

Spatial-Temporal Sequential Hypergraph Network for Crime Prediction with Dynamic Multiplex Relation Learning

Lianghao Xia^{1*}, Chao Huang^{2*}, Yong Xu^{1,3,4†}, Peng Dai⁵, Liefeng Bo⁵,
Xiyue Zhang¹, Tianyi Chen¹

¹South China University of Technology, China

²University of Hong Kong, Hong Kong

³Communication and Computer Network Laboratory of Guangdong, China

⁴Peng Cheng Laboratory, China

⁵JD Finance America Corporation, USA

{cslianghao.xia,zhang.xiyue,csttychen}@mail.scut.edu.cn, yxu@scut.edu.cn
chao Huang75@gmail.com, {peng.dai,liefeng.bo}@jd.com

Abstract

Crime prediction is crucial for public safety and resource optimization, yet is very challenging due to two aspects: i) the dynamics of criminal patterns across time and space, crime events are distributed unevenly on both spatial and temporal domains; ii) time-evolving dependencies between different types of crimes (*e.g.*, Theft, Robbery, Assault, Damage) which reveal fine-grained semantics of crimes. To tackle these challenges, we propose Spatial-Temporal Sequential Hypergraph Network (ST-SHN) to collectively encode complex crime spatial-temporal patterns as well as the underlying category-wise crime semantic relationships. In specific, to handle spatial-temporal dynamics under the long-range and global context, we design a graph-structured message passing architecture with the integration of the hypergraph learning paradigm. To capture category-wise crime heterogeneous relations in a dynamic environment, we introduce a multi-channel routing mechanism to learn the time-evolving structural dependency across crime types. We conduct extensive experiments on two real-word datasets, showing that our proposed ST-SHN framework can significantly improve the prediction performance as compared to various state-of-the-art baselines. The source code is available at <https://github.com/akaxlh/ST-SHN>.

1 Introduction

Criminal activities (*e.g.*, Burglary, Robbery and Assault) have become a major social problem due to their adverse effect on public safety and economic development [Wortley and others, 2016]. For example, accurate crime prediction results could facilitate the decision making process of governments and prevent crimes happening for public safety and

emergency management [Zhao and Tang, 2017]. In recent years, the availability of crime data has enabled the development of data-driven methods for forecasting the occurrences of crimes [Wang *et al.*, 2016; Huang *et al.*, 2019].

One solution in predicting crimes is to uncover effective features for each category of crimes based on hand-engineering domain-specific features (*e.g.*, regional transportation information [Wei *et al.*, 2019], meteorological conditions [Zheng *et al.*, 2017]). However, without domain-specific expert knowledge (*e.g.*, criminology and anthropology), we may not have sufficient and accurate external data sources to model their correlations with different types of crimes (*e.g.*, burglary, felony assault) precisely [Zhang *et al.*, 2017a]. Therefore, a general crime data learning framework is a necessity to reduce the effort of hand-crafted feature engineering for the crime prediction task.

Among various spatial-temporal prediction methods, deep learning-based methods stand out owing to the feature representation effectiveness of neural network architecture. There exist many recently proposed forecasting frameworks focusing on modeling the time-evolving regularities over the temporal dimension and the underlying region-wise geographical dependencies over the spatial dimension, such as attentional neural methods [Yao *et al.*, 2019; Huang *et al.*, 2018], convolution-based learning approach [Zhang *et al.*, 2017b], spatial relation encoder with graph neural networks [Zheng *et al.*, 2020; Zhang *et al.*, 2021].

Despite their effectiveness, we argue that current spatial-temporal prediction models fall short in addressing unique challenges that are specifically for multi-dimensional crime data. Specifically, due to the crime data heterogeneity, there exist explicit and implicit dependencies among different categories of crimes. The intrinsic design of current methods, making them incapable of capturing cross-type crime influences in a fully dynamic scenario, with the integration of both spatial and temporal patterns.

Contribution. In light of the challenges, we propose to study the crime prediction with the goal of effectively encoding crime-type dependent representations, capturing spatial-

*Both authors contribute equally to this work

†Corresponding author: Yong Xu

temporal dynamics with the awareness of inherent cross-type crime influences. In this work, we present the Spatial-Temporal Sequential Hypergraph Network (ST-SHN) crime prediction architecture. Technically, we develop a multi-channel routing mechanism to model the cross-type crime influences at a fine-grained level of semantics under the graph neural network framework. Additionally, the developed hierarchical spatial dependency encoder via the hypergraph learning paradigm, endows the ST-SHN framework with the capability of exploring region-wise relations from locally to globally. To handle the time-wise crime patterns across different types, a graph temporal shift mechanism is developed to inject the time-varying spatial-temporal crime patterns into the representation process. ST-SHN reveals multi-view crime patterns for better model capacity and interpretability. In a nutshell, the dynamic multi-relational modeling of crimes allows us to encode relation dependencies and semantics corresponding to time, location and crime type dimensions. We conduct extensive experiments on two datasets, to demonstrate the advantages of our model for the effectiveness of crime prediction, modeling of latent cross-type crime influence, and interpretability of multi-view relations.

2 Preliminaries

We begin with some important notations and then introduce the formalized crime prediction problem. Considering a typical spatial-temporal forecasting scenario with R urban regions, T historical time slots (*e.g.*, days), and C crime categories (*e.g.* theft, assault, criminal damage), we define the crime tensor $\mathbf{X} \in \mathbb{R}^{R \times T \times C}$ corresponding to the spatial, temporal and categorical dimension, respectively, in which each entry $\mathbf{X}_{r,c}^t$ represents the number of c -th category of crime occurrences reported from region r at the t -th time slot.

Problem Statement. With the aforementioned definitions, the crime prediction task can be formalized as follows: learning a prediction function that takes the historical crime tensor \mathbf{X} as the input, and yields predictions $\mathbf{X}^{T+1} \in \mathbb{R}^{R \times C}$ on future crime occurrences. In this work, we tackle the tasks of both the binary crime occurrence prediction and forecasting the quantitative number of crime cases.

3 Methodology

3.1 Spatial Dependency Encoder

ST-SHN proposes to comprehensively capture the complex spatial dependencies across regions from different category views. We first generate the region graph $G = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} and \mathcal{E} represents the set of regions ($[r_1, \dots, r_R]$) and their spatially adjacent relations. Here, we use $S \times S$ grid scale to define geographical neighbors between region r_i and r_j .

Type-aware Crime Embedding Layer

With the consideration of cross-type crime influences in a dynamic environment, we describe a region r with multiple time-aware embeddings $\mathbf{E}_{r,c}^t \in \mathbb{R}^d$ corresponding to the c -th type of crime and t -th time slot. In particular, we first normalize the crime vector \mathbf{X}_r^t in tensor \mathbf{X} using the mean μ and std $\sigma \in \mathbb{R}^C$ value over all R regions and T time slots, *i.e.*, $\bar{\mathbf{X}}_r^t = (\mathbf{X}_r^t - \mu)/\sigma$. Formally, the time- and type-aware

region embedding $\mathbf{E}_{r,c}^t$ is generated as: $\mathbf{E}_{r,c}^t = \bar{\mathbf{X}}_{r,c}^t \cdot \mathbf{e}_c$, where $\mathbf{e}_c \in \mathbb{R}^d$ denotes the global embedding vector under the crime type of c over all time periods, and d is the hidden state size. Hence, we associate each region r with C different type-specific embeddings at each time slot in our dynamic multiplex spatial-temporal relation learning.

Spatial Message Propagation Paradigm

Based on the generated region embedding $\mathbf{E}_{r,c}^t$ and spatial region graph G , we generalize our spatial relation encoder with multiplex message propagation paradigm over G :

$$m_{i \leftarrow j}^t = \bar{\mathbf{A}}_{i,j} \cdot \text{Propagate}(\{\mathbf{E}_{i,c}^t, \mathbf{E}_{j,c}^t : c = 1, \dots, C\}) \quad (1)$$

where $m_{i \leftarrow j}^t$ represents the message passed from region r_j to r_i at t -th time slot. The region-wise geographical relations are reflected by the spatial adjacent matrix $\mathbf{A} \in \mathbb{R}^{R \times R}$, in which $\mathbf{A}_{i,j} = 1$ if edge $(r_i, r_j) \in \mathcal{E}$. To debias the influence of node degree in information propagation, we further incorporate the normalization factor as $\bar{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$, where \mathbf{D} denotes the diagonal degree matrix. Additionally, the embedding propagation function $\text{Propagate}(\cdot)$ is determined by both the source and target region, with the awareness of latent influence across different crime types (*e.g.*, Burglary, Robbery).

Multi-Channel Routing Mechanism

To comprehensively characterize the type-aware crime influences across different regions, we develop a routing mechanism under multiple channels as our information encoding function $\text{Propagate}(\cdot)$, which is formally presented as follows:

$$m_{i \leftarrow j}^{t,c} = \text{MC-Rout}(\mathbf{E}_{i,c}^t, \mathbf{E}_{j,c}^t) = \prod_{h=1}^H \sum_{c'=1}^C \alpha_{c,c'}^h \cdot \mathbf{V}^h \mathbf{E}_{j,c'}^t \quad (2)$$

In our routing mechanism, we design a multiplex mutual attention network with the learnable weight $\alpha_{c,c'}^h$. In specific, $\alpha_{c,c'}^h$ captures the explicit relevance between the c -th and c' -th crime types under the h -th representation subspace. $\mathbf{V}^h \in \mathbb{R}^{d/H \times d}$ represents the c -channel transformation. \parallel denotes the concatenation operation. Inspired by the architecture of Transformer network [Vaswani *et al.*, 2017], we propose to respectively map the target region r_i and source region r_j into query and key vector projection:

$$\bar{\alpha}_{c,c'}^h = \frac{(\mathbf{Q}^h \mathbf{E}_{i,c}^t)^\top (\mathbf{K}^h \mathbf{E}_{j,c'}^t)}{\sqrt{d/H}}; \alpha_{c,c'}^h = \frac{\exp(\bar{\alpha}_{c,c'}^h)}{\sum_{c'} \exp(\bar{\alpha}_{c,c'}^h)} \quad (3)$$

where $\mathbf{Q}^h, \mathbf{K}^h \in \mathbb{R}^{d/H \times d}$ are the h -channel transformation to acquire the query and key information.

Iterative Aggregation for High-order Connectivity

With the generated propagated message $m_{i \leftarrow j}^t \in \mathbb{R}^{C \times d}$, ST-SHN conducts aggregation to smooth the neighboring nodes and refine new region representation with the summation as: $\sum_{j=1}^R m_{i \leftarrow j}^t$. To extract the long-range spatial connectivity between regions, we further stack multiple spatial embedding propagation and aggregation blocks. We denote the initial region embeddings as $\mathbf{E}^{(0)}$ of size $R \times T \times C \times d$, then the

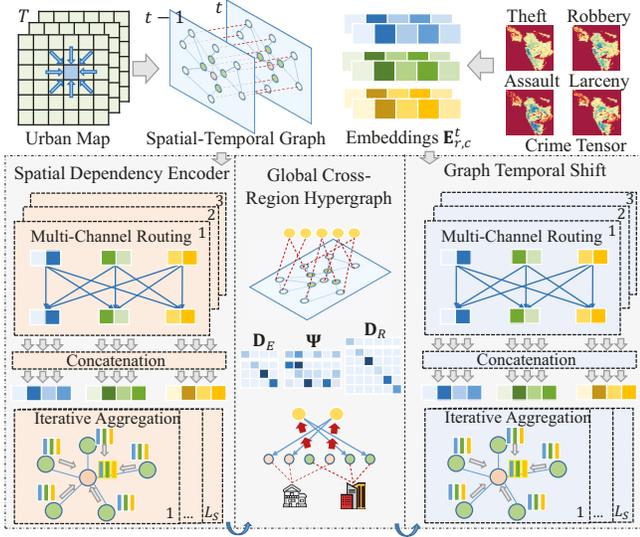


Figure 1: The model architecture of ST-SHN framework.

high-order message aggregation can be formalized as:

$$\mathbf{E}_i^{(l+1)} = \sigma\left(\sum_{j=1}^R \bar{\mathbf{A}}_{i,j} \cdot \text{Propagate}(\mathbf{E}_i^{(l)}, \mathbf{E}_j^{(l)})\right) \quad (4)$$

where $\sigma(\cdot)$ denotes the ReLU activation function with the incorporation of non-linear feature interactions. $\mathbf{E}_i^{(L)}$ is obtained with L iterations and is able to receive crime information from regions of L -hop distance.

3.2 Cross-Region Hypergraph Relation Learning

In addition to modeling spatial dependency between regions with location-aware information, we further enhance the cross-region relation learning without the limitation of adjacent connections by devising a hypergraph neural framework. With the incorporation of hypergraph neural network, we endow ST-SHN with the expressive power of global relation representation. Specifically, we first define a trainable adjacent matrix $\Psi \in \mathbb{R}^{E \times R}$, where E is a hyperparameter denoting the number of hyperedges. ST-SHN performs the embedding propagation along Ψ between R regions and E hyperedges. During this process, hyperedges serve as the intermediate information hubs which correlate not only locally connected but also far away regions in the entire urban space. Formally, with adjacency Ψ and regional embeddings $\bar{\mathbf{E}}$, the hypergraph message passing can be defined as follows:

$$\tilde{\mathbf{E}}^{(l+1)} = \sigma(\mathbf{D}_R^{-1/2} \Psi^\top \mathbf{D}_E^{-1/2} \sigma(\mathbf{D}_E^{-1/2} \Psi \mathbf{D}_R^{-1/2} \tilde{\mathbf{E}}^{(l)})) \quad (5)$$

where $\sigma(\cdot)$ denotes the activation function. $\mathbf{D}_R \in \mathbb{R}^{R \times R}$, $\mathbf{D}_E \in \mathbb{R}^{E \times E}$ denotes the diagonal degree matrices.

3.3 Graph Temporal Shift Mechanism

We further propose a graph temporal shift mechanism to encode i) intra-region type-aware crime temporal dependencies; ii) inter-region type-aware crime influences across different time slots. Towards this end, we construct a spatial-temporal

crime graph between R regions with the new adjacent matrix $\Gamma \in \mathbb{R}^{R \times R}$, with the former dimension referring to R regions at the t -th time slot, and the latter one corresponding to R regions at the $(t+1)$ -th time slot. In particular, to reflect the intra-region type-aware crime dependencies over the temporal dimension, we set $\Gamma_{i,i} = 1$ for region r_i . To enable the joint learning of inter-region and inter-type crime influence in a dynamic scenario, we set $\Gamma_{i,j} = 1$ between spatial adjacent region r_i and r_j in our graph temporal shift mechanism.

With the spatial-temporal adjacent matrix Γ and two multi-typed embedding tensors of adjacent time slots $\bar{\mathbf{E}}^{(0),t}$, $\bar{\mathbf{E}}^{(0),t+1}$ with the size $R \times C \times d$, ST-SHN recursively conducts the embedding recalibration and aggregation in a similar manner as our spatial relation encoder, as follows:

$$\bar{\mathbf{E}}_i^{(l+1),t+1} = \sigma\left(\sum_{j=1}^R \bar{\Gamma}_{i,j} \cdot \text{Propagate}_T(\bar{\mathbf{E}}_i^{(l),t+1}, \bar{\mathbf{E}}_j^{(l),t})\right) \quad (6)$$

where Propagate_T represents the temporal message passing paradigm which is built upon the designed multi-channel routing mechanism. $\bar{\Gamma} = \mathbf{D}^{-\frac{1}{2}} \Gamma \mathbf{D}^{-\frac{1}{2}}$ is the normalized adjacent matrix where \mathbf{D} is the corresponding degree matrix.

3.4 Model Prediction and Optimization

We integrate order-specific embeddings $(\mathbf{E}^{(0)}, \dots, \mathbf{E}^{(L)})$ with summation. At last, ST-SHN sums up the embeddings along the time dimension to generate the final embeddings $\mathbf{A} \in \mathbb{R}^{R \times C \times d}$ for making predictions of different types of crimes as: $\hat{\mathbf{X}}_{r,c}^{T+1} = \phi(\mathbf{w}_c^\top \mathbf{A}_{r,c})$, where $\phi(\cdot)$ is Sigmoid for classification (crime occurrence) and $\phi(x) = x$ for regression (quantitative number of crime cases). For the classification and the regression tasks, ST-SHN is optimized by minimizing the following two loss functions, respectively:

$$\begin{aligned} \mathcal{L}_c &= -\sum_t \delta(\mathbf{X}^t) \log \hat{\mathbf{X}}^t + \bar{\delta}(\mathbf{X}^t) \log(1 - \hat{\mathbf{X}}^t) + \lambda \|\Theta\|_2^2 \\ \mathcal{L}_r &= \sum_t \|\mathbf{X}^t - \hat{\mathbf{X}}^t\|_2^2 + \lambda \cdot \|\Theta\|_2^2 \end{aligned} \quad (7)$$

where t iterates the time slots, and $\delta(\cdot), \bar{\delta}(\cdot)$ denotes the element-wise positive and negative indicator functions, respectively. The later terms are the regularization terms and λ denotes the decay weight.

Model Complexity Analysis

The spatial and the temporal pattern encoding share similar complexity: $O((|\mathbf{A}|L_S + |\Gamma|L_T)(C+d)TCd)$, where $|\cdot|$ denotes the number of edges in the graph, and L_S, L_T denotes the number of stacked graph layers for spatial and temporal modeling, respectively. The hypergraph model costs $O(REL_H TCd)$ complexity, which is relatively minor in practice. Therefore, our model achieves the comparable time complexity as compared to other graph attentive frameworks.

4 Evaluation

This section aims to answer the following research questions:

- **RQ1:** How does our developed *ST-SHN* framework perform as compared to various state-of-the-art methods?

Data	NYC-Crimes		Chicago-Crimes	
Time Span	Jan, 2014 to Dec, 2015		Jan, 2016 to Dec, 2017	
Category	Burglary	Robbery	Theft	Battery
Instance #	31,799	33,453	124,630	99,389
Category	Assault	Larceny	Damage	Assault
Instance #	40,429	85,899	59,886	37,972

Table 1: Statistics of Experimented Datasets.

- **RQ2:** How do different components (*e.g.*, spatial context encoder, hypergraph relation learning, temporal shift mechanism) affect the results of *ST-SHN*?
- **RQ3:** What is the influence of parameters in *ST-SHN*?
- **RQ4:** Can *ST-SHN* provide the model interpretability *w.r.t* spatial, temporal and semantic dimensions?

4.1 Experimental Setup

Datasets

We evaluate our model on two real-world datasets collected from the crime reporting platforms on New York City and Chicago. Each record is formatted as (crime category, timestamp and geographical coordinates). Descriptive statistics for those datasets are presented in Table 1. Each dataset contains different categories of crimes, *e.g.*, Burglary and Robbery in NYC crime data; Theft and Damage in Chicago crime data. The reported prediction performance is averaged over 90 consecutive days during the test period of time. The grid-based mapping strategy is applied to partition NYC and Chicago into disjoint geographical regions with $3km \times 3km$ spatial unit. We use one day as the time interval to map the crime records into series. We generate the training and test set with the ratio of 7:1 and use the crimes of the last month in the training set as the validation data.

Evaluation Metrics

We adopt two types of metrics for performance evaluation. Specifically, to validate the prediction results of crime occurrences, *Macro-F1* and *F1-score* [Huang *et al.*, 2018] are used for evaluating the cross-type overall performance and type-specific forecasting accuracy, respectively. To evaluate the performance in predicting quantitative number of crimes, we utilize the *Mean Absolute Error (MAE)* and *Mean Absolute Percentage Error (MAPE)* [Geng *et al.*, 2019] as metrics.

Alternative Baselines

We compare our model with various state-of-the-art baselines which can be summarized into 5 groups. 1) conventional time series analysis techniques (ARIMA, SVM); 2) convolution-based learning model (ST-ResNet); 3) attentional spatial data prediction methods (STDN, DeepCrime, STtrans); 4) spatial-temporal prediction with graph neural networks (DCRNN, ST-GCN, GMAN); 5) deep hybrid spatial-temporal predictive solutions (UrbanFM, ST-MetaNet).

- **ARIMA** [Pan *et al.*, 2012]: it is a time series analysis method which models the temporal structures with kernel-based variable regression on future values.
- **SVM** [Chang and Lin, 2011]: it is another time series prediction technique which transforms data into feature space.

- **ST-ResNet** [Zhang *et al.*, 2017b]: it enhances the convolution neural network-based traffic prediction with the incorporation of residual network for model training efficiency.
- **STDN** [Yao *et al.*, 2019]: it models the spatial similarity and long-term periodic temporal pattern with the designed flow gating mechanism and attention mechanism.
- **DeepCrime** [Huang *et al.*, 2018]: it is an attentive recurrent neural model to encode the periodicity of crime data.
- **STtrans** [Wu *et al.*, 2020]: it leverages the self-attention-based transformer network to as the information encoder for predicting crime events.
- **DCRNN** [Li *et al.*, 2018a]: it designs bidirectional random walks to capture spatial correlations and sampling-based encoder-decoder for temporal pattern modeling.
- **ST-GCN** [Yu *et al.*, 2018a]: it proposes to use graph convolutional layers on the graph-structured time series data to model the corresponding spatial and temporal similarities.
- **GMAN** [Zheng *et al.*, 2020]: GMAN is built upon the graph-based attention network for aggregating information from both spatial and temporal dimensions.
- **UrbanFM** [Liang *et al.*, 2019]: it designs the convolutional feature extractor to consider local region-wise dependencies and a diffusion network to fuse meteorological factors.
- **ST-MetaNet** [Pan *et al.*, 2019]: this is a meta-learning prediction approach which employs the meta knowledge from geo-graph attributes for spatial correlation modeling.

Parameter Settings

We implement *ST-SHN* using TensorFlow and adopt Adam as the optimizer. We present the default hyperparameter settings of *ST-SHN* as follows: the hidden state size d is set as 16. The adjacent spatial relation scale is configured with the 3×3 grid unit. In addition, the depth of our spatial path aggregation layers is set as 2. We integrate long-range temporal dependencies by stack 7 temporal shift layers. The number of latent channels is set as 4 in our multi-channel dynamic routing mechanism. We consider the number hyperedges as 128. The learning rate is initialized as 0.001 with 0.96 decay rate. We search the weight decay factor of regularization term in the range of $\{0, 0.0001, 0.001, 0.01\}$.

4.2 Performance Comparison (RQ1)

The empirical results of all methods for predicting crime occurrences (measured by F1-score and Macro-F1) and quantitative number of crime instances (measured by MAE and MAPE) are reported in Table 2 and Table 3, respectively. We can find that *ST-SHN* consistently outperforms all baselines on two datasets in terms of all measures. We attribute these improvements to: i) by uncovering type-aware crime influence between regions, *ST-SHN* could preserve the holistic semantics of people crime behaviors; ii) benefiting from our hypergraph relational learning, *ST-SHN* is able to characterize the region-wise dependencies under global urban space, which results in more powerful representations of regions;

Model	New York City						Chicago					
	Crime Categories				Overall		Crime Categories				Overall	
	Burglary	Larceny	Robbery	Assault	Micro-F1	Macro-F1	Theft	Battery	Assault	Damage	Micro-F1	Macro-F1
ARIMA	0.4054	0.4878	0.5323	0.4109	0.4591	0.4629	0.4054	0.4878	0.5323	0.4109	0.4591	0.4629
SVM	0.4490	0.5649	0.5526	0.4262	0.4982	0.5049	0.6535	0.6434	0.5074	0.6314	0.6089	0.6068
ST-ResNet	0.4868	0.6019	0.5638	0.5321	0.5461	0.5497	0.5551	0.6760	0.5944	0.6817	0.6268	0.6339
DCRNN	0.5327	0.6242	0.6102	0.5187	0.5715	0.5773	0.6834	0.6469	0.5978	0.6788	0.6517	0.6507
STGCN	0.5467	0.6182	0.6141	0.5100	0.5722	0.5771	0.6857	0.6894	0.6294	0.7123	0.6752	0.6755
STtrans	0.5482	0.6067	0.6230	0.5289	0.5767	0.5792	0.7069	0.7106	0.6353	0.5475	0.6501	0.6498
DeepCrime	0.4681	0.6470	0.6362	0.5357	0.5717	0.5820	0.6240	0.7050	0.6409	0.7312	0.6653	0.6717
STDN	0.5255	0.5424	0.5876	0.4711	0.5316	0.5355	0.6290	0.6622	0.5864	0.6342	0.6279	0.6261
UrbanFM	0.5877	0.4658	0.6385	0.5603	0.5631	0.5695	0.6885	0.6694	0.5819	0.6459	0.6464	0.6420
ST-MetaNet	0.5375	0.5528	0.5354	0.4948	0.5301	0.5316	0.6983	0.6191	0.6558	0.7261	0.6748	0.6765
GMAN	0.5536	0.5621	0.6256	0.4653	0.5517	0.5570	0.6162	0.7077	0.6407	0.7248	0.6723	0.6759
<i>ST-SHN</i>	0.6173	0.6503	0.6721	0.5954	0.6341	0.6348	0.7154	0.7378	0.6743	0.7293	0.7142	0.7128

 Table 2: Crime occurrence prediction performance on NYC and CHI dataset in terms of *F1-score* and *Macro-F1*

Model	New York City		Chicago	
	MAE	MAPE	MAE	MAPE
ARIMA	0.8786	0.5509	0.8595	0.5056
SVM	1.0862	0.6831	0.9899	0.5862
ST-ResNet	0.9964	0.5799	0.9091	0.4500
DCRNN	0.6535	0.4514	0.6807	0.4711
STGCN	0.6203	0.4131	0.6822	0.4624
STtrans	0.6307	0.4038	0.6890	0.4559
DeepCrime	0.6231	0.4149	0.6772	0.4551
STDN	0.8977	0.4714	0.9530	0.5046
UrbanFM	0.8095	0.5742	0.8865	0.6125
ST-MetaNet	0.6243	0.4183	0.7581	0.4795
GMAN	0.6606	0.4691	0.7592	0.5438
<i>ST-SHN</i>	0.5833	0.3572	0.6425	0.3961

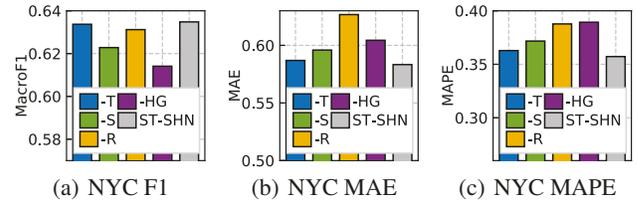
Table 3: Forecasting results of quantitative number of crimes.

iii) with the design of graph temporal shift mechanism, *ST-SHN* could better inject the dynamism of both intra- and inter-region dependencies across various crime categories underlying the spatial-temporal crime graph.

Compared with graph neural network-based spatial-temporal prediction methods (ST-GCN, GMAN, DCRNN), we can see that although they utilize location-based region graph to guide the learning of region embeddings, these approaches fail to model crime dependencies based on multiple crime categories, results in lower performance than *ST-SHN*. Besides, we can observe that although DeepCrime and STtrans utilize attention networks to model category-aware crime patterns, *ST-SHN* still achieves better results than them. The reason lies in that they fall short in exploring the time-evolving implicit influence among different types of crime patterns. The obvious performance gap between *ST-SHN* and methods: ST-ResNet and ST-MetaNet shed light on the limitation of modeling stationary spatial dependencies.

4.3 Model Ablation Study (RQ2)

We consider four contrast models to evaluate the effectiveness of sub-modules in *ST-SHN*: 1) “-S”: *ST-SHN* without the spatial relation encoder with the location-based contextual signals; 2) “-HG”: *ST-SHN* without the hypergraph learning


 Figure 2: Ablation studies of sub-modules in *ST-SHN*.

component to capture global region-wise relation; 3) “-T”: *ST-SHN* without the graph temporal shift mechanism; 4) “-R”: *ST-SHN* without the multi-channel routing mechanism. Figure 2 shows the comparison results. It is lean that the full version of *ST-SHN* achieves the best performance compared to other variants. Therefore, leaving cross-type crime influences across either the spatial or temporal dimension unexplored, will downgrade the prediction performance.

4.4 Hyperparameter Study (RQ3)

Dimensionality of Hidden State. The model performance is evaluated through varying the hidden dimensionality d from 4 to 32. We can notice that the best performance is achieved with the channel embedding size of 16. When we conduct feature representation with latent dimensionality > 16 , the performance degrades due to the overfitting issue.

Depth of Spatial-Temporal Graph Neural Model. We search the number of graph layers for spatial, temporal dimension in the range of $[0,8]$ and $[0,14]$. For spatial dimension, *ST-SHN* with 2 layers outperforms the others. For temporal dimension, the best performance is achieved with the propagation path of length 8. Clearly, increasing the model depth is capable of endowing our predictive model with better representation ability. Stacking more embedding propagation layers may involve noise in refining learned representations.

4.5 Explainability of *ST-SHN* (RQ4)

We offer several examples to give an intuitive impression of our model explainability. Specifically, Figure 4 visualizes i) different types of crime distribution (*i.e.* heat-map-based

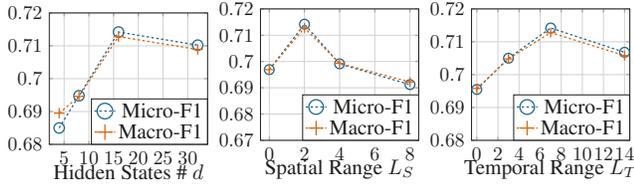


Figure 3: Hyperparameter study of *ST-SHN*

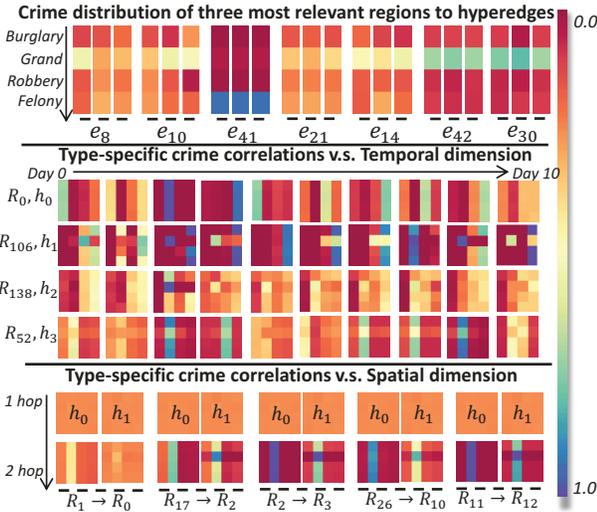


Figure 4: Visualization for learned relevance weights from the hypergraph learning, temporal and spatial relation learning. Bars represent the distribution of different types of crimes over regions. Squares depict the pair-wise correlations between crime categories.

statistics) from top-3 regions ranked by their learned relevance values with hyperedges (e.g., e_8, e_{41}); ii) learned type-specific crime influence scores for message propagation in our multi-channel routing mechanism over spatial and temporal dimensions. *First*, we can see that *ST-SHN* enables the dynamic modeling of multiplex spatial-temporal dependencies for different types of crimes, which is helpful to provide relation interpretations for policy makers. *Second*, the observed similar distributions among the most relevant regions, indicates that we endow our *ST-SHN* with the capability of identifying dependent regions which are either spatially neighbors or not, by capturing local and global geographical relation structures. In addition, we visualize the relevance matrix between a specific hyperedge and all regions in Figure 5(a) and Figure 5(b). By jointly analyzing the visualized relevance matrix and the actual crime spatial distributions across different regions, we can notice that the consistent distribution patterns suggest that *ST-SHN* could well differentiate regions with their underlying crime patterns.

5 Related Work

Spatial-Temporal Prediction. Many prediction methods have been proposed to uncover the implicit dependencies between future data points and historical observations from various spatial-temporal data [Li *et al.*, 2020]. Specifically,

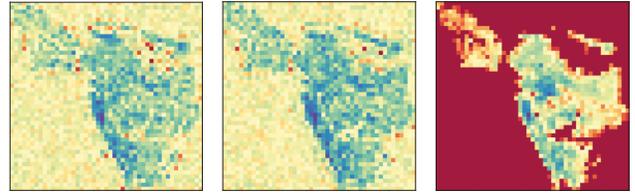


Figure 5: Heat maps for the encoded relations between all regions and two sampled hyperedges with the joint analysis of actual crime burglary spatial distribution in NYC.

attention mechanism has been used to enhance the temporal representation of recurrent neural network [Yao *et al.*, 2019; Feng *et al.*, 2018; Huang *et al.*, 2020]. Furthermore, convolution-based sequential models propose to capture the spatial-temporal information simultaneously [Feng *et al.*, 2020; Zhang *et al.*, 2017b]. In this work, we aim to predict crimes by exploring inherent dependencies across time, location and semantic dimensions.

Graph Neural Networks for Spatial-Temporal Data.

The objective of graph neural networks is to project nodes from the generated graph into a low-dimensional vector space [Wang *et al.*, 2019; Hu *et al.*, 2020]. It is worth mentioning that several recent efforts provide insights into graph neural networks for mining spatial-temporal data [Song *et al.*, 2020; Diao *et al.*, 2019]. For example, *ST-GCN* [Yu *et al.*, 2018b] takes inspiration from graph convolution network to model inter-region spatial dependency over the generated geography graph. *DCRNN* [Li *et al.*, 2018b] captures the spatial correlations by performing the bidirectional random walks on the region graph. However, these methods fall short in encoding region-wise relations in a dynamic manner, due to their utilization of static spatially adjacency matrices.

6 Conclusion

In this paper, we propose a novel Spatial-Temporal Sequential Hypergraph Network (*ST-SHN*), which approaches better crime pattern learning from spatial, temporal and semantic dimensions. *ST-SHN* is equipped with multi-channel routing mechanism to capture region-wise relations at a fine-grained level and exhibit the explainable semantics. Furthermore, with the relational path-aware message aggregation, *ST-SHN* refines the spatial-temporal representations under the context of long-range and global dependencies. Experiments demonstrate that the proposed method significantly outperforms various baselines. In the future, we would like to extend our model for real-time crime prediction with streaming data.

Acknowledgments

We thank the anonymous reviewers for their constructive feedback and comments. This work is supported by National Nature Science Foundation of China (62072188), Major Project of National Social Science Foundation of China (18ZDA062), Science and Technology Program of Guangdong Province (2019A050510010).

References

- [Chang and Lin, 2011] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *TIST*, 2(3):27, 2011.
- [Diao *et al.*, 2019] Zulong Diao, Xin Wang, et al. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. In *AAAI*, pages 890–897, 2019.
- [Feng *et al.*, 2018] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. Deepmove: Predicting human mobility with attentional recurrent networks. In *WWW*, pages 1459–1468, 2018.
- [Feng *et al.*, 2020] Jie Feng, Ziqian Lin, et al. A sequential convolution network for population flow prediction with explicitly correlation modelling. In *IJCAI*, 2020.
- [Geng *et al.*, 2019] Xu Geng, Yaguang Li, Leye Wang, et al. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *AAAI*, 2019.
- [Hu *et al.*, 2020] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *WWW*, pages 2704–2710, 2020.
- [Huang *et al.*, 2018] Chao Huang, Junbo Zhang, Yu Zheng, and Nitesh V Chawla. Deepcrime: Attentive hierarchical recurrent networks for crime prediction. In *CIKM*, pages 1423–1432. ACM, 2018.
- [Huang *et al.*, 2019] Chao Huang, Chuxu Zhang, Jiashu Zhao, Xian Wu, et al. Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting. In *WWW*, pages 717–728, 2019.
- [Huang *et al.*, 2020] Chao Huang, Chuxu Zhang, Peng Dai, and Liefeng Bo. Cross-interaction hierarchical attention networks for urban anomaly prediction. In *IJCAI*, 2020.
- [Li *et al.*, 2018a] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. In *ICLR*, 2018.
- [Li *et al.*, 2018b] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *ICLR*, 2018.
- [Li *et al.*, 2020] Ting Li, Junbo Zhang, et al. Autost: Efficient neural architecture search for spatio-temporal prediction. In *KDD*, pages 794–802, 2020.
- [Liang *et al.*, 2019] Yuxuan Liang, Kun Ouyang, Lin Jing, Sijie Ruan, et al. Urbanfm: Inferring fine-grained urban flows. In *KDD*, pages 3132–3142. ACM, 2019.
- [Pan *et al.*, 2012] Bei Pan, Ugur Demiryurek, et al. Utilizing real-world transportation data for accurate traffic prediction. In *ICDM*, pages 595–604. IEEE, 2012.
- [Pan *et al.*, 2019] Zheyi Pan, Yuxuan Liang, et al. Urban traffic prediction from spatio-temporal data using deep meta learning. In *KDD*. ACM, 2019.
- [Song *et al.*, 2020] Chao Song, Youfang Lin, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *AAAI*, volume 34, pages 914–921, 2020.
- [Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, et al. Attention is all you need. In *NIPS*, pages 5998–6008, 2017.
- [Wang *et al.*, 2016] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. Crime rate inference with big data. In *KDD*, pages 635–644, 2016.
- [Wang *et al.*, 2019] Xiao Wang, Houye Ji, et al. Heterogeneous graph attention network. In *WWW*, 2019.
- [Wei *et al.*, 2019] Hua Wei, Chacha Chen, Guanjie Zheng, Kan Wu, Vikash Gayah, et al. Presslight: Learning max pressure control to coordinate traffic signals in arterial network. In *KDD*, pages 1290–1298, 2019.
- [Wortley and others, 2016] Richard Wortley et al. *Environmental criminology and crime analysis*. Taylor & Francis, 2016.
- [Wu *et al.*, 2020] Xian Wu, Chao Huang, Chuxu Zhang, and Nitesh V Chawla. Hierarchically structured transformer networks for fine-grained spatial event forecasting. In *WWW*, pages 2320–2330, 2020.
- [Yao *et al.*, 2019] Huaxiu Yao, Xianfeng Tang, Hua Wei, et al. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In *AAAI*, 2019.
- [Yu *et al.*, 2018a] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *IJCAI*, 2018.
- [Yu *et al.*, 2018b] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *IJCAI*, 2018.
- [Zhang *et al.*, 2017a] Chao Zhang, Keyang Zhang, Quan Yuan, and etc. Regions, periods, activities: uncovering urban dynamics via cross-modal representation learning. In *WWW*, pages 361–370. ACM, 2017.
- [Zhang *et al.*, 2017b] Junbo Zhang, Yu Zheng, and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, 2017.
- [Zhang *et al.*, 2021] Xiyue Zhang, Chao Huang, Yong Xu, Lianghao Xia, et al. Traffic flow forecasting with spatial-temporal graph diffusion network. In *AAAI*, 2021.
- [Zhao and Tang, 2017] Xiangyu Zhao and Jiliang Tang. Modeling temporal-spatial correlations for crime prediction. In *CIKM*, pages 497–506, 2017.
- [Zheng *et al.*, 2017] Guanjie Zheng, Susan L Brantley, et al. Contextual spatial outlier detection with metric learning. In *KDD*, pages 2161–2170, 2017.
- [Zheng *et al.*, 2020] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. Gman: A graph multi-attention network for traffic prediction. In *AAAI*, pages 1234–1241, 2020.