table 2: Overall prediction performance of our model in comparison with compared algorithms on Shanghai and Nanjing datasets.

	Shanghai Dataset			Nanjing Dataset		
Model	**MAE**	**RMSE**	R^2	**MAE**	**RMSE**	R^2
SVR	0.2092	0.3018	0.7479	0.2316	0.3336	0.7510
ARIMA	0.2058	0.3041	0.7499	0.2328	0.3275	0.7797
GAT	0.1984	0.2650	0.5547	0.3574	0.4830	0.5735
GraphSAGE	0.2138	0.2979	0.7418	0.2354	0.3467	0.7483
T-GCN	0.1908	0.2694	0.7990	0.2516	0.3519	0.8178
STGCN	0.1996	0.2785	0.7767	0.2537	0.3642	0.8058
DeepTP	0.1869	0.2610	0.7991	0.2322	0.3327	0.8196
GMAN	<u>0.1807</u>	<u>0.2554</u>	<u>0.8078</u>	<u>0.2209</u>	<u>0.3237</u>	<u>0.8237</u>
our model	**0.1478**	**0.2211**	**0.8577**	**0.1908**	**0.2832**	**0.8658**
Improv.	18.20%	13.42%	6.17%	13.62%	12.51%	5.11%

5.2 Overall Performance

In Table 2, we display the overall results of our model, temporal model (SVR, ARIMA), spatial model (GAT, GraphSAGE), and spatial-temporal model (DeepTP, STGCN, T-GCN, GMAN) to predict the next time stamp (30 minutes) in Shanghai Datasets and Nanjing Datasets. We list three metrics of all methods. From the results, we have the following findings:

- **Our framework steadily achieves the best performance.** Our model gets superior results on both datasets and performances better than other compared algorithms. For example, the R^2 improvement of our model compared with the second-best performance model (GMAN), is around 5.11% to 6.17%. The MAE reduction of our model is about 13% to 18%.
- **Spatial models perform poorly in the mobile traffic prediction task.** Spatial models are commonly used to analyze spatial data such as geographic patterns and location-based information. However, these models may not have the necessary modules to model time series data or capture temporal features. As a result, their performance may be inferior to models that incorporate temporal information. By incorporating temporal components into spatial models, we can achieve more accurate predictions and better performance in real-world applications.
- **It is essential to model various semantic relationships and environmental information.** Modeling various semantic relationships and environmental information is crucial as it not only facilitates the capture of spatial features but also enhances the performance of models. Besides, we can see that the Urban Knowledge Graph could mine more environmental features that could be useful for mobile traffic prediction than



(a) Shanghai Dataset (b) Nanjing Dataset

Figure 4: Ablation study.

Table 3: Prediction results of the effectiveness of the augmented spatial knowledge graph.

	Shanghai Dataset			Nanjing Dataset		
Graph	**MAE**	**RMSE**	R^2	**MAE**	**RMSE**	R^2
our model	**0.1478**	**0.2211**	**0.8577**	**0.1908**	**0.2832**	**0.8658**
w/o category	0.1481	0.2213	0.8566	0.1915	0.2836	0.8654
w/o POI	0.1494	0.2224	0.8554	0.1948	0.2866	0.8625
w/o BA	0.1507	0.2243	0.8512	0.2019	0.2980	0.8514
w/o region	0.1518	0.2251	0.8502	0.2037	0.2995	0.8499

the distance matrix. Compared with STGCN and T-GCN, we can obtain that our augmented spatial knowledge graph, consisting of four semantic relations and environmental information, can capture more spatial and temporal features than only the distance matrix. Compared with all spatial-temporal models, the improvement of our model is around 7.33% to 10%. The MAE reduction of our model is about 20% to 25%.

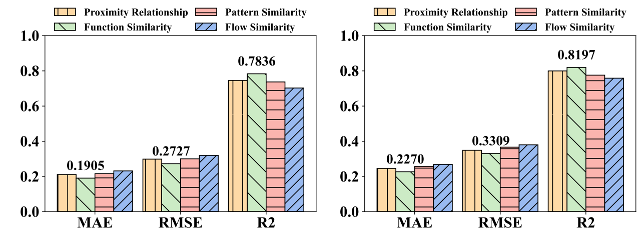
5.3 Ablation Study

To gain a deeper understanding of each component of our model, we conducted a series of ablation experiments. Firstly, we removed the augmented spatial knowledge graph, followed by removing only the RGCN layer within this modelling. Next, we eliminated the traffic encoder.

The results of the ablation study are presented in Figure 4. We observe that removing the augmented spatial knowledge graph, which involves deleting the environmental context modelling, results in a decrease in performance as the model lacks environmental information. Eliminating the *RGCN* layer and using the pre-trained embeddings from TuckER may not be suitable for this task as the embeddings are static. In semantic relationship modelling, the transformer can better model time series and capture more temporal features. Without the traffic encoder, the model fails to capture the temporal dependency, resulting in a rapid decrease in performance of around 10% to 17%.

5.4 Sensitivity Study

• **Effectiveness of the Urban Knowledge Graph.** To showcase the effectiveness of each entity in the augmented spatial knowledge graph, we remove an entity and the relationship connected to this entity each time. We conducted experiments on the Shanghai and Nanjing datasets, and the results are presented in Table 3. The results indicate that the intact Urban Knowledge Graph outperforms the sub-graphs, demonstrating the importance of leveraging the full extent of the graph for optimal performance. Furthermore, we observed that regions and business areas, outperform POIs and categories. This can be attributed to the fact that POIs contain more micro-geographic information and are more challenging to capture in terms of spatial structure and environmental information. In contrast, regions and business areas provide a more macro-level perspective and are easier to model and analyze within the context of the augmented spatial knowledge graph.



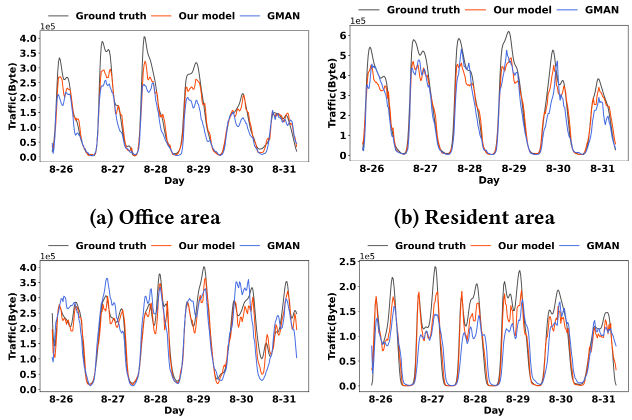
(a) Shanghai Dataset (b) Nanjing Dataset

Figure 5: Prediction results of single semantic relation between base stations.

• **Performance of Semantic Relationships between Base Stations.** To evaluate the effectiveness of different semantic relations between base stations, we conducted experiments by selecting one relation at a time and incorporating it into the semantic relations between base stations modelling. We then tested the performance on both the Shanghai and Nanjing datasets. The results of these experiments are presented in Table 5. Our findings indicate that incorporating all four semantic relations simultaneously leads to better performance than using only a single relation. This suggests that each relation captures different aspects of the relationships between base stations, and utilizing all of them leads to a more comprehensive understanding of the behaviour of base stations in the urban environment. Of the four individual relations, the Similar traffic pattern relation performs the best, while the Similar POI distribution relation performs the worst. The success of the Similar traffic pattern relation can be attributed to the fact that the inputs of traffic encoders are temporal features such as local shapes, which are well-suited to capture this type of relationship. In contrast, the Similar POI distribution relation is a geographically similar relationship that may not be as easily captured by the model. These results provide insights into the effectiveness of different semantic relations between base stations and can inform decisions related to the selection of relations in future models.

5.5 Case Study

• **Comparison in Predicting Base Station Traffic Patterns.** In our study, we conducted a performance comparison between our proposed model and the second-best model, GMAN, in predicting the values of four different types of base stations, each of which is located in different areas and has its own traffic patterns [35]. To evaluate the performance of the models, we plotted the predicted values against the actual values for each model, as shown in Figure 6.



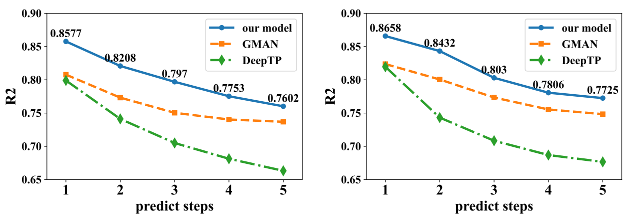
(a) Office area

(b) Resident area

(c) Entertainment area

(d) Transport area

Figure 6: Prediction vs. the ground truth in four different patterns.



(a) Office area

(b) Resident area

Figure 7: Prediction on Multi-steps.

Our results demonstrate that our proposed model outperforms GMAN in terms of prediction accuracy, producing more reliable and low-latency predictions. Specifically, our model displays superior performance in distinguishing the different traffic trends of base stations on weekdays and weekends. We attribute this to the fact that our model utilizes an augmented spatial knowledge graph, which integrates various environmental information. This allows our model to capture more nuanced and complex traffic patterns in the urban environment, leading to better prediction accuracy.

• **Multi-step Prediction.** To further evaluate the effectiveness of our proposed model, we conducted a multi-step prediction experiment to test its long-term prediction performance. In this experiment, we compared the performance of our model with GMAN and DeepTP, over a period of five prediction steps. The results of this experiment are presented in Figure 7.

Our results demonstrate that as the number of prediction steps increases, all models experience a decline in performance. However, our proposed model continues to outperform all others, showcasing its ability to effectively capture both short and long-term time series features. This is a crucial advantage for predicting traffic patterns in urban environments, where accurate forecasting over extended periods is essential for effective resource allocation and planning.

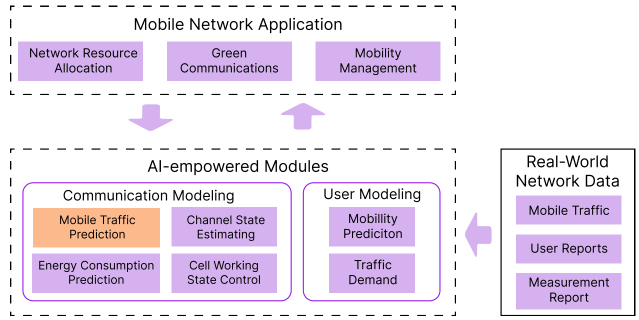


Figure 8: The framework JiuTian AI platform.

5.6 Application on Real-world System

Our cellular traffic prediction model has been deployed on the Jiutian Artificial Intelligence (AI) Platform[1]. Jiutian Artificial Intelligence (AI) Platform is China Mobile's self-developed AI innovation platform, providing intelligent decision-making support for mobile networks. As shown in Figure 8, our model acts as a key part in the AI-empowered modules for communication modelling, which support mobile network applications, including network resource allocation, green communications, and mobility management.

6 CONCLUSION

Our research leverages the power of spatial knowledge graphs to provide comprehensive semantic relationships between base stations, allowing us to capture various types of spatial information. In addition, we augment the spatial knowledge graph with other critical components of urban areas, such as regions, business areas, POIs, and categories. This enables the knowledge graph to be a more powerful tool for capturing environmental information. By leveraging the augmented spatial knowledge graph, we can incorporate multiple types of spatial data and capture a wide range of spatial dependencies and environmental factors. To fully utilize the spatial knowledge graph and capture the complex relationships between different urban components, we propose a Multi-relational Knowledge Graph Convolutional Network, which is designed to integrate both spatial and semantic information to predict future traffic volumes accurately. The RGCN component of the model captures environmental information, while the transformer encoder and causal convolution model semantic relationships. Finally, an inter-attentional mechanism combines these features to make the prediction. In future work, we aim to extend our approach and evaluate its effectiveness in diverse settings by applying it to different cities and regions, assessing its generalizability and identifying any limitations or challenges in varying mobile network environments. We also plan to investigate the transferability of our model to 5G networks, adapting it to their unique requirements and characteristics for network optimization and management. Lastly, we plan to collaborate with industry partners to integrate our model into existing mobile network management systems. This will enable us to enhance network performance and user experience by providing more accurate predictions and refined allocation and proactive scheduling of network resources.

[1] https://jiutian.10086.cn/portal/

REFERENCES

[1] Davide Andreoletti, Sebastian Troia, Francesco Musumeci, Silvia Giordano, Guido Maier, and Massimo Tornatore. 2019. Network Traffic Prediction based on Diffusion Convolutional Recurrent Neural Networks. In *IEEE INFOCOM WKSHPS*. 246–251. https://doi.org/10.1109/INFCOMW.2019.8845132

[2] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. *Syntax, Semantics and Structure in Statistical Translation* (2014), 103.

[3] Zulong Diao, Guanxin Wang, Dafang Zhang, Yingru Liu, Kun Xie, and Shaoyao He. 2019. Dynamic Spatial-Temporal Graph Convolutional Neural Networks for Traffic Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (07 2019), 890–897. https://doi.org/10.1609/aaai.v33i01.3301890

[4] Harris Drucker, Christopher Burges, Linda Kaufman, Alexander Smola, and V. Vapnik. 1997. Support Vector Regression Machines. *Adv Neural Inform Process Syst* 28 (01 1997), 779–784.

[5] Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S Weld. 2008. Open Information Extraction From the Web. *Commun. ACM* 51, 12 (2008), 68–74.

[6] Luoyang Fang, Xiang Cheng, Haonan Wang, and Liuqing Yang. 2018. Mobile Demand Forecasting via Deep Graph-Sequence Spatiotemporal Modeling in Cellular Networks. *IEEE Internet of Things Journal* 5, 4 (2018), 3091–3101. https://doi.org/10.1109/JIOT.2018.2832071

[7] Jie Feng, Xinlei Chen, Rundong Gao, Ming Zeng, and Yong Li. 2018. DeepTP: An End-to-End Neural Network for Mobile Cellular Traffic Prediction. *IEEE Network* 32, 6 (2018), 108–115. https://doi.org/10.1109/MNET.2018.1800127

[8] Rui Fu, Zuo Zhang, and Li Li. 2016. Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. 324–328. https://doi.org/10.1109/YAC.2016.7804912

[9] S. Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. In *AAAI Conference on Artificial Intelligence*.

[10] William Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. (06 2017).

[11] Qing He, Arash Moayyedi, György Dán, Georgios P. Koudouridis, and Per Tengkvist. 2020. A Meta-Learning Scheme for Adaptive Short-Term Network Traffic Prediction. *IEEE Journal on Selected Areas in Communications* 38, 10 (2020), 2271–2283. https://doi.org/10.1109/JSAC.2020.3000408

[12] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-term Memory. *Neural computation* 9 (12 1997), 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735

[13] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. 2021. Knowledge Graphs. *ACM Comput. Surv.* 54, 4, Article 71 (jul 2021), 37 pages. https://doi.org/10.1145/3447772

[14] Wei-Chiang Hong. 2010. Application of Seasonal SVR with Chaotic Immune Algorithm in Traffic Flow Forecasting. *Neural Computing and Applications* 21 (2010), 583 – 593.

[15] F. Itakura. 1975. Minimum Prediction Eesidual Principle Applied to Speech Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 23, 1 (1975), 67–72. https://doi.org/10.1109/TASSP.1975.1162641

[16] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014).

[17] Rongpeng Li, Zhifeng Zhao, Xuan Zhou, Jacques Palicot, and Honggang Zhang. 2014. The prediction analysis of cellular radio access network traffic: From entropy theory to networking practice. *IEEE Communications Magazine* 52, 6 (2014), 234–240. https://doi.org/10.1109/MCOM.2014.6829969

[18] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. 2019. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. *Advances in neural information processing systems* 32 (2019).

[19] Yu Liu, Jingtao Ding, Yanjie Fu, and Yong Li. 2023. UrbanKG: An Urban Knowledge Graph System. *ACM Trans. Intell. Syst. Technol.* 14, 4, Article 60 (may 2023), 25 pages. https://doi.org/10.1145/3588577

[20] Laisen Nie, Zhaolong Ning, Mohammad S. Obaidat, Balqies Sadoun, Huizhi Wang, Shengtao Li, Lei Guo, and Guoyin Wang. 2021. A Reinforcement Learning-Based Network Traffic Prediction Mechanism in Intelligent Internet of Things. *IEEE Transactions on Industrial Informatics* 17, 3 (2021), 2169–2180. https://doi.org/10.1109/TII.2020.3004232

[21] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2009. The Graph Neural Network Model. *IEEE Transactions on Neural Networks* 20, 1 (2009), 61–80. https://doi.org/10.1109/TNN.2008.2005605

[22] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*. Springer, 593–607.

[23] Yantai Shu, Minfang Yu, Jiakun Liu, and O.W.W. Yang. 2003. Wireless Traffic Modeling and Prediction Using Seasonal ARIMA Models. In *IEEE International Conference on Communications, 2003. ICC '03.*, Vol. 3. 1675–1679 vol.3. https://doi.org/10.1109/ICC.2003.1203886

[24] Feiyang Sun, Pinghui Wang, Junzhou Zhao, Nuo Xu, Juxiang Zeng, Jing Tao, Kaikai Song, Chao Deng, John C.S. Lui, and Xiaohong Guan. 2022. Mobile Data Traffic Prediction by Exploiting Time-Evolving User Mobility Patterns. *IEEE Transactions on Mobile Computing* 21, 12 (2022), 4456–4470. https://doi.org/10.1109/TMC.2021.3079117

[25] Yaohua Sun, Mugen Peng, Yangcheng Zhou, Yuzhe Huang, and Shiwen Mao. 2019. Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues. *IEEE Communications Surveys Tutorials* 21, 4 (2019), 3072–3108. https://doi.org/10.1109/COMST.2019.2924243

[26] Ilya O. Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, and Alexey Dosovitskiy. 2021. MLP-Mixer: An all-MLP Architecture for Vision. In *Neural Information Processing Systems*.

[27] Ledyard R Tucker. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* 31, 3 (1966), 279–311.

[28] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. *Advances in neural information processing systems* 30 (2017).

[29] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio', and Yoshua Bengio. 2017. Graph Attention Networks. *ArXiv* abs/1710.10903 (2017).

[30] Jing Wang, Jian Tang, Zhiyuan Xu, Yanzhi Wang, Guoliang Xue, Xing Zhang, and Dejun Yang. 2017. Spatiotemporal Modeling and Prediction in Cellular Networks: A Big Data Enabled Deep Learning Approach. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*. 1–9. https://doi.org/10.1109/INFOCOM.2017.8057090

[31] Pengyang Wang, Kunpeng Liu, Lu Jiang, Xiaolin Li, and Yanjie Fu. 2020. Incremental Mobile User Profiling: Reinforcement Learning with Spatial Knowledge Graph for Modeling Event Streams. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2020).

[32] Xu Wang, Zimu Zhou, Zheng Yang, Yunhao Liu, and Chunyi Peng. 2017. Spatio-Temporal Analysis and Prediction of Cellular Traffic in Metropolis. In *2017 IEEE 25th International Conference on Network Protocols (ICNP)*. 1–10. https://doi.org/10.1109/ICNP.2017.8117559

[33] Billy Williams and Lester Hoel. 2003. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering* 129 (11 2003), 664–672. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)

[34] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In *International Joint Conference on Artificial Intelligence*.

[35] Fengli Xu, Yong Li, Huandong Wang, Pengyu Zhang, and Depeng Jin. 2017. Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment. *IEEE/ACM Transactions on Networking* 25, 2 (2017), 1147–1161. https://doi.org/10.1109/TNET.2016.2623950

[36] Yue Xu, Wenjun Xu, Feng Yin, Jiaru Lin, and Shuguang Cui. 2017. High-Accuracy Wireless Traffic Prediction: A GP-Based Machine Learning Approach. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*. 1–6. https://doi.org/10.1109/GLOCOM.2017.8254808

[37] Fan Yang, Yun Jiang, Tian Pan, and Xinhua E. 2018. Traffic Anomaly Detection and Prediction Based on SDN-Enabled ICN. In *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. 1–5. https://doi.org/10.1109/ICCW.2018.8403693

[38] Ting Yu, Haoteng Yin, and Zhanxing Zhu. 2017. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In *International Joint Conference on Artificial Intelligence*.

[39] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. 2014. Recurrent Neural Network Regularization. *ArXiv* abs/1409.2329 (2014).

[40] Chaoyun Zhang and Paul Patras. 2018. Long-term Mobile Traffic Forecasting Using Deep Spatio-Yemporal Neural Networks. In *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*. 231–240.

[41] Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. 2020. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems* 21, 9 (2020), 3848–3858. https://doi.org/10.1109/TITS.2019.2935152

[42] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. 2020. GMAN: A Graph Multi-Attention Network for Traffic Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 01 (Apr. 2020), 1234–1241. https://doi.org/10.1609/aaai.v34i01.5477

A DETAILS OF AUGMENTED SPATIAL KNOWLEDGE GRAPH

A.1 Entities

- **Region.** According to the road network, the urban space is divided into regions, which aim to reflect the organizational structure of the urban space and the location choice of people in the city for social production and life.
- **Business Areas (Ba).** Business areas are spaces within a city that result from the clustering of economic and social activities. They are formed through the process of social production and daily life, but do not have clear physical boundaries.
- **Points of Interest (POI).** POIs describe the basic functional units and places in the city, such as schools, hospitals, shopping malls, etc., which are fine-grained places where people carry out social production and life in the city.
- **Categories (Cate).** Categories describe the functional attributes and categories of POI entities and combine expert knowledge to divide all POI entities into 14 categories.

A.2 Relationships

In the augmented spatial knowledge graph, the relationship between urban elements is distinguished into three aspects, spatial relationship, subordination relationship and functional relationship.

Spatial relationships model the spatial location and attribute knowledge between urban elements.

- **borderBy.** The *borderBy* relationship is a symmetrical association between two regions that share a common boundary, which describes the relationship between regions that are adjacent to each other.
- **nearBy.** The *nearBy* relationship is a symmetrical association between two regions that are in spatial proximity to each other, which describes the relationship between two region entities that are within a certain distance of each other.
- **locateAt.** The *locateAt* relationship describes the spatial relationship between a POI and a region, which associates the POI with the region where its physical location is situated.
- **baseLocateAt.** The *baseLocateAt* relationship associates a base station with the region where it is physically located, describing their spatial relationship.

Subordination relationships model the affiliation between urban elements at the cognitive level.

- **belongTo.** The *belongTo* relationship describes the association between a POI and the business area it belongs to, which links the POI with the corresponding business area in the knowledge graph. As the coverage of different business areas may overlap, a POI can be associated with multiple business areas through this relationship.
- **CateOf.** The *CateOf* relationship describes the category attribute information of the POI.
- **baseBelongTo.** The *baseBelongTo* relationship describes the association between a base station and the business area it belongs to, which links the base station with the corresponding business area in the knowledge graph. As the coverage of different business areas may overlap, a base station entity can be associated with multiple business areas through this relationship.

Table 4: Statistics of relationships in Shanghai and Nanjing-datasets.

Semantic Info	Relation	Shanghai	Nanjing
Spatial	borderBy	13550	5498
	nearBy	39312	12658
	locateAt	85018	51264
	baseLocateAt	4505	8000
Subordinate	belongTo	67760	42813
	CateOf	85018	51264
	baseBelongTo	4012	7069
Function	BaServe	13256	6457
	competitive	508	-
	coCheckIn	13621	-
	similarFunction	5152	-
	flowTransition	650	-
	POIserverBy	352137	228923

Functional relationships model specific functional-related knowledge between urban elements from different functional perspectives.

- **BaServe.** The *BaServe* relationship inks business areas and regions, connecting the business area to the region within its service range. As service ranges may overlap, a region can be linked to multiple business areas through this relationship.
- **competitive.** The *competitive* relationship describes the competition between POIs in terms of their economic attributes and spatial locations, which associates two competitive POIs with each other in the knowledge graph.
- **coCheckIn.** The *coCheckIn* relationship describes the correlation between two POIs in terms of user access, which associates two correlation POIs with each other in the knowledge graph.
- **similarFunction.** The *similarFunction* relationship describes the similarity between regions in terms of urban functions and associates two regions with similar urban functions in the knowledge graph.
- **flowTransition.** The *flowTransition* relationship describes the transfer of people flow between two regions, which are referred to as the source region and the destination region, which associates the two regions with each other in the knowledge graph.
- **POIserverBy.** The *POIserverBy* relationship describes the service dependency between base stations and POIs, which associates base station with POIs within their service radiation range. As the service radiation ranges of different base stations may overlap, a POI can be associated with multiple base stations through this relationship.

Table 4 shows the statistics of relationships in Shanghai and Nanjing datasets. Due to the absence of user access data, the Nanjing dataset lacks *competitive*, *coCheckIn* *similarFunction* and *flowTransition* relationships.