# Fine-grained Spatio-Temporal Distribution Prediction of Mobile Content Delivery in 5G Ultra-Dense Networks

Shaoyuan Huang, Heng Zhang, Xiaofei Wang<sup>\*</sup>, *Senior Member, IEEE*, Min Chen, *Fellow, IEEE*, Jianxin Li, *Member, IEEE*, and Victor C. M. Leung, *Life Fellow, IEEE* 

**Abstract**—The 5G networks have extensively promoted the growth of mobile users and novel applications, and with the skyrocketing user requests for a large amount of popular content, the consequent content delivery services (CDSs) have been bringing a heavy load to mobile service providers. As a key mission in intelligent networks management, understanding and predicting the distribution of CDSs benefits many tasks of modern network services such as resource provisioning and proactive content caching for content delivery networks. However, the revolutions in novel ubiquitous network architectures led by ultra-dense networks (UDNs) make the task extremely challenging. Specifically, conventional methods face the challenges of insufficient spatio precision, lacking generalizability, and complex multi-feature dependencies of user requests, making their effectiveness unreliable in CDSs prediction under 5G UDNs. In this paper, we propose to adopt a series of encoding and sampling methods to model CDSs of known and unknown areas at a tailored fine-grained level. Moreover, we design a spatio-temporal-social multi-feature extraction framework for CDSs hotspots prediction, in which a novel edge-enhanced graph convolution block is proposed to encode dynamic CDSs networks based on the social relationships and the spatio features. Besides, we introduce the Long-Short Term Memory (LSTM) to further capture the temporal dependency. Extensive performance evaluations with real-world measurement data collected in two mobile content applications demonstrate the effectiveness of our proposed solution, which can improve the prediction area under the curve (AUC) by 40.5% compared to the state-of-the-art proposals at a spatio granularity of 76m, with up to 80% of the unknown areas.

Index Terms—Mobile Content Delivery Prediction, Spatio Fine-grained, Spatio-Temporal-Social Features Extraction, Graph Convolution Network.

## 1 INTRODUCTION

W ITH the popularization of mobile networks, user requests for mobile application content [1], [2] are growing explosively. When it comes to the 5G era, as the number of users and novel applications like Virtual Reality and cloud games soar, user content requests, and more importantly, the consequent content delivery services (CDSs), will account for a significant portion of mobile traffic in the networks [3].

More complex user requests and lower latency requirements place greater pressure on operators and mobile service providers. To address these challenges, significant efforts are devoted to optimize the mobile network. While at the architecture level, to meet the characteristics of proximity transmission due to the high-frequency band of 5G, ubiquitous networks are proposed to push the network resources (e.g., access points (APs)/content allocation/computing) to

- Shaoyuan Huang, Heng Zhang and Xiaofei Wang are with Tianjin Key Laboratory of Advanced Networking, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China. E-mail: {hsy\_23, hengzhang, xiaofeiwang}@tju.edu.cn
- Min Chen is with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China. E-mail: minchen@ieee.org
- Jianxin Li is with the School of Information Technology, Deakin University, Burwood, VIC 3220, Australia. E-mail: jianxin.li@deakin.edu.au
- Victor C. M. Leung is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China E-mail: vleung@ieee.org

the edge of users to reduce the latency of CDSs. On the other hand, at the algorithm level, more network intelligence technologies are proposed to identify user requests and the corresponding CDSs automatically. Compared to reacting to users' requests passively, such network intelligence technology enables operators and mobile service providers to proactively improve their resource provision or cache strategies to achieve higher network throughput and lower service latency [4].

#### 1.1 CDSs Prediction under 5G Networks

Revolutions in novel ubiquitous network architectures have brought paradigm-level changes to mobile networks. Specifically, the Ultra-Dense Networks (UDNs) empower 5G tremendous access capability, composed of extensive densedeployed small cell base stations (SBSs). Additionally, with the assistance of Mobile Edge Computing (MEC), UDNs are capable of providing intelligent and efficient CDSs for user requests, which is implemented by SBSs edge servers. By contrast, the network intelligence technologies for CDSs prediction in 5G UDNs are still at an early stage and facing the following challenges:

(*i*) **Insufficient spatio precision.** The APs in the ubiquitous networks are becoming extremely dense, making each AP serve a smaller range and the distribution of CDSs become spatio fine-grained. In UDNs, each SBSs serves user requests within 100m [5], while MEC's edge servers are also

<sup>\*</sup> Xiaofei Wang is corresponding author.



Fig. 1: The Pipeline of the Designed Network Intelligence.

deployed within the nearest hop of the network from users [6]. However, existing studies on the analysis and prediction of CDSs are scoped to regions or even entire cities [7]–[10]. These spatio coarse-grained approaches are not compatible with the service granularity of 5G UDNs and provide very limited guidance for APs under UDNs or MEC. To make network intelligence techniques compatible with the needs of the 5G network ecology, a method that can perform spatio fine-grained prediction of CDSs is necessary.

(*ii*) Lack of generalizability for predicting unknown areas. Compared to traditional networks, user mobility has a greater impact on CDSs in ubiquitous networks. UDNs and MEC aim to provide closer communication and computational resources to hotspots of CDSs such as transportation stations, shopping malls and factories. However, such hotspots may migrate from old locations to adjacent areas as users move, requiring the methods that can accurately capture CDSs for these unknown areas (*without historical data coverage*). However, existing studies mostly focus on predicting CDSs under a fixed area [11]–[13] and cannot predict unknown areas with the possibility of CDSs appearing.

(*iii*) Complex multi-feature dependencies of user requests. Extensive studies and dataset analysis have demonstrated that users' mobile content requests and the consequent CDSs have spatio, temporal and social dependencies. A comprehensive features capture model is needed to improve the CDSs prediction accuracy of network intelligence. While there are models that capture the temporal or spatio dependency of CDSs [7], [14], [15], few studies consider more than two dependency features (especially social dependency, which has a significant impact on CDSs.)

To fill the gap between the network intelligence techniques and 5G ubiquitous networks, in this paper, we address the problem of spatio fine-grained prediction of mobile CDSs.

We use Fig. 1 to show the pipeline of our designed network intelligence. First, there will be a dense distribution of edge servers (or SBSs) in the cities covered by 5G UDNs. These servers generally cover tens of meters and can handle and record a mass of mobile users' heterogeneous CDSs. The network intelligence deployed in the central cloud servers of the core network will then process the records data and train the prediction model that identifies how the spatio and temporal distribution of CDSs will evolve.

To make the problem more focused, we propose that

the output of the prediction model is not just the location of CDSs but the **hotspots** (as defined in Section 3) for a specific category of content as shown in the red area in Fig. 1. The hotspots reflect the areas with the most CDSs over time, and large-scale hotspots analysis can be used by the core network to guide SBSs optimization at various areas throughout the city.

#### 1.2 Paper Contributions

In this paper, we propose a network intelligence oriented 5G UDNs. Specifically, we propose a spatio-temporal graphs modeling algorithm to model CDSs into CDSs networks at the specific spatio granularity. Besides, we propose a multi-feature extraction framework tailored to solve the spatio fine-grained hotspots prediction with high accuracy and spatio generalizability. We summarize our contributions into three aspects:

(*i*) **Spatio fine-grained encoding with high generalizability based on Geohash.** To refine the spatio granularity and model mobile CDSs of arbitrary areas (including known and adjacent unknown areas), we introduce the Geohash algorithm to encode the GPS information with specific granularity. After that, we propose a spatio-temporal graphs algorithm that can model CDSs networks for the CDSs records collected from mobile devices and embed the realworld spatio-temporal attributes.

(*ii*) Spatio-temporal-social driven hotspots prediction model with edge enhanced block. To maximize the predictor's accuracy, we analyze the main features affecting the distribution of CDSs from datasets. After that, we propose a hybrid model named Temporal-Edge Enhanced Graph Convolution Network (T-EEGCN) that simultaneously captures spatio, temporal, and social features to predict CDSs hotspots. The core lies in an elaborate Edge-Enhanced Graph Convolution Block (EEGCB), which is designed to encode the dynamic CDSs networks based on the extraction of spatio-social features. The EEGCB solves two major problems of incorporating real-world spatio features in virtual social networks and handling dynamic graphs (which is difficult for traditional GCNs). After that, we introduce LSTM to learn the CDSs networks encoding at different times and extract the temporal dependency of CDSs.

(*iii*) Extensive performance evaluations with a large number of measurements collected in two real-world CDSs applications demonstrates that our approach outperforms a broad benchmarks based on traditional and state-of-theart prediction models. Besides, we perform ablation studies to analyze the role of different components of T-EEGCN and the dependencies of CDSs. Finally, we compared the performance of the proposed approach under different time scales settings to validate its temporal scalability.

The rest of this paper is organized as follows. In Section 2, we introduce the related works and highlight the novelty of our approaches. Formal definitions of key concepts and the problem are given in Section 3. Section 4 presents the proposed spatio-temporal graphs modeling and the T-EEGCN framework in detail. In Section 5, we evaluate the effectiveness of the solution via extensive experiments. We finally draw conclusions in Section 6.

## 2 RELATED WORK

In this section, we review previous approaches regarding network intelligence for CDSs prediction and spatiotemporal prediction models.

Network Intelligence for CDSs Prediction. 1) Content popularity prediction. Collaborative filtering (CF) [16] is considered an effective way to uncover user demands for content and products. Previous CF algorithms only consider the user's interest in the content (ratings), which is effective but has significant limitations in accuracy. To improve the accuracy of CF, lots of studies try to incorporate spatio, temporal or social correlation considerations into the algorithm. Study [17] explores collaborative filtering with temporal dynamics. Study [18] does collaborative filtering with user mobility and context. Nevertheless, the CF algorithms still face the problems of cold start, sparse data, and poor scalability, which means the CF cannot be applied to complex and large-scale mobile CDSs prediction.

2) Traffic/capacity forecasting. There are two types of commonly used predicting methods: classic linear methods such as AutoRegressive Integrated Moving Average (ARIMA) [19] and the other is machine learning based non-linear methods. In fact, non-linear methods gradually achieve better performance [20], [21]. Study [13] presents a hybrid deep learning model on the basis of autoencoder and Long-Short Term Memory networks (LSTM) to simultaneously capture the spatial and temporal dependency among different cells. Besides, the city-scale wireless traffic predictions are also investigated in [22]-[24], in which the authors introduce novel prediction frameworks by modeling spatio-temporal dependency over cross-domain datasets. In the latest study [25], the authors propose a novel wireless traffic prediction framework, by which a high-quality prediction model based on dual attention scheme is trained collaboratively by multiple edge clients.

Regardless of forms, a major drawback of existing methods is that they can only predict CDSs in fixed areas and cannot address the new CDSs in unknown areas caused by user mobility. In addition, these methods are spatio coarsegrained (city-scale), making it difficult to predict CDSs in specific fine-grained areas.

Spatio-temporal Prediction Methods. 1) Spatio-temporal analysis and feature extraction models. From the spatio domain, [8] analyzes the spatio importance of telecommunication hotspots. The authors model the hotspots as nodes in a graph and then apply node centrality metrics that quantify the importance of each node. From the time domain, study [26] finds the traffic dynamics of CDSs can be well captured by Markov models. Realizing the spatio-temporal dependency of mobile CDSs, traditional methods [27], [28] and machine learning models for spatio-temporal features extraction have been investigated in order to achieve accurate prediction of CDSs. [29] propose a novel deep generative adversarial network to address the crowdsourcingbased urban cellular traffic prediction. In [30], the authors introduce a multivariate and spatio-temporal approach to predict cellular traffic, which considers cell features, peak hours and handover traffic.

In the latest studies, more novel deep learning based spatio-temporal predictors (e.g., Transformer) are proposed for user traffic prediction [31], passenger demand prediction [32], user mobility prediction [33], [34], and other fields.

2) Graph Convolution Networks (GCNs) for spatio-temporal prediction. In recent studies, GCNs [35] have been widely used in the spatio-temporal modeling of graph structure due to their capability of handling Non-Euclidean data. [11] propose a graph-based deep learning approach to model intower and inter-tower traffic and learn long-distance spatio correlations for mobile traffic prediction. Both [36] and [37] propose a convolutional neural network fusing GCN and GLU to capture spatio features of the roads network and temporal features in the transportation traffic domain.

Nonetheless, traditional GCNs assume a fixed structure among networks (e.g., BSs and roads networks), so they compute graph structure features once and set them as constant coefficients. In this study, the CDSs records of different time slots are modeled as a series of dynamic graphs (with different nodes and graph structures), so static graph convolution methods do not apply to our scenario.

All aforementioned studies mainly focus on spatiotemporal features but ignore the social relations among users. Since users share more frequently in current mobile applications, recent studies have found that social relations have a significant impact on mobile content requests [38]– [40]. In this study, we exploit graph convolution to model user social relations. Moreover, the exploitation of edge features and novel information propagation/aggregation methods are implemented for capturing spatio dependency and addressing dynamic graphs, respectively.

# **3 PRELIMINARIES**

We proceed to introduce the background settings and formalize the hotspots prediction problem. Frequently used notations are summarized in Table 1.

**TABLE 1: Summary of Notations** 

Notation	Description
R	The spatio granularity.
Ι	The CDSs intensity, which is set to the num-
	ber of contents requests.
M	The hotspots matrix.
au	Hotspots threshold.
${\cal G}^i_i$	The CDSs network of area $i$ at time $j$ .
$\check{G^i}$	The series of CDSs networks of area $i$ .
X	The node features.
$\mathcal{P}$	The dimension of node features.
T	The size of time windows of $G$ .
V	Node set of $\mathcal{G}$ .
E	Edge set of $\mathcal{G}$ .
ε	The edge features.
$geo_i^j$	The $j$ bits Geohash encoding of node $i$ .
$a_{ij}$	The edge weights of $edge_{ij}$ .
$h_i^l$	The feature embedding of node $i$ in layer $l$ .
$m_{ij}$	The information propagation to node <i>i</i> from
0	node <i>j</i> .
$\mathcal{V}$	The CDSs network embedding.
0	The output of LSTM.

#### 3.1 Settings

**Definition 1 (Spatio Fine-grained).** The spatio granularity R refers to the radius of the predicted unit area. In the 5G UDNs architecture, the range served by any edge server or cell is greatly reduced, which requires the spatio granularity of the model's prediction to be changed from the previous coarse granularity (city-level) to fine granularity ( $R \le 100m$  [5]).

**Definition 2 (Unknown Areas).** Unknown areas are the areas with no CDSs appearing in the historical data

We call the areas with CDSs appearing in the historical data known areas, and we predict the volume of CDSs in these areas and determine whether they become hotspots. In addition, we concern all unknown areas adjacent to known areas. Though no historical CDSs occur in these areas, we argue that they also have the potential to become hotspots in the foreseeable future due to user mobility. The unknown areas prediction is particularly important for service scheduling and device deployment in 5G UDNs.

**Definition 3 (Hotspots).** *Hotspots are areas with very high number of CDSs relative to others.* 

To calculate the hotspots in all areas, we propose the hotspots delineation method based on the overall threshold with reference to [8]. Specifically, if the number of CDSs of an area are greater than the over all threshold  $\tau$ , we set the area as a hotspot. The threshold  $\tau$  is calculated as follows:

$$\tau = \frac{1}{N} \cdot \sum_{a=1}^{N} I_a + (Max(\vec{I}) - \frac{1}{N} \cdot \sum_{a=1}^{N} I_a) \cdot P$$
(1)

where N is the number of areas,  $I_a$  is the num of CDSs in area a, and P is the parameter to determine the cutoff threshold.

**Definition 4 (Hotspots Matrix).** A hotspots matrix M is a 0/1 matrix which represents the distribution of hotspots and non-hotspots in a large space with N areas.

**Definition 5 (CDSs Network).** CDSs network G is the social network built from the CDSs records. Each node v in CDSs networks represents once **CDSs**, e.g., video streaming delivery. Each edge e represents the social relation between users of twice CDSs, i.e., the two users have a friendly relationship.

## 3.2 Problem Statement

We use a series of snapshots  $G = \{\mathcal{G}_{\kappa-T+1}, ..., \mathcal{G}_{\kappa}\}$  to represent different CDSs networks of T time windows. Each snapshot  $\mathcal{G}_t = (V_t, E_t)$  is a directed graph with different node set V and edge set E. We use  $X \in \mathbb{R}^{T \times |V_t| \times \mathcal{P}}$  as node features and  $\mathcal{P}$  as the feature dimension (will be discussed in Section 5.2). Besides, we use  $\mathcal{E} \in \mathbb{R}^{T \times |V_t| \times |V_t|}$  as edge features. The hotspots prediction problem can be formally described below:

$$[G; X; \mathcal{E}] \xrightarrow{f(\cdot)} \hat{M}_{\kappa+1} \tag{2}$$

where  $f(\cdot)$  is the model that we need to construct. We can summarize the hotspots prediction task to a classification

TABLE 2: *R* of Different Geohash Encoding Lengths.

Geohash length	lat bits	lng bits	R(m)
5	12	13	2400
6	15	15	610
7	17	18	76
8	20	20	19

problem based on sequential graph data and the problem solving process can be formalized as follows.

$$\hat{M}_{\kappa+1} = \operatorname*{arg\,max}_{M_{\kappa+1}} \log P\left(M_{\kappa+1} \mid G, X, \mathcal{E}\right) \tag{3}$$

$$\mathbf{s.t.} \forall a \in M : R(a) <= 100m \tag{4}$$

$$\|G\| = T \tag{5}$$

where the objective of CDSs hotspots prediction problem is to find the most likelihood CDSs hotspots matrix  $\hat{M}_{\kappa+1}$ according to the history CDSs networks *G*, node features *X*, and edge features  $\mathcal{E}$ .

There are two main constraints to limit the problem sovling. In the first constraint (eq. 4), we limit the spatio granularity R of every area a to less than 100m, since we should guarantee that the model can accurately distinguish the distribution of CDSs in different fine-grained areas over a wide range of regions. Besides, we limit the periods length of the historical data to T since we need to trade off the impact of historical information on the prediction and the temporal granularity.

## 4 FINE-GRAINED HOTSPOTS PREDICTION

We proceed to introduce an end-to-end deep learning model, called temporal-edge enhanced graph convolution network (T-EEGCN), to predict spatio fine-grained CDSs hotspots. First, we introduce a spatio information encoding method, based on which we propose a spatio-temporal subgraph sampling algorithm, the complete encoding and algorithm process is shown in Fig. 2. Then, we present the details of T-EEGCN.

## 4.1 Spatio-Temporal Graphs Modeling

The GPS of CDSs are key information to model the space, which can be automatically obtained by the mobile terminal when users make requests. However, since the GPS encoding is global-level, its spatio granularity is much larger than our problem concerns, so we need a method to further



Fig. 2: Spatio-Temporal Graphs Modeling

refine the encoding of spatio information to help distinguish between fine-grained areas.

In addition, we need to access the adjacency between areas and equally encode these areas, which is a prerequisite for predicting unknown areas.

We introduce the Geohash encoding algorithm <sup>1</sup> to encode GPS information of CDSs into short strings of letters and digits. Geohash is a hierarchical spatio data structure that subdivides space into buckets of grid shapes. It allows areas of arbitrary precision to be encoded by increasing the encoding length. Considering Geohash's accuracy and the requirement of spatio fine-grained (as shown in the Table 2), we set the encoding length of Geohash to **7bits**, i.e., the concerned spatio granularity R = 76m.

Another feature of Geohash is the hierarchical encoding, which guarantees that the longer a shared prefix between two geohashes is, the spatio closer they are together. Therefore, we can obtain the encoding of adjacent areas of known areas based on the prefixes of existing codes. Although Geohash encodes grid shape areas, we argue that in our modeling process, Geohash can be replaced by other shape geocoding algorithms as long as it can encode areas of arbitrary precision and encode adjacent areas.

After determining the spatio granularity and the encoding of adjacent areas, we propose the spatio-temporal graphs modeling, i.e., the algorithm 1 to transform CDSs records into CDSs networks. The spatio-temporal graphs modeling algorithm achieves the following three main goals:

(*i*) **Construction of global CDSs network** *GG*. As shown in steps 2-7, the first step is to build the global CDSs network *GG* based on CDSs and users' social relations. Specifically, we create a node for each CDS and generate an edge between two nodes if a social connection exists between the two users (possibly friends or the CDS is a content share). Note that the edges in the CDSs network are all directed edges obtained according to the time sequence.

(*ii*) Generation of high-dimensional features. As shown in step 19, another goal of the algorithm is to translate the Geohash encoding of CDSs into high-dimensional features of nodes and generate edge features. We propose the node-area diffusion method to calculate node features. Specifically, when calculating the features of node *i*, we consider the impact of CDSs within its surrounding  $\mathcal{P}$  areas. Therefore, the node features *X* will be obtained as follows:

$$X[i][j] = \begin{cases} 0, & geo_i[\text{last}] \neq j, \\ I, & geo_i[\text{last}] = j. \end{cases}$$
(6)

where  $i \in [0, n)$  is the node id,  $j \in [0, \mathcal{P})$  is the area id, and *I* is the CDSs intensity, which is set to the number of successfully completed content requests, e.g., the number of application downloads or video transfers. Since Geohash uses 32-decimal encoding, we set *last* to the decimal form of the last digit of the encoding, e.g., gbsuv[last] = 28.

The approach to generate node features based on adjacent areas has two benefits: first, the distribution of CDSs under the  $\mathcal{P}$  areas can be rebuilt easily from all node features. Second, the adjacent areas that have never been recorded with CDSs can be represented and modeled.

1. http://geohash.org/site/tips.html

Algorithm 1: Spatio-Temporal Graphs Modeling					
Input: CDSs Records Set:Reqs_Set					
<b>Output:</b> Spatio-Temporal Graphs Set: <i>G_Set</i>					
1 Init Global Graph: $GG = \emptyset$ , $G\_Set = \emptyset$					
2 for $record_i$ in $Reqs\_Set$ do					
3 Add $node_i$ $(t_i, geo_i^7)$ to $GG$					
4 <b>if</b> $user_j$ is a friend to existing node's user $user_i$ then					
5 Add $edge_{ji}$ ( $\mathcal{E}_{ji}$ ) to $GG$					
6 end					
7 end					
s for $node_i$ in $GG$ do					
9 // Avoid modeling duplicate areas					
10 <b>if</b> $geo_i^{o}$ hasn't appear then					
11 get all $nodes_j$ that $geo_j^0 == geo_i^0$					
12 // Sub-area extraction					
13 get $subgraph_i$ all $nodes_j$					
14 Init $G^a = \emptyset$					
15 <b>for</b> $T_i$ in Time Windows <b>do</b>					
16 // Time extraction					
get all $nodes_j$ in $subgraph_i$ that $t_j$ in $T_i$					
18 get $\mathcal{G}_T^a$ contains all $nodes_j$					
19 get X, $\mathcal{E}$ of $\mathcal{G}_T^a$					
20 Add $\mathcal{G}_T^a$ to $G^a$					
21 end					
22 add $G^a$ to $STG\_Set$					
23   end					
24 end					
25 return STG Set					

We then set the spatio distances between nodes (CDSs) as edge features  $\mathcal{E}$ :

$$\Delta lng = lng_i - lng_j, \Delta lat = lat_i - lat_j$$
  

$$d = sin(\Delta lat/2)^2 + cos(lat_i)cos(lat_j)sin(\Delta lng/2)^2$$
  

$$\mathcal{E}_{ij} = 2dsin(\sqrt{d}) * r$$
(7)

where  $lng_i$  and  $lat_i$  are the longitude and latitude of node i, and r = 6371km is the radius of earth.

(*iii*) Sub-area extraction and time segmentation on *GG*. As shown in steps 8-23, the last function of the algorithm is to perform sub-area extraction and time extraction on the global graph. The main purpose of sub-area extraction is to reduce the dimension of node features. As stated in **ii**, the feature dimension  $\mathcal{P}$  of node *i* is equal to the number of its adjacent areas. Without sub-area extraction, the node dimension that the model needs to handle would be unacceptably large. We use the encoding prefix of Geohash to extract all  $\mathcal{P}$  adjacent areas (called an **unit**) of an area (as shown in steps 11-13). Each new bit encoded by Geohash will divide the previous area into 32 sub-areas, so we set  $\mathcal{P} = 32$ . The purpose of time extraction is to distinguish CDSs networks under different time windows, which is common in time-series prediction problems.

## 4.2 Multi-feature Based Prediction Framework

As Fig. 3 shows, the proposed Temporal-Edge Enhanced Graph Convolution Network (T-EEGCN) consists of four components: input and edge attribute normalization, Edge-Enhanced Graph Convolution Block (EEGCB) for spatio-



Fig. 3: T-EEGCN Framework

social features extraction, Long Short Time Memory (LSTM) for temporal dependency extraction, and fully connected layer for result formalization. In the flowing parts, we will explain how the T-EEGCN is adopted to realize the spatio fine-grained hotspots prediction task.

## 4.2.1 Edge-Enhanced Graph Convolution Block for Extracting Spatio-Social Features

As shown in Fig. 4, we design two graph convolution layers in EEGCB and introduce three types of learnable parameters (**with bold font**) to adjust the propagation and aggregation functions of the graph convolution, enabling EEGCB to handle dynamic graphs flexibly. The first edgeweighted convolution layer is used to update node embedding based on social relations and spatio distances. The second edge-enhanced convolution layer will further enhance the EEGCB's capability to learn the spatio features and handle dynamic graphs.

**Edge-Weighted Graph Convolution.** The topology structure of the CDSs network represents the **social relations** between users who led to these CDSs. GCN can naturally learn this structure feature through the message



Fig. 4: Edge-Enhanced Graph Convolution Block

passing and aggregation of neighbors' information of the central node [35]. However, in our problem, the node embedding vectors  $H = \{h_0, \ldots, h_n\} \in \mathbb{R}^{n \times \mathcal{P}'}$  in convolution (*n* is the number of nodes and *P'* is the feature dimension) represents the spatio distribution of CDSs. Therefore, we need to learn how the social relations will affect the associated CDSs' spatio distribution, i.e. the weight of node features influenced by neighbors' features.

Since the CDSs networks are embedded in real geographic space, a deserved way is to add spatio distance as edge feature  $A \in \mathbb{R}^{n \times n}$ . We design an edge-weighted graph convolution layer to use the **spatio distance** to adjust the influence of neighboring nodes  $\mathcal{N}(i)$  on the features of the central node *i*. The layer-wise propagation rule can be defined as follows:

$$h_i^{l+1} = \sigma(b^l + \sum_{j \in \mathcal{N}(i)} a_{ij} h_j^l W_\alpha^l) \tag{8}$$

where *l* represents the layer of the convolution,  $h_i$  is the feature of node *i*,  $a_{ij}$  is the edge weight between node *i* and *j*,  $W_{\alpha}$  is the first learnable parameter used to adjust the information aggregation rules based on the dynamic graph structures while the propagation rules are determined (i.e., weights on edges), and  $\sigma$  is an activation function.

Referring to the subsequent data analysis and study [41], we know that the social closeness will decrease with the spatio distance increase. To this end, we set the edge weight  $a_{ij}$  to the negative normalized Euclidean distance with bias:

$$a_{ij} = 1 - \operatorname{Norm}(\mathcal{E}_{ij}) = 1 - \frac{\mathcal{E}_{ij} - \min(\mathcal{E}_{i:})}{\max(\mathcal{E}_{i:}) - \min(\mathcal{E}_{i:})} \quad (9)$$

where  $\mathcal{E}$  is the input edge attribute.

Users' requests are affected by friends and the entire group. The influence will decrease with the increase of distance. The edge-weighted graph convolution will capture the two types of dependency for each node through information propagation and aggregation.

Edge-Enhanced Graph Convolution. Although exploiting the property that social closeness decreases as distance increases, the intuitive weights cannot wholly represent the relationship between CDSs. Due to user mobility and online friendships, sometimes there are *long-distance* spatio dependency between CDSs. Besides, with the structures and features of the **dynamic graphs** change, both information propagation and aggregation rules need to be adjusted to enhance the learning of the dynamic graphs.

Therefore, we design an edge-enhanced graph convolution layer to enhance the learning of **spatio features** and the flexibility of graph convolution. The edge-enhanced graph convolution layer also accepts external inputs as edge features. Here we input the normalized distance Norm( $\mathcal{E}$ ), but do not directly treat it as the edge weight. We introduce the second learnable parameter  $W_{\phi}$  to dynamically adjust the influence of neighboring node *j* at different distances on the node being updated (adjust the information propagation rules):

$$m_{ij}^{l} = W_{\phi}(h_{j}^{l} * \operatorname{Norm}(\mathcal{E}_{ij})) + b_{\phi}$$
(10)

where  $m_{ij}$  represents the update information to node *i* from node *j*,  $b_{\phi}$  are the bias of the linear layer, in this way, *m* is not only influenced by the distance between nodes but also varies with the features of neighboring nodes.

Since the central node i has already encoded the neighbors' information from the first convolution layer, we introduce the third learnable parameter  $W_{\theta}$  to preserve the learned node features  $h_i$  dynamically (adjust the information aggregation rules):

$$h_i^{\prime l} = W_\theta(h_i^l) + b_\theta \tag{11}$$

where  $b_{\theta}$  are the bias. Finally, the propagation rule for the edge-enhanced graph convolution layer can be formalized as follows:

$$h_i^{l+1} = \max_{j \in \mathcal{N}(i)} \operatorname{ReLU}(m_{ij}^l + h_i'^l)$$
(12)

where max is the max pooling. At the end of T-EEGCN, to obtain the embedding representation  $\mathcal{V}$  of the graph, we average the feature vectors of all nodes as the output, i.e.  $\mathcal{V}_t = \sum_{i \in V_t} h_i / |V_t|$ .

## 4.2.2 Long Short Term Memory for Extracting Temporal Features

The occurrence of CDSs is often affected by users' historical behavior, e.g., watching regularly updated videos and downloading games during holidays. This temporal dependency leads to the correlations of CDSs networks under different time windows, and GCN cannot capture this distinction and connection.

Currently, the most widely used neural network model for temporal dependency extraction is the Recurrent Neural Network (RNN). However, due to defects such as gradient disappearance and explosion, the traditional RNN has limitations for long-term prediction. As a variant of RNN, Long Short Term Memory (LSTM) [42] has been proven to solve the above problems through the gates function.

Here, we adopt LSTM to capture the **temporal dependency** between CDSs networks at different time windows. The input of LSTM is a sequential graph embedding vector  $\mathcal{V} = \{\mathcal{V}_{\kappa-T+1}, \ldots, \mathcal{V}_{\kappa}\}$ , the compute process for each layer can be described as follows:

$$f_{t} = \sigma(W_{f} \cdot [h_{t-1}, x_{t}] + b_{f})$$

$$i_{t} = \sigma(W_{i} \cdot [h_{t-1}, x_{t}] + b_{i})$$

$$C_{t} = f_{t} * C_{t-1} + i_{t} * tanh(W_{C} \cdot [h_{t-1}, x_{t}] + b_{C})$$

$$o_{t} = \sigma(W_{o} \cdot [h_{t-1}, x_{t}] + b_{o})$$

$$h_{t} = o_{t} * tanh(C_{t})$$
(13)

where *t* is the time step in terms of the length of time windows,  $h_t, c_t, x_t$  are the hidden state, cell state, and input at time *t* ( $x_t = V_{g_t}$ ), and  $f_t, i_t, o_t$  are respectively the forget gate, input gate and output gate.

## 4.2.3 Calculation Process and Result Formalization

The calculation process of T-EEGCN can be seen in Fig. 3. The graph topology, node features, and edge features will be input into the Edge-Enhanced Graph Convolution Block to produce a sequence of graph representation vectors  $\boldsymbol{\mathcal{V}}$ .

After that,  $\mathcal{V}$  will be input into the LSTM for temporal dependency extraction (equation.17), and finally the output of LSTM will be passed to FC layer and sigmoid function (equation.18) to generate the hotspots matrix  $\hat{M}$ , which can be formalized as follows:

$$X_t' = \sigma(AX_t W_\alpha) \tag{14}$$

$$\mathcal{V}_t = mean(\sigma(W_\theta X_t' W_\phi)) \tag{15}$$

$$\boldsymbol{\mathcal{V}} = \{\mathcal{V}_{\kappa-T+1}, ..., \mathcal{V}_{\kappa}\}$$
(16)

$$O = \mathcal{L}(\boldsymbol{\mathcal{V}}, W_L) \tag{17}$$

$$\hat{M} = \text{sigmoid}(WO_i + b) \tag{18}$$

where A is the edge weight matrix,  $W_L$  represents the trainable parameters of the LSTM, W and b represent the parameter matrix and bias of FC, respectively. We use the BCELoss to measure the performance of our model which is defined as follows,

$$L(\hat{M}, M) = -W[Mlog(\hat{M}) + (1 - M)log(1 - \hat{M})]$$
(19)

where W is the weight matrix.

We summarize the main characteristics of T-EEGCN in the following,

- T-EEGCN is a universal framework for processing dynamic graphs. It can not only tackle CDSs networks modeling and prediction issues but also be applied to more general spatio-temporal sequence learning tasks.
- We introduce three learnable parameters in EEGCB and guide the update of the parameters based on the dynamic graph structures and edge features. The design allows EEGCB to adjust the node embedding update rules based on social and spatio features. EEGCB does not depend on the structure information of a specific graph and can be applied to dynamic CDSs networks coding.
- The output layer of EEGCB is tightly connected to the input layer of LSTM. While the hybrid model is being trained, the parameters of EEGCB and LSTM can be jointly updated. Thus, the whole framework can be trained end-to-end.

#### 4.2.4 Computational Complexity

In the analysis of computational complexity, we assume that the number of hidden units in EEGCB is constant. The size of the time window T is a fixed number during the training and evaluation process. These parameters have no relationship with the input scale, so we don't consider them. Then the computational complexity of T-EEGCB can be represented as  $O(|V| \cdot |E| \cdot (\mathcal{P} + |\mathcal{E}|))$ , where:

- |V| represents the number of nodes in every CDSs network \$\mathcal{G}\_{j}^{i}\$. Since we want to aggregate the nodes in a CDSs network, we should traverse every node in \$\mathcal{G}\_{j}^{i}\$.
- |*E*| represent the number of edges in every CDSs network *G*<sup>i</sup><sub>j</sub>. For every node in *V*, we use node sampling to get the node sets. In this study, we sample



Fig. 5: Spatio-Temporal-Social Dependencies Analysis

the K = 2 depth of neighbor nodes. However, the K = 2 neighbor nodes are uncertain but there exists an upper bound |E| of the neighbor nodes, in which |E| represents the number of edges.

P + |ε| represents the cost of aggregation and edge enhance, in which P represents the dimension of node features and |ε| represents the dimension of edge features. Since aggregating is applied to node features and edge enhance is applied to edge features, and only simple operations in the aggregation and edge enhance the process, the complexity of the two processes can be denoted O(P) and O(|ε|) respectively.

## 5 EXPERIMENTS

## 5.1 Dataset Description

We use two large-scale real-world mobile content datasets: point-to-point (P2P) file transmission records **OPPST** and online places check-in sharing data **Gowalla**<sup>2</sup> in this paper.

**OPPST** comes from the log files of a P2P file sharing app, whose transferred contents include not only files and pictures but also videos and apps. The trace is 843GB, including 30485335 users and 443440043 sharing records, ranging from 08/01/2016 to 10/30/2016.

**Gowalla** is a location-based social network created in 2009. The users check in at places through their mobile devices. Check-ins are shared with friends, and the dataset contains a total of 6,442,890 check-ins of users over the period of Feb. 2009 - Oct. 2010.

#### 5.2 Data Analysis

As shown in Fig. 5, we examined the spatio correlation, social distance distribution, and temporal dependency of

DataSets	Cate.	Nodes	Edges
OPPST	Game(OPP1)	1,791,310	2,133,529
	Tools(OPP2)	1,072,962	1,306,140
	Video(OPP3)	1,521,456	1,746,311
	Comm.(OPP4)	1,305,664	1,472,994
Gowalla	Coffee(Gow1)	3,999,797	10,614,556
	Sand.(Gow2)	1,927,685	4,341,135
	Theatre(Gow3)	917,141	1,393,035

CDSs in both datasets. Our main observations are summarized as follows.

**Observation 1: The distribution of CDSs are spatio correlated.** Fig. 5(a) shows the probability distribution of the Pearson correlation coefficient of CDSs between different areas (within 76m). It can be seen that the correlation coefficients of the OPPST are mainly concentrated in 0.1, showing a certain non-zero correlation in the spatio domain. On the other hand, the correlation coefficients of Gowalla are higher, indicating that its spatio features have a more substantial influence on CDSs.

**Observation 2: The distribution of CDSs is intensely temporal dependent.** Fig. 5(c) shows that the CDSs of both datasets have a solid seasonal periodic feature within a year, especially Gowalla, which has a significant cyclical variation, which proves that the distribution of CDSs has a solid temporal dependency.

**Observation 3: Social dependency decreases with distance.** It can be seen from Fig. 5(b) that the social distances of users among most CDSs are short (<1000km), which is consistent with our setting in Section 4.2.1. However, there are also long-distance social relationships that exceed 10000km, so dynamic adjustment of the edge features is needed. Since OPPST is a D2D-type transmission dataset, its social distance is much smaller than that of the Gowalla, which is shared online.

The above analysis demonstrates the spatio-temporalsocial dependencies of CDSs and validates the T-EEGCN design's rationality from the perspective of data regularity.

#### 5.3 Experimental Settings

We set the number of periods for training to 9 (T = 9) and the number for prediction to 1. Thus, the length of each time window is seven days in OPPST and thirty days in Gowalla.

**Chosen Categories.** Different content categories of CDSs significantly affect their spatio-temporal distribution patterns [9]. To better focus on the effects of spatio-temporal-social dependency, we select four major categories of app files under the OPPST and three categories of most frequent check-in places under the Gowalla for experiments.

**Networks and Spatio Information.** Table 3 illustrates the total number of nodes and edges of the spatio-temporal graphs utilized in the experiment, which are sufficient to support the training of end-to-end models. Table 4 illustrates the number of units (refer to §4.1 #iii), the total number of areas, the number of known areas, and the percentage of hotspots in the time window to be predicted, respectively.

**Evaluation Metric & Baselines.** We adopted Precision (Prec), Recall, F1, and Area Under the ROC Curve (AUC) to evaluate the performance. The baselines are as follows:

TABLE 4: Spatio Information

Cate.	Units	Areas	Known Areas	Hot Rate( $\tau$ )
OPP1	5,756	184,224	39,279	9.1%(3)
OPP2	1,926	61,632	9,864	8.5%(5)
OPP3	4,255	136,160	26,748	10.4%(5)
OPP4	2,205	70,560	13,974	10.0%(3)
Gow1	10,285	329,120	46,486	14.6%(3)
Gow2	5 <i>,</i> 519	176,608	24,939	9.1%(3)
Gow3	12,149	388,768	56,188	6.6%(5)

- SVM [43]: Support Vector Machine method, widely used in classification tasks.
- DeepInf [38] A GCN/GAT-based framework for learning social relations of users' social networks. Here we use its GCN form to capture the social features of the CDSs networks.
- D2D-LSTM [9] A LSTM-based recurrent network takes multiple dimensions into account, including time, geography, and category.
- T-GCN [44] A temporal GCN for static road network traffic prediction, which is combined with the GCN and the gated recurrent unit (GRU).
- Macro Temporal GCN (MT-GCN) [45] A temporal GCN for social influence prediction, which can encode the dynamic social networks and capture the temporal feature.

## 5.4 Evaluation

**Fine-grained hotspots prediction accuracy.** As shown in Table 5 and Table 6<sup>3</sup>, T-EEGCN performs well in fine-grained hotspots prediction for four evaluation metrics. Specifically, it performs better than the five baselines on the different categories and the average values of the two datasets. Numerically, it significantly outperforms other models at the 0.01 level. T-EEGCN performs very well on Gowalla for *long-term* prediction, achieving an average F1 of up to 0.86 and an average AUC of 0.91.

Unlike T-EEGCN, DeepInf cannot handle dynamic graphs. It deals with the network of all CDSs in the first nine periods and computes the graph structure feature once to update the final node embedding. Besides, DeepInf does not exploit edge features to adjust the update rules. The drawbacks of these two mechanisms make DeepInf inappropriate for extracting spatio-temporal features, leading to its poor performance (0.49/0.24 for average F1/AUC in OPPST and 0.11/0.07 for average F1/AUC in Gowalla compared to T-EEGCN). It indicates that the single social feature has an insufficient influence on predicting CDSs hotspots. The real-world spatio-temporal features need to be considered.

D2D-LSTM achieves sub-optimal performance in most content categories, which indicates that CDSs have a strong temporal dependency, i.e., users are more likely to request mobile content in areas where they have historically performed requests. T-EEGCN does not outperform D2D-LSTM in Prec and AUC in Gow1. It is because Gow1 is the LBSN of category **Coffee**. As analyzed in Fig. 5(c), CDSs under Gowalla has a strong temporal periodicity, so the model's ability to extract temporal features plays a greater role (at which D2D-LSTM is good). However, T-EEGCN outperforms D2D-LSTM in all other datasets, especially in Gow3 (Theatre), T-EEGCN exceeded D2D-LSTM by 0.06, 0.13, 0.11, and 0.06 for the four metrics. It is because CDSs about Theatre are less affected by temporal features compared to daily communication behaviors, e.g., checking in at Coffee (Gow1) and Sandwich shop (Gow2).

Compared with D2D-LSTM, T-EEGCN can capture spatio-social features through the encoding of CDSs networks by EEGCB. Although D2D-LSTM introduces the influence of spatio features by embedding geographic location, it does not consider the **social relations** between users, making its prediction ability weaker than T-EEGCN (0.14/0.10 for average F1/AUC in OPPST and 0.05/0.03 for average F1/AUC in Gowalla).

The metrics show that MT-GCN outperforms T-GCN, it is because the GCN of T-GCN is applied to the static graph, which computes the graph structure once and updates the node embedding according to the fixed graph structure. However, the GCN of MT-GCN is trained to handle dynamic social networks. It adjusts node embedding rules according to different graph structures—the above results prove the necessity for dynamic graph learning.

Although both MT-GCN and T-EEGCN are combinations of GCNs and recurrent neural networks, compared to MT-GCN, T-EEGCN has significantly improved the prediction performance on all datasets (0.25/0.21 for average F1/AUC in OPPST and 0.25/0.16 for average F1/AUC in Gowalla). Compared to the GCN of MT-GCN, the EEGCB of T-EEGCN can update node embedding by utilizing both dynamic graph structures and edge features, enabling EEGCB to capture fine-grained spatio dependency features, thus encoding CDSs networks more accurately.

Generalizability analysis for unknown areas prediction. From Table 4 we can calculate the percentage of known areas in different datasets, where the average percentage of known areas in OPPST accounts for 19%, while Gowalla is 14%. In addition, due to the truncation of the hotspots threshold  $\tau$ , the Hot Rate under the last time window will be smaller than the percentage of known areas (9.5% for OPPST and 10.1% for Gowalla).

According to our setting  $\mathcal{P} = 32$ , the percentage of areas with CDSs (known areas) should be about 3.13%. However, the actual percentage of known areas and the Hot Rate for each dataset are much larger than 3.13%. This phenomenon verifies that part of the areas adjacent to known areas is showing up with CDSs, even up to the size of the number of hotspots threshold.

We argue that this phenomenon occurs mainly due to the migration of CDSs caused by user mobility, as stated in *challenge ii*. The large number of unknown areas poses a challenge to the spatio generalization of the model.

By combining the performance of OPP1-OPP4/Gow1-Gow3 and the corresponding percentage of unknown areas, we find that the prediction of unknown areas increases the problem's difficulty. Specifically, the higher the percentage of unknown areas in the dataset, the lower the prediction accuracy of the models tends to be. For example, OPP2 has 84% of unknown areas, the highest percentage among the four data categories, and correspondingly all models

<sup>3.</sup> The bold in Tables 5 and 6 means the optimum value in a column, and underlining represents the suboptimal value, and we compare the 0.001 level when the two values are equal.

Model	OPPST(OPP1/OPP2/OPP3/OPP4)						
Widdei	Prec	Recall	F1	AUC			
SVM	0.35 / 0.31 / 0.31 / 0.36	0.40 / 0.35 / 0.40 / 0.39	0.37 / 0.33 / 0.35 / 0.37	0.63 / 0.61 / 0.61 / 0.63			
DeepInf	0.42 / 0.33 / 0.44 / 0.38	0.30 / 0.23 / 0.24 / 0.19	0.35 / 0.27 / 0.31 / 0.26	0.63 / 0.59 / 0.59 / 0.58			
D2D-LSTM	<u>0.59</u> / <u>0.61</u> / 0.56 / <u>0.61</u>	<u>0.52</u> / <u>0.55</u> / <u>0.48</u> / <u>0.55</u>	<u>0.56</u> / <u>0.58</u> / <u>0.52</u> / <u>0.58</u>	<u>0.74</u> / <u>0.76</u> / <u>0.72</u> / <u>0.75</u>			
T-GCN	0.51 / 0.55 / 0.55 / 0.51	0.37 / 0.32 / 0.28 / 0.43	0.43 / 0.40 / 0.37 / 0.47	0.63 / 0.62 / 0.62 / 0.64			
MT-GCN	0.54 / 0.59 / <u>0.57</u> / 0.56	0.39 / 0.35 / 0.29 / 0.43	0.46 / 0.44 / 0.38 / 0.49	0.63 / 0.62 / 0.63 / 0.65			
<b>T-EEGCN</b>	0.76 / 0.76 / 0.76 / 0.76	0.64 / 0.63 / 0.64 / 0.65	0.70 / 0.68 / 0.69 / 0.70	0.84 / 0.85 / 0.84 / 0.84			
Model	Gowalla(Gow1/Gow2/Gow3)						
Widdei	Prec	Recall	F1	AUC			
SVM	0.75 / 0.76 / 0.77	0.46 / 0.46 / 0.50	0.57 / 0.58 / 0.61	0.62 / 0.62 / 0.65			
DeepInf	0.87 / 0.78 / 0.82	0.72 / 0.70 / 0.65	0.79 / 0.74 / 0.72	0.85 / 0.84 / 0.82			
D2D-LSTM	<b>0.95</b> / 0.87 / 0.84	<u>0.80</u> / <u>0.74</u> / <u>0.69</u>	0.87 / 0.80 / 0.75	<b>0.91</b> / <u>0.86</u> / <u>0.85</u>			
T-GCN	0.81 / 0.874 / 0.88	0.43 / 0.43 / 0.53	0.56 / 0.58 / 0.66	0.75 / 0.71 / 0.76			
MT-GCN	0.82 / <u>0.88</u> / <u>0.89</u>	0.44 / 0.44 / 0.54	0.57 / 0.59 / 0.67	0.76 / 0.72 / 0.77			
T-EEGCN	<u>0.94</u> / 0.88 / 0.90	0.82 / 0.80 / 0.82	0.87 / 0.84 / 0.86	<u>0.91</u> / <b>0.91</b> / <b>0.91</b>			

TABLE 6: Average Performance Comparison

Model	OPPST_AVG.				
Widdei	Prec	Recall	F1	AUC	
DeepInf	0.39	0.24	0.30	0.60	
D2D-LSTM	0.59	0.53	0.56	0.74	
T-GCN	0.53	0.35	0.42	0.63	
MT-GCN	0.56	0.37	0.44	0.63	
T-EEGCN	0.76	0.64	0.69	0.84	
Madal	Gowalla_AVG.				
Wouei	Prec	Recall	F1	AUC	
DeepInf	0.82	0.69	0.75	0.84	
D2D-LSTM	0.89	0.74	0.81	0.88	
T-GCN	0.85	0.46	0.60	0.74	
MT-GCN	0.86	0.47	0.61	0.75	
T-EEGCN	0.91	0.81	0.86	0.91	

perform worse on this dataset (except D2D-LSTM and our T-EEGCN), since D2D-LSTM is the only model that does not consider user CDSs networks, unknown areas due to user mobility have no effect on its performance.

Table 5 and Table 6 show that for OPPST, T-EEGCN successfully predicts 76% of the hotspots and recalls 65% of the hotspots despite the high percentage of unknown areas of 80%. As for Gowalla (unknown areas of 86%), TEEGCN successfully predicts 91% of the hotspots and recalls 81% of the hotspots, demonstrating the strong generalization of T-EEGCN toward unknown areas. This is due to the proposed EEGCB, which learns user spatio location features, while the incorporated edge weights capture the impact of user mobility on CDSs by adapting the node update process through the spatio distance between CDSs.

**Content Categories Robustness Analysis.** We calculate the models prediction AUC STDEV (standard deviation, *the smaller, the better*) for different categories of contents. As seen in Fig. 6, the AUC STDEV of T-EEGCN is much lower than those of baselines in both datasets. The lowest AUC STDEV demonstrates the robustness of T-EEGCN on different categories of content, indicating that it can obtain stable results for different contents, which is not available for other models focusing on limited features.



Fig. 6: AUC STDEV of Different Approaches on the Datasets OPPST and Gowalla.

#### 5.5 Multi-feature Extraction Performance Analysis

In this section, we perform ablation studies of T-EEGCN based on the AUC performance on both datasets as a demonstration to illustrate the role of different components and features in implementing T-EEGCN. Table 7 demonstrates the multiple ablation variants of T-EEGCN in which various components and features are deactivated.

TABLE 7: Feature extraction for different models

Mode	Dynamic	E.W.	E.E.	LSTM	Social	Spatio	Temporal
T-EEGCN#1	-	~	~	-	√		-
T-EEGCN#2	$\checkmark$	-	$\checkmark$	$\checkmark$	$\checkmark$	-	$\checkmark$
T-EEGCN#3	weak	$\checkmark$	-	$\checkmark$	$\checkmark$	weak	$\checkmark$
T-EEGCN	enhanced	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	enhanced	$\checkmark$

**AUC performance analysis of ablation models.** Fig. 7 shows the AUC performance of ablation models and our T-EEGCN.

• The starting points of the curves in Fig. 7 shows the AUC of **T-EEGCN#1** (with static graph and deactivate LSTM). It can be seen that #1 performs the worst among all models. Since #1 updates the node embedding based on the static graph structure, the prediction performance on unknown nodes/graphs is not guaranteed. Besides, #1 lacks the mechanism to capture temporal features effectively. In contrast, T-EEGCN has a significant increase in performance,



Fig. 7: AUC Performance of Ablation Models and T-EEGCN.

indicating that the learning of the **dynamic graph** structures (adjusting the three learnable parameters based on the dynamic graph structures) and LSTM are key mechanisms to extract **temporal features**.

- **T-EEGCN#2** deactivate **Edge Weighted** from T-EEGCN. Therefore it only considers the social and temporal features and ignores the effect of spatio features on CDSs. The improvement of T-EEGCN compared to #2 illustrates the necessity to consider **spatio features.**
- **T-EEGCN#3 (deactivate Edge Enhanced learning)** is the best among the three ablation models. However, its performance is still worse than T-EEGCN, indicating that exploiting spatio distance as negative weights to adjust node information propagation is not always effective. The additional learnable parameters  $W_{\phi}$  and  $W_{\theta}$  introduced by T-EEGCN enhance the model's capability to extract spatio features and handle dynamic graphs.

The ablation studies demonstrate that spatio-temporal and social features play a key role in the prediction (since we cannot deactivate the graph convolution from the T-EEGCN, we analyze the importance of social features by comparing T-EEGCN with D2D-LSTM in Section 5.4). In terms of model performance, we can infer that in both OPPST and Gowalla services, the relationship between the influence of features on the prediction model is **temporal** > **spatio** > **social features**. The finding makes sense because our problem is typically a spatio-temporal prediction problem.

However, we cannot deny the importance of social features. Firstly, T-EEGCN significantly improves prediction performance compared with the D2D-LSTM model that only considers spatio-temporal features (25%/12% for average F1/AUC in OPPST and 6%/4% for average F1/AUC in Gowalla). Second, the extraction of spatio features by T-EEGCN is built on the extraction of social features. EEGCB trains the learnable parameters and adjusts the node embedding rules by fusing spatio distance and dynamic graph structures. Hence, the structures of CDSs networks (i.e., social features) is indispensable.

# 5.6 Temporal Scalability Analysis

Besides the optimization of existing APs (content caching or service orchestration), the growing CDSs often require the frequent deployment of new APs to cope with the tremendous data traffic. Optimization for existing APs is aimed at



Fig. 8: Comparison with Baselines on OPPST Varying the Time Scales



Fig. 9: Comparison with Baselines on Gowalla Varying the Time Scales

dissolving the congestion in a short period. However, deploying new APs must consider the potential CDSs for multiple days or even months, requiring network intelligence with strong temporal scalability. Therefore, we perform the temporal scalability test to evaluate the effectiveness of various approaches in handling long-short term predictions.

Fig. 8 and Fig. 9 show the performances with various prediction time windows sizes T, and time scales (days in one time window). Specifically, different experiment settings are designed for performance tests on the prediction of CDSs hotspots within one day (T = 6), one week (T = 3), and one month (T = 2). Note that to avoid the bias of categories on the model, we use all categories under the corresponding dataset in the test.

From Fig. 8 and Fig. 9, we can conclude that:

(i) As the prediction time scale becomes larger and the time window size becomes shortened, the prediction accuracy of each model decreases. One reason is that as the time scale becomes larger, the number of CDSs also becomes larger, and the CDSs networks become more complex, which will increase the difficulty of GCN encoding. The other reason is that when the time window size decreases, fewer historical features are available to be captured by the LSTM.

(ii) Although the model suffers from two limitations, T-EEGCN achieves the highest prediction accuracy at all time scales, which proves the effectiveness of the proposed multifeature capture. The above temporal scalability tests demonstrate that the proposed network intelligence framework can effectively predict the CDSs hotspots for long-short periods, which helps to optimize various services of the 5G UDNs.

(iii) The proposed approach outperforms baselines and stat-of-the-art on day-granularity predictions, while the model update time on the central cloud is much less than one day. Since we performed experiments on the server with 16 CPUs, 256GB RAM,  $2 \times NVIDIA$  GeForce RTX 3090,



Fig. 10: CDSs heatmaps of ground truth and different models' prediction during Sep. 2010 - Oct. 2010

and the training/inference time of the prediction model with day granularity was 2h30min/97ms on average. Thus the proposal guarantees instantaneous at the smallest time granularity currently considered.

#### 5.7 Case Study

We choose the CDSs distribution for the last 30 days (September 2010 to October 2010) in the range of 1216m\*1216m (16\*16 areas) in New York under Gowalla to plot the CDSs hotspots heat map. Specifically, we calculate the ground truth and the prediction results to form a ground-truth matrix and three prediction matrices (in the form of a probability matrix). The ground truth matrix only has values 0: non-hotspot and 1: hotspot, while the continuous value of the probability matrix represents the predicted hotspot probability.

The results are reported in Fig. 10. In the heatmap, blue represents the hotspot's negative instance, while white represents the positive instance. Moreover, the color depth indicates the hotspot probability, in which the color close to dark blue indicates low probability while that close to white means high likelihood.

As illustrated in Fig. 10, the two baselines are deficient in terms of prediction accuracy (predicted hotspots distribution does not match the real distribution) and hotspots discovery capability (low confidence in discovering true hotspots with shallow whites). In contrast, our framework can provide predictions consistent with the distribution of real CDSs and possesses a very high confidence level for true hotspots, which are sufficient for guiding real-world decisions.

# 6 CONCLUSION

We study the spatio fine-grained and generalized CDSs hotspots prediction in 5G UDNs. We first adopt the Geohash encoding and propose a spatio-temporal subgraph sampling

algorithm for fine-grained and arbitrary areas CDSs modeling. Then, we propose a novel multi-feature extraction framework T-EEGCN, in which we propose a novel edgeenhanced graph convolution block to accurately encode dynamic CDSs networks by considering spatio-social features. Besides, we adopt the LSTM to capture the temporal dependency. The analysis and experiments on two realworld CDSs datasets verify the influence of spatio-temporalsocial features and the effectiveness of T-EEGCN.

## ACKNOWLEDGEMENT

This work is partially supported by the National Key Research and Development Program of China (Grant No. 2019YFB2101901), the National Science Foundation of China (Grant No.62072332), and Australian Research Council Linkage Project (LP180100750).

#### REFERENCES

- X. Chen, Q. Xiang, L. Kong, H. Xu, and X. Liu, "Learning from fm communications: Toward accurate, efficient, all-terrain vehicle localization," *IEEE/ACM Transactions on Networking*, pp. 1–16, 2022.
- [2] F.-H. Tseng, Y.-M. Jheng, L.-D. Chou, H.-C. Chao, and V. C. Leung, "Link-aware virtual machine placement for cloud services based on service-oriented architecture," *IEEE Transactions on Cloud Computing*, vol. 8, no. 4, pp. 989–1002, 2020.
  [3] J. Qiao, Y. He, and X. S. Shen, "Proactive caching for mobile video
- [3] J. Qiao, Y. He, and X. S. Shen, "Proactive caching for mobile video streaming in millimeter wave 5g networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 7187–7198, 2016.
- [4] J. Park, S. Samarakoon, M. Bennis, and M. Debbah, "Wireless network intelligence at the edge," *Proceedings of the IEEE*, vol. 107, no. 11, pp. 2204–2239, 2019.
- [5] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," IEEE Communications Surveys Tutorials, vol. 18, no. 4, 2016.
- [6] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2018.
- [7] 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*. Association for Computing Machinery, 2018.
  [8] E. Mededovic, V. G. Douros, and P. Mähönen, "Node centrality
- [8] E. Mededovic, V. G. Douros, and P. Mähönen, "Node centrality metrics for hotspots analysis in telecom big data," in *INFOCOM WKSHPS*, 2019.
- [9] H. Zhang, X. Wang, J. Chen, C. Wang, and J. Li, "D2d-lstm: Lstmbased path prediction of content diffusion tree in device-to-device social networks," AAAI, vol. 34, 2020.
- [10] H. Wang, J. Ding, Y. Li, P. Hui, J. Yuan, and D. Jin, "Characterizing the spatio-temporal inhomogeneity of mobile traffic in large-scale cellular data networks," in *Proceedings of the 7th International Workshop on Hot Topics in Planet-scale mObile computing and online Social neTworking*, 2015, pp. 19–24.
  [11] X. Wang, Z. Zhou, Z. Yang, Y. Liu, and C. Peng, "Spatio-temporal
- [11] X. Wang, Z. Zhou, Z. Yang, Y. Liu, and C. Peng, "Spatio-temporal analysis and prediction of cellular traffic in metropolis," in 2017 IEEE 25th International Conference on Network Protocols (ICNP), 2017.
- [12] D. Bega, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, "Deepcog: Cognitive network management in sliced 5g networks with deep learning," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019.
- [13] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, and D. Yang, "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in *IEEE INFOCOM* 2017 - *IEEE Conference on Computer Communications*, 2017.
- [14] C. Huang, C. Chiang, and Q. Li, "A study of deep learning networks on mobile traffic forecasting," in 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017.
- [15] R. Li, Z. Zhao, J. Zheng, C. Mei, Y. Cai, and H. Zhang, "The learning and prediction of application-level traffic data in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, 2017.

#### IEEE TRANSACTIONS ON MOBILE COMPUTING

- [16] Y. Koren, S. Rendle, and R. Bell, "Advances in collaborative filtering," *Recommender systems handbook*, pp. 91–142, 2022.
- [17] Y. Lu, Y. He, Y. Cai, Z. Peng, and Y. Tang, "Time-aware neural collaborative filtering with multi-dimensional features on academic paper recommendation," in 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2021, pp. 1052–1057.
- [18] J. Zeng, H. Tang, Y. Zhao, M. Gao, and J. Wen, "Pr-rcuc: A poi recommendation model using region-based collaborative filtering and user-based mobile context," *Mob. Networks Appl.*, vol. 26, pp. 2434–2444, 2021.
- [19] W. W. Wei, "Time series analysis," in The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2, 2006.
- [20] M. Abbasi, A. Shahraki, and A. Taherkordi, "Deep learning for network traffic monitoring and analysis (ntma): A survey," *Computer Communications*, vol. 170, pp. 19–41, 2021.
- [21] A. Azari, P. Papapetrou, S. Denic, and G. Peters, "Cellular traffic prediction and classification: A comparative evaluation of lstm and arima," in *Discovery Science: 22nd International Conference*, *DS 2019, Split, Croatia, October 28–30, 2019, Proceedings.* Berlin, Heidelberg: Springer-Verlag, 2019, p. 129–144.
- [22] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on crossdomain big data," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1389–1401, 2019.
- [23] X. Zhou, Y. Zhang, Z. Li, X. Wang, J. Zhao, and Z. Zhang, "Largescale cellular traffic prediction based on graph convolutional networks with transfer learning," *Neural Computing and Applications*, vol. 34, no. 7, pp. 5549–5559, 2022.
- [24] 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*, vol. 21, no. 4, pp. 1244–1256, 2022.
- [25] C. Zhang, S. Dang, B. Shihada, and M.-S. Alouini, "Dual attentionbased federated learning for wireless traffic prediction," in *IEEE INFOCOM 2021-IEEE conference on computer communications*. *IEEE*, 2021, pp. 1–10.
- [26] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling internet traffic dynamics of cellular devices," ACM SIGMETRICS Performance Evaluation Review, vol. 39, no. 1, pp. 265– 276, 2011.
- [27] D. Lee, S. Zhou, X. Zhong, Z. Niu, X. Zhou, and H. Zhang, "Spatial modeling of the traffic density in cellular networks," *IEEE Wireless Communications*, vol. 21, no. 1, pp. 80–88, 2014.
- [28] Y. Shu, M. Yu, J. Liu, and O. Yang, "Wireless traffic modeling and prediction using seasonal arima models," in *IEEE International Conference on Communications*, 2003. ICC '03., vol. 3, 2003, pp. 1675– 1679 vol.3.
- [29] J.-H. Duan, W. Li, X. Zhang, and S. Lu, "Forecasting finegrained city-scale cellular traffic with sparse crowdsourced measurements," *Computer Networks*, vol. 214, p. 109156, 2022.
- [30] E. Tuna and A. Soysal, "Multivariate spatio-temporal cellular traffic prediction with handover based clustering," in 2022 56th Annual Conference on Information Sciences and Systems (CISS), 2022, pp. 55–59.
- [31] 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*, vol. 20, no. 12, pp. 3453–3466, 2021.
- [32] J. Chu, X. Wang, K. Qian, L. Yao, F. Xiao, J. Li, and Z. Yang, "Passenger demand prediction with cellular footprints," *IEEE Transactions on Mobile Computing*, vol. 21, no. 1, pp. 252–263, 2022.
- [33] Q. Hu, S. Wang, X. Cheng, J. Zhang, and W. Lv, "Cost-efficient mobile crowdsensing with spatial-temporal awareness," *IEEE Transactions on Mobile Computing*, vol. 20, no. 3, pp. 928–938, 2021.
- [34] H. Li, F. Lin, X. Lu, C. Xu, G. Huang, J. Zhang, Q. Mei, and X. Liu, "Systematic analysis of fine-grained mobility prediction with ondevice contextual data," *IEEE Transactions on Mobile Computing*, vol. 21, no. 3, pp. 1096–1109, 2022.
- [35] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017.
- [36] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *IJCAI*, 2018.
- [37] R. Huang, C. Huang, Y. Liu, G. Dai, and W. Kong, "Lsgcn: Long short-term traffic prediction with graph convolutional networks," in *Proceedings of the Twenty-Ninth International Joint Conference on*

Artificial Intelligence, IJCAI-20. International Joint Conferences on Artificial Intelligence Organization, 7 2020.

- [38] J. Qiu, J. Tang, H. Ma, Y. Dong, K. Wang, and J. Tang, "Deepinf: Social influence prediction with deep learning," in ACM SIGKDD, 2018.
- [39] Y. Xiao, C. Song, and Y. Liu, "Social hotspot propagation dynamics model based on multidimensional attributes and evolutionary games," *Communications in Nonlinear Science and Numerical Simulation*, vol. 67, no. FEB., pp. 13–25, 2019.
- [40] C. H. Liu, J. Xu, J. Tang, and J. Crowcroft, "Social-Aware Sequential Modeling of User Interests: A Deep Learning Approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 11, pp. 2200–2212, 2019.
- [41] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, "Socio-spatial properties of online location-based social networks." *ICWSM*, vol. 11, pp. 329–336, 2011.
- [42] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 07 2019. [Online]. Available: https://doi.org/10.1162/neco\_a\_01199
- [43] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, 1999.
- [44] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2020.
- [45] Y. Liu, X. Shi, L. Pierce, and X. Ren, "Characterizing and forecasting user engagement with in-app action graph: A case study of snapchat," in ACM SIGKDD, 2019.



Shaoyuan Huang Shaoyuan Huang received a B.S. degree from Tianjin University, Tianjin, China, in 2020. He is currently pursuing an M.S. degree in the College of Intelligence and Computing, Tianjin University. His current research interests include spatial-temporal prediction, recommend systems and graph convolution networks.



Heng Zhang Heng Zhang is currently pursuing a PhD degree from the School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, China. His current research interests include D2D content propagation, recommend system and edge computing



Xiaofei Wang [S'06, M'13, SM'18] (Senior Member, IEEE) received the B.S. degree from Huazhong University of Science and Technology, China, and received M.S. and Ph.D. degrees from Seoul National University, Seoul, South Korea. He was a Postdoctoral Fellow with The University of British Columbia, Vancouver, Canada, from 2014 to 2016. He is currently a Professor with the Tianjin Key Laboratory of Advanced Networking, College of Intelligence and Computing, Tianjin University, Tianjin, China.

Focusing on the research of edge computing, edge intelligence, and edge systems, he has published more than 150 technical papers in IEEE JSAC, TCC, ToN, TWC, IoTJ, COMST, TMM, INFOCOM, ICDCS and so on. In 2017, he was the recipient of the "IEEE ComSoc Fred W. Ellersick Prize", and in 2022, he received the "IEEE ComSoc Asia-Pacific Outstanding Paper Award".



**Min Chen** (Fellow, IEEE) has been a Full Professor with the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST), since February 2012. He is the Director of the Embedded and Pervasive Computing Laboratory and the Director of the Data Engineering Institute, HUST. Before he joined HUST, he was an Assistant Professor with the School of Computer Science and Engineering, Seoul National University. His Google Scholar citations reached over 34,450

with an H-index of 89. His top paper was cited 3,825 times. He was selected as a Highly Cited Researcher from 2018 to 2021. He received the IEEE Communications Society Fred W. Ellersick Prize in 2017 and the IEEE Jack Neubauer Memorial Award in 2019. He is the Chair of the IEEE Globecom 2022 eHealth Symposium. He is the Founding Chair of the IEEE Computer Society Special Technical Communities on Big Data.



Jianxin Li received a Ph.D. degree in computer science from the Swinburne University of Technology, Melbourne, VIC, Australia, in 2009. He is an Associate Professor with the School of Information Technology, Deakin University, Burwood, VIC, Australia. His current research interests include database query processing and optimization, social network analytics, and traffic network data processing.



Victor C. M. Leung (Life Fellow, IEEE) is a Distinguished Professor of Computer Science and Software Engineering with Shenzhen University, Shenzhen, China. He is also an Emeritus Professor of Electrical and Computer Engineering and the Director of the Laboratory for Wireless Networks and Mobile Systems, University of British Columbia, Vancouver, BC, Canada. His research is in the broad areas of wireless networks and mobile systems. Dr. Leung received the IEEE Vancouver Section Centennial

Award, the 2011 UBC Killam Research Prize, the 2017 Canadian Award for Telecommunications Research, and the 2018 IEEETCGCC Distinguished Technical Achievement Recognition Award.He is named in the current Clarivate Analytics list of Highly Cited Researchers. He is serving on the editorial boards of IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE NETWORK, and several other journals. He coauthored papers that won the 2017 IEEE ComSoc Fred W. Ellersick Prize, the 2017 IEEE Systems JournalBest Paper Award, the 2018 IEEE CSIM Best Journal Paper Award, and the 2019 IEEE TCGCC Best Journal Paper Award.He is a Fellow of the Royal Society of Canada, Canadian Academy of Engineering, and Engineering Institute of Canada.