Spatial-Temporal Cellular Traffic Prediction for 5G and Beyond: A Graph Neural Networks-based Approach

Zi Wang, Jia Hu*, Geyong Min*, Zhiwei Zhao, Zheng Chang, and Zhe Wang

Abstract—**During the past decade, Industry 4.0 has greatly promoted the improvement of industrial productivity by introducing advanced communication and network technologies in the manufacturing process. With the continuous emergence of new communication technologies and networking facilities, especially the rapid evolution of cellular networks for 5G and beyond, the requirements for smarter, more reliable, and more efficient cellular network services have been raised from the Industry 5.0 blueprint. To meet these increasingly challenging requirements, proactive and effective allocation of cellular network resources becomes essential. As an integral part of the cellular network resource management system, cellular traffic prediction faces severe challenges with stringent requirements for accuracy and reliability. One of the most critical problems is how to improve the prediction performance by jointly exploring the spatial and temporal information within the cellular traffic data. A promising solution to this problem is provided by Graph Neural Networks (GNNs), which can jointly leverage the cellular traffic in the temporal domain and the physical or logical topology of cellular networks in the spatial domain to make accurate predictions. In this paper, we present the spatial-temporal analysis of a real-world cellular network traffic dataset and review the state-of-the-art researches in this field. Based on this, we further propose a time-series similarity-based graph attention network, TSGAN, for the spatial-temporal cellular traffic prediction. The simulation results show that our proposed TSGAN outperforms three classic prediction models based on GNNs or GRU on a real-world cellular network dataset in short-term, mid-term, and long-term prediction scenarios.**

Index Terms—**Cellular Network Traffic Prediction, Graph Neural Networks, 5G/6G, Industry 5.0.**

I. Introduction

RECENTLY, based on the successful advancement of Industry 4.0 during the past decade, the European Commission proposes Industry 5.0 to further improve the cooperation between humans and smart devices. Industry 5.0 has its roots in the concept of Industry 4.0 which focuses on exploiting new technologies to improve the efficiency and flexibility of production [1]. With the emergence of the fifth-generation networks (5G), the provisioning of such a remarkable cellular network system is essential for the evolution to Industry 5.0. According to the latest investigation, 5G connections will grow over 100-fold from about 13 million in 2019 to 1.4 billion by 2023 [2]. This spectacular growing trend will continue in the future sixth-generation networks (6G) and Industry 5.0 era with the vision of the Internet of Everything. Based on these massive and reliable communication facilities and the ultra-low-latency services they support, the advanced cellular network can improve the data and system interoperability which is one of the major enabling technologies in Industry 5.0 [3]. The sources of cellular traffic consist of various devices from daily life and industry, especially mobile devices in these scenarios. Many novel and advanced researches that exploit such large amount of cellular traffic from massive devices have emerged in recent years, including mobile traffic classification [4], mobile traffic prediction and characterisation [5], etc. In order to proactively prepare resources to provide consumers or industrial devices with high-quality services, an accurate cellular traffic prediction becomes more and more important to help the network management system cope with unprecedented challenges. As illustrated in the conceptual framework in Fig.1, many devices in various cellular network applications are connected to the base stations at cellular edge to acquire the fast and reliable network services. Their cellular traffic usage data are collected by base stations at the edge. By analysing the real-world traffic usage requirements, the cellular management centre can utilise the pre-processed cellular traffic data to predict the future traffic trends of consumers and industrial devices. Based on accurate and reliable prediction results, the cellular management centre is able to proactively allocate the cellular resources, which can be deployed at the edge in advance to meet the traffic demands of various services under different conditions and time spans.

Artificial intelligence has been introduced as a promising solution to improve the cellular traffic prediction performance. In recent years, deep learning methods such as long short-term memory (LSTM) [6] and convolutional neural network (CNN) [7] have been demonstrated to be effective in time-series prediction. For cellular traffic, all traffic usage data is recorded in the form of time-series in the temporal domain. Compared with the conventional statistic-based models, extensive researches using deep learning models have significantly improved the cellular traffic prediction performance. However, due to the limitations of neural network functionalities, most of the existing work can only process cellular traffic data in the temporal domain. Recently, some researches (e.g. [8]–[11])

Z. Wang is with the School of Computer Science and Technology, University of Electronic Science and Technology of China, and the College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK. (Email: zw300@exeter.ac.uk) J. Hu, G. Min and Z. Wang are with the College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK. (Email: {j.hu, g.min, zw329}@exeter.ac.uk) Z. Zhao and Z. Chang are with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China (e-mail: {zzw, zheng.chang}@uestc.edu.cn)

*Corresponding Author

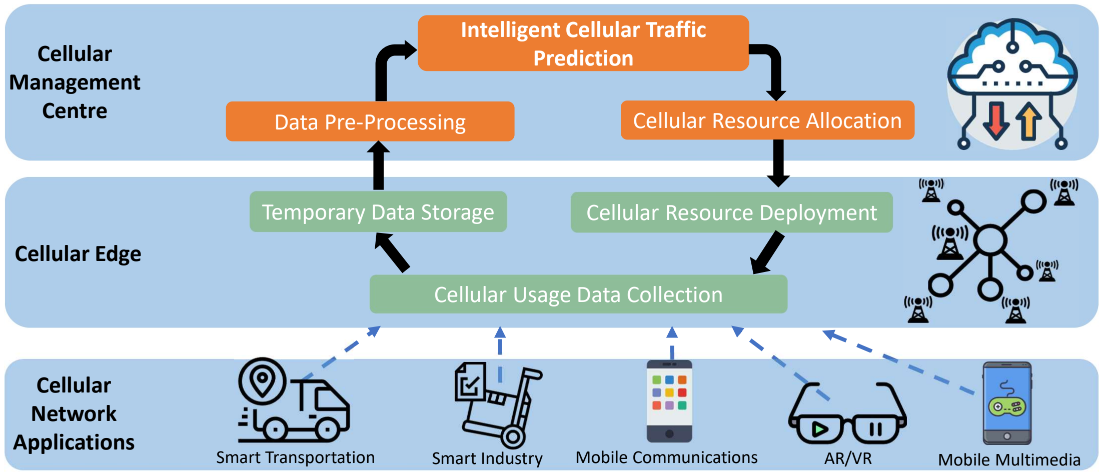


Fig. 1. The illustration of the intelligent cellular network framework for 5G and beyond.

have made attempts to integrate the variants of convolution and recurrent operation to deal with the cellular traffic data in both temporal and spatial domain. Nevertheless, they proceed the spatial-temporal analysis by using coarse-grained grid-based Euclidean structure data of cellular base stations to extract the spatial information. Considering the real-world topology or spatial information within traffic of base stations at the edge is the non-Euclidean structure data [12], those researches cannot comprehensively capture spatial-temporal characteristics of cellular traffic data across all base stations in the network.

To address these issues, utilising the non-Euclidean graph structure has attracted more and more research attention in spatial-temporal cellular traffic prediction. This is because the real-world cellular network includes a set of base stations and their physical or logical connections can be denoted as nodes and edges in a graph (e.g., cellular edge in Fig. 1) which can present more comprehensive spatial information than grid-based structure. As an emerging deep learning method, graph neural network (GNN) [13] was introduced for the graph processing in the spatial domain. It is widely used in many research areas involving graph data such as physical and biological analysis. Therefore, researchers start to put effort in using GNNs to improve the cellular traffic prediction accuracy by exploiting the spatial-temporal characteristics of cellular networks.

In this regards, we propose a novel time-series similarity-based graph attention network, TSGAN, for the GNNs-empowered spatial-temporal cellular traffic prediction in this paper. TSGAN combines dynamic time warping (DTW) and graph attention networks, which can extract and utilise the spatial-temporal dependencies and characteristics of cellular traffic data and consider the differentiated influence between adjacent cells and core cells.

The main contributions of this paper are summarised as follows:

- We propose a novel time-series similarity-based graph attention network (TSGAN) for accurate and reliable spatial-temporal cellular traffic prediction. This model can extract and utilise the spatial-temporal dependencies and characteristics of cellular traffic and consider the differentiated influence among cells to perform the accurate and reliable prediction.
- We analyse the spatial-temporal dependencies and characteristics of a real-world cellular traffic dataset and review the state-of-the-art researches on GNNs-empowered spatial-temporal cellular traffic prediction.
- We evaluate our proposed TSGAN by comparing it with three classic prediction models based on GNNs or GRU on a real-world cellular network dataset in short-term, mid-term, and long-term prediction scenarios. We also discuss several exciting research opportunities for using GNN-empowered cellular traffic prediction in future 6G and Industry 5.0 scenarios.

The remainder of this paper is organised as follows. Section II analyses a real-world citywide cellular traffic dataset and introduces the fundamental evolution of graph neural networks for using GNNs-empowered models in cellular traffic prediction. Section III reviews the state-of-the-art researches in GNNs-empowered spatial-temporal cellular traffic prediction. Section IV describes our proposed time-series similarity-based graph attention network (TSGAN) for the spatial-temporal cellular traffic prediction. Section V presents the evaluation and result analysis among TSGAN and three referencing models on the real-world citywide cellular traffic dataset in short-term, mid-term, and long-term prediction scenarios. Finally, Section VI concludes this paper and provides and discusses several exciting research opportunities in future work. The acronyms shown throughout the paper are summarised in Table I.

II. Preliminaries

A. Analysis of Spatial-Temporal Characteristics of Real-world Cellular Network Traffic

A large European communication services operator, Telecom Italia, published a real-world cellular traffic dataset of the city of Milan [14] which can be used to analyse and study the spatial-temporal characteristics in the cellular traffic data. Specifically, Table II shows examples including 8 features (SquareID, Datetime, SmsIn, SmsOut, CallIn, CallOut, and

TABLE I
ACRONYMS

5G	The fifth-generation networks	**UAVs**	Unmanned aerial vehicles
6G	The sixth-generation networks	**GRUs**	Gated recurrent units
GNNs	Graph neural networks	**MSE**	Mean square error
LSTM	Long short-term memory	**MAE**	Mean absolute error
CNN	Convolutional neural network	**MAPE**	Mean absolute percentage error
CDRs	Call Detail Records	**RMSE**	Root mean square error
BS	Base station	**GCGRN**	Graphic convolution gated recurrent unit network
ChebNet	Chebyshev spectral convolutional neural networks	**DTW**	Dynamic time warping
GCN	Graph convolutional networks	**TSGAN**	Time-series Similarity-based Graph Attention Network
GAT	Graph attention networks	**GDPR**	General Data Protection Regulation
GCLSTM	Graph convolutional LSTM	**FL**	Federated learning
ITW	Idle time windows	**STACN**	Spatial-temporal attention convolution network
TGCN	Temporal graph convolutional networks	**STGCN-HO**	Spatio-temporal graph convolutional networks incorporating handover information
STHGCN	Spatio-temporal hybrid graph convolutional network		

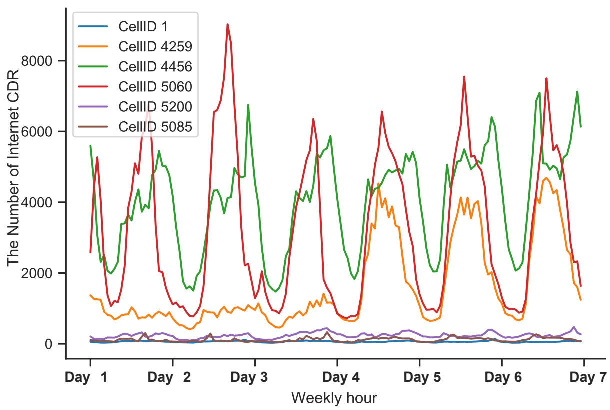


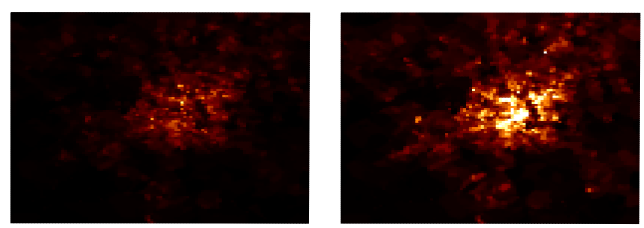
Fig. 2. Temporal distributions of internet activities at different cells in Milan. An Internet CDR is generated each time a user starts or ends an Internet connection.

Internet) of the CDR which records the call details produced by cellular network users. In this paper, 3 features (SquareID, Datetime, and Internet) are directly related to the cellular network traffic prediction. This public dataset contains millions of Call Detail Records (CDRs) of internet activities during 62 days at a 10-minute sampling interval, which provides sufficient time-series samples for temporal analysis. As for the spatial analysis, the map of the city of Milan is divided into 100 × 100 areas and each area is 0.05 km^2 The CDRs for each area in the city are spatially aggregated by calculating the coverage area of the processing base station.

TABLE II
CDR FEATURES OF DATASET

SquareID	Datetime	CountryCode	SmsIn	SmsOut	CallIn	CallOut	Internet
1	1383260400000	39	0.141864	0.156787	0.160938	0.052275	11.028366
1	1383261000000	39	0.278452	0.119926	0.188777	0.133637	11.100963
…	…	…	…	…	…	…	…
9999	1383864600000	39	0.184105	0.144481	0.319243	0.085995	21.650467
…	…	…	…	…	…	…	…

For the ease of observation, we pick up a fragment of the whole records to present the temporal and spatial distributions of the cellular traffic of Milan city. Fig. 2 presents the temporal distribution of internet connection services at certain areas in Milan for 7 days. In the record of cellular traffic, during the same connection, a CDR is generated if the connection lasts for more than 15 mins or the user transferred more than 5 MB [14]. It is clear that internet activities present strong daily-periodic and weekly-periodic characteristics from the temporal domain. Besides, different areas show diverse usage patterns according to their different functionalities in the city. For example, compared with the place with ID 4456 which is active at night, the place with ID 5060 is more active during the day.



(a) 03:00 - 04:00 network activities (b) 15:00 - 16:00 network activities

Fig. 3. The one-hour heat map for spatial distributions of internet activities in Milan.(Black: 0 CDR, White: 5000 CDRs)

As for the observation of spatial characteristics, we illustrate two one-hour heat maps of internet activities of the whole city of Milan in Fig. 3. We can clearly see from this figure that the cellular traffic is distributed unevenly and dynamically among the whole city. It indicates that the cellular traffic demands are obviously strong at daytime than the demands at night. Although both of the heat maps show that the cellular traffic demands are mainly concentrated in the city centre, there are still signs of widespread cellular traffic usage in the suburban areas in the daytime.

In this regard, cellular traffic prediction can rely on the spatial-temporal characteristics of cellular traffic data to improve prediction accuracy and reliability. The efficiency of proactive cellular resources allocation can be further enhanced to meet the dynamic demands in different periods, which will bring outstanding benefits to greatly improve cellular network management for 5G and beyond.

B. Graph Neural Networks

With the accumulation of massive cellular traffic data and the advances in deep learning techniques, deep learning-based cellular traffic prediction methods have established themselves as strong competitors to conventional statistic-based models

[8]. Since various important machine learning problems involve tasks on graph structure data, researchers have developed a family of deep learning models called graph neural networks (GNNs) that naturally handle non-Euclidean graph structure data.

Chebyshev Spectral Convolutional Neural Networks (ChebNet) [15] was proposed to consider that the value of the convolution kernel in the spectral domain is a function related to the eigenvalue and then approximate the filter by Chebyshev polynomials of the diagonal matrix of eigenvalues. Formally, the graph convolution $\star_G$ of ChebNet can be formulated by

$$x \star_G g_\theta = U g_\theta U^T x = \sum_{k=0}^{K} \beta_k T_k(\hat{L}) x \tag{1}$$

where x is the input signal and β_k is the parameter to be learned and $T_k(\hat{L})$ is the Chebyshev polynomial of order k evaluated at the scaled Laplacian matrix $\hat{L}$.

Following the ChebNet, the most popular graph neural network model, Graph Convolutional Networks (GCN) [16] was proposed, which improved the ChebNet by limiting the Chebyshev expansion to only the first order and simplifying the convolution kernel with a reduced number of parameters. Formally, the graph convolution $\star_G$ of GCN can be formulated by

$$\begin{aligned} x \star_G g_\theta &= \theta(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2}) x \\ \tilde{D}_{ij} &= \sum_i \tilde{A}_{ij} \\ \tilde{A} &= A + I_n \end{aligned} \tag{2}$$

where x is the input signal, g_θ is the convolution kernel, θ is the learned parameter, A is the adjacency matrix, D is the node degree matrix, and I_n is an n-order identity matrix. By focusing on the first-order approximation of spectral graph convolutions, this model significantly reduces the number of parameters. However, the reduction of the model complexity may limit its ability to handle complex tasks.

III. Related Work

In this section, we will analyse the state-of-the-art researches using the GNNs-empowered models to improve the cellular traffic prediction performance.

A. Spatial-Temporal Cellular Traffic Prediction On Undirected Graph

Different from researches that focused on all grid-divided base stations of the cellular network, Fang *et al.* [17] focused on the per-cell demand forecasting because the irregular spatial distribution of cells in the real-world setting limits the applications to future cellular network management that requires variable spatial granularity. They first used a dependency graph to model the spatial relevancy among base stations so retained the spatial granularity without any data aggregation. Their proposed model graph convolutional LSTM (GCLSTM) mainly replaced the matrix multiplication in LSTMs with the graph convolution in each gates, which can utilise the spatial and temporal information and reduce the number of trainable parameters.

Moreover, Fang *et al.* [18] also studied from the network subscribers' demand and mobility behaviours observed by network operators to predict the idle time windows (ITW) for cellular networks in the energy aspect. Moreover, their proposed temporal graph convolutional networks (TGCN), a graph-sequence representation network model, combined the temporal convolutional networks and graph convolutional networks to learn high-level spatial-temporal patterns for the ITW prediction.

Zhao *et al.* [19] modelled the traffic network as an undirected graph. Dut to the lack of geographic topology of BSs in the dataset, they explored the spatial information based on the correlation between time-series of cellular traffic by using the Anseline local Moran's I statistic measure, which is proved to be an effective solution in this work. Besides, they also considered the impact of external factors on prediction performance. Therefore, their proposed model, spatial-temporal attention convolution network (STACN), can adaptively aggregate the external factors and output the hourly, daily and weekly temporal components of cellular traffic.

B. Spatial-Temporal Cellular Traffic Prediction On Directed Graph

Wang *et al.* [20] decomposed cellular traffic into in-tower and inter-tower traffic to characterise the spatial dependency among cell towers. As they jointly considered temporal and spatial dependencies among cell towers, their proposed GNN model achieved fine-grained traffic prediction in a metropolitan range.

To further exploit the spatial information, Kalander *et al.* [21] identified three categories of spatial dependencies: spatial proximity, functional similarity, and recent trend similarity. Based on these categories, they proposed a hybrid GCN called Spatio-Temporal Hybrid Graph Convolutional Network (STHGCN) to explore the spatial information. The STHGCN used the ChebyNet to obtain the spatial dependency and then consecutively obtained the temporal dependency by connecting with gated recurrent units (GRUs). In the evaluation datasets, their model can present a more consistent prediction compared with six baselines. They also claimed that their model was highly suitable for traffic prediction on 5G networks.

In order to guarantee the service continuity for users, the cellular network is also required to consider the effect of handover on the spatial characteristics of the traffic. Most of researches lacked of solving problems under this circumstance, therefore, Zhao *et al.* [22] built a directed, weighted handover graph for base stations based on the handover frequencies presented by the transition probability matrix to improve the prediction accuracy. Their proposed spatio-temporal graph convolutional networks incorporating handover information (STGCN-HO) model used a stacked residual neural network structure incorporating graph convolutions and gated linear units to capture both spatial and temporal information of the cellular traffic.

In some cases, researchers may model the network traffic patterns as probabilistic distributions or stochastic processes.

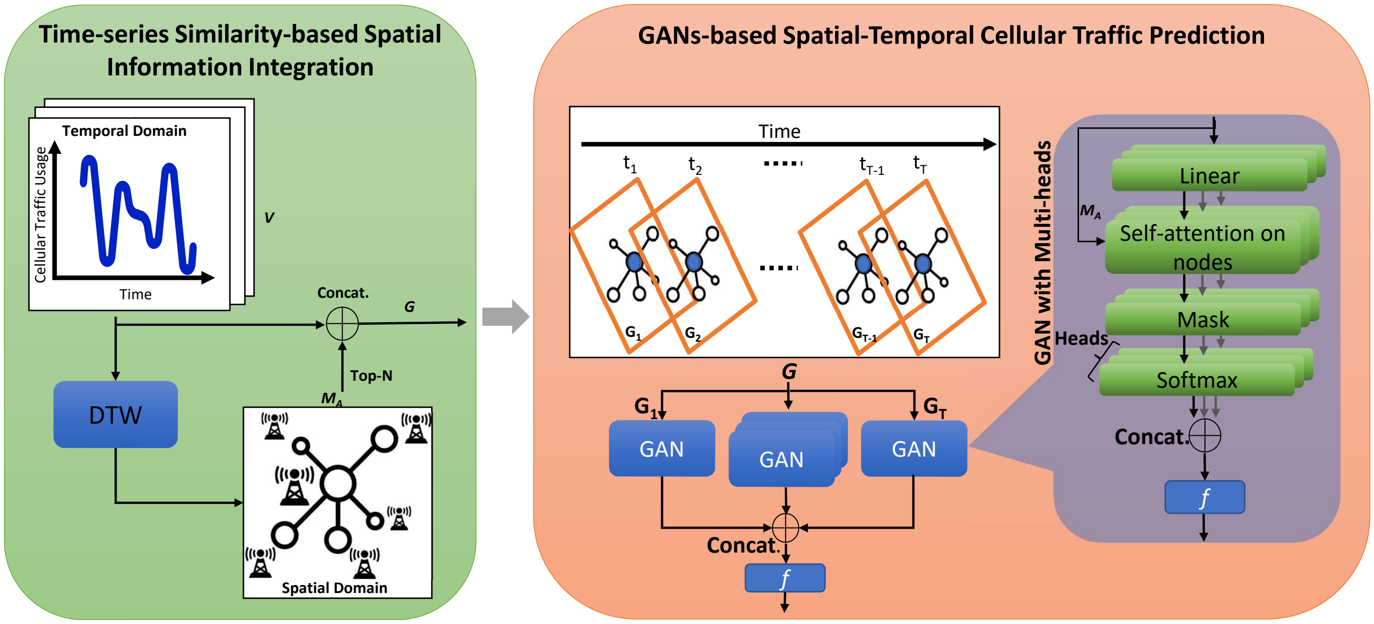


Fig. 4. The system model of TSGAN consists of time-series similarity-based spatial information integration and graph attention networks-based spatial-temporal cellular traffic prediction.

This makes assumptions in their work not in line with practices. Yu *et al.* [23] conducted massive in-field experiments with the help of 10 volunteers who contributed their cellular usage data on daily routines. Their proposed spatial-temporal fine-granular user traffic prediction model, graphic convolution gated recurrent unit network (GCGRN), was developed by a combination of the graph convolution network and gated recurrent units. Different from the [21], their model can provide two outputs that are the user's future data rate for each of online apps and the set of base stations that the user is going to visit, respectively.

Recently, attention mechanism [24] becomes an outstanding development in deep learning that has had success in challenging tasks such as machine translation and image processing. The attention mechanism could reflect the different impacts on neighbours around a single node. Unlike the graph convolution that uses the same weight for neighbouring nodes, the attention mechanism sets different and appropriate weights for each node. Graph Attention Networks (GAT) [25] were the first model to incorporate the attention mechanism in GNNs. Motivated by the outstanding performance of the attention mechanism and the topological defects of the real-world cellular network datasets, we propose to combine the time-series similarity-based spatial information integration method and graph attention networks to conduct the spatial-temporal cellular traffic prediction.

IV. Time-series Similarity-based Graph Attention Network

In this section, we propose a time-series similarity-based graph attention network, TSGAN, for spatial-temporal cellular traffic prediction, which includes a time-series similarity-based spatial information integration and graph attention networks for cellular traffic prediction.

The cellular traffic usage information can be measured by each base station and then the amount of cellular traffic consumption will be recorded as time-series in a feature vector. Traditional deep learning models such as LSTM lose the important spatial information because of learning from time-series only for traffic prediction. Taking full advantage of the spatial and temporal dependencies and characteristics within the cellular network data will explore more potential benefits of cellular traffic. To this end, we propose TSGAN, a deep graph neural network to conduct the spatial-temporal cellular traffic prediction.

As shown in Fig. 4, the proposed model consists of two sub-modules, the time-series similarity-based spatial information integration and GANs-based Spatial-Temporal Cellular Traffic Prediction. Different from the traditional statistical or deep learning methods that only analyse and learn from the temporal domain of cellular traffic, TSGAN solves the traffic prediction problem by utilising the features of the cellular traffic usage from both spatial and temporal domains among all base stations. TSGAN first extracts the spatial topology of the target cellular networks from the time-series of the cellular traffic usage and integrates the most relevant connection information in spatial domain with the time-series cellular traffic in temporal domain as a graph. Then, the cellular traffic usage data with spatial information integration can be regarded as the continuous slices containing completed spatial-temporal information of cellular network traffic. Next, the cellular traffic usage slices will become the input of the graph attention networks for spatial-temporal cellular traffic prediction. The following subsections will explain the proposed model in details.

A. Time-series Similarity-based Spatial Information Integration

For the GNNs-empowered model, the adjacency matrix is required to indicate the non-Euclidean spatial information of cellular traffic among cellular cells. In the TSGAN, we adopt the DTW [26] to calculate the time-series similarity between the network traffic of every two cells to present the spatial information. Specifically, the DTW distance $DTW(i,j)$ between time-series of two cells can be calculated by

$$\begin{aligned} DTW(i,j) = D(i,j) + min\{&DTW(i-1,j), \\ &DTW(i,j-1), \\ &DTW(i-1,j-1)\} \end{aligned} \tag{3}$$

where i and j are two time instances' index from the time-series of two cells, and $D(i,j)$ presents the distance between the matched pair of indices i and j based on the absolute differences between their values. By iterating all cells in the cellular network, we obtain the adjacency matrix M_A of the network by filtering top-N neighbours of each cell to indicate the most relevant spatial connection of each cell. We define $G = \{(V, M_A)\}$ as a directed graph set of cellular traffic usage slices where V is the set of all time-series cellular traffic of the network. During a time period of T, $G = \{G_1, G_2, ..., G_T\}$ where $G_t = (V_t, M_A)$ is a directed graph of cellular traffic usage slice at time t.

B. Graph Attention Networks for Spatial-Temporal Cellular Traffic Prediction

Inspired by the GCN, the graph attention neural network uses the first-order neighbours to make coefficients across nodes, which significantly reduces the number of parameters of the network and improves the efficiency [25]. Moreover, the attention mechanism can be regarded as implicitly assigning different convolution kernel parameters to each node, avoiding the limitation of the learning ability of the prediction model if all adjacent nodes share the same convolution kernel parameters like GCN. Therefore, TSGAN performs the spatial-temporal prediction for the traffic data based on the graph attention neural network considering that the targeted cell has different similarities to its cells in adjacency in the cellular network.

Spatial-temporal cellular traffic prediction can be viewed as the problem of extending the GNN model to the temporal domain. By receiving the G as the input for the spatial-temporal cellular traffic prediction, TSGAN aims to predict the cellular traffic usage in the next Δt period, $V_{T+\Delta t} = F_{TSGAN}(\{V_T, M_A\})$. To achieve this, TSGAN applies the stacked graph attention neural networks which work for the cellular traffic usage slices in G. TSGAN concatenates from each output of graph attention neural networks as the final output of the model.

Specifically, in each graph attention neural network of TSGAN, the graph attention layer, the attention mechanism, and attention coefficients can be formulated by

$$\begin{aligned} v_i' &= f(\sum_j \alpha_{ij} W v_j) \\ \alpha_{ij} &= \text{sotfmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})} \\ e_{ij} &= \alpha(Wv_i', Wv_j') = \alpha^T [Wv_i' || Wv_j'] M_A \end{aligned} \tag{4}$$

where f is a non-linearity which is LeakyReLU in TSGAN, $v_i, v_j \in V$ are the traffic flows of ith and jth cells in a graph slice, W is learned weight matrix, and $||$ is the concatenation operation. Since zero in the result of the attention coefficient means that there is no relationship between nodes, TSGAN performs the masked operation by replacing those zeros with negative infinity for the next softmax function.

The TSGAN also incorporates the multi-head attention mechanism, which has the important advantage of capturing more features with efficient computation of the node-neighbour pairs in parallel. Specifically, for H heads of independent attention mechanism described in equation 4, the final output of each graph attention neural network is formulated by

$$v_i' = ||_{h=1}^{H} f(\sum_j \alpha_{ij}^h W^h v_j) \tag{5}$$

where $||$ is the concatenation operation.

C. Loss Function

For training the TSGAN, the model aims to minimise the differences between the raw cellular traffic and the prediction results. Therefore, mean square error (MSE) is utilised as the loss function in the TSGAN training. Specifically, let y_t represent the target time-series observation and $\hat{y}_t$ denote the prediction result of TSGAN. The loss value can be obtained by the following equation:

$$Loss_{MSE} = \frac{1}{N}\frac{1}{W}\frac{1}{T}\sum_{i=1}^{N}\sum_{w=1}^{W}\sum_{t=1}^{T}(\hat{y}_t^{w,i} - y_t^{w,i})^2 \tag{6}$$

where N, W and T denote the number of nodes, time windows and time steps respectively.

V. Evaluation

In this section, the evaluation experiments are conducted within 100 cells of the cellular traffic dataset in Milan that we analysed in section II-A. The evaluations aim to present 1) the validation and the effectiveness of TSGAN in spatial-temporal cellular traffic prediction and 2) the performance improvement compared with the models used in existing mainstream researches.

A. Experiment Settings

We choose a 10*10 cell range in the centre of Milan city and extract continuous 30 days cellular traffic CDRs as time-series for evaluation. The cellular traffic data from the first 24 days are used as the training set and the rest 6 days are used as the test set. The length of the time window is set as 7 time instances and we use slicing-window method to expand the training samples. For the TSGAN, the adjacency matrix is generated by filtering top-5 neighbours of cells to indicate the spatial information in DTW-based spatial information integration. TSGAN sets six dimensions for the input layer, six dimensions for the hidden layer and one dimension for the output layer. The number of multi-head attention is set as two. In the training phase, we set 32 samples as batch size and choose Adam [27] as the optimisers for the loss convergence. The early stop mechanism is applied from the J.COp DL open-source library.

B. Compared References

We compare the performance between the TSGAN and the following three classic prediction models based on GNNs or GRU:

- ST-GCN: A GNN-based model use GCN described in section II-B as the core module for spatial-temporal traffic prediction, as in [19], [21], [23].
- ST-ChebNetGNN: A spatial-temporal Chebyshev graph neural network model proposed by Yan *et al.* [28] for intelligent transportation traffic prediction.
- GRU: A conventional deep learning model, which is a recurrent neural network variant, achieves higher learning efficiency with fewer parameters [29].

C. Evaluation Metrics

According to the state-of-the-art researches, the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are common metrics to represent the performance of prediction.

Specifically, we define the N is the time steps, $\hat{y}$ is the set of prediction values, and y is the set of target values.

- RMSE indicates the differences between the predicted value and the real value:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{7}$$

- MAE indicates the average of absolute differences between prediction value and real value.

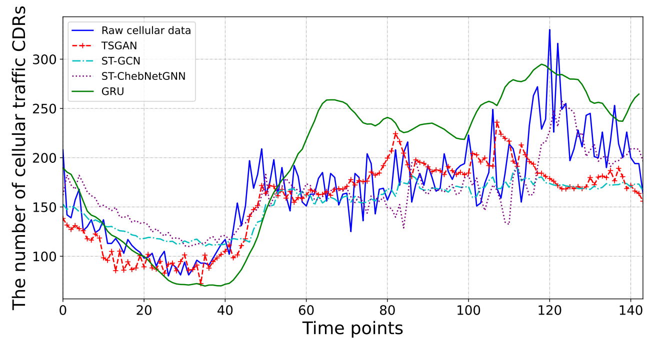
$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{8}$$

- MAPE is an average for absolute percentage errors. The smaller the MAPE, the closer the model is to perfect.

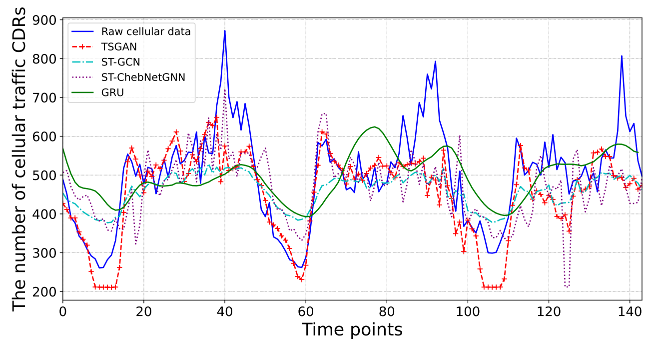
$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}|\frac{\hat{y}_i - y_i}{y_i}| \tag{9}$$

D. Evaluation Results

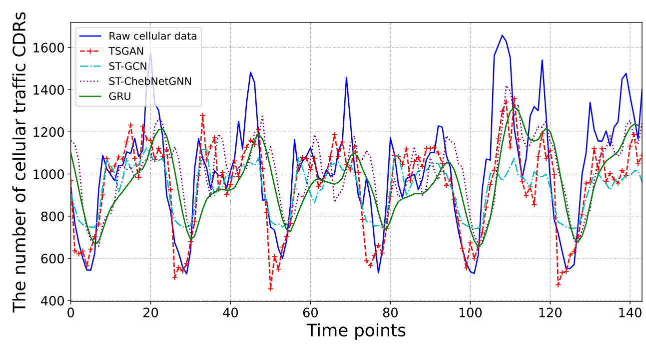
The evaluation results include the numerical comparisons in MAE, MAPE, and RMSE and performance displays among TSGAN and other three prediction models based on the cellular traffic dataset of Milan. We compare the prediction in three scenarios: short-term, mid-term, and long-term predictions with 10 minutes, 30 minutes, and 60 minutes



((a)) Shor-term Prediction: 10 minutes sampling interval



((b)) Mid-term Prediction: 30 minutes sampling interval



((c)) Long-term Prediction: 60 minutes sampling interval

Fig. 5. Prediction performance display on cell No.50 for different sampling interval

sampling intervals. To illustrate the prediction performance of TSGAN over different sampling intervals of cellular traffic, we randomly choose a cell (ID:50) to display the prediction results on the test dataset. Fig. 5 intuitively illustrates the prediction comparisons of last 144 time points which stand for short-term (1 day), mid-term (3 days), and long-term (6 days) prediction in the chosen cell. Moreover, the numerical comparison results of the above prediction performance are shown in Table III.

According to the Fig. 5 and Table III, the TSGAN achieves more accurate cellular traffic prediction performance than the ST-GCN, ST-ChebNetGNN, and GRU models in the Milan dataset from the perspectives of lower RMSE, MAE, and MAPE. Specifically, performance improvements of TSGAN for each scenario and metric are calculated and described below.

In the short-term prediction scenario, compared with the ST-GCN, ST-ChebNet, and GRU, the TSGAN presents performance improvements around 4.17%, 1.71%, and 24.13% in

TABLE III
NUMERICAL COMPARISONS OF PREDICTION PERFORMANCE.

Model	**TSGAN (ours)**			ST-GCN			ST-ChebNet			GRU		
Metrics / *Sampling Interval*	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*	*RMSE*
10-min	**25.75**	**0.13**	**35.94**	26.87	0.15	37.09	26.2	0.16	32.81	33.94	0.18	40.75
30-min	**66.16**	**0.12**	**98.67**	78.59	0.16	100.19	85.89	0.17	107.58	102.69	0.21	124.78
60-min	**141.18**	**0.13**	**185.71**	161.42	0.15	214.09	188.73	0.2	229.19	229.28	0.24	272.09

MAE; 13.33%, 18.75%, and 27.78% in MAPE; 3.1%, 1.72%, and 11.8% in RMSE, respectively.

In the mid-term prediction scenario, TSGAN improves prediction performance around 15.81%, 22.97%, and 35.57% in MAE; 25%, 29.41%, and 42.86% in MAPE; 1.52%, 8.28%, and 20.92% in RMSE, compared with ST-GCN, ST-ChebNet, and GRU, respectively.

In the long-term prediction scenario, TSGAN achieves around 12.54%, 25.19%, and 38.42% in MAE; 13.33%, 35%, and 45.83% in MAPE; 13.26%, 18.97%, and 31.75% in RMSE for the performance improvements, compared with ST-GCN, ST-ChebNet, and GRU, respectively.

The effectiveness of TSGAN in spatial-temporal cellular traffic prediction is demonstrated through the analysis of numerical comparisons among these four models. The attention mechanism brings the advantage of focusing on the important nodes with less focus on others, which can help the model to explore the practical spatial-temporal information of cellular traffic and to improve the prediction performance. Moreover, the comparisons further indicate that this kind of advantage presents different degrees in different scenarios. The TSGAN has the better expressive ability with the increase of sampling interval, which means the more temporal information that the cellular traffic has (e.g. periodical patterns), the more accurate result the model predicts.

E. Ablation Studies

To further investigate the effectiveness of modules in TSGAN, we conduct evaluations for three TSGAN variants: (a) TSGAN w/o DTW w/ Multi-heads, (b) TSGAN w DTW w/o Multi-heads, and (c) TSGAN w/o DTW w/o Multi-heads. We replace the adjacency matrix provided by DTW with the adjacency matrix of the ordinary one-hop geographic distance between cells of the cellular network to evaluate the effectiveness of time-series similarity-based spatial information integration. We remove the multi-head mechanism of GANs-based spatial-temporal prediction to evaluate the effectiveness of capturing more features with efficient computation in parallel. The evaluations is conducted within all cells in the selected network area. The results are calculated on average, and the prediction performance comparisons of TSGAN variants are shown in Table IV. We can observed that compared with the prediction performance of complete TSGAN, the prediction performance of all variants drop. This demonstrates that the complete architecture of TSGAN is effective in spatial-temporal cellular traffic prediction. Moreover, the time-series similarity-based spatial information can provide more features in spatial domain to help improve the prediction performance than the ordinary geographic information. Furthermore, the multi-head mechanism employed in TSGAN can effectively improve the learning ability and help to learn more plentiful features and information from different representation subspaces. Through the above evaluation, we can further confirm the effectiveness of TSGAN and the importance of exploiting both spatial and temporal characteristics of cellular traffic for more accurate predictions.

VI. CONCLUSION AND FUTURE WORK

In this paper, we first presented the spatial-temporal analysis of a real-world cellular network traffic dataset of the city of Milan. Then, we reviewed the existing state-of-the-art researches using GNN-empowered models to improve the accuracy and reliability of the cellular network traffic prediction by exploiting the practical spatial and temporal information of the cellular network traffic data. Based on our analysis and review, we proposed a time-series similarity-based graph attention network (TSGAN) model, which includes a time-series similarity-based spatial information integration and graph attention networks for cellular traffic prediction. To validate the effectiveness and the performance improvement of TSGAN, comparison experiments were conducted and the results were analysed based on the Milan dataset over three classic prediction models based on GNNs or GRU. The experiment results demonstrated that our proposed TSGAN outperformed these referencing models on all metrics in short-term, mid-term, and long-term prediction scenarios.

In the future, GNN-empowered spatial-temporal prediction models not only can improve the prediction performance under current 5G environments but also show the great potentials in adapting emerging applications and scenarios such as the era of 6G and Industry 5.0. For example, under the pressure of strict data protection regulations such as General Data Protection Regulation (GDPR) which become the core requirements in the design of 6G and Industry 5.0, service providers may not be able to centrally store user data to train the GNNs-empowered models for cellular traffic prediction in the

TABLE IV
PREDICTION PERFORMANCE COMPARISONS OF ABLATION STUDIES OF TSGAN.

Sampling Interval	10-min			30-min			60-min		
Model \ *Metrics*	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*	*RMSE*
TSGAN	**73.9**	**0.31**	**110.32**	**196.34**	**0.28**	**332.6**	**399.46**	**0.31**	**670.6**
w/o DTW, w/ Multi-heads	185.8	0.67	217.49	278.61	0.37	382.73	524.07	0.33	751.04
w/ DTW, w/o Multi-heads	171.1	0.99	224.92	449.67	0.53	595.25	638.86	0.42	897.28
w/o DTW, w/o Multi-heads	208.75	0.77	237.34	639.24	1.06	812.59	888.81	0.54	1124.33

future. In this case, federated learning (FL) as an alternative communication-efficient learning paradigm that pushes model training to the devices where data generated [30] would be a promising solution cooperated with the GNN-based models using the attention mechanism to preserve the data privacy. Moreover, in the vision of sixth generation networks (6G) and Industry 5.0, emerging facilities such as unmanned aerial vehicles (UAVs) and communication satellites will be deployed and become essential kinds of service provision platforms to offer fast and seamless communication and networking services. According to time scales of service requirements and changes of connections among these dynamic facilities, the management of dynamic networking will take a non-negligible part in this system. In this regard, there are potential research opportunities for GNNs-empowered spatial-temporal cellular traffic prediction under this promising scenario to be solved before dynamic non-terrestrial platforms can be effectively used in future 6G and Industry 5.0 scenarios.

ACKNOWLEDGEMENT

This work was partially supported by the EU Horizon 2020 INITIATE project under the Grant Agreement No. 101008297, the Royal Society International Exchanges project - IEC\NSFC\211460, and the China Scholarship Council (No. 201806070140). The European Union Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] E. Commission, D.-G. for Research, Innovation, M. Breque, L. De Nul, and A. Petridis, *Industry 5.0 : towards a sustainable, human-centric and resilient European industry*. Publications Office, 2021.

[2] Cisco, "Cisco annual internet report (2018–2023)," 2020, https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html, Last accessed on March 9, 2020.

[3] J. Müller, "Enabling technologies for industry 5.0, results of a workshop with europe's technology leaders," *Directorate-General for Research and Innovation*, 2020.

[4] A. Rago, G. Piro, G. Boggia, and P. Dini, "Multi-task learning at the mobile edge: An effective way to combine traffic classification and prediction," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 10 362–10 374, 2020.

[5] G. Aceto, G. Bovenzi, D. Ciuonzo, A. Montieri, V. Persico, and A. Pescapé, "Characterization and prediction of mobile-app traffic using markov modeling," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 907–925, 2021.

[6] A. Dalgkitsis, M. Louta, and G. T. Karetsos, "Traffic forecasting in cellular networks using the lstm rnn," in *Proceedings of the 22nd Pan-Hellenic Conference on Informatics*, 2018, pp. 28–33.

[7] C. Zhang, H. Zhang, D. Yuan, and M. Zhang, "Citywide cellular traffic prediction based on densely connected convolutional neural networks," *IEEE Communications Letters*, vol. 22, no. 8, pp. 1656–1659, 2018.

[8] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1389–1401, 2019.

[9] C. Zhang and P. Patras, "Long-term mobile traffic forecasting using deep spatio-temporal neural networks," in *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. Mobihoc '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 231–240. [Online]. Available: https://doi.org/10.1145/3209582.3209606

[10] K. He, Y. Huang, X. Chen, Z. Zhou, and S. Yu, "Graph attention spatial-temporal network for deep learning based mobile traffic prediction," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[11] K. He, X. Chen, Q. Wu, S. Yu, and Z. Zhou, "Graph attention spatial-temporal network with collaborative global-local learning for citywide mobile traffic prediction," *IEEE Transactions on Mobile Computing*, 2020.

[12] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.

[13] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner *et al.*, "Relational inductive biases, deep learning, and graph networks," *arXiv preprint arXiv:1806.01261*, 2018.

[14] G. Barlacchi, M. De Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland, and B. Lepri, "A multi-source dataset of urban life in the city of milan and the province of trentino," *Scientific data*, vol. 2, no. 1, pp. 1–15, 2015.

[15] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016.

[16] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

[17] L. Fang, X. Cheng, H. Wang, and L. Yang, "Mobile demand forecasting via deep graph-sequence spatiotemporal modeling in cellular networks," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3091–3101, 2018.

[18] ——, "Idle time window prediction in cellular networks with deep spatiotemporal modeling," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1441–1454, 2019.

[19] N. Zhao, Z. Ye, Y. Pei, Y.-C. Liang, and D. Niyato, "Spatial-temporal attention-convolution network for citywide cellular traffic prediction," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2532–2536, 2020.

[20] X. Wang, Z. Zhou, F. Xiao, K. Xing, Z. Yang, Y. Liu, and C. Peng, "Spatio-temporal analysis and prediction of cellular traffic in metropolis," *IEEE Transactions on Mobile Computing*, vol. 18, no. 9, pp. 2190–2202, 2018.

[21] M. Kalander, M. Zhou, C. Zhang, H. Yi, and L. Pan, "Spatio-temporal hybrid graph convolutional network for traffic forecasting in telecommunication networks," *arXiv preprint arXiv:2009.09849*, 2020.

[22] S. Zhao, X. Jiang, G. Jacobson, R. Jana, W.-L. Hsu, R. Rustamov, M. Talasila, S. A. Aftab, Y. Chen, and C. Borcea, "Cellular network traffic prediction incorporating handover: A graph convolutional

approach," in *2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2020, pp. 1–9.

[23] L. Yu, M. Li, W. Jin, Y. Guo, Q. Wang, F. Yan, and P. Li, "Step: A spatio-temporal fine-granular user traffic prediction system for cellular networks," *IEEE Transactions on Mobile Computing*, 2020.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[25] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[26] M. Müller, "Dynamic time warping," *Information retrieval for music and motion*, pp. 69–84, 2007.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[28] B. Yan, G. Wang, J. Yu, X. Jin, and H. Zhang, "Spatial-temporal chebyshev graph neural network for traffic flow prediction in iot-based its," *IEEE Internet of Things Journal*, 2021.

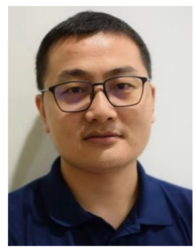
[29] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.

[30] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

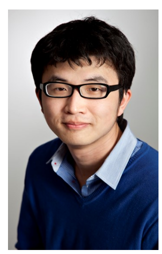
Zi Wang is currently a Ph.D. candidate in Computer Science at the College of Engineering, Mathematics and Physical Science, University of Exeter, UK. He received his M.Eng. and B.Eng. in Computer Science from the University of Electronic Science and Technology of China (UESTC) in 2018 and 2015, respectively. His research interests focus on deep learning, applied machine learning, cloud and edge computing and computer networks.

Jia Hu is a Senior Lecturer in Computer Science at the University of Exeter. He received his Ph.D. degree in Computer Science from the University of Bradford, UK, in 2010, and M.Eng. and B.Eng. degrees in Electronic Engineering from Huazhong University of Science and Technology, China, in 2006 and 2004, respectively. His research interests include mobile and ubiquitous computing, resource optimization, applied machine learning, and network security.

Geyong Min is a Professor of High Performance Computing and Networking in the Department of Computer Science within the College of Engineering, Mathematics and Physical Sciences at the University of Exeter, United Kingdom. He received the PhD degree in Computing Science from the University of Glasgow, United Kingdom, in 2003, and the B.Sc. degree in Computer Science from Huazhong University of Science and Technology, China, in 1995. His research interests include Computer Networks, Wireless Communications, Parallel and Distributed Computing, Ubiquitous Computing, Multimedia Systems, Modelling and Performance Engineering.



Zhiwei Zhao is currently a full professor at the School of Computer Science and Engineering in University of Electronic Science and Technology of China. He received his PhD degree in Computer Science at the College of Computer Science, Zhejiang University in 2015. His research interests include edge computing, IoT systems, low-power computing, etc. He is a member of IEEE, ACM and CCF.



Zheng Chang [SM] received Ph.D degree in Communications Engineering from the University of Jyväskylä, Jyväskylä, Finland in 2013 . He has published over 130 papers in journals and conferences, and received best paper awards from IEEE TCGCC and APCC in 2017 and has been awarded as the 2018 IEEE Best Young Research Professional for EMEA and 2021 IEEE MMTC Outstanding Young Researcher. He has been served as symposium chair, publicity chair and workshop chair and also participated in organizing workshops and special sessions for many IEEE flagship conferences, such as Infocom, ICC and Globecom. He is an editor of Springer Wireless Networks, International Journal of Distributed Sensor Networks, and IEEE Wireless Communications Letters. He was the exemplary reviewer of IEEE Wireless Communication Letters in 2018. He also acts as a guest editor of IEEE Communications Magazine, IEEE Wireless Communications, IEEE Networks, IEEE Internet of Things Journal and IEEE Transactions on Industrial Informatics. His research interests include IoT, cloud/edge computing, security and privacy, vehicular networks, and green communications.

Zhe Wang is a Ph.D. student in Computer Science at the University of Exeter. He received both B.Eng. and M.Eng. degree in Computer Science from the University of Electronic Science and Technology of China (UESTC), Chengdu, China in 2016 and 2019, respectively. His research interests include deep reinforcement learning, cloud and edge computing, and computer system optimization.