Accepted: 08 May 2024
Revised: 23 April 2024
Received: 30 November 2022

RESEARCH-ARTICLE

HDM-GNN: A Heterogeneous Dynamic Multi-view Graph Neural Network for Crime Prediction

BINBIN ZHOU, Hangzhou City University, Hangzhou, Zhejiang, China

HANG ZHOU, Zhejiang University, Hangzhou, Zhejiang, China

WEIKUN WANG, Zhejiang University, Hangzhou, Zhejiang, China

LIMING CHEN, Dalian University of Technology, Dalian, Liaoning, China

JIANHUA MA, Hosei University, Tokyo, Japan

ZENGWEI ZHENG, Hangzhou City University, Hangzhou, Zhejiang, China

Open Access Support provided by:

Hangzhou City University

Dalian University of Technology

Hosei University

Zhejiang University

ACM Transactions on Sensor Networks
https://doi.org/10.1145/3665141
EISSN: 1550-4867

HDM-GNN: A Heterogeneous Dynamic Multi-view Graph Neural Network for Crime Prediction

BINBIN ZHOU, School of Computer and Computing Science, Hangzhou City University, Hangzhou, China
HANG ZHOU, College of Computer Science and Technology, Zhejiang University, Hangzhou, China
WEIKUN WANG, College of Computer Science and Technology, Zhejiang University, Hangzhou, China
LIMING CHEN, School of Computer Science and Technology, Dalian University of Technology, Dalian, China
JIANHUA MA, Hosei University, Chiyoda-ku, Japan
ZENGWEI ZHENG, School of Computer and Computing Science, Hangzhou City University, Hangzhou, China

Smart cities have drawn a lot of interest in recent years, which employ Internet of Things (IoT)-enabled sensors to gather data from various sources and help enhance the quality of residents' life in multiple areas, e.g. public safety. Accurate crime prediction is significant for public safety promotion. However, the complicated spatial-temporal dependencies make the task challenging, due to two aspects: 1) spatial dependency of crime includes correlations with spatially adjacent regions and underlying correlations with distant regions, e.g. mobility connectivity and functional similarity; 2) there are near-repeat and long-range temporal correlations between crime occurrences across time. Most existing studies fall short in tackling with multi-view correlations, since they usually treat them equally without consideration of different weights for these correlations. In this paper, we propose a novel model for region-level crime prediction named as Heterogeneous Dynamic Multi-view Graph Neural Network (HDM-GNN). The model can represent the dynamic spatial-temporal dependencies of crime with heterogeneous urban data, and fuse various types of region-wise correlations from multiple views. Global spatial dependencies and long-range temporal dependencies can be derived by integrating the multiple GAT modules and Gated CNN modules. Extensive experiments are conducted to evaluate the effectiveness of our method using several real-world datasets. Results demonstrate that our method outperforms state-of-the-art baselines. All the code are available at https://github.com/ZJUDataIntelligence/HDM-GNN.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**; • **Information systems → Data mining**; • **Applied computing → Law, social and behavioral sciences**.

Additional Key Words and Phrases: crime prediction, graph neural network, data fusion, spatio-temporal prediction

*Both authors contributed equally to this research.
†Corresponding Author

Authors' Contact Information: Binbin Zhou, School of Computer and Computing Science, Hangzhou City University, Hangzhou, Zhejiang, China; e-mail: bbzhou@hzcu.edu.cn; Hang Zhou, College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang, China; e-mail: hangz@zju.edu.cn; Weikun Wang, College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang, China; e-mail: weikunwang@zju.edu.cn; Liming Chen, School of Computer Science and Technology, Dalian University of Technology, Dalian, China; e-mail: limingchen0922@dlut.edu.cn; Jianhua Ma, Hosei University, Chiyoda-ku, Tokyo, Japan; e-mail: jianhua@hosei.ac.jp; Zengwei Zheng, School of Computer and Computing Science, Hangzhou City University, Hangzhou, Zhejiang, China; e-mail: zhengzw@hzcu.edu.cn.

ACM 1550-4867/2024/5-ART
https://doi.org/10.1145/3665141

1 INTRODUCTION

Smart cities have drawn a lot of interest in recent years, which employ Internet of Things (IoT)-enabled sensors to gather data from various sources and help inform and enhance the quality of residents' life in multiple areas, such as public safety, resource allocation and so on. The collected big urban data contributes to safer and more efficient smart cities. Accurate crime prediction is significant for public safety promotion. Crime has been a major issue all throughout the world. According to data [4], there were 398.5 violent crime offenses per 100,000 people in the United States in 2020, up from 380.8 in 2019, causing great devastation, such as physical hurt, economic losses, and immeasurable psychological distress. The general public's concern of crime is growing. According to a Gallup poll [10] conducted from October 1 to 19, 2021, 51% of Americans feel crime in their own local area has grown, up from 38% in 2020. Thus, it necessitates the crime-related research to explore potential patterns and influential factors, which can be helpful for crime reduction and prevention.

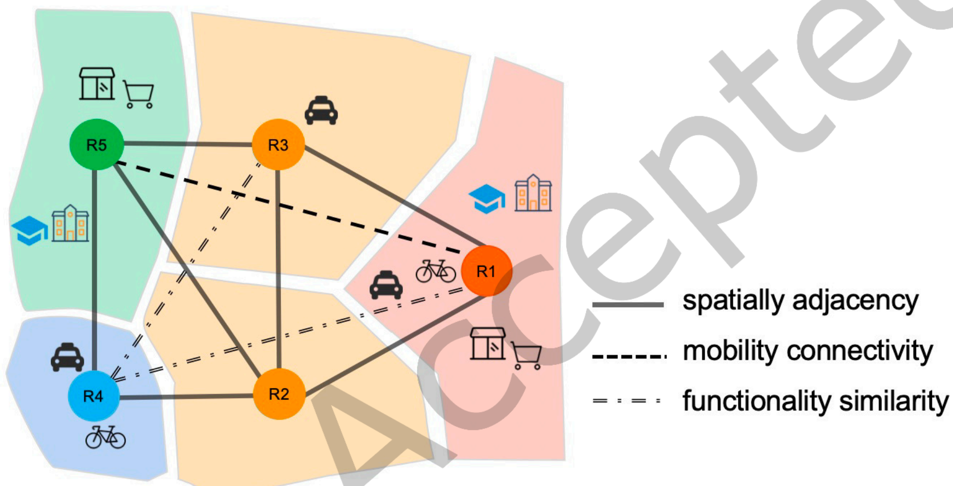

Fig. 1. An illustration of various types of region-wise spatial correlations. For crime prediction in region 1, spatially adjacent regions 2 and 3, mobility-connected region 4, and functionality-similar region 5 should be considered for spatial correlations.

In this paper, we study an important task: region-level crime prediction. Here, a region can be defined as a neighborhood, a grid or a community. In the past decades, with the advent of ubiquitous sensing technologies, various data has been collected, such as Point-of-Interests (PoIs), taxi mobility data, meteorological data and so forth. With the rapid development of machine learning and data-driven methods in recent years [71, 73], such large amount of urban data provide unprecedented opportunity for data-driven crime-related studies. The objective of our crime prediction at the regional level is to forecast future crime in regions using historical data. Accurate crime prediction can assist police resource schedules be optimized, residents' travel plans be adjusted, and public safety be enhanced.

The critical challenge for this crime prediction task is the complex spatial and temporal dependencies, which is two-fold. First, the spatial dependency is complicated that the crime situation would be affected by both crime events nearby and various external factors. Specifically, the crime situation in a region would be influenced by the spatially adjacent regions, e.g. suspects fleeing from surrounding regions. It is also correlated with distant regions that have similar functionalities, social-economic environments, or mobility connectivity. An example is presented in Fig.1. For region 1, it correlates with spatially adjacent regions such as regions 2 and 3, and maybe also correlate with mobility-connected region 4 due to the frequent traffic mobility between two regions. It may also be influenced by region 5 due to their similar functionalities, e.g. POI distributions. For example, drunken quarrels and subsequent brawls are more likely to occur in regions with more bars. Second, the temporal dependency is non-linear and non-stationary. Crime occurrences would be time-varying and pertinent with historical data, such as seasonal correlation and the typical near-repeat phenomenon of crime domains.

Many studies have been conducted over the last few decades modeling temporally time-evolving dynamics and spatial geographical dependencies to solve prediction tasks effectively, e.g. traffic flow and crime prediction [24, 35, 46, 64, 69]. However, due to the complicated impact of external urban data on crime, we are unable to directly apply these methods to our problem. In particular, crime would be dynamically affected by heterogeneous external urban data, such as parks with more crowds may witness more larceny events. Actually, knowledge distillation [67, 68, 75] and federated learning methods [66, 70, 74] provide promising to enhance the learning from external data. Despite these results of these spatial-temporal prediction methods and other app, we argue that current models may fall short when it comes to tackling one significant issue as multi-source external data. Most previous spatio-temporal prediction models were single-view based with one type of data. For example, in traffic flow prediction, the prediction was based only on historical records [24, 35], and in flashover prediction, the prediction was based only on temperature information [46]. As little investigation has been conducted into the potential of multiple sources of external data (e.g., human mobility data), its performance may be limited.

We attempt to explore crime prediction study modeling spatial-temporal dynamics with multi-source urban data. An intuitive method is to devise hand-crafted or learn high-level spatial and temporal features from various urban data, and then feed into a well-designed model for final prediction. However, the complicated spatial-temporal dependencies and heterogeneities makes the challenge more difficult. Some previous studies model spatial dependencies from multi-view and then simply concatenate or stack them together as feature representation where each single view data is treated equally [61]. This may be not the optimal choice as data from different views can have varying effects on the model.

To address these challenges, we propose a novel model called Heterogeneous Dynamic Multi-view Graph Neural Network (HDM-GNN) for region-level crime prediction. Motivated by multi-view joint representation [60], we propose a novel method for graph construction modeling dynamic spatial-temporal dependencies using heterogeneous urban data, with multi-view edge representations. These graphs are effectively fused so as to capture the complex and long-range spatial dependencies by introducing multiple graph attention mechanisms. The non-linear and long-range temporal dependencies can be captured through dilated convolutional neural networks. Finally, we introduce a fusion layer for crime prediction. The main contributions of this study are as follows.

- We construct a multi-view graph with heterogeneous urban data to model region-wise correlations, including spatial adjacency, mobility connectivity and functional similarity. We as well formulate the problem of crime prediction with heterogeneous urban data based on the graph.
- We propose an effective spatial-temporal framework for region-level crime prediction. Specifically, we present a multi-GAT module to incorporate multi-view region-wise correlations to capture spatial dependencies. With multiple layers, global spatial dependencies can be extracted. Moreover, a Gated Dilated CNN module is introduced to capture temporal dependencies.
- We evaluate the effectiveness of our method using the real-world graph datasets which we collected and constructed based on massive public open city data. Experimental results demonstrate that our method can outperform state-of-the-art baseline methods.

2 RELATED WORK

2.1 Spatial-temporal Prediction

The spatial-temporal prediction problem is a significant research topic. Many spatial-temporal prediction studies have been undertaken in multiple domains over the last few decades, including transportation, environment, public safety, economics, and so on [18, 28, 32, 72]. [49] used a deep convolutional recurrent model with spatial and temporal contexts to make region recommendation. [23] proposed an adaptive spatial-temporal graph attention networks for traffic flow forecasting. [48] constructed spatial and temporal graphs for citywide crowd flow prediction. [57] designed a spatial attentive and temporal dilated GCN for skeleton-based action recognition. And [27] solved the route recommendation problem based on a temporal-spatial metric.

Through well-designed models, these research showed latent spatial and temporal correlations between historical observations and predictive data. Recurrent neural network-based models, in particular, have been proposed to capture temporal correlations, which can be strengthened by the attention mechanism [8, 17, 50]. GAN-based models are proposed for high-level representations [76]. Meanwhile, convolution-based models for extracting spatial representations have been devised [58].

However, these spatio-temporal models do not effectively deal with the specific characteristics of crime forecasting, such as correlation with multiple factors, long distance dependence and cyclicality. In this study, we aim to investigate at crime prediction utilizing heterogeneous multi-type urban data and spatial-temporal modeling.

2.2 Graph Neural Networks for Spatial-temporal Prediction

Many studies on spatial-temporal prediction problems have focused on graph neural networks, which are capable of mapping nodes of graphs into low-dimensional vectors [16, 35, 46, 78]. [62] made a multimodal pedestrian trajectory prediction based on a relative interactive spatial-temporal graph. [56] predicted social events with multimodal fusion of spatial and temporal dynamic graph representations. [36] proposed a spatial temporal graph neural network model for predicting flashover in arbitrary building floorplans, while [25] solved the human-related anomalous event detection via spatial-temporal graph convolutional autoencoder with embedded long short-term memory network. Besides, [20] designed a dynamical spatial-temporal graph neural network for traffic demand prediction. Graph convolution network (GCN) is one representative work that has been used in a variety of fields due to its powerful capabilities, such as node classification, link prediction, and graph representation [22, 26, 52, 59]. Spectral GCNs [3, 5, 22] usually smooth nodes' input utilizing graph spectral filters. The other important type of GNN is spatial GCNs, which prefer to represent high-level features of nodes by aggregating neighboring nodes' information. The weight of neighboring nodes' weight can be learned either through fixed relationships [14] or via attention mechanisms dynamically [39]. By modeling spatial correlations of graphs, STSGCN [35] and STFGNN [24] incorporated spatial and temporal correlations simultaneously for further spatial-temporal prediction.

However, these methods are single-view, based on only one type of data. So they may fall limited when it comes to dealing with dynamic multi-view correlations generated by various external urban data.

2.3 Crime Prediction

In the past decades, there have been many studies related to crime prediction. From a characteristic perspective, there are many articles that have examined the relationship between crime and various characteristics, including historical criminal records, educational background, race, income level, and unemployment [6, 9, 29, 31, 43, 65, 77]. In recent years based on data mining studies, some have used more types of external data, such as crime-related urban data or social network data. For example, there are studies that use Twitter data to predict crime [12, 44] and studies that use cell phone communication data to assess crime risk on a large scale [1, 38]. Overall, existing work on crime prediction can be divided into three types in terms of the features selected.

The first type is the approach centered on temporal features. The focus of this work is on the temporal dimension of crime events. For example, the literature [30] uses a self-exciting point process to model crime and provides an in-depth analysis of temporal trends in burglary rates. In the literature [33], the authors investigate the temporal constraints on criminal behavior and propose a model of offender travel and opportunity and confirm the rule that a portion of crime is driven by opportunities that arise in the offender's daily life.

The second type of approach is centered on location characteristics. Predicted crime locations are often referred to as hotspots, and there is a large body of research on the exploration of crime hotspots. For example, historical crime records are used in the literature [37] to identify spatio-temporal patterns at multiple scales. This literature uses a variety of mathematical and physically quantitative methods to identify significant correlations in crime behavior data at both spatial and temporal scales. The literature [34] uses a simple random wandering model to study the changing patterns of crime hotspots and to identify stable hotspots. The literature [1] uses behavioral data consisting of call network data and demographic data, as well as publicly available crime data, to predict crime hotspots. The literature [40] aims to use point-of-interest information to infer the crime rate of a city. The literature [53] developed an augmented clustering-based algorithm to identify crime hotspots.

The third category is demographic feature-centric approaches. Most of the existing methods in this category use demographic data to solve crime prediction problems. Characteristics such as the racial composition of the population and the poverty level of the population are considered [2, 7, 13]. However, due to the relative stability of demographic characteristics, such methods are unable to capture dynamic crime patterns in cities.

From a methodological perspective, earlier literature [41] used regression models based on static characteristics for region-level crime rate prediction. The literature [19] used GRU-based time-series models for crime prediction. Literature [54] and literature [63] combined spatial and temporal features for spatio-temporal prediction of crime. With the rise of graph neural networks in recent years, there are also many studies using graph neural network-based spatio-temporal prediction models for more accurate crime prediction. For example, the literature [47] proposed spatio-temporal sequential hypergraph networks to encode complex spatio-temporal patterns of crime as well as potential class-level semantic relationships of crime to better handle spatio-temporal characteristics in long-term and global settings.

These traditional methods of crime prediction, however, tend to be based on simplified assumptions and therefore make it difficult to comprehensively consider the correlation of crime with massive external factors and the spatial and temporal correlation between urban data.

3 PRELIMINARIES AND FRAMEWORK OVERVIEW

3.1 Preliminaries

We begin with definitions and then formulate the region-level crime prediction problem. We use $\mathcal{G} = (V, E, A)$ to represent a spatial graph, where V is the finite set of nodes $|V| = N$, corresponding to N disjoint regions in the city, $V = \{r_1, r_2, ..., r_N\}$; E is the set of edges. Note that since there are multiple relationships between urban areas, there are multiple types of edges corresponding to them. So the graph is actually a multi-view graph; $A \in \mathbb{R}^{N \times N \times c}$ is the spatial correlation matrices representing c-type correlations between regions and each type of correlation here is represented as a type of edge in graph.

Denote $X^t_{\mathcal{G}}$ as the observations in all regions of the graph $\mathcal{G}$ at the t-th time-step, where each element X^t_r is the crime occurrences in region r at time t. Thus, the crime prediction problem can be formulated as learning a function $f : \mathbb{R}^{N \times T} \rightarrow \mathbb{R}^N$, mapping historical T observations to predict the future observations of this graph $\mathcal{G}$.

$$[X^{t-T+1}_{\mathcal{G}}, ..., X^t_{\mathcal{G}}] \xrightarrow{f(\cdot)} X^{t+1}_{\mathcal{G}}$$

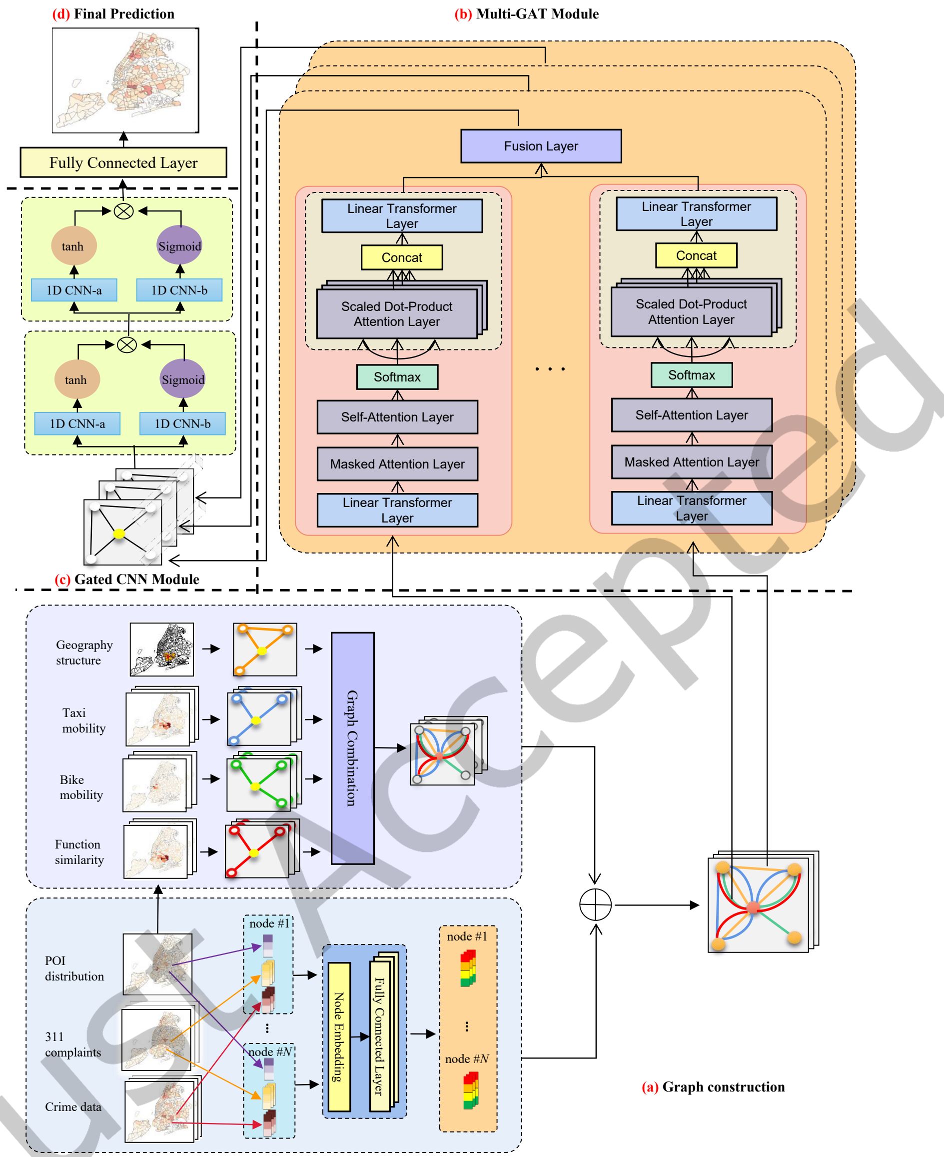

Fig. 2. Architecture of the proposed heterogeneous dynamic multi-view graph neural network (HDM-GNN). (a) is the graph construction, in which we represent regions and region-wise correlations from a variety of perspectives (e.g. geographical proximity, mobility connectivity, and functional similarity) by leveraging multi-source urban data, followed by hierarchical combination operations. (b) is the process of spatial dependency modeling using multiple GAT blocks on multiple graphs. (c) models temporal dependency using multiple Gated CNN layers, with the results connected to the subsequent layer. (d) is the output layer, which transforms the final representation into the prediction.

3.2 Framework Overview

The architecture of our proposed heterogeneous dynamic multi-view graph neural network (HDM-GNN) is shown in Fig.2. First, we provide a graph construction layer to generate multi-graphs modelling dynamic spatio-temporal dependencies using heterogeneous urban data, where nodes encode local representations of regions and edges represent region-wise relationships between regions from different views. We then use Graph Attention Network (GAT) layers based on the multi-graphs to aggregate information from spatially surrounding regions by propagating them with different learned weights based on the different region-wise edges. Specifically, for each view corresponding to a particular edge type, we use an independent GAT module to extract the node features of that view. Then a fusion layer is used to aggregate the multi-view representations. After that, Gated CNN layers with parallel 1D dilated convolution layers are used to obtain insights from a temporal perspective. Finally, fully connected layers are adopted, followed by an activation layer to map the learned features into the output space. The output vector can then be used to make final predictions. We will elaborate on the details in the following.

4 HDM-GNN: A HETEROGENEOUS DYNAMIC MULTI-VIEW GRAPH NEURAL NETWORK

In this section, we first introduce the method of constructing multi-view graphs from multi-source heterogeneous urban data, including the construction of node features and the construction of edges. Next, the method for extracting spatio-temporal features is presented. Specifically, we introduce the GAT-based model for spatial feature extraction in multi-view graph, the Gated CNN-based model for time series feature representation and the final results prediction.

4.1 Graph construction

We intend to construct spatial-temporal graphs representing nodes' characteristics and various impact of each node on its neighbors dynamically leveraging heterogeneous multi-source urban data. The detailed graph construction information are as follows.

4.1.1 Node Representation.

Using relevant urban data, inherent region characteristics can be extracted, which can indicate the social nature of regions. To represent regions' local characteristics, we use three types of urban data: crime statistics, 311 complaint records, and POI distribution data. Specifically, we employ POI distribution data to indicate each region's functionality, with the relevance of each type of POI being extracted as region features. Furthermore, we use crime and 311 data to encode each region's social features, which reflects the region's dynamic social stability. We define $\mathcal{P}$ as the region attributes as follows.

$$\mathcal{P} = \{\vec{p}_1, \vec{p}_2, .., \vec{p}_N\}, \vec{p} \in \mathbb{R}^D, \forall \vec{p} \in \mathcal{P} \tag{1}$$

where $\vec{p}_i$ is the encoded features of region r_i and D is the number of feature dimensions.

4.1.2 Edge Representation.

We then model various types of region-wise correlations based on the node representation. Urban regions would be affected by other regions from multiple perspectives, as illustrated in Fig.1. Previous studies have shown that human mobility plays significant role in region correlations [11, 42, 51, 60]. In this study, we construct three types of region correlations with graphs leveraging multiple urban data (e.g human mobility and geographical network), i.e. spatial adjacency graph $\mathcal{G}_G$, mobility connectivity graph $\mathcal{G}_M$, and functional similarity graph $\mathcal{G}_F$. Note that graph $\mathcal{G}_G$ and $\mathcal{G}_F$ are static graph, which means these graphs at different time steps are the same one, because the spatial adjacency relationship and the POI distribution of different regions do not varies over time. On the contrary, graph $\mathcal{G}_M$ is a dynamic graph which means it can be different graph at different time steps.

A. Spatial Adjacency Graph

We use the graph $\mathcal{G}_G = (V, E_G, A_G)$ to represent the spatially adjacent correlations between regions using the graphical network data. $A_G \in \mathbb{R}^{N \times N}$.

$$A_{G,ij} = \begin{cases} 1 & r_i \text{ and } r_j \text{ are geographically connected} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

B. Mobility Connectivity Graph

The graph $\mathcal{G}_M = (V, E_M, A_M)$ denotes the mobility connectivity between distant regions using human mobility data, including taxi mobility and bike mobility. As aforementioned, human mobility can exhibit underlying correlations, especially between distant regions. For example, a pair of source and destination regions with frequent human flow interactions would have a close tie from the human mobility perspective. We define a set of human mobility trip records as $\mathcal{M}$. Each record m denotes a pair of source and destination regions, $m = (r_s, r_d)$, $\forall m \in \mathcal{M}$. For taxi and bike mobility, we have $\mathcal{M}^T$ and $\mathcal{M}^B$ respectively. For human mobility data $\mathcal{M}^T$, the type correlation w^T_{sd} between regions r_s and r_d is defined to consider both the connectivity from r_s to r_d and from r_d and r_s.

Its calculation is as follows, $| \cdot |$ denotes the set size.

$$\mu^T_{ij} = |\{(r_s, r_d) \in \mathcal{M}^T | r_s = r_i, r_d = r_j\}| \tag{3}$$

$$w^T_{ij} = \begin{cases} 1 & \mu^T_{ij} \geq \delta \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

For the human mobility data $\mathcal{M}^B$, the correlation computation is similar. Note that, the mobility correlations are dynamic features. Finally, we encode $A_{M,ij}$ as follows. $A_M \in \mathbb{R}^{N \times N \times 2}$.

$$A_{M,ij} = \{w^T_{ij}, w^B_{ij}\} \tag{5}$$

C. Functional Similarity Graph

We define $\mathcal{G}_F = (V, E_F, A_F)$ to represent the similarity between regions, which may share the similar social and economic environment. For example, distant regions with similar POI distribution (e.g. clubs, entertainment places) may have similar functionality of the regions. Based on the node representation, all regions can be encoded as $\mathcal{P} = \{\vec{p}_i\}^N_{i=1}$. The similarity can be computed as follows:

$$A_{F,ij} = sim(\vec{p}_i, \vec{p}_j) \tag{6}$$

Here, $sim(\cdot)$ denotes the cosine similarity computation. $A_F \in \mathbb{R}^{N \times N}$. Since the region attributes are not dynamically represented, the similarity features are not dynamic.

Based on these types region-wise correlations, we represent edges from multiple views. $E = E_G \cup E_M \cup E_F$, the adjacent matrices $A = || (A_G, A_M, A_F), A \in \mathbb{R}^{N \times N \times 4}$, $||$ denotes the concatenation operation. Combined with the node representation, we construct multiple spatial-temporal graphs for further spatial and temporal dependency modeling.

4.2 Spatial Dependency Modeling

With multiple constructed graphs, we then model the spatial dependencies and represent the central regions with aggregation of spatial neighbor regions' information. We use graph attention network (GAT) [39] for spatial dependency modeling, which operates on graph-structured data using an attention mechanism. The use of GAT is motivated by the fact that a region is varying in its relevance to different neighboring regions, and this difference in relevance can be well modeled by the attention mechanism. Using the proposed multi-GAT module, the central node's representation would be updated through propagating various types of neighbor nodes' information with different specified weights.

Based on node representation $\mathcal{P} = \{\vec{p}_1, \vec{p}_2, .., \vec{p}_N\}, \vec{p} \in \mathbb{R}^D$ and edge representation with $E = E_G \cup E_M \cup E_F$ and $A = || (A_G, A_M, A_F)$, we first adopt multiple GAT blocks for different edge representations respectively. More specifically, for each type edge, we employ a linear transformer layer to map the input data of all regions into higher-level features by a weight matrix $\mathcal{W}$, $\mathcal{W} \in \mathbb{R}^{D \times D'}$. The weight matrix here is a learnable parameter matrix that serves to enable the multi-head attention mechanism. For each attention head, there is a corresponding weight matrix that maps the node features to a new space so that different attention heads can capture different potential attention relationships. Then, to improve the efficiency, a masked attention layer is used to drop the information from less correlated regions and compute the coefficients between the one-hop spatial neighbor regions $\mathcal{N}_i$ and the central region r_i. For the computation of these coefficients e_{ij}, we apply a self-attention mechanism a on all the regions with a feedforward neural network.

$$e_{ij} = a(\mathcal{W}\vec{p}_i, \mathcal{W}\vec{p}_j) \tag{7}$$

The coefficients e_{ij} can represent the significance of region r_j on region r_i from the spatial view. We then employ the softmax function to normalize them.

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} exp(e_{ik})} \tag{8}$$

The normalized attention coefficients are used for further output generation. We introduce a multi-head attention mechanism to obtain a more stable learning. The multi-head attention mechanism can learn and express inter-regional correlations from a wider range of perspectives. K attention mechanisms are employed to learn the attentional representation of regions with a non-linearity operation σ. And then, these learned features would be concatenated to generate a comprehensive region representation $\mathcal{P}' = \{\vec{p'}_1, \vec{p'}_2, .., \vec{p'}_N\}, \vec{p'} \in \mathbb{R}^{D'}$.

$$\vec{p'}_i = \mathop{||}_{k=1}^{K} \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k \mathcal{W}^k \vec{p}_j\right) \tag{9}$$

After that, a fusion layer is used for fusing these multiple region representations. Specifically, we average the node features from multiple views in the fusion step. In this way, the long-range spatial dependencies can be aggregated for each central region. After the fusion layer, the fused multi-view graph can be viewed as a single-view graph, and the output is a single node feature matrix at each time step corresponing to this synthetic single view.

4.3 Temporal Dependency Modeling

Following the spatial dependence modeling, we use the Gated CNN structure [24, 45] to capture our graphs' temporal dependency. We utilize dilated causal convolution [55] as the convolution layer in this case since it is capable of capturing long-range dependency in a non-recursive manner. A high rate of dilation results in a broad receptive field capable of capturing long-range temporal dependency. A gated mechanism is critical for handling sequence data and is effective at controlling output flow across layers. In this way, we aim to capture temporal dependency both in the short range and long-range, with different dilation rates.

Given the temporal sequence data of the updated graphs $X = [X_{\mathcal{G}}^{t-T+1}, ..., X_{\mathcal{G}}^{t}] \in \mathbb{R}^{N \times T \times D'}$, the parallel computation is as follows:

$$Y = g(\Theta_1 * X + \mathbf{a}) \odot h(\Theta_2 * X + \mathbf{b}) \tag{10}$$

where $\odot$ refers to a Hadamard product, $g(\cdot)$ denotes an activation function, and $h(\cdot)$ is the control function to determine the ratio passed to the next layer. Θ_1 and Θ_2 are the 1D convolution operations which can be processed parallelly. $\mathbf{a}$ and $\mathbf{b}$ are the model parameters.

The objective function can be defined as

$$L(X_{\mathcal{G}}^{t+1};\Theta) = \frac{1}{N \times D'} \sum_{i=1}^{N} \sum_{j=1}^{D'} l(\hat{X}_{\mathcal{G}}^{t+1} - X_{\mathcal{G}}^{t+1}) \tag{11}$$

Here, $l(\cdot)$ denotes the loss function.

Through multiple Gated CNN layers, the short-range and long-range temporal dependency can be captured, connected to following fully connected layers. After the linear transformation, final prediction results of the graph can be generated.

5 EXPERIMENTS

In this section, we first describe the datasets utilized in this research, as well as the workflow and evaluation metrics. Multiple baseline are also introduced. The performance on crime prediction is then evaluated. We also conduct hyperparameter studies to see how different parameters affect prediction tasks.

Table 1. Dataset Statistics

Datasets	NYC2020	NYC2018
#Crime Records	404,892	453,603
#311 Records	2,592,840	2,747,960
#POIs	56,449	51,192
#Taxi Trips	39,927,680	158,216,451
#Bike Trips	19,506,857	17,548,339
Datasets	NYC2017	NYC2016
# Crime Records	466,977	477,788
#311 Records	2,492,050	2,391,367
# POIs	42,294	37,052
#Taxi Trips	267,687,655	212,850,949
# Bike Trips	16,364,657	13,845,655

5.1 Datasets

We collect four real-world datasets of New York from NYC open data website[1] and OpenStreetMap[2]. Table 1 shows the dataset statistics. All datasets contain the geographical information, e.g.,the crime datasets include detailed crime records in 4 years with time, type and location information.The 311 records contains all 311 service requests

[1]https://opendata.cityofnewyork.us/
[2]https://www.openstreetmap.org/

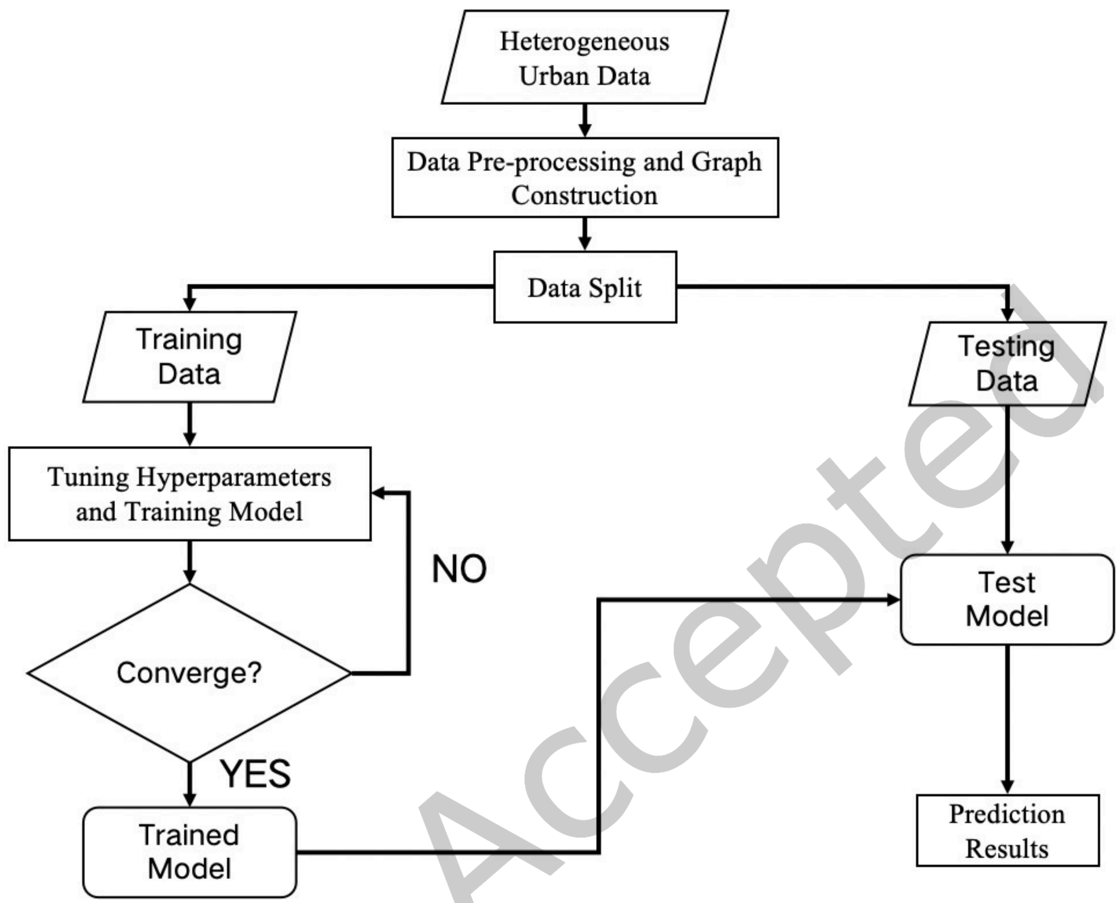

Fig. 3. Workflow of crime prediction

in each year. The POI datasets contains over 40 thousands POI records of more than 100 classes. We choose all 263 New York City neighborhoods as target regions and map each record to its associated region based on its location information.

5.2 Implementation and Evaluation Metrics

We present the workflow of our region-level crime prediction, as demonstrated in Fig.3, to present the specific implementation steps for region crime prediction. Firstly, we collect the heterogeneous urban data from various data sources and conduct the data pre-processing work. Secondly, we construct the multi-view temporal graphs and divide the whole dataset into the training dataset and the testing dataset. After that, we train our model through many epochs and obtain a set of hyperparameters for the model with best prediction performance by fine-tuning work until the model converges and achieves good performance. Finally, we use the well-trained model to make predictions and conduct the evaluation on the testing dataset.

We split all datasets with ratio 3:1 into training sets and test sets. Note that we are not doing a semi-supervised graph node classification problem, but a spatio-temporal prediction problem. Therefore, instead of following the division as in the GCN, we refer to the works on supervised spatio-temporal prediction for datasets division such as STFGCN [24]. The last 7 days' historical data is used for the next day's prediction. Experiments are repeated 5 times on all datasets. We implement our HDM-GNN model on the environment with one AMD EPYC 7502P CPU @ 3.35GHZ and NVIDIA RTX3090 24GB card. In our study, the hyperparameters are determined by the model's performance. Specifically, we use the tangent hyperbolic function as $g(\cdot)$, sigmoid function as $h(\cdot)$ and Huber loss as $l(\cdot)$ motivated by [45]. The multi-GAT module is composed 4 GAT blocks, each with 4 heads of scaled dot-product attention layers. The average operation is specified for the fusion layer. Adam optimizer is utilized in the model training with a learning rate 1e-4.

For evaluation metrics, we use the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The two metrics can be calculated through Equation 12 and 13.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{12}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{13}$$

Here, n denotes the total number of test data, $\hat{y}_i$ denotes the predicted value and y_i is the ground truth.

5.3 Baseline

We compare our HDM-GNN with several baselines, and categorize them into the following 2 groups, i.e. single view methods and multi-view methods. For the spatio-temporal designed for traffic flow forecasting, we processed the crime data into an input format consistent with the original paper. For the single-view model, we used only criminal history record data as inputs.

I. Single view methods

- AUTO-ARIMA [21]: an improved version of the effective time series forecasting model ARIMA, which get rid of the time-consuming parameter tuning procedure.
- LSTM [15]: the well-acknowledged long short-term memory network which uses forget gate, input gate and output gate to capture the long-term and short-term elements in the sequence, and solves the problems of gradient disappearance and gradient explosion in the long sequence training process.
- STGCN [52]: a spatial-temporal graph convolution network which applies graph convolution into graph-structured series data, which includes two ST-Conv Block and an Output Layer. As the core of the network structure, the spatial-temporal convolution block is composed of two temporal-gated convolutions and a spatial graph convolution in the middle, just like a sandwich. Convolution neural network takes less time because of its better parallel training, and it is not limited by the prediction data of previous time points, so that it can better capture the drastic changes of data.
- STFGNN [24]: a spatial-temporal fusion graph neural network which fuse various spatial and temporal graphs for spatial-temporal dependency learning. Based on the data-driven method, it constructs a new graph structure to preserve the hidden temporal and spatial features. This data-driven adjacency matrix can extract the correlation features that can't be extracted from the given space map in advance. It can capture spatio-temporal correlation synchronously.

II. Multi-view methods

Table 2. Performance comparison of HDM-GNN and baseline methods

Dataset	NYC2016		NYC2017		NYC2018		NYC2020	
Metric	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE
AUTO-ARIMA	56.66	1.99	56.62	1.96	56.12	1.92	58.34	1.81
LSTM	100.48	4.74	100.44	4.58	100.87	4.66	100.34	4.38
STGCN	49.85	1.82	49.26	1.81	50.01	1.77	51.54	1.71
STFGNN	51.61	**1.77**	52.71	1.78	55.53	1.74	52.66	1.65
ST-MGCN	92.52	4.24	53.53	1.84	53.46	1.75	53.52	1.70
MVJGPL	99.37	3.36	92.44	3.07	85.39	2.81	90.83	2.67
HMD-GNN	**46.74**	1.78	**48.09**	**1.77**	**50.01**	**1.71**	**50.42**	**1.63**

- ST-MGCN [11]: a spatial-temporal multi-graph convolution network which can represent non-Euclidean pair-wise correlations. The data of each frame is represented by three graphs, including (1) adjacency graph, which encodes spatial proximity, (2) regional functional similarity graph, which encodes the similarity of points of interest around the region, and (3) traffic connectivity graph, which encodes the connectivity between distant regions. In the prediction of time dimension, the information of T historical time steps is fused into one graph, while in the prediction of space dimension, the graphs of different correlations at one moment are combined into one graph.
- MVJGPL [60]: a multi-view joint graph representation learning model which can learn different region correlations. Specifically, they design a cross-view information sharing layer to boost the learning of individual view and a fusion layer to effectively combine multiple views.

5.4 Experimental Results and Analysis

5.4.1 Baseline Comparison.

Table 2 presents the comprehensive comparison results between our model and baselines. From the table, we observe that our HDM-GNN outperforms all state-of-the-art baseline methods from the perspective of MAPE, with an averaged improvement of 14.26%, 51.44%, 2.69%, 8.10%, 17.97%, and 46.72% respectively on all the four datasets. In terms of the metric MAE, on all the four datasets, our method gets an averaged improvement of 10.70%, 62.64%, 3.56%, 1.19%, 17.50%, and 42.15% compared with baseline methods, respectively. Note that on the dataset NYC2016, our method HDM-GNN obtains a comparable prediction performance with baseline method STFGNN in MAE value, which is slightly greater by HDM-GNN than that by STFGNN. These results indicates the superiority of our proposed method.

We observe that LSTM has the lowest prediction performance, with higher MAPE value and MAE value on all the four datasets. The reason for this could be that this method solely considers temporal dependency without taking spatial dependency into account. We notice that AUTO-ARIMA achieves better performance than LSTM, which is more suitable for handling stationary time series data, such as region-level crime data. Meanwhile, this method does not consider generalization learning among regions.

The baseline methods STGCN and STFGNN have been modeling spatial and temporal dependencies for further prediction tasks. Within them, STFGNN is able to extract long-range dependencies. But they are lack of consideration of heterogeneous external urban data which would be beneficial for the improvement of model performance. Our model incorporates a variety of external data and models various spatial relations. After the multi-view fusion, our method achieves better performance.

ST-MGCN and MVJGPL model multiple relations among regions leveraging multiple urban data. MGJPGL, in particular, does not consider node representation that it uses POI and Check-in data for edge correlations computation. They also do not take temporal dependency into consideration. ST-MGCN uses graph convolution networks for spatial aggregation. Our method takes both the node and edge representation into account, and adopts graph attention layers to learn spatial dependency dynamically.

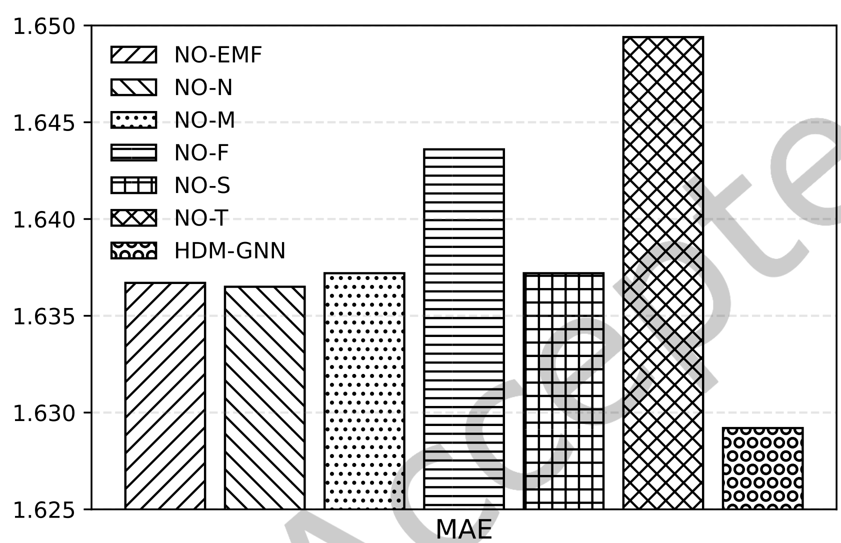

Fig. 4. Ablation study of HDM-GNN

5.4.2 *Ablation Study.*

To further investigate the effect of different components of our HDM-GNN, we devise the following variants of the HDM-GNN model.

- Ours (NO-EMF): This model only involves crime data and geographically adjacent relationship, without consideration of external urban data and spatial correlations (e.g. the mobility connectivity and function similarity).
- Ours (NO-N): This model disables the node representation module in which the features of region characteristics are concatenated together to represent node features.
- Ours (NO-M): This model does not consider the mobility connectivity edges.
- Ours (NO-F): This model does not consider the function similarity edges.
- Ours (NO-S): This model treats multiple edges equally in the spatial aggregation.
- Ours (NO-T): This model disables the Gated CNN module. A simple single layer linear feed-forward neural network is used for prediction instead of Gated CNN.

We compare these six variants with the HDM-GNN model on the NYC2020 dataset. Fig.4 presents the comparison results. It is obvious that the complete version of HDM-GNN yields the best results compared with all these variants. Compared with the complete model, the performance of NO-EMF, NO-N, NO-M, NO-F, NO-S and NO-T decreased by 0.49%, 0.48%, 0.51%, 0.92%, 0.52% and 1.23% respectively. The results of the ablation experiment demonstrate the role of multiple external data for crime prediction. For example, the risk of crime spreads as criminals flee, and mobility data can provide information about this. Areas with a lot of entertainment facilities, such as bars, tend to have a higher risk of crime, and POI data can provide the information about the functioning of the area. Thus the integration of crime-related external data can be effective in improving the accuracy of crime prediction. Moreover, we observe the importance of multiple graphs construction in terms of improving performance through mobility connectivity and function similarity edges. Furthermore, HDM-GNN outperforms NO-S by a significant margin, demonstrating the necessity of the multi-view fusion of different types of edges.

5.4.3 Hyperparameter Study.

In order to investigate the influence of different hyperparameters of our model on the final crime prediction results, we evaluate the performance of models on the NYC2020 dataset by varying three significant hyperparameters, including the node feature dimension, the number of GAT heads and the learning rate.

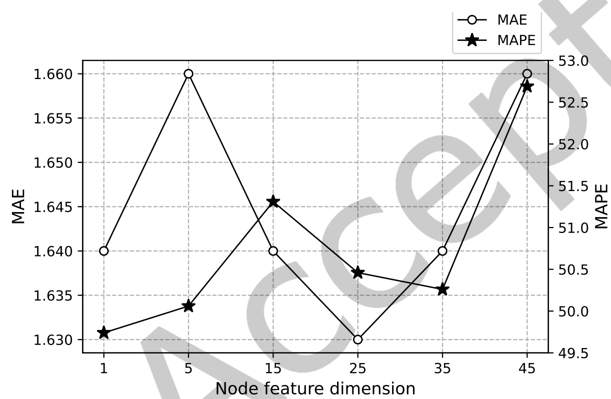

Fig. 5. Hyperparameter study of node feature dimension

Node Feature Dimension
We present the hyperparameter study on node feature dimension in Fig.5. From the figure, we observe that the best performance obtained when the node feature dimension equals to 25. Too low dimension indicates insufficient ability, and features can't be spread evenly in dimension space; too high a dimension introduces more redundant information of training sets, and it is easy to over-fit, which leads to the reduction of generalization ability of the model. An appropriate value of node feature dimension helps the model capture node information more effectively.
Number of GAT Heads
We show the hyperparameter study on number of GAT heads in Fig.6. Multi-head allows the model to focus on information from different representation subspaces in different positions, that is, it allows the network to capture richer information. The performance of the model is assessed by varying the number of heads n from 1 to 6. We notice the best results are obtained when $n = 4$. When $n < 4$, the performance worsens due to the lack of representation ability.
Learning Rate
We demonstrate the hyperparameter study on learning rate in Fig.7. Our model is optimized by random gradient descent algorithm, which is one of the most important parameters affecting the convergence of model performance.

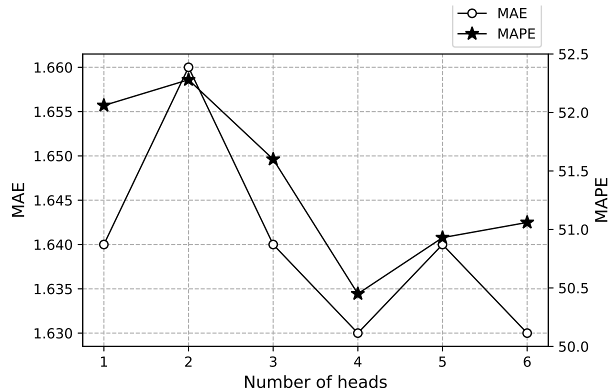

Fig. 6. Hyperparameter study of number of GAT heads

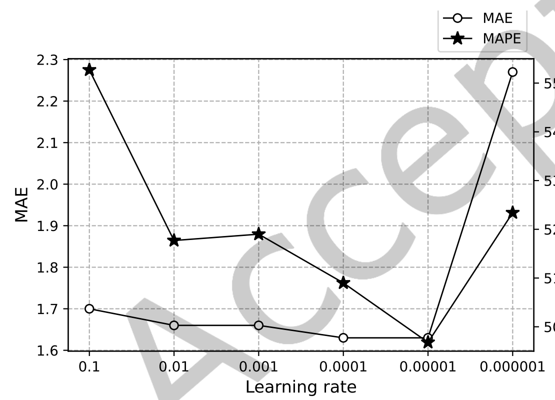

Fig. 7. Hyperparameter study of learning rate

The learning rate determines the step of weight change, so it is a very sensitive parameter. Its influence on the performance of the model is reflected in two aspects: the first is the initial learning rate, and the second is the change scheme of the learning rate. Adam optimizer is adopted to adjust the learning rate in the training process. For the initial learning rate, we search the learning rate from 0.000001 to 0.1, and confirm the value with the best performance in our model.

6 CONCLUSION

We propose a model which can model dynamic spatial-temporal dependencies leveraging heterogeneous urban data and effectively fuse various types of region-wise correlations from multiple views. Global spatial and long-range temporal dependencies can be derived. Extensive experiments on real-world datasets demonstrate that our method is superior to state-of-the-art methods. Our model is also helpful to enhance the public safety and contribute to the researches related to the smart city. We also release all the code in a GitHub repository: https://github.com/ZJUDataIntelligence/HDM-GNN.

Overall, there are still some limitations to this work. For example, in this work we only experimented on a dataset with different years in one city, and the transferability between different cities is a remaining challenge. Besides, many other factors related to crime have not been explored, such as social media data. Our model is trained on fixed historical data, so the real-time crime prediction with live data streams is still a challenge. Although we can update the model periodically, it would be a better solution to combine our model with some continuous learning techniques to enhance its practical application. In the future, we will try to incorporate more external data to achieve more accurate predictions using richer scenario information. In addition, more fine-grained crime prediction (e.g., at street level) is also a valuable research topic. It would be also interesting to predict crime in different cities based on some transfer learning approach in the future. Furthermore, due to the crime-specific multi-scale correlation between regions, designing more flexible and expressive GNN models for crime prediction scenarios is also an interesting research problem.

7 ACKNOWLEDGMENTS

This work is partially supported by the Zhejiang Provincial Natural Science Foundation (No. LTGG24F020002), the National Natural Science Foundation of China (No. 62102349, No. 62072402) and Research Project of Science and Education Innovation Complex of Hangzhou City University (No.22FHXM04). The authors would like to acknowledge the Supercomputing Center of Hangzhou City University, for the support of the advanced computing resources.

REFERENCES

[1] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, and Alex Pentland. 2014. Once upon a crime: towards crime prediction from demographics and mobile data. In *Proceedings of the 16th international conference on multimodal interaction*. 427–434.

[2] John Braithwaite et al. 1989. *Crime, shame and reintegration.* Cambridge University Press.

[3] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2014. Spectral networks and locally connected networks on graphs. In *ICLR*.

[4] CDE. 2021. Crime Data Explorer-Trend of Violent Crime from 2010 to 2020. https://crime-data-explorer.app.cloud.gov/pages/explorer/crime/crime-trend.

[5] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *NIPS*.

[6] Isaac Ehrlich. 1975. On the relation between education and crime. In *Education, income, and human behavior*. NBER, 313–338.

[7] Coral Featherstone. 2013. Identifying vehicle descriptions in microblogging text with the aim of reducing or predicting crime. In *2013 International Conference on Adaptive Science and Technology*. IEEE, 1–8.

[8] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. 2018. Deepmove: Predicting human mobility with attentional recurrent networks. In *WWW*. 1459–1468.

[9] Richard B Freeman. 1999. The economics of crime. *Handbook of labor economics* 3 (1999), 3529–3571.

[10] Gallup. 2021. GALLUP POLL SOCIAL SERIES: CRIME. https://news.gallup.com/file/poll/357119/211110LocalNationalCrime.pdf.

[11] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *AAAI*.

[12] Matthew S Gerber. 2014. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems* 61 (2014), 115–125.

[13] Eva GT Green, Christian Staerkle, and David O Sears. 2006. Symbolic racism and Whites' attitudes towards punitive and preventive crime policies. *Law and Human Behavior* 30, 4 (2006), 435–454.

[14] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *NIPS*.

[15] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[16] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. 2020. Heterogeneous graph transformer. In *WWW*. 2704–2710.

[17] Chao Huang, Chuxu Zhang, Peng Dai, and Liefeng Bo. 2021. Cross-interaction hierarchical attention networks for urban anomaly prediction. In *IJCAI*. 4359–4365.

[18] Chao Huang, Junbo Zhang, Yu Zheng, and Nitesh V Chawla. 2018. DeepCrime: attentive hierarchical recurrent networks for crime prediction. In *CIKM*.

[19] Chao Huang, Junbo Zhang, Yu Zheng, and Nitesh V Chawla. 2018. DeepCrime: Attentive hierarchical recurrent networks for crime prediction. In *Proceedings of the 27th ACM international conference on information and knowledge management*. 1423–1432.

[20] Feihu Huang, Peiyu Yi, Jince Wang, Mengshi Li, Jian Peng, and Xi Xiong. 2022. A dynamical spatial-temporal graph neural network for traffic demand prediction. *Inf. Sci.* 594 (2022), 286–304. https://doi.org/10.1016/j.ins.2022.02.031

[21] Rob J Hyndman and Yeasmin Khandakar. 2008. Automatic time series forecasting: the forecast package for R. *Journal of statistical software* 27, 1 (2008), 1–22.

[22] Thomas N Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *ICLR*.

[23] Xiangyuan Kong, Jian Zhang, Xiang Wei, Weiwei Xing, and Wei Lu. 2022. Adaptive spatial-temporal graph attention networks for traffic flow forecasting. *Appl. Intell.* 52, 4 (2022), 4300–4316. https://doi.org/10.1007/s10489-021-02648-0

[24] Mengzhang Li and Zhanxing Zhu. 2021. Spatial-Temporal Fusion Graph Neural Networks for Traffic Flow Forecasting. In *AAAI*.

[25] Nanjun Li, Faliang Chang, and Chunsheng Liu. 2022. Human-related anomalous event detection via spatial-temporal graph convolutional autoencoder with embedded long short-term memory network. *Neurocomputing* 490 (2022), 482–494. https://doi.org/10.1016/j.neucom.2021.12.023

[26] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2018. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *ICLR*.

[27] Feng Liang, Honglong Chen, Kai Lin, Junjian Li, Zhe Li, Huansheng Xue, Vladimir V. Shakhov, and Hannan Bin Liaqat. 2022. Route recommendation based on temporal-spatial metric. *Comput. Electr. Eng.* 97 (2022), 107549. https://doi.org/10.1016/j.compeleceng.2021.107549

[28] Hao Liu, Qiyu Wu, Fuzhen Zhuang, Xinjiang Lu, Dejing Dou, and Hui Xiong. 2021. Community-Aware Multi-Task Transportation Demand Prediction. In *AAAI*.

[29] George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E Tita. 2011. Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* 106, 493 (2011), 100–108.

[30] George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E Tita. 2011. Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* 106, 493 (2011), 100–108.

[31] E Britt Patterson. 1991. Poverty, income inequality, and community crime rates. *Criminology* 29, 4 (1991), 755–776.

[32] Huiling Qin, Songyu Ke, Xiaodu Yang, Haoran Xu, Xianyuan Zhan, and Yu Zheng. 2021. Robust Spatio-Temporal Purchase Prediction via Deep Meta Learning. In *AAAI*.

[33] Jerry H Ratcliffe. 2006. A temporal constraint theory to explain opportunity-based spatial offending patterns. *Journal of Research in Crime and Delinquency* 43, 3 (2006), 261–291.

[34] Martin B Short, Maria R D'orsogna, Virginia B Pasour, George E Tita, Paul J Brantingham, Andrea L Bertozzi, and Lincoln B Chayes. 2008. A statistical model of criminal behavior. *Mathematical Models and Methods in Applied Sciences* 18, supp01 (2008), 1249–1267.

[35] Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan. 2020. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *AAAI*.

[36] Wai Cheong Tam, Eugene Yujun Fu, Jiajia Li, Xinyan Huang, Jian Chen, and Michael Xuelin Huang. 2022. A spatial temporal graph neural network model for predicting flashover in arbitrary building floorplans. *Eng. Appl. Artif. Intell.* 115 (2022), 105258. https://doi.org/10.1016/j.engappai.2022.105258

[37] Jameson L Toole, Nathan Eagle, and Joshua B Plotkin. 2011. Spatiotemporal correlations in criminal offense records. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 4 (2011), 1–18.

[38] Martin Traunmueller, Giovanni Quattrone, and Licia Capra. 2014. Mining mobile phone data to investigate urban crime theories at scale. In *International Conference on Social Informatics*. Springer, 396–411.

[39] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph attention networks. In *ICLR*.

[40] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 635–644.

[41] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 635–644.

[42] Hongjian Wang and Zhenhui Li. 2017. Region representation learning via mobility flow. In *CIKM*.

[43] Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri. 2013. Learning to detect patterns of crime. In *Joint European conference on machine learning and knowledge discovery in databases*. Springer, 515–530.

[44] Xiaofeng Wang, Matthew S Gerber, and Donald E Brown. 2012. Automatic crime prediction using events extracted from twitter posts. In *International conference on social computing, behavioral-cultural modeling, and prediction*. Springer, 231–238.

[45] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph wavenet for deep spatial-temporal graph modeling. In *IJCAI*.

[46] Lianghao Xia, Chao Huang, Yong Xu, Peng Dai, Liefeng Bo, Xiyue Zhang, and Tianyi Chen. 2021. Spatial-Temporal Sequential Hypergraph Network for Crime Prediction with Dynamic Multiplex Relation Learning. In *IJCAI*. 1631–1637.

[47] Lianghao Xia, Chao Huang, Yong Xu, Peng Dai, Liefeng Bo, Xiyue Zhang, and Tianyi Chen. 2021. Spatial-Temporal Sequential Hypergraph Network for Crime Prediction with Dynamic Multiplex Relation Learning.. In *IJCAI*. 1631–1637.

[48] Jintao Xing, Xiangyuan Kong, Weiwei Xing, Xiang Wei, Jian Zhang, and Wei Lu. 2022. STGs: construct spatial and temporal graphs for citywide crowd flow prediction. *Appl. Intell.* 52, 11 (2022), 12272–12281. https://doi.org/10.1007/s10489-021-02939-6

[49] Hengpeng Xu, Wenjian Ding, Wei Shen, Jun Wang, and Zhenglu Yang. 2022. Deep convolutional recurrent model for region recommendation with spatial and temporal contexts. *Ad Hoc Networks* 129 (2022), 102545. https://doi.org/10.1016/j.adhoc.2021.102545

[50] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, and Zhenhui Li. 2019. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In *AAAI*.

[51] Zijun Yao, Yanjie Fu, Bin Liu, Wangsu Hu, and Hui Xiong. 2018. Representing urban functions through zone embedding with human mobility patterns. In *IJCAI*.

[52] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2018. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *IJCAI*.

[53] Chung-Hsien Yu, Wei Ding, Ping Chen, and Melissa Morabito. 2014. Crime forecasting using spatio-temporal pattern with ensemble learning. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 174–185.

[54] Chung-Hsien Yu, Wei Ding, Ping Chen, and Melissa Morabito. 2014. Crime forecasting using spatio-temporal pattern with ensemble learning. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 174–185.

[55] Fisher Yu and Vladlen Koltun. 2016. Multi-scale context aggregation by dilated convolutions. In *ICLR*.

[56] Guoshuai Zhang, Jiaji Wu, Mingzhou Tan, and Hong Han. 2022. Predicting Social Events with Multimodal Fusion of Spatial and Temporal Dynamic Graph Representations. *Big Data* 10, 5 (2022), 440–452. https://doi.org/10.1089/big.2021.0270

[57] Jiaxu Zhang, Gaoxiang Ye, Zhigang Tu, Yongtao Qin, Qianqing Qin, Jinlu Zhang, and Jun Liu. 2022. A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition. *CAAI Trans. Intell. Technol.* 7, 1 (2022), 46–55. https://doi.org/10.1049/cit2.12012

[58] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*.

[59] Muhan Zhang and Yixin Chen. 2018. Link prediction based on graph neural networks. In *NeurIPS*.

[60] Mingyang Zhang, Tong Li, Yong Li, and Pan Hui. 2020. Multi-view joint graph representation learning for urban region embedding. In *IJCAI*.

[61] Yunchao Zhang, Yanjie Fu, Pengyang Wang, Xiaolin Li, and Yu Zheng. 2019. Unifying inter-region autocorrelation and intra-region structures for spatial embedding via collective adversarial learning. In *KDD*.

[62] Duan Zhao, Tao Li, Xiangyu Zou, Yaoyi He, Lichang Zhao, Hui Chen, and Minmin Zhuo. 2022. Multimodal Pedestrian Trajectory Prediction Based on Relative Interactive Spatial-Temporal Graph. *IEEE Access* 10 (2022), 88707–88718. https://doi.org/10.1109/ACCESS.2022.3200066

[63] Xiangyu Zhao and Jiliang Tang. 2017. Modeling temporal-spatial correlations for crime prediction. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 497–506.

[64] Binbin Zhou, Longbiao Chen, Sha Zhao, Fangxun Zhou, Shijian Li, and Gang Pan. 2023. Spatio-temporal analysis of urban crime leveraging multisource crowdsensed data. *Personal and Ubiquitous Computing* (2023), 1–14.

[65] Binbin Zhou, Longbiao Chen, Fangxun Zhou, Shijian Li, Sha Zhao, Sajal K Das, and Gang Pan. 2020. Escort: Fine-grained urban crime risk inference leveraging heterogeneous open data. *IEEE Systems Journal* (2020).

[66] Jingwen Zhou, Shantanu Pal, Chengzu Dong, and Kaibin Wang. 2024. Enhancing quality of service through federated learning in edge-cloud architecture. *Ad Hoc Networks* (2024), 103430.

[67] Wujie Zhou, Yangzhen Li, Juan Huang, Yuanyuan Liu, and Qiuping Jiang. 2024. MSTNet-KD: Multilevel Transfer Networks Using Knowledge Distillation for the Dense Prediction of Remote-Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing* (2024).

[68] Xiaokang Zhou, Wang Huang, Wei Liang, Zheng Yan, Jianhua Ma, Yi Pan, I Kevin, and Kai Wang. 2024. Federated distillation and blockchain empowered secure knowledge sharing for Internet of medical Things. *Information Sciences* 662 (2024), 120217.

[69] Xiaokang Zhou, Wei Liang, Akira Kawai, Kaoru Fueda, Jinhua She, I Kevin, and Kai Wang. 2024. Adaptive Segmentation Enhanced Asynchronous Federated Learning for Sustainable Intelligent Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems* (2024).

[70] Xiaokang Zhou, Wei Liang, I Kevin, Kai Wang, Zheng Yan, Laurence T Yang, Wei Wei, Jianhua Ma, and Qun Jin. 2023. Decentralized P2P federated learning for privacy-preserving and resilient mobile robotic systems. *IEEE Wireless Communications* 30, 2 (2023), 82–89.

[71] Xiaokang Zhou, Wei Liang, I Kevin, Kai Wang, and Laurence T Yang. 2020. Deep correlation mining based on hierarchical hybrid networks for heterogeneous big data recommendations. *IEEE Transactions on Computational Social Systems* 8, 1 (2020), 171–178.

[72] Xiaokang Zhou, Qiuyue Yang, Qiang Liu, Wei Liang, Kevin Wang, Zhi Liu, Jianhua Ma, and Qun Jin. 2024. Spatial–Temporal Federated Transfer Learning with multi-sensor data fusion for cooperative positioning. *Information Fusion* 105 (2024), 102182.

[73] Xiaokang Zhou, Qiuyue Yang, Xuzhe Zheng, Wei Liang, I Kevin, Kai Wang, Jianhua Ma, Yi Pan, and Qun Jin. 2024. Personalized federation learning with model-contrastive learning for multi-modal user modeling in human-centric metaverse. *IEEE Journal on Selected Areas in Communications* (2024).

[74] Xiaokang Zhou, Xiaozhou Ye, I Kevin, Kai Wang, Wei Liang, Nirmal Kumar C Nair, Shohei Shimizu, Zheng Yan, and Qun Jin. 2023. Hierarchical federated learning with social context clustering-based participant selection for internet of medical things applications. *IEEE*

Transactions on Computational Social Systems (2023).

[75] Xiaokang Zhou, Xuzhe Zheng, Xuesong Cui, Jiashuai Shi, Wei Liang, Zheng Yan, Laurance T Yang, Shohei Shimizu, I Kevin, and Kai Wang. 2023. Digital twin enhanced federated reinforcement learning with lightweight knowledge distillation in mobile networks. *IEEE Journal on Selected Areas in Communications* (2023).

[76] Xiaokang Zhou, Xuzhe Zheng, Tian Shu, Wei Liang, I Kevin, Kai Wang, Lianyong Qi, Shohei Shimizu, and Qun Jin. 2023. Information theoretic learning-enhanced dual-generative adversarial networks with causal representation for robust OOD generalization. *IEEE Transactions on Neural Networks and Learning Systems* (2023).

[77] Qing Zhu, Fan Zhang, Shan Liu, Lin Wang, and Shouyang Wang. 2022. Static or dynamic? Characterize and forecast the evolution of urban crime distribution. *Expert Systems with Applications* 190 (2022), 116115.

[78] Yuwen Zhu and Lei Yu. 2023. Key Node Identification Based on Vulnerability Life Cycle and the Importance of Network Topology. *International Journal of Digital Crime and Forensics* 15, 1 (2023), 1–16.

Received 30 November 2022; revised 23 April 2024; accepted 8 May 2024