KGDA: A Knowledge Graph Driven Decomposition Approach for Cellular Traffic Prediction

JIAHUI GONG, TONG LI, HUANDONG WANG, and YU LIU, Tsinghua University, Beijing, China
XING WANG, ZHENDONG WANG, CHAO DENG, and JUNLAN FENG, China Mobile Research Institute, Beijing, China
DEPENG JIN and YONG LI, Tsinghua University, Beijing, China

Understanding and accurately predicting cellular traffic data is vital for communication operators and device users, as it facilitates efficient resource allocation and ensures superior service quality. However, large-scale cellular traffic data forecasting remains challenging due to intricate temporal variations and complex spatial relationships. This article proposes a Knowledge Graph Driven Decomposition Approach (KGDA) for precise cellular traffic prediction. The KGDA breaks down the impact of static environmental factors and dynamic autocorrelations of cellular traffic time series, enabling the capture of overall traffic changes and understanding of traffic dependence on past values. Specifically, we propose an urban knowledge graph to capture the static environmental context of base stations, mapping these entities into the same latent space while retaining static environmental knowledge. The cellular traffic is divided into a regular pattern and fluctuating residual components, with the KGDA comprising four modules: a Knowledge Graph Representation Learning model, a traffic regular pattern prediction module, a traffic residual dynamic prediction module, and an attentional fusion module. The first leverages graph neural networks to extract spatial contexts and predict regular patterns, the second utilizes the Bi-directional Long Short-Term Memory (Bi-LSTM) model to capture autocorrelations of traffic time series, and the final module integrates the patterns and residuals to produce the final prediction result. Comprehensive experiments demonstrate that our proposed model outperforms state-of-the-art models by more than 10% in forecasting cellular traffic.

CCS Concepts: • **Networks** → **Mobile networks**; • **Information systems** → *Mobile information processing systems*;

Additional Key Words and Phrases: knowledge graph, mobile traffic prediction, graph neural networks

This research has been supported in part by the National Key Research and Development Program of China under Grant 2022ZD0116402; in part by the National Natural Science Foundation of China under Grant U22B2057, Grant U21B2036, Grant U20B2060; in part by the joint project of China Mobile Research Institute and Tsinghua University; and in part by the grant from the Guoqiang Institute, Tsinghua University under 2021GQG1005.
Authors' Contact Information: Jiahui Gong, Tsinghua University, Beijing, China; e-mail: gjh22@mails.tsinghua.edu.cn; Tong Li, Tsinghua University, Beijing, China; e-mail: t.li@connect.ust.hk; Huandong Wang, Tsinghua University, Beijing, China; e-mail: wanghuandong@tsinghua.edu.cn; Yu Liu (corresponding author), Tsinghua University, Beijing, China; e-mail: liuyu2419@126.com; Xing Wang, China Mobile Research Institute, Beijing, China; e-mail: wangxing@chinamobile.com; Zhendong Wang, China Mobile Research Institute, Beijing, China; e-mail: wangzhendongai@chinamobile.com; Chao Deng, China Mobile Research Institute, Beijing, China; e-mail: dengchao@chinamobile.com; Junlan Feng, China Mobile Research Institute, Beijing, China; e-mail: fengjunlan@chinamobile.com; Depeng Jin, Tsinghua University, Beijing, China; e-mail: jindp@tsinghua.edu.cn; Yong Li, Tsinghua University, Beijing, China; e-mail: liyong07@tsinghua.edu.cn.

ACM 2157-6912/2024/11-ART123
https://doi.org/10.1145/3690650

ACM Reference format:
Jiahui Gong, Tong Li, Huandong Wang, Yu Liu, Xing Wang, Zhendong Wang, Chao Deng, Junlan Feng, Depeng Jin, and Yong Li. 2024. KGDA: A Knowledge Graph Driven Decomposition Approach for Cellular Traffic Prediction. *ACM Trans. Intell. Syst. Technol.* 15, 6, Article 123 (November 2024), 22 pages.
https://doi.org/10.1145/3690650

1 Introduction

Accurately forecasting cellular traffic is essential for providing reliable and high-quality Internet services while optimizing resource utilization [20, 40, 47, 49]. This capability benefits both communication operators and smartphone users [13, 27, 35]. For example, communication operators can turn off base stations during expected decreases in traffic to save resources and energy consumption [25, 28] and, conversely, activate base stations in anticipation of increased traffic to avoid potential network congestion [52]. Furthermore, accurate cellular traffic prediction enables network detectors to monitor unusual user activities [26], allowing the identification of unexpectedly high cellular traffic volumes and abnormal devices [7, 46].

Researchers have invested considerable effort in predicting cellular traffic. Traditional time series prediction models, such as **Support Vector Regression (SVR)** [9], ARIMA [42], and SARIMA [34], have been used by treating cellular traffic prediction as a general time series forecasting issue. Additionally, **Recurrent Neural Network (RNN)** [50], **Long Short-Term Memory (LSTM)** [18], and **Gated Recurrent Unit (GRU)** [4] have been utilized to enhance forecasting accuracy. In recent years, several models [5, 48, 51, 54] have used **Graph Neural Networks (GNNs)** [33] to capture spatial information, e.g., the distance relationship between regions and base stations. Current models are limited to analyzing the autocorrelations within traffic time series data and do not consider the different environmental factors that may affect the base stations. This shortfall reduces the efficiency of current methods.

Several existing studies on mining cellular traffic data have shown that particular traffic patterns in base stations are affected by contextual information in the surrounding environment [38, 53]. For instance, base stations situated in commercial districts witnessed a notable increase in traffic during typical office hours. Likewise, those in residential zones exhibited peaks in traffic during the early morning and late evening hours, indicative of residents being at home and engaging with their devices for leisure or work purposes. By considering and modeling these static environmental factors, valuable insights into the general trends of cellular traffic can be gained, enabling more accurate predictions for future traffic patterns.

This article aims to incorporate static environmental factors and propose a **Knowledge Graph Driven Decomposition Approach (KGDA)** for precise cellular traffic prediction. Our approach involves explicitly breaking down the impact of static environmental factors and dynamic autocorrelations of cellular traffic time series. This enables us to capture the overall direction of traffic changes and understand how traffic depends on its past values. We propose an urban knowledge graph aimed at encapsulating static environmental information. The graph comprises diverse environmental elements, including **Points of Interest (POIs)**, regions, business districts, and base stations, each depicted as distinct entities interconnected to illustrate their relationships. We can efficiently map these entities into a unified latent space by employing a knowledge graph embedding model, thereby preserving the static environmental knowledge crucial for base stations. Furthermore, we break down the cellular traffic into regular patterns and fluctuating residual components. We discover that regular traffic patterns of base stations primarily depend on static environmental factors while the varying residuals can be predicted by modeling the autocorrelations of the traffic's historical data. The KGDA we proposed comprises three components: a traffic regular pattern

prediction module, a traffic residual dynamic prediction module, and an attentional fusion module. The regular pattern prediction model employs GNNs to extract spatial contexts of base stations from the urban knowledge graph, leveraging these contexts to forecast the regular patterns of base stations. In contrast, the traffic residual dynamic prediction module utilizes the **Bidirectional Long Short-Term Memory (Bi-LSTM)** model. This model adeptly captures autocorrelations within traffic time series, where the current value of a variable relies on its past values. Finally, the attentional fusion module integrates the patterns and residuals to produce the final prediction result.

Below are the summarized major contributions of our work:

—We are the first to use the urban knowledge graph to capture environmental information and decompose cellular traffic series into patterns and residuals. The proposed decomposition approach allows us to identify the general trend of traffic changes and gain insight into how traffic is influenced by its previous values.
—We propose the KGDA. This approach comprises four components: Knowledge Graph Representation Learning using the TuckER model, regular pattern estimation using **Graph Convolutional Network (GCN)**, residual dynamic estimation using Bi-LSTM, and attention fusion for integrating predictions. KGDA extracts spatial information from the urban knowledge graph to estimate base station regular patterns, predicts residual traffic dynamic, and generates precise predictions by combining patterns and residuals.
—We conducted comprehensive experiments on two real-world datasets. Our proposed model demonstrates superior performance, surpassing state-of-the-art approaches by over 10% in precision, highlighting its efficiency and effectiveness. Additionally, we conducted an in-depth analysis further to illustrate the effectiveness of the urban knowledge graph.

This article is structured as follows. We initially present the preliminaries in Section 2, which includes an introduction to cellular traffic patterns, the urban knowledge graph, and the problem definition. We then outline the entire framework of our model, the KGDA, along with the specifics of each part of KGDA in Section 3. We describe evaluation environments in Section 4. Related studies are discussed in Section 5. Finally, we draw conclusions in Section 6.

2 Preliminaries

2.1 Cellular Traffic Patterns

In order to improve the accuracy of predicting cellular traffic, we suggest creating models that incorporate the consistent patterns of base stations and their surrounding environment. In this section, we will demonstrate how environmental factors impact the usage of base station traffic by visualizing the distribution of temporal traffic. We will be using real-world data from 4,505 base stations located in Shanghai.

Each base station has its own stable traffic patterns. Figure 1 illustrates the similarity between cellular traffic temporal distributions across 3 weeks and the regular pattern, with reduced traffic during sleeping hours and increased traffic during the daytime [44]. On a weekly basis, a base station's cellular traffic oscillates around its regular pattern, with pronounced differences between weekdays and weekends [8].

The location of the base station significantly influences its traffic pattern [6]. Areas such as commercial centers, residential areas, and entertainment places encounter different levels of user activity at different times of the day. This leads to unique surges in cellular traffic at various times. To identify these patterns, we used hierarchical clustering methods and an unsupervised machine learning algorithm with Euclidean distance as the distance metric. By integrating the specific locations of each base station cluster, we have identified four primary cellular traffic patterns, as

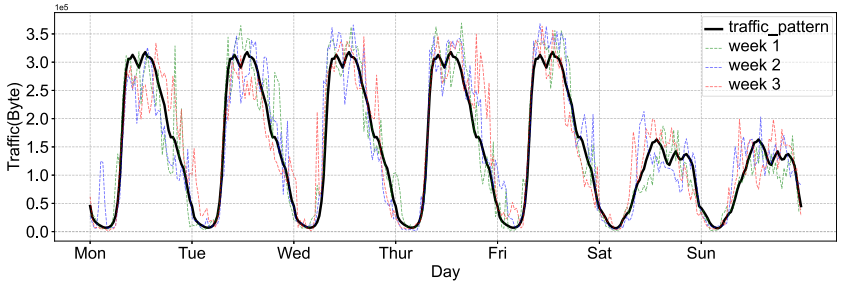
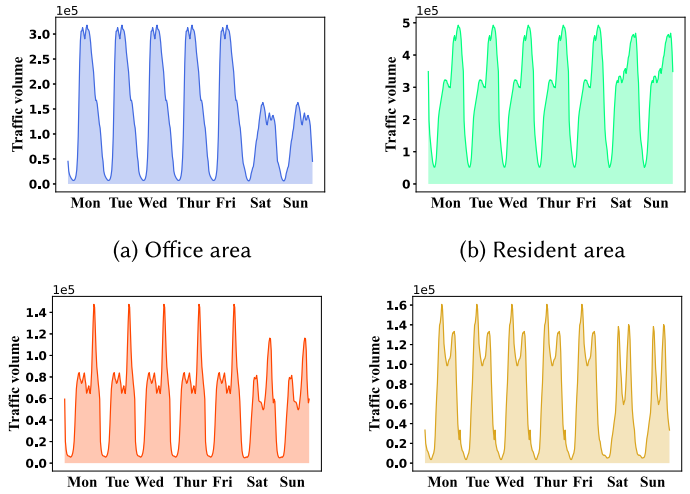


Fig. 1. An example of traffic pattern of the base station. A base station has its own traffic pattern and varies greatly between weekdays and weekends.



(a) Office area (b) Resident area

(c) Entertainment area (d) Transport area

Fig. 2. Different traffic patterns in different areas.

shown in Figure 2. These four patterns vary based on peak traffic value times and disparities in traffic value during a week. For instance, base stations in residential areas typically witness higher traffic volumes at night, attributed to residents commuting to work in the morning and returning home at night. Conversely, base stations situated in office areas experience peak traffic during daytime hours. Additionally, weekend traffic volumes generally decrease compared to weekdays, reflecting reduced work-related commuting. Base stations near the subways and stations observe heightened traffic loads during rush hours. Consequently, the environmental features surrounding base stations significantly shape their traffic flows. Stations with similar environmental features tend to display analogous traffic patterns, and vice versa. Hence, we regard environmental contexts as crucial exogenous factors in shaping traffic profiles for base stations.

2.2 Urban Knowledge Graph

Constructing an urban knowledge graph entails aggregating diverse data from various sources, encompassing road network data, POI data, business area data, category data, and base station data. Next, we introduce the construction process of each dataset in detail.

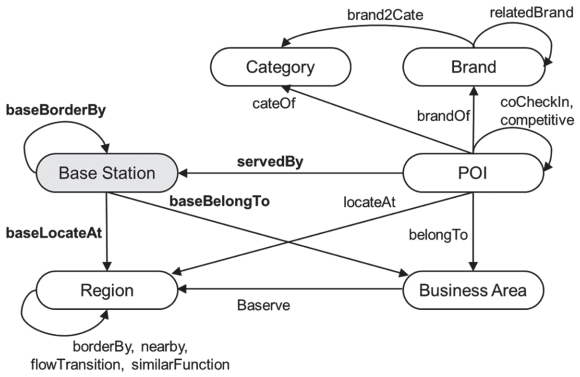


Fig. 3. The overview of the urban knowledge graph.

—*Road network dataset*: The road network dataset delineates cities based on their road infrastructure, serving as a fundamental source for depicting urban transportation networks. This dataset is compiled by extracting data from Baidu Maps.
—*POI dataset*: The POI dataset provides comprehensive information about various POI within the city, offering insights into urban functions and the spatial arrangement of infrastructure. This dataset is collected and constructed in cooperation with Internet application service providers.
—*Business area dataset*: The business area dataset pinpoints key zones characterized by bustling commercial and economic activities, essential for understanding the city's economic landscape. This dataset is collected and constructed in cooperation with Internet application service providers.
—*Category dataset*: The category dataset amalgamates expert insights to categorize and define the functional attributes of different urban facilities, enriching our understanding of the city's diverse infrastructure.
—*Base station dataset*: The base station dataset furnishes details about the distribution of base stations across the city, generously provided by mobile operators.

Integrating data from these disparate sources makes the urban knowledge graph a robust framework, empowering comprehensive insights into urban dynamics and structures.

To more comprehensively characterize the environment surrounding the base station, we propose building an urban knowledge graph. This graph is crafted to model the semantic relations between urban elements, including business areas, POIs, regions, brands, and categories. We enhance the urban knowledge graph by integrating base station entities and establishing connections with their adjacent regions, business areas, and POIs. Additionally, we link neighboring base station entities, drawing upon the spatial correlation in cellular traffic. Figure 3 showcases the schema of our refined urban knowledge graph, which effectively incorporates the influence of the urban environment on cellular traffic.

Formally, we define the urban knowledge graph as $G = (E, R, \mathcal{F})$, where E, R, and F are the sets of entities, relationships, and facts therein. The fact set includes triplets on factual knowledge, i.e., $\mathcal{F} = \{(h, r, t)|h, t \in \mathcal{E}, r \in \mathcal{R}\}$.

2.3 Problem Definition

We delineate the cellular traffic prediction task. In the framework of a base station network, the task involves utilizing historical cellular traffic series associated with each one. The aim of cellular traffic forecasting is to train a model to forecast future cellular traffic volumes.

Table 1. Summary of Notations

Notations	Definition
b, B	A base station and the set of base stations
x_i^t	The traffic volume of base station bs_i at timestamp t
I	The input length and the patch length
s_i^t	The traffic sequence of bs_i at t in the past I steps
h, r, t	The head entity, relationship, and the tail entity in the knowledge graph
M^t	The regular traffic pattern
R^t	The residual traffic series
E_K	The node embedding of knowledge graph

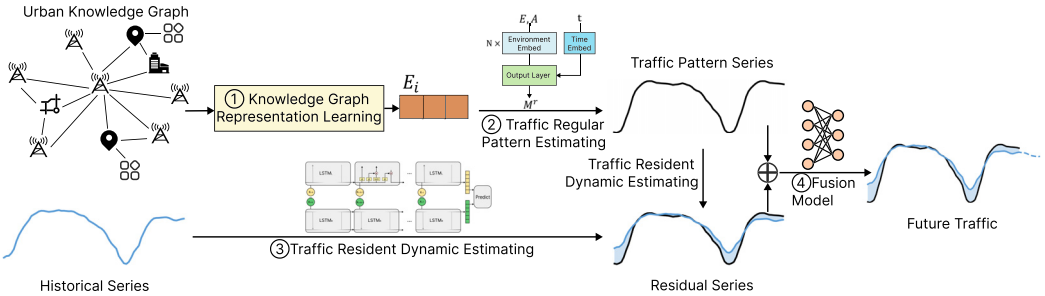


Fig. 4. Overview of our KDGA model.

Mathematically, let $B = b_1, b_2, ..., b_N$ denote the set of base stations, where N represents the total number of base stations. Each $x_i^t \in \mathbb{R}$ signifies the traffic volumes of base station b_i at timestamp t. Furthermore, let $s_i^t = \left[x_i^{t-T+1}, x_i^{t-T+2}, ..., x_i^t\right] \top \in \mathbb{R}^T$ denote the past T timesteps of base station b_i at time t. Consequently, we define $S^t = \left[s_1^t, s_2^t, ..., s_N^t\right] \in \mathbb{R}^{N \times T}$ as the historical traffic matrix encompassing all base stations at time t. For a detailed explanation of the notations, please refer to Table 1.

Cellular traffic prediction endeavors to anticipate future cellular traffic by leveraging past traffic data. This task entails training a mapping function, denoted as f, using historical traffic series S^t to forecast the traffic volumes for the subsequent timestamp, which can be succinctly expressed as:

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

3 Methods

3.1 Framework Overview

The KGDA framework is depicted in Figure 4. With the urban knowledge graph G, timestamp t, and historical cellular traffic matrix S^t as input, KGDA forecasts the future cellular traffic volumes for all nodes in the network. Specifically, the KGDA model consists of four parts: the Knowledge Graph Representation Learning model, the Traffic Regular Pattern Estimating model, the traffic residual dynamics estimating model, and the Attentional Fusion Model.

First, we utilize Knowledge Graph Representation Learning to generate knowledgeable representations for network nodes in the urban knowledge graph, which captures both the spatial structure of the node network and the functional similarity of nodes therein. Second, the Traffic Regular Pattern Estimating model takes the node embedding E, the adjacency matrix of the base station network A, and the timestamp t as the input. The model estimates the cellular traffic patterns M^t

on weekdays and weekends. Third, we use the historical cellular traffic as the input, and the Traffic Residual Dynamics Estimation model estimates the difference between the true volumes and the historical mean volumes at the current moment. Finally, the Attentional Fusion model takes the mean traffic value M^t, which is the output of the Traffic Regular Pattern Estimating model and the residual traffic value, the output of the Traffic Residual Dynamics Estimating model, as the input and forecasts the future cellular traffic volumes of each node in the mobile network.

3.2 Knowledge Graph Representation Learning

To harness the full potential of urban knowledge graphs for cellular traffic prediction, we initially employ Knowledge Graph Representation Learning to acquire low-dimensional vectors (embeddings) for base stations within the graph. In particular, we utilize Tucker model [2] and assess the plausibility of a triplet $(h, r, t) \in \mathcal{F}$:

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

where $\mathcal{W} \in \mathbf{R}^{d \times d \times d}$ is the key tensor in Tucker decomposition [36], $\boldsymbol{e}_h, \boldsymbol{e}_t, \boldsymbol{r} \in \mathbf{R}^d$ are embeddings for entities and relation. Besides, d signifies the dimension of embedding, and $\times_i$ denotes the tensor product along the ith mode.

From the observed triplets within the urban knowledge graph, we compute plausibility scores using the specified scoring function. Following this, we formulate cross-entropy loss functions for parameter learning, with the aim of prioritizing valid triplets over invalid ones. Ultimately, the acquired embeddings for base station entities can be seamlessly integrated into the subsequent traffic prediction module, effectively leveraging the captured urban knowledge.

3.3 Traffic Regular Pattern Estimating

The Traffic Regular Pattern Estimating model takes the node embedding E, which we obtained from the previous model, the adjacency matrix A, and the timestamp t as the input. The model estimates the cellular traffic patterns on weekdays and weekends and outputs the mean cellular traffic matrix M.

First, the adjacency matrix A is calculated from the distances among base stations in the network. The adjacency matrix A can be formed as

$$a_{i,j} = \begin{cases} exp\left(\frac{-d_{i,j}}{\sigma^2}\right), & exp\left(\frac{-d_{i,j}}{\sigma^2}\right) \geq \epsilon, \\ 0, & exp\left(\frac{-d_{i,j}}{\sigma^2}\right) < \epsilon, \end{cases} \tag{3}$$

where $a_{i,j}$ is the value in the adjacency matrix and is decided by $d_{i,j}$ the distance between node v_i and node v_j. σ^2 and ϵ are thresholds to control the distribution and sparsity of the adjacency matrix.

The adjacency matrix A represents the interconnections between nodes in the network. These connections effectively capture the distance relationships among the nodes, reflecting the relevance of cellular traffic volume between them. For instance, when a smartphone user moves out of the coverage area of a particular base station, they are more likely to connect to a nearby base station and transfer their traffic volume accordingly. Thus, capturing the distance relationships among nodes in the network proves valuable for predicting cellular traffic patterns.

In addition, we incorporate the embedding matrix E_K obtained from the Knowledge Graph Representation Learning model as the feature matrix in our approach. The Knowledge Graph Representation Learning model enables us to capture the topological similarities among nodes within the urban knowledge graph. On the other hand, the embedding matrix E_K reflects the similarity of the scenes covered by base stations. When two base stations have similar covered

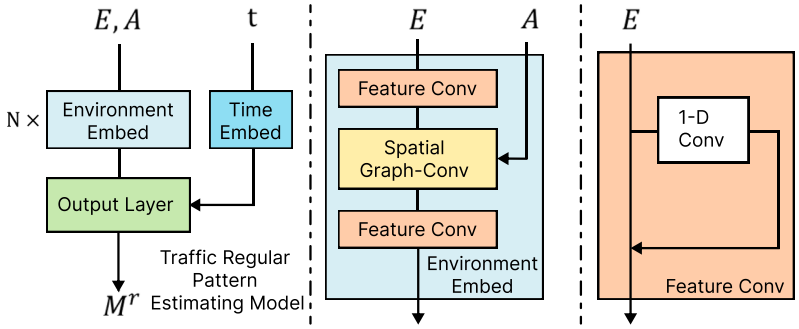


Fig. 5. Traffic Regular Pattern Estimating model.

scenes, it is more likely that their cellular traffic patterns will exhibit similarities as well. For instance, if a base station predominantly covers educational institutions, the cellular traffic will likely increase during the daytime on weekdays and decrease during nighttime or weekends. Conversely, if the base station primarily covers transportation hubs, cellular traffic may peak during rush hours and decline during other times. Hence, considering the similarity of covered scenes among base stations proves beneficial for predicting cellular traffic patterns.

In this subsection, we study the above two similarities. The detail of the Traffic Regular Pattern Estimating model is shown in Figure 5. To capture the topology similarity among the nodes in the network, we employ the **Convolution Neural Network (CNN)** to capture the features of the embedding vector, which can be expressed as

$$E_C = (H * K)(x) = \sum_{i=1}^{k} H(x+i)K(i), \tag{4}$$

where $*$ denotes the convolution operation, k is the filter size. The filter multiple k elements by its k parameters and outputs the matrix E_C.

To capture the distance relation among nodes in the network, we use a GCN [23], which can be expressed as

$$E_G = f(E_C, A) = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}EW^l), \tag{5}$$

where the $\tilde{A}$ denotes the adjacency matrix which we calculated above, with self-loop, and the $\tilde{D}$ denotes the degree matrix of $\tilde{A}$.

Next, we add an additional CNN layer to process and extract features further. This step aids the model in comprehending global feature patterns more effectively, thereby enhancing its understanding of the overall data structure and information.

As for the timestamp t, we let them pass through the embedding layer and obtain the time embedding E^t. Ultimately, we proposed a **Multi-Layer Perception (MLP)** as the output layer. The MLP can be formed by,

$$M^t = f(S, E^t) = W_2(\sigma(W_1(S||E_t) + b_1)) + b_2, \tag{6}$$

where $||$ denotes the concatenation operation. The M is the mean cellular traffic patterns and the Traffic Regular Pattern Estimating model output. The W and the b are the weight matrix and bias matrix, respectively, which are trainable parameters.

3.4 Traffic Residual Dynamics Estimating

To capture the temporal dependences from cellular traffic series, we proposed a Traffic Residual Dynamics Estimating model, which takes the historical traffic matrix S^t and the future traffic pattern

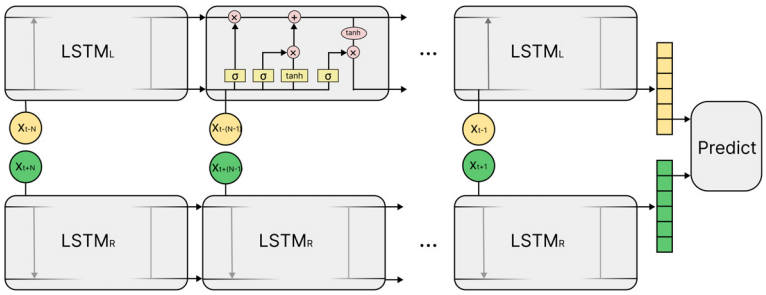


Fig. 6. Traffic Residual Dynamic Estimating model.

matrix M^t as the input and outputs the difference between the true volumes and the historical mean volumes at the current moment. As shown in Figure 6, we use the Bi-LSTM model [14], which could extract temporal features from the historical traffic sequence and future pattern sequence. In this context, the residual represents the difference between the actual traffic volumes and the pattern volumes. And the dynamic embodied in the underlying variations in the traffic data that are not fully captured by the traffic patterns, including short-term fluctuations, sudden spikes or drops, and so forth. The model can be formed as,

$$R^t = Bi - LSTM(s^t, M^t). \tag{7}$$

The Traffic Residual Dynamics Estimating model can estimate the temporal dependency and predict the residual cellular traffic matrix R.

3.5 Fusion Model

The Fusion model takes the mean cellular traffic matrix M^t, which is the output of the Traffic Regular Pattern Estimating model, and the residual cellular traffic matrix R^t, which is the output of the Traffic Residual Dynamics Estimating model. We employ an MLP model to fuse the mean and residual matrices. The model can be formed as

$$\hat{x}_1^{t+1}, \hat{x}_2^{t+1}, ..., \hat{x}_N^{t+1} = MLP(M^t, R^t). \tag{8}$$

The model is trained in batches to minimize the disparity between predicted volumes $\left[\hat{x}_1^{t+1}, \hat{x}_2^{t+1}, ..., \hat{x}_N^{t+1}\right]$ and their true volumes $\left[x_1^{t+1}, x_2^{t+1}, ..., x_N^{t+1}\right]$. Throughout the training procedure, we employ the MSE loss function to optimize the model's parameters, which is expressed as follows:

$$\mathcal{L} = \sum_{i=1}^{N} \left\| \hat{x}_i^{t+1} - x_i^{t+1} \right\|_2. \tag{9}$$

4 Evaluation

4.1 Experimental Settings

In this subsection, we will introduce our read-world datasets, metrics, baselines, and parameter settings.

Datasets.

— *Shanghai Dataset.* The Shanghai Dataset comprises anonymous cellular traffic data gathered by China Mobile in Shanghai throughout August 2014. These datasets encompass records from 4,505 base stations and more than 150,000 users. Each entry in the dataset comprises an anonymous device ID, starting and ending timestamps of the data record, anonymous

Table 2. Statistics of the Datasets Used in Our Experiments

Dataset	Shanghai	Nanjing
Collection Duration	1–31 August 2014	2 February to 31 March 2021
Time Interval	30 minutes	
Covered Users	≥150,000	≥450,000
Covered BSs	4,505	8,000
Covered Area	6,340 km^2	6,587 km^2
Flow Records	8.65×10^8	8.18×10^8
Regions	2,579	1,022
Business Areas	280	228
POIs	85,018	51,264
Categories	14	14
Relationships	17	14
Nodes	92,296	60,528
Edges	≥ 1 M	≥ 600 K

base station ID, and the volume of data transmitted during the connection. We contributed a total of 1.96 billion entries corresponding to the 4,505 base stations in Shanghai, logged at 30-minute intervals. Furthermore, the Shanghai urban knowledge graph incorporates 5 entity types and 14 relationship types, housing 92,396 nodes and over one million edges, making it a substantial urban knowledge graph.

—*Nanjing Dataset.* The Nanjing Dataset comprises anonymous cellular traffic data collected by China Mobile in Nanjing spanning from 2 February to 31 March 2021. This dataset surpasses the Shanghai Dataset in scale and encompasses data from 8,000 base stations. We facilitated data collection for all 8,000 base stations in Nanjing at 30-minute intervals. Additionally, the Nanjing urban knowledge graph comprises 5 entity types and 14 relationship types, featuring 60,528 nodes and over six hundred thousand edges.

Metrics. We elaborately select the following three metrics to evaluate the performance of cellular traffic prediction:

—*RMSE*

$$RMSE = \sqrt{\frac{1}{N \cdot T} \sum_{i=1}^{N \cdot T} (Y_i - \hat{Y}_i)^2}, \tag{10}$$

where N denotes the number of base stations in the network and T is the timestamps of test data.

—*Mean Absolute Error (MAE)*

$$MAE = \frac{1}{N \cdot T} \sum_{i=1}^{N \cdot T} \left| Y_i - \hat{Y}_i \right|. \tag{11}$$

—*Coefficient of Determination (R^2)*

$$R^2 = 1 - \frac{\sum_{i=1}^{N \cdot T} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{N \cdot T} (Y_i - \bar{Y})^2}, \tag{12}$$

where $\bar{Y}$ denotes the average value of Y, i.e., $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$

Compared Algorithms.

—*SVR* [9]. SVR constitutes a crucial branch within the **Support Vector Machine (SVM)** framework, initially designed for binary classification tasks. In regression applications, SVR utilizes a linear SVM approach.
—*ARIMA* [42]. ARIMA stands as a prevalent technique in time series forecasting. It integrates **auto-regressive (AR)** and **moving average (MA)** components, where "p" signifies the count of AR terms and "q" denotes the number of MA terms. The parameter "d" indicates the number of differences applied to achieve temporal smoothness in the series.
—**Graph Attention Network (GAT)** [37]. In response to the limitations of GCN regarding dynamic graphs and varying learning weights for neighbors, researchers devised the GAT. GAT incorporates a masked graph attention mechanism, empowering each node to interact attentively with its neighbors and aggregate their outcomes. By adopting this approach, GAT adeptly surmounts these constraints, bolstering the learning process within dynamic graph scenarios.
—**Graph Sample and Aggregate (GraphSAGE)** [16]. The core of the GraphSAGE algorithm revolves around optimizing the sampling process, specifically focusing on sampling the current neighboring node rather than the entire graph. This optimization allows for efficient training on large-scale graphs and leads to enhanced performance, surpassing that of GCN even with smaller parameters and faster processing speed.
—*DeepTP* [11]. DeepTP represents an end-to-end deep learning architecture designed to precisely forecast spatially correlated cellular traffic over extended durations. It employs a flexible feature extraction mechanism to capture spatial dependencies and incorporate external data. Additionally, DeepTP integrates a sequential module to capture intricate temporal patterns, facilitating comprehensive traffic predictions.
—**Spatio-Temporal Graph Convolutional Networks (STGCN)** [48]. STGCN combines GCN and gated CNN to capture spatial and temporal dependencies. GCN extracts spatial dependencies by analyzing the graph's topological structure represented by the adjacency matrix. Gated CNN explores dynamic features of cellular traffic to detect temporal dependencies. The model predicts future cellular traffic using an output layer. In STGCN, the adjacency matrix is replaced with an embedding matrix to capture spatial structure more accurately, resulting in STGCN (emb).
—**Temporal Graph Convolutional Network (T-GCN)** [54]. In the T-GCN, GCN and GRU are combined to capture temporal dependencies in dynamic cellular traffic. GCN is utilized to model time series and explore the changing node attributes, while GRU is utilized to model the dynamic nature of the data. Additionally, the network's hidden state is initialized using an embedding matrix to enhance the capture of topology similarity. This variant is referred to as T-GCN (emb).
—**Graph Multi-Attention Network (GMAN)** [55]. In its architecture, GMAN utilizes an encoder–decoder framework with spatiotemporal attention blocks to accurately represent the impact of spatial and temporal parameters on traffic conditions. The encoder handles input traffic characteristics, while the decoder forecasts the output timestamp series.
—*PatchTST* [31]. PatchTST integrates two significant components: first, the time series segmentation into patches at the sub-series level, which serves as input tokens, and second, channel independence, where each channel comprises a distinct time sequence sharing the same embedding and model weights.
—**Spatio-Temporal Self-Supervised Learning (ST-SSL)** [22]. ST-SSL framework enhances the traffic pattern representations to be reflective of both spatial and temporal heterogeneity, with auxiliary self-supervised learning paradigms.

Table 3. Values of the Hyper-Parameters

Modules	Hyper-Parameters	Value
Global Variables	Epoch	50
	Optimizer	Adam
	Learning Rate	0.01
	Schedular	Step Learning Rate Scheduler
	γ	0.9
	Loss	MSE Loss
	Input Length	12
	Dimension of Embedding	32
	Dropout Rate	0.5
Knowledge Graph Representation Learning	Epoch	50
	Dimension of Embedding	32
Traffic Regular Pattern Estimating	Dimension of Embedding	32
	Dropout Rate	0.5

Parameter Settings. We set the learning rate as 0.01 and apply a Step Learning Rate scheduler to decrease the learning rate per 10 epochs set $\gamma = 0.9$ and apply the MSE loss to the train model. To balance the efficiency and the performance, in the Knowledge Graph Representation Learning model, we set the dimension of the embedding as 32, and the epoch as 50. As for the Traffic Regular Pattern Estimating model, we set the output time of the embedding size of the embedding layer as 8. We set the length of the input historical traffic series and the length of future traffic patterns as 12, which has both higher accuracy and faster speed. In order to prevent overfitting the model, we set the dropout rate as 0.5. We apply it to the outputs of both the Traffic Regular Pattern Estimating model and the Attentional Fusion model. Our experiments use an early stopping strategy with a patience of 5 epochs on the validation dataset. The dataset is divided into three components: training, validation, and test, with a ratio of 0.7:0.05:0.25. We provide detailed values of the hyper-parameters in Table 3 for reproducibility. The experiments are performed on the "Jiutian"[1] artificial intelligence platform.

4.2 Overall Performance

In Table 4, we show the overall performances of our model KGDA, temporal prediction models (SVR, ARIMA, and PatchTST), spatial prediction model (GAT and GraphSAGE), and spatial-temporal prediction models (DeepTP, STGCN, T-GCN, and GMAN) to forecast the futural traffic volumes in the dataset. We list three metrics of all algorithms. The RMSE and MAE are log-normalized. According to the result, we have the following findings:

— *Our framework steadily achieves the best performance.* Our model consistently outperforms other compared algorithms on both datasets. For instance, compared with the second-best performing model (GMAN), our model demonstrates an improvement in R^2 ranging from approximately 4.7% to 5.8%. Moreover, our model achieves a reduction in MAE of approximately 16.7% to 19.3%.
— *Geographic prediction models perform poorly in the cellular traffic prediction task.* Spatial models are frequently utilized in analyzing spatial patterns and location-based data. Nonetheless, these algorithms overlook the crucial components for modeling time sequences and capturing temporal information. Consequently, their efficacy might be limited compared to models

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

Table 4. Overall Prediction Performance of Our Model KGDA in Comparison with Compared Algorithms on Our 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
STGCN	0.1996	0.2785	0.7767	0.2537	0.3642	0.8058
STGCN (emb)	0.1978	0.2767	0.7795	0.2505	0.3619	0.8083
T-GCN	0.1908	0.2694	0.7990	0.2516	0.3519	0.8178
T-GCN (emb)	0.1893	0.2674	0.8011	0.2497	0.3502	0.8190
DeepTP	0.1869	0.2610	0.7991	0.2322	0.3327	0.8196
PatchTST	0.1963	0.2708	0.7791	0.2634	0.3728	0.7989
ST-SSL	0.1811	0.2567	0.8061	0.2251	0.3286	0.8220
GMAN	0.1807	0.2554	0.8078	0.2209	0.3237	0.8237
KGDA	**0.1507**	**0.2234**	**0.8548**	**0.1934**	**0.2857**	**0.8624**

Bold values indicate the best performance across all methods. Underlined values indicate the second-best performance across all methods.

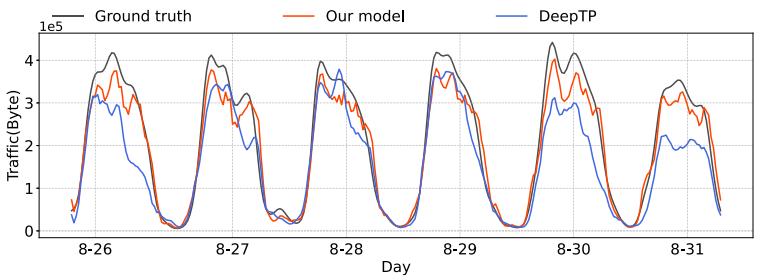


Fig. 7. Prediction vs. the ground truth for a sample base station.

seamlessly integrating temporal features. Integrating temporal elements into geographic models can lead to more precise predictions and notably improved performance in real-world scenarios.

—*It is essential to capture various environmental information.* STGCN and T-GCN solely rely on the distance adjacency matrix and fail to fully exploit environmental information, resulting in compromised model performance. However, by substituting the distance adjacency matrix with the embedding matrix derived from the urban knowledge graph, we were able to enhance the performance of the baseline model and demonstrate the effectiveness of leveraging the urban knowledge graph.

We depict the predicted values of KGDA against the real values, alongside those of the second-best model, DeepTP, on the testing datasets in Figure 7. The plot demonstrates that KGDA produces predictions with greater stability and lower latency compared to DeepTP.

4.3 Case Study

—*Effectiveness of Embedding Matrix of Urban Knowledge Graph.* To showcase the capabilities of the Traffic Regular Pattern Estimating model and the effectiveness of the embedding matrix E, we

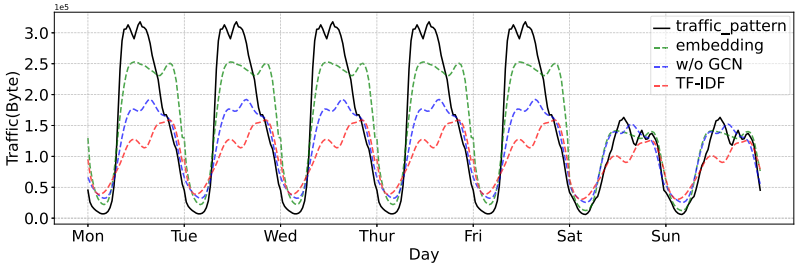


Fig. 8. Prediction vs. the ground truth for the pattern of sample base station.

conducted an experiment. First, we removed the GCN from the Traffic Regular Pattern Estimating model to exclude the capture of the graph's topology structure. This modification allowed the model to focus solely on capturing the intrinsic features of its embedding vector. Additionally, we created a new embedding matrix using the **Term Frequency-Inverse Document Frequency (TF-IDF)** method, based on the relationship between base stations and POI. TF measures the frequency of a term within a file, while IDF evaluates the importance of a term, which can be formed as

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}, \tag{13}$$

$$idf_i = log\frac{|D|}{\left|j+1 : t_i \in d_j\right|}, \tag{14}$$

$$tf - idf_{i,j} = tf_{i,j} * idf_i, \tag{15}$$

where $tf_{i,j}$ denotes the term frequency of term i in the file j, $n_{i,j}$ denotes the number of appearances of term i in the file j, idf_i means the inverse document frequency of term i. The $|D|$ denotes the number of the files and $\left|j+1 : t_i \in d_j\right|$ means the number of the files which contains the term i. In order to prevent zero, we plus one. And the $tf - idf_{i,j}$ denotes the term frequency-inverse document frequency of term i in the file j.

We present the visualizations of the output from the Traffic Regular Pattern Estimating model and the aforementioned methods. Figure 8 demonstrates that our embedding matrix E, which incorporates a wider range of entities, outperforms the TF-IDF matrix in capturing the topology structure of the urban knowledge graph. Comparing the output of the Traffic Regular Pattern Estimating model with and without GCN, we observe that the model with GCN, capable of capturing the graph's topology structure and propagating information across nodes, provides more accurate traffic pattern estimations.

—*Sensitivity of Length of Historical Traffic Series and Futural Pattern Series.* To assess the performance of the Traffic Residual Dynamics Estimating model with various hyper-parameters, we conducted experiments using different traffic series lengths, ranging from 3 (equivalent to 90 minutes) to 18 (equivalent to 9 hours). The results are presented in Table 5, where the future traffic pattern series length is fixed at 12 (equivalent to 6 hours). Furthermore, Table 6 displays the performances of different future traffic pattern series lengths while keeping the historical traffic series fixed at 12.

—*Transferability of Different Urban Area.* In order to evaluate the transferability of the model for different urban functional areas, we conduct a transfer experiment, training on one dataset and testing on another and compare the prediction results of the model in different functional areas, respectively. The results are presented in Table 7. According to the results, our model achieves good transferability. Besides, the traffic prediction results in the office area and resident area are

Table 5. Prediction Results w.r.t Different Length of Historical Traffic Series

	Shanghai Dataset			Nanjing Dataset		
Historical Length	**MAE**	**RMSE**	R^2	**MAE**	**RMSE**	R^2
3	0.1531	0.2273	0.8482	0.1978	0.2901	0.8587
6	0.1517	0.2253	0.8534	0.1959	0.2886	0.8601
9	0.1516	0.2244	0.8539	0.1946	0.2870	0.8618
12	0.1507	0.2234	0.8548	0.1934	0.2857	0.8624
15	0.1492	0.2229	0.8556	0.1920	0.2843	0.8635
18	**0.1490**	**0.2230**	**0.8562**	**0.1903**	**0.2835**	**0.8641**

Bold values indicate the best performance across all methods. Underlined values indicate the second-best performance across all methods.

Table 6. Prediction Results w.r.t Different Length of Future Traffic Pattern Series

	Shanghai Dataset			Nanjing Dataset		
Historical Length	**MAE**	**RMSE**	R^2	**MAE**	**RMSE**	R^2
3	0.1525	0.2289	0.8475	0.1978	0.2893	0.8578
6	0.1505	0.2241	0.8539	0.1962	0.2874	0.8594
9	0.1497	0.2232	0.8550	0.1949	0.2866	0.8606
12	0.1507	0.2234	0.8548	0.1934	0.2857	0.8624
15	0.1503	0.2230	0.8554	0.1927	0.2849	0.8632
18	**0.1498**	**0.2224**	**0.8563**	**0.1918**	**0.2838**	**0.8637**

Bold values indicate the best performance across all methods. Underlined values indicate the second-best performance across all methods.

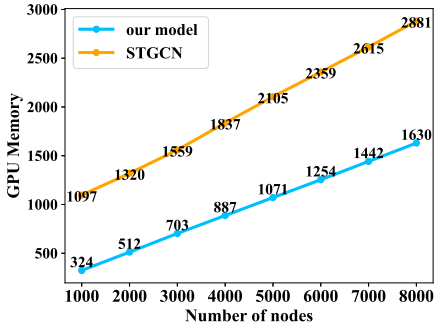
Table 7. Transfer Experiment Results on Different Urban Areas

	Shanghai Dataset			Nanjing Dataset		
Urban Areas	**MAE**	**RMSE**	R^2	**MAE**	**RMSE**	R^2
Office Area	0.1508	0.2228	0.8540	0.1945	0.2835	0.8626
Transport Area	0.1537	0.2268	0.8525	0.1983	0.2885	0.8575
Entertainment Area	0.1525	0.2261	0.8512	0.1963	0.2879	0.8603
Resident Area	**0.1496**	**0.2229**	**0.8549**	**0.1927**	**0.2829**	**0.8629**
Overall	**0.1513**	**0.2239**	**0.8541**	**0.1943**	**0.2867**	**0.8608**

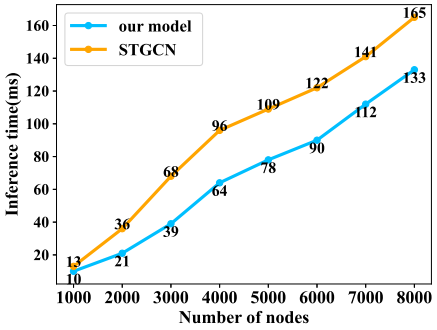
Bold values indicate the best performance across all methods. Underlined values indicate the second-best performance across all methods.

more accurate, while the prediction results in the transportation area and entertainment area are worse. This could be caused by transportation and entertainment areas experiencing sudden spikes or drops in traffic due to events, holidays, or shifts in consumer behavior, making predictions less accurate and more challenging.

—*Efficient of Model Inference.* We conducted a comparison of GPU memory usage and inference time between our method and STGCN, which is the spatiotemporal model with the smallest GPU memory consumption in the baseline. As illustrated in Figure 9, the results show that the GPU memory requirements and inference time of STGCN increase rapidly as the number of



(a) As the number of nodes in the graph increases, the changes in GPU memory usage of our method and baseline.



(b) As the number of nodes in the graph increases, the changes in inference time of our method and baseline.

Fig. 9. Comparison of computational cost.

Table 8. Ablation Study

	Shanghai Dataset			Nanjing Dataset		
	MAE	**RMSE**	R^2	**MAE**	**RMSE**	R^2
KGDA	**0.1507**	**0.2234**	**0.8548**	**0.1934**	**0.2857**	**0.8624**
LSTM	0.1582	0.2373	0.8361	0.2022	0.2986	0.8468
w/o Regular Pattern	0.1575	0.2360	0.8364	0.2003	0.2971	0.8453
w/o Residual Dynamics	0.2925	0.3818	0.4638	0.3515	0.4462	0.5142
Patterns Prediction	0.1511	0.2291	0.8519	0.1953	0.2891	0.8531
Residual Prediction	0.1524	0.2304	0.8514	0.1967	0.2904	0.8516

Bold values indicate the best performance across all methods.

nodes increases, while the GPU memory usage and inference time of our method increase to a smaller extent. Experiments prove that our model requires less computing resources and shorter experiments and can better meet practical needs.

4.4 Ablation Study

In order to better understand each part of our model KGDA, we perform the following ablation experiments. We first changed the Bi-LSTM model in the Traffic Residual Dynamics Estimating model to the original LSTM. We also remove the Traffic Regular Pattern Estimating model. Third, we remove the Traffic Residual Dynamics Estimating model. Moreover, we test the performance output of the Traffic Regular Pattern Estimating model and Traffic Residual Dynamics Estimating model with its true traffic pattern and traffic residual series. The prediction results of different variants are presented in Table 8. We observe that the Traffic Regular Pattern Estimating model utilizes a static embedding matrix as input, which limits its ability to estimate real cellular traffic accurately. However, when coupled with the Traffic Regular Pattern Estimating model, KGDA demonstrates improved performance.

—*Effectiveness of Urban Knowledge Graph.* We systematically conducted experiments to demonstrate the effectiveness of individual entities within the urban knowledge graph by iteratively removing one entity along with its associated relationships. These results are summarized in Table 9. The findings highlight the superior performance of the complete urban knowledge graph compared to its partial counterparts, emphasizing the importance of utilizing the entire graph for

Table 9. Ablation Study on the Urban Knowledge Graph

	Shanghai Dataset			Nanjing Dataset		
	MAE	**RMSE**	R^2	**MAE**	**RMSE**	R^2
Our Model	**0.1507**	**0.2234**	**0.8548**	**0.1934**	**0.2857**	**0.8624**
w/o Category	0.1523	0.2241	0.8558	0.1956	0.2858	0.8610
w/o POI	0.1534	0.2272	0.8488	0.1981	0.2910	0.8584
w/o BA	0.1554	0.2282	0.8525	0.2068	0.3042	0.8467
w/o Region	0.1555	0.2284	0.8535	0.2073	0.3022	0.8462

Bold values indicate the best performance across all methods.

optimal results. Additionally, our analysis revealed that regions and business areas consistently outperformed POIs and categories. This disparity can be attributed to the highly detailed micro-geographic data associated with POIs, which poses challenges in capturing spatial structures and environmental nuances accurately. Conversely, regions and business areas offer a broader, macro-level perspective that is more conducive to modeling and analysis within the urban knowledge graph framework.

5 Related Work

5.1 Cellular Traffic Prediction

Initially, cellular traffic forecasting was viewed as a generic time series prediction issue, and significant efforts and models were employed to promote the performance of traffic forecasting in communication networks. Hong et al. [19] utilized SVR to model short-term cellular traffic series. Although SVR is commonly used in time series forecasting, it struggles to capture rapid variations in traffic flow as it relies on the average volume of historical traffic series. Furthermore, it fails to model non-linear relationships, requiring substantial time and effort to identify appropriate parameters for achieving satisfactory performance. Li et al. [24] and Xu et al. [45] focus on predicting short-term cellular traffic series for a single base station in cellular networks. However, as the prediction window is expanded, the performance of these models deteriorates rapidly.

Recently, there has been increasing interest in improving the precision of cellular traffic forecasting through spatial structure analysis. Several spatial models have been proposed for this purpose. Cao et al. [3] examine the distribution of hotspots over space and time, categorizing various types of hot regions. Feng et al. [11] propose a deep-learning-based end-to-end model that captures spatial-dependent and long-term cellular traffic patterns. This model highlights the influence of environmental factors, such as the distribution of POIs, on the basic traffic demands within a given area. Wang et al. [39] integrate spatial-temporal modeling and prediction in cellular networks. They introduce an innovative deep model that combines autoencoder-based spatial modeling with LSTM for temporal modeling. These models employ a grid-based method for cellular traffic prediction. However, they do not differentiate between base stations within the same grid. Wang et al. [41] introduce a novel decomposition method for in-cell and inter-cell data traffic. They also propose a graph-based deep learning approach for precise cellular traffic forecasting. This method considers the relationships between base stations and leverages graph-based techniques to accurately capture spatial dependencies.

5.2 Spatial-Temporal Traffic Forecasting

Fundamentally, urban traffic originates from the actions of citizens, rendering them inherently similar. However, there are still some differences.

For the road traffic, its spatiotemporal dependence is significant. Vehicles can only transfer from one node to its neighbor nodes. During the transfer process, if the source node decreases by one, the target node will increase by one. Road traffic changes follow simpler patterns than mobile traffic. Fang et al. [10] propose a method to model spatial dependencies among base stations using a dependency graph constructed based on spatial distances between cells. This preserves the spatial granularity of the data, with edges in the graph representing spatial relationships between nodes. Yu et al. [48] introduce STGCN, a deep learning model combining GCN and CNN. The GCN component captures the topology structure of the graph through an adjacency matrix calculated based on node distances, enabling spatial dependency capture. Meanwhile, the gated CNN component explores dynamic features in cellular traffic, capturing temporal dependencies. Guo et al. [15] propose ASTGCN, comprising three significant parts to capture specific temporal features of cellular traffic. These components' outputs are weighted and combined to produce the final prediction. Zhao et al. [54] present T-GCN, merging GCN and GRU. GCN captures graph topology similarity using an adjacency matrix based on node distances for spatial dependency capture, while GRU models dynamic changes in cellular traffic at each node for temporal dependency capture. Wu et al. [43] propose GraphWaveNet, incorporating a trainable adaptive dependency matrix via node embedding and stacked dilated convolution to broaden the receptive field, capturing a broader range of spatial and temporal information for spatial-temporal dependency modeling. Diao et al. [5] introduce DGCNN, utilizing a dynamic Laplacian matrix estimator to track dynamic spatial dependencies. This adeptly captures enduring global temporal-spatial traffic relationships and transient local traffic fluctuations.

Predicting the network traffic entails forecasting forthcoming traffic patterns by leveraging network structure and historical traffic data obtained from routers. Davide et al. [1] utilize DCRNN to predict traffic levels and congestion incidents. DCRNN employs a graph-based machine learning approach to learn individual node representations, considering their intrinsic properties and the network's structure. Laisen et al. [30] introduce a method based on reinforcement learning to address the traffic forecasting challenge, treating it as a Markov decision process. Their approach integrates a residual-based dictionary learning algorithm to extract temporal feature sets relevant to the prediction task. He et al. [17] introduce a meta-learning framework comprising a suite of predictors, each tailored to forecast specific traffic types. Furthermore, they devise a master policy trained to dynamically select the most suitable predictor based on individual performance metrics.

Notably, other traffic prediction models are primarily based on modeling spatiotemporal correlation and failure to incorporate urban environmental factors into their models, resulting in subpar performance.

5.3 Urban Knowledge Graph

A knowledge graph, also known as a semantic network, represents a network of real-world entities and illustrates their relationships. Similarly, an urban knowledge graph represents urban elements such as base stations, POIs, and regions as entities, modelling spatial and semantic dependencies as relationships between them. Liu et al. [29] present Urban KG, an urban knowledge graph system, which is built upon the data layer, the system further develops the multiple layers of construction, storage, algorithm, operation, and applications, which achieve knowledge distillation and support various functions to the users. Ning et al. [32] introduce UUKG, a unified urban knowledge graph dataset designed for knowledge-enhanced urban spatial-temporal predictions, which uncovers diverse high-order structural patterns that can be leveraged to improve downstream Unsupervised Spatio-Temporal Prediction tasks.

Recently, there has been increasing interest in applying urban KG to traffic prediction and traffic generation in cellular networks. Gong et al. [12] propose a multi-relational knowledge GCNmodel for

mobile traffic prediction, which captures spatial information from the augmented spatial knowledge graph using Tucker decomposition and relational graph convolutional network. Zhang et al. [52] propose ADAPTIVE, a deep transfer learning framework for city-scale cellular traffic generation through the urban knowledge graph. Hui et al. [21] propose a knowledge-enhanced GAN with multi-periodic patterns to generate large-scale cellular traffic based on the urban environment. Therefore, we are motivated to develop urban knowledge graphs to represent the urban environment for cellular traffic prediction.

6 Conclusion

In this article, we examine the inherent traffic patterns of base stations and identify four key traffic patterns influenced by location, using hierarchical clustering. We introduce a KGDA to predict future cellular traffic. This approach breaks down the impact of static environmental factors and dynamic autocorrelations, thereby capturing the overall direction of traffic changes and understanding the dependence of traffic on past values. Our urban knowledge graph encapsulates the static environmental context of base stations. Extensive experiments, conducted on a real-world dataset, corroborate the efficiency and accuracy of our approach.

In future work, we plan to extend our proposed approach to evaluate its effectiveness in diverse settings by applying it to other cities and regions. This will help us assess the generalizability of our model and identify any limitations or challenges in applying it to various mobile network environments. Moreover, it is interesting to investigate the transferability of our model to different types of mobile networks, such as sensor networks and vehicle networks, which have unique characteristics and requirements. We can explore how our model can be adapted to these networks and applied in optimization and management. Additionally, we plan to sustain our collaboration with industry partners to integrate our model into existing mobile network management systems. This collaborative effort is aimed at enhancing network performance and augmenting user experience.

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 *Proceedings of the IEEE INFOCOM WKSHPS*. DOI: https://doi.org/10.1109/INFCOMW.2019.8845132

[2] Ivana Balažević, Carl Allen, and Timothy M. Hospedales. 2019. Tucker: Tensor Factorization for Knowledge Graph Completion. arXiv:1901.09590. Retrieved form https://doi.org/10.18653/v1/D19-1522

[3] Aoxiang Cao, Yuanyuan Qiao, Kewu Sun, Hao Zhang, and Jie Yang. 2018. Network Traffic Analysis and Prediction of Hotspot in Cellular Network. In *Proceedings of the 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC '18)*. IEEE, 452–456.

[4] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In *Proceedings of the 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103.

[5] 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 (July 2019), 890–897. DOI: https://doi.org/10.1609/aaai.v33i01.3301890

[6] Jingtao Ding, Yong Li, Pengyu Zhang, and Depeng Jin. 2017. Time Dependent Pricing for Large-Scale Mobile Networks of Urban Environment: Feasibility and Adaptability. *IEEE Transactions on Services Computing* 13, 3 (2017), 559–571.

[7] Jingtao Ding, Xihui Liu, Yong Li, Di Wu, Depeng Jin, and Sheng Chen. 2016. Measurement-Driven Capability Modeling for Mobile Network in Large-Scale Urban Environment. In *Proceedings of the 2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS '16)*. IEEE, 92–100.

[8] Jingtao Ding, Rui Xu, Yong Li, Pan Hui, and Depeng Jin. 2017. Measurement-Driven Modeling for Connection Density and Traffic Distribution in Large-Scale Urban Mobile Networks. *IEEE Transactions on Mobile Computing* 17, 5 (2017), 1105–1118.

[9] Harris Drucker, Christopher Burges, Linda Kaufman, Alexander Smola, and V. Vapnik. 1997. Support Vector Regression Machines. *Advances in Neural Information Processing Systems* 28 (Jan. 1997), 779–784.

[10] 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. DOI: https://doi.org/10.1109/JIOT.2018.2832071

[11] 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. DOI: https://doi.org/10.1109/MNET.2018.1800127

[12] Jiahui Gong, Yu Liu, Tong Li, Haoye Chai, Xing Wang, Junlan Feng, Chao Deng, Depeng Jin, and Yong Li. 2023. Empowering Spatial Knowledge Graph for Mobile Traffic Prediction. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23)*, 1–11.

[13] Jiahui Gong, Qiaohong Yu, Tong Li, Haoqiang Liu, Jun Zhang, Hangyu Fan, Depeng Jin, and Yong Li. 2023. Scalable Digital Twin System for Mobile Networks with Generative AI. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*, 610–611.

[14] Alex Graves, Santiago Fernández, and Jürgen Schmidhuber. 2005. Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. In *Proceedings of the Artificial Neural Networks: Formal Models and Their Applications (ICANN '05)*. Wodzisaw Duch, Janusz Kacprzyk, Erkki Oja, and Sawomir Zadrożny (Eds.), Springer, Berlin, 799–804.

[15] S. Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 1–6.

[16] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems* 30 (2017), 1–11.

[17] 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* (2020). DOI: https://doi.org/10.1109/JSAC.2020.3000408

[18] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9 (Dec. 1997), 1735–1780. DOI: https://doi.org/10.1162/neco.1997.9.8.1735

[19] Wei-Chiang Hong. 2012. Application of Seasonal SVR with Chaotic Immune Algorithm in Traffic Flow Forecasting. *Neural Computing and Applications* (April 2012). DOI: https://doi.org/10.1007/s00521-010-0456-7

[20] Wenzhen Huang, Tong Li, Yuting Cao, Zhe Lyu, Yanping Liang, Li Yu, Depeng Jin, Junge Zhang, and Yong Li. 2023. Safe-NORA: Safe Reinforcement Learning-Based Mobile Network Resource Allocation for Diverse User Demands. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 885–894.

[21] Shuodi Hui, Huandong Wang, Tong Li, Xinghao Yang, Xing Wang, Junlan Feng, Lin Zhu, Chao Deng, Pan Hui, Depeng Jin, and Yong Li. 2023. Large-Scale Urban Cellular Traffic Generation via Knowledge-Enhanced GANs with Multi-Periodic Patterns. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*. ACM, New York, NY. DOI: https://doi.org/10.1145/3580305.3599853

[22] Jiahao Ji, Jingyuan Wang, Chao Huang, Junjie Wu, Boren Xu, Zhenhe Wu, Junbo Zhang, and Yu Zheng. 2023. Spatio-Temporal Self-Supervised Learning for Traffic Flow Prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37, 4356–4364.

[23] Thomas N. Kipf and Max Welling. 2016. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the International Conference on Learning Representations*, 1–14.

[24] 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. DOI: https://doi.org/10.1109/MCOM.2014.6829969

[25] Tong Li and Yong Li. 2023. Artificial intelligence for reducing the carbon emissions of 5G networks in China. *Nature Sustainability* 6, 12 (2023), 1522–1523.

[26] Tong Li, Yong Li, Mohammad Ashraful Hoque, Tong Xia, Sasu Tarkoma, and Pan Hui. 2020. To What Extent We Repeat Ourselves? Discovering Daily Activity Patterns across Mobile App Usage. *IEEE Transactions on Mobile Computing* 21, 4 (2020), 1492–1507.

[27] Tong Li, Tong Xia, Huandong Wang, Zhen Tu, Sasu Tarkoma, Zhu Han, and Pan Hui. 2022. Smartphone App Usage Analysis: Datasets, Methods, and Applications. *IEEE Communications Surveys & Tutorials* 24, 2 (2022), 937–966.

[28] Tong Li, Li Yu, Yibo Ma, Tong Duan, Wenzhen Huang, Yan Zhou, Depeng Jin, Yong Li, and Tao Jiang. 2023. Carbon Emissions of 5G Mobile Networks in China. *Nature Sustainability* 6, 12 (2023), 1620–1631.

[29] Yu Liu, Jingtao Ding, Yanjie Fu, and Yong Li. 2023. UrbanKG: An Urban Knowledge Graph System. *ACM Transactions on Intelligent Systems and Technology* 14, 4, Article 60 (May 2023), 25 pages, 2157–6904. DOI: https://doi.org/10.1145/3588577

[30] 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* (2021). DOI: https://doi.org/10.1109/TII.2020.3004232

[31] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2022. A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. arXiv:2211.14730. Retrieved from https://openreview.net/forum?id=Jbdc0vTOcol

[32] Yansong Ning, Hao Liu, Hao Wang, Zhenyu Zeng, and Hui Xiong. 2024. UUKG: Unified Urban Knowledge Graph Dataset for Urban Spatiotemporal Prediction. *Advances in Neural Information Processing Systems* 36 (2024), 1–15.

[33] 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. DOI: https://doi.org/10.1109/TNN.2008.2005605

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

[35] 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. DOI: https://doi.org/10.1109/TMC.2021.3079117

[36] Ledyard R. Tucker. 1966. Some Mathematical Notes on Three-Mode Factor Analysis. *Psychometrika* 31, 3 (1966), 279–311.

[37] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio', and Yoshua Bengio. 2017. Graph Attention Networks. arXiv:1710.10903. Retrieved form https://openreview.net/forum?id=rJXMpikCZ

[38] Huandong Wang, Fengli Xu, Yong Li, Pengyu Zhang, and Depeng Jin. 2015. Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment. In *Proceedings of the 2015 Internet Measurement Conference*, 225–238.

[39] 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 *Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM '17)*, 1–9. DOI: https://doi.org/10.1109/INFOCOM.2017.8057090

[40] Xu Wang, Zimu Zhou, Fu Xiao, Kai Xing, Zheng Yang, Yunhao Liu, and Chunyi Peng. 2018. Spatio-Temporal Analysis and Prediction of Cellular Traffic in Metropolis. *IEEE Transactions on Mobile Computing* 18, 9 (2018), 2190–2202.

[41] Xu Wang, Zimu Zhou, Zheng Yang, Yunhao Liu, and Chunyi Peng. 2017. Spatio-Temporal Analysis and Prediction of Cellular Traffic in Metropolis. In *Proceedings of the 2017 IEEE 25th International Conference on Network Protocols (ICNP '17)*, 1–10. DOI: https://doi.org/10.1109/ICNP.2017.8117559

[42] 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 (Nov. 2003), 664–672. DOI: https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)

[43] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1907–1913.

[44] Fengli Xu, Yong Li, Min Chen, and Sheng Chen. 2016. Mobile Cellular Big Data: Linking Cyberspace and the Physical World with Social Ecology. *IEEE Network* 30, 3 (2016), 6–12. DOI: https://doi.org/10.1109/MNET.2016.7474338

[45] Yue Xu, Wenjun Xu, Feng Yin, Jiaru Lin, and Shuguang Cui. 2017. High-Accuracy Wireless Traffic Prediction: A GP-Based Machine Learning Approach. In *2017 IEEE Global Communications Conference (GLOBECOM '17)*, 1–6. DOI: https://doi.org/10.1109/GLOCOM.2017.8254808

[46] Fan Yang, Yun Jiang, Tian Pan, and Xinhua E. 2018. Traffic Anomaly Detection and Prediction Based on SDN-Enabled ICN. In *Proceedings of the 2018 IEEE International Conference on Communications Workshops (ICC Workshops '18)*, 1–5. DOI: https://doi.org/10.1109/ICCW.2018.8403693

[47] Qiaohong Yu, Huandong Wang, Tong Li, Depeng Jin, Xing Wang, Lin Zhu, Junlan Feng, and Chao Deng. 2022. Network Traffic Overload Prediction with Temporal Graph Attention Convolutional Networks. In *Proceedings of the 2022 IEEE International Conference on Communications Workshops (ICC Workshops '22)*. IEEE, 885–890.

[48] Ting Yu, Haoteng Yin, and Zhanxing Zhu. 2017. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3634–3640.

[49] Yuan Yuan, Chenyang Shao, Jingtao Ding, Depeng Jin, and Yong Li. 2024. Spatio-Temporal Few-Shot Learning via Diffusive Neural Network Generation. In *Proceedings of the 12th International Conference on Learning Representations*, 1–28.

[50] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. 2014. Recurrent Neural Network Regularization. arXiv:1409.2329. Retrieved from https://arxiv.org/pdf/1409.2329

[51] Mingyang Zhang, Tong Li, and Pan Hui. 2023. CityNeuro: Towards Location and Time Prediction for Urban Abnormal Events. *IEEE Transactions on Knowledge and Data Engineering* 35, 12 (Dec. 2023), 12125–12138.

[52] Shiyuan Zhang, Tong Li, Shuodi Hui, Guangyu Li, Yanping Liang, Li Yu, Depeng Jin, and Yong Li. 2023. Deep Transfer Learning for City-Scale Cellular Traffic Generation through Urban Knowledge Graph. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4842–4851.

[53] Ying Zhang and Ake Årvidsson. 2012. Understanding the Characteristics of Cellular Data Traffic. In *Proceedings of the 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design*, 13–18.

[54] 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. DOI: https://doi.org/10.1109/TITS.2019.2935152

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

Received 15 September 2023; revised 10 June 2024; accepted 8 August 2024