Hawkes-Enhanced Spatial-Temporal Hypergraph Contrastive Learning Based on Criminal Correlations

Ke Liang[1], Sihang Zhou[2*], Meng Liu[1] Yue Liu[1], Wenxuan Tu[1], Yi Zhang[1], Liming Fang[3], Zhe Liu[4], Xinwang Liu[1*]

[1]School of Computer, National University of Defense Technology, Changsha, China
[2]School of Intelligence Science and Technology, National University of Defense Technology, Changsha, China
[3]Nanjing University of Aeronautics and Astronautics, Nanjing, China
[4]Zhejiang Lab, Hangzhou, China

Abstract

Crime prediction is a crucial yet challenging task within urban computing, which benefits public safety and resource optimization. Over the years, various models have been proposed, and spatial-temporal hypergraph learning models have recently shown outstanding performances. However, three correlations underlying crime are ignored, thus hindering the performance of previous models. Specifically, there are two spatial correlations and one temporal correlation, *i.e.,* (1) co-occurrence of different types of crimes (type spatial correlation), (2) the closer to the crime center, the more dangerous it is around the neighborhood area (neighbor spatial correlation), and (3) the closer between two timestamps, the more relevant events are (hawkes temporal correlation). To this end, we propose **H**awkes-enhanced Spatial-Temporal Hypergraph **C**ontrastive **L**earning framework (HCL), which mines the aforementioned correlations via two specific strategies. Concretely, contrastive learning strategies are designed for two spatial correlations, and hawkes process modeling is adopted for temporal correlations. Extensive experiments demonstrate the promising capacities of HCL from four aspects, *i.e.,* superiority, transferability, effectiveness, and sensitivity.

Introduction

Crimes, as a crucial social problem, are seen worldwide. If poorly managed, such activities will lead to severe adverse effects on urban safety (Wortley and Townsley 2016). Thus, the government has put forward higher requirements for effective crime prevention and combat. With the development of artificial intelligence (AI) technologies (Liu et al. 2023d,c; Chen et al. 2022a,b; Li et al. 2023; Tian et al. 2020; Peng et al. 2023; Luo et al. 2023; Liu et al. 2023b,a; Liu and Liu 2021; Liu, Wu, and Liu 2022) and the availability of crime data (Wang et al. 2016; Huang et al. 2019), various crime prediction models have emerged, which will assist governments in allocating police resources rationally based on spatial-temporal characteristics of different crimes and further contribute to the construction of the smart city in the future (Su, Li, and Fu 2011; Angelidou 2014).

Early attempts for crime prediction leverage convolutional neural networks (CNNs) to model the region-wise correlations and temporal properties (Feng et al. 2020; Zhang, Zheng, and Qi 2017). Later on, recurrent neural networks (RNNs) are adopted to capture the temporal information (Huang et al. 2018; Wu et al. 2020a). After that, graph neural networks (GNNs) have shown their promising capacities for spatial dependency modeling among different locations. For example, DCRNN (Li et al. 2017) and STGCN (Yu, Yin, and Zhu 2017) leverage spectral graph-based message-passing for information aggregation. Besides, GMAN (Zheng et al. 2020) and ST-MetaNet (Pan et al. 2019) adopt the graph attention network (GAT) for spatial feature aggregation between different region blocks. More recently, hypergraph models, *i.e.,* ST-HSL (Li et al. 2022), ST-SHN (Xia et al. 2022c), have shown their huge advantages on crime prediction, relying on more accurate and sufficient spatial-temporal modeling. However, they are still at an early stage. Specifically, they fail to leverage three correlations within crime data, including two spatial and one temporal correlation, thus hindering their capacity.

At first, we will introduce two spatial correlations, *i.e.,* type spatial correlations and neighbor spatial correlations, as follows. **(1) Type Spatial Correlation.** Different types of crimes may co-occur in the same region block, which is indicated in Fig. 1 (b). For example, robbery usually occurs along with the assault simultaneously. There are two main reasons for it. (i) Criminals often commit various crimes simultaneously unconsciously due to poor moral character and lack of legal awareness. (ii) Regions with poor law and order provide fertile soil for all kinds of crimes. **(2) Neighbor Spatial Correlation.** When crimes occur in a block, the government will usually warn the people to be careful when passing by its neighborhood areas. It reveals a typical characteristic of crime data, *i.e.,* the farther from the crime center, the safer the region is; the closer, the more dangerous. Previous hypergraph crime prediction models ignore the spatial information underlying the aforementioned correlations. Motivated by the promising ability of CL techniques on mining information underlying data itself, especially for graph data (Thakoor et al. 2021; Lee, Lee, and Park 2022; Liang et al. 2023a), specific contrastive learning (CL) strategies are designed in this work to leverage the spatial correlations for better prediction performance.

Besides the spatial correlations, temporal characteristics

*Corresponding author

[1]www.nyc.gov/site/nypd/stats/crime-statistics/historical.page

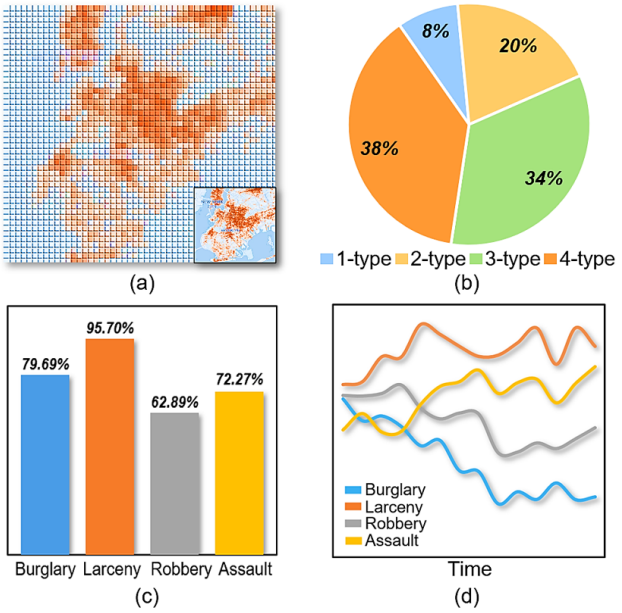

Figure 1: Characteristics of crime data. (a) is an example of used crime data. (b) describes the proportion of co-occurrence of different crimes in the same block. (c) presents the probability of the same crime occurring in both a block and its 1-hop neighbors. (d) reflects the dynamics[1] of different crimes along the timeline.

also play an essential role in crime prediction. Previous spatial-temporal hypergraph models (Li et al. 2022; Xia et al. 2022c) only use temporal relation encoding for dynamic information mining. Unlike these models, we designed a specific hawkes-enhanced spatial-temporal hypergraph model for feature encoding. Specifically, hawkes process modeling (Hawkes 1971), a classic yet effective technique, is adopted to enhance the dynamic interactions underlying the typical temporal correlations, *i.e.,* the closer two moments are, the more relevant events are, which is named as **Hawkes Temporal Correlations** in this paper.

Building upon the above ideas, we propose **H**awkes-enhanced Spatial-Temporal Hypergraph **C**ontrastive **L**earning (HCL), which mines the aforementioned correlations via two specific strategies, *i.e.,* contrastive learning and hawkes process modeling for spatial and temporal correlations, respectively. Concretely, the spatial criminal correlation extractor is designed to capture type and neighbor spatial correlations, further contributing to constructing contrastive pairs. Then, an alignment self-supervised loss, *i.e.,* MSE loss (Ermolov et al. 2021), is adopted for learning and optimization. Besides, hawkes process is utilized for dynamic interaction modeling within specific historical scopes. To evaluate HCL, experiments are carried out from four aspects, *i.e.,* superiority, transferability, effectiveness, and sensitivity. Our contributions are in three aspects:

- **Problem.** To the best of our knowledge, we are the first to point out that three criminal correlations, *i.e.,* type and neighbor spatial correlations and hawkes temporal correlation, should be exploited to benefit crime prediction.
- **Algorithm.** We propose a **H**awkes-enhanced spatial-temporal hypergraph **C**ontrastive **L**earning framework (HCL), which mines spatial and temporal correlations via contrastive learning and hawkes process modeling.
- **Evaluation.** We conduct experiments on two typical datasets, *i.e.,* NYC and CHI, between seventeen state-of-the-art crime prediction models. The results demonstrate promising capacities of HCL from four aspects, *i.e.,* superiority, transferability, effectiveness, and sensitivity.

Related Work

GNN-based Crime Prediction Models Spatial-temporal scenarios are important in real-world applications, such as urban computing, which benefits public safety and resource optimization. Various models (Yu et al. 2023a,b,c; Shao et al. 2023) are developed based on it. Multiple GNN-based models are proposed for crime prediction due to their promising capacities to model spatial dependency among different locations (Wang et al. 2020; Shao et al. 2022; Ji et al. 2023b; Tang et al. 2023; Tang, Xia, and Huang 2023). For example, DCRNN (Li et al. 2017) and STGCN (Yu, Yin, and Zhu 2017) leverage spectral graph-based message-passing for information aggregation. Then, GMAN (Zheng et al. 2020) and ST-MetaNet (Pan et al. 2019) adopt the graph attention network (GAT) for spatial message communication for feature aggregation between correlated regions. More recently, more works are developed based on hypergraph learning models (Wu et al. 2024b; Wu, Yan, and Ng 2023; Wu et al. 2024a). Meanwhile, **spatial-temporal hypergraph learning models**, *i.e.,* ST-HSL (Li et al. 2022), ST-SHN (Xia et al. 2022c), have shown their huge advantage on crime prediction, relying on more accurate and sufficient spatial-temporal modeling. However, they are still at an early stage. Specifically, they fail to leverage three correlations within crime data, including two spatial and one temporal correlation, thus hindering their capacity.

Graph Contrastive Learning Graph neural networks show their promising capacities in graph structural data (Yang et al. 2023c; Lee, Lee, and Park 2022; Zhao, Zhang, and Wang 2022). Contrastive learning (CL) is a simple yet effective way to enhance the representation capability of graph neural networks (GNN) (Liu et al. 2022a; Ji et al. 2023a; Xia et al. 2022a,b; Liu et al. 2022b; Liang et al. 2023b,c; Yang et al. 2022, 2023a,b). There are usually two essential elements for CL, *i.e.,* correlated view and contrastive loss. CL methods usually mine information within the data itself by optimizing specific contrastive loss to reach an agreement between correlated views (Li et al. 2022), *i.e.,* pull together the same samples across correlated views (positive samples) and push away the others (negative samples). GCL techniques are widely applied to various applications, such as recommendation (Wei et al. 2023, 2022; Wei, Xia, and Huang 2023; Xia et al. 2023), More recently, more researchers have believed that CL methods should pursue the precision of contrastive pairs rather than overemphasizing the completeness of positive and negative samples. Thus, many negative-free CL methods (Thakoor et al. 2021; Lee,

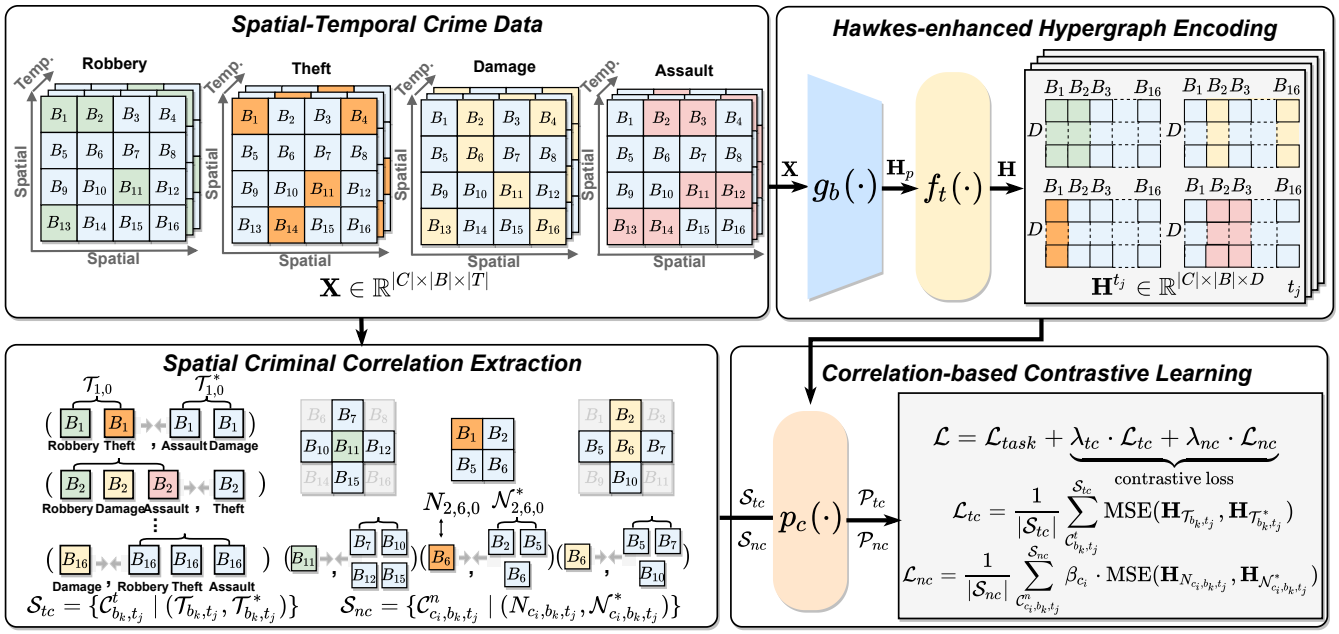

Figure 2: Framework of HCL. HCL contains three main procedures, *i.e.,* (a) hawkes-enhanced hypergraph encoding, (b) spatial criminal correlation extraction, and (c) correlation-based contrastive learning. The appeared notations can be found in Tab. 1.

Lee, and Park 2022; Liang et al. 2023a) are proposed, which also achieve promising performances. When it comes to CL-based spatial-temporal data prediction, few works exist. In particular, only two typical works are there for crime prediction. ST-HSL (Li et al. 2022) performs cross-view CL between local and global patterns and AutoST (Zhang et al. 2023) proposes an automated graph-level CL model to address the data noise and distribution diversity issues. Compared to them, HCL is the first negative-free spatial-temporal hypergraph CL framework based on three aforementioned criminal correlations.

Method

Preliminary

In this section, we introduce preliminaries *i.e.,* problem formulation and notation summary to understand our methods better. The urban space is generally divided into different geographical region blocks with grid-based map segmentation in urban computing (Xia et al. 2022c). Following these settings, we assume that there are $|B|$ urban region blocks, $|T|$ timestamps, and $|C|$ crime categories in the data. Based on it, we further define the initial crime feature as $\mathbf{X} \in \mathbb{R}^{|C|\times|B|\times|T|}$, which describes the occurrence of crimes. With such crime tensor as input, we aim to get the predictions $\mathbf{X}^{T+1} \in \mathbb{R}^{|C|\times|B|}$ on future crime occurrences. Moreover, we summarize the notations in Tab. 1.

Hawkes-enhanced Hypergraph Encoding

This procedure is mainly designed to compensate for the information underlying the hawkes temporal correlations. We leverage hawkes process modeling to enhance the dynamic interactions between different representations in this procedure. Specifically, given the initial criminal feature $\mathbf{X} \in \mathbb{R}^{|C|\times|B|\times|T|}$, the hawkes-enhanced representation $\mathbf{H}$ will be generated by two substeps, *i.e.,* primitive hypergraph encoding, and hawkes process enhancement.

Primitive Hypergraph Encoding As elaborated before, hypergraph learning paradigms can achieve better performance on crime prediction based on more accurate spatial-temporal modeling and more expressive representations. To inherit such good attributes, HCL is also developed based on hypergraph learning, and the adopted hypergraph backbones $\boldsymbol{g}_b(\cdot)$ will generate primitive spatial-temporal representations $\mathbf{H}_p$ by Eq. 1.

$$\mathbf{H}_p = \boldsymbol{g}_b(\mathbf{X}), \tag{1}$$

where different hypergraph learning encoders can be selected as candidate backbones,*e.g.,* ST-HSL (Li et al. 2022), and ST-SHN (Xia et al. 2022c).

Hawkes Process Enhancement Hawkes process modeling (Hawkes 1971), a simple yet effective technique, is adopted to enhance the dynamic interactions underlying the typical temporal correlations (hawkes temporal correlations), *i.e.,* the closer two moments are, the more relevant events are. Motivated by it, the procedure of hawkes process enhancement $\boldsymbol{f}_t(\cdot)$ aims to enhance the modeling of the temporal information underlying such correlations. As shown in Eq. 2, the primitive representation $\mathbf{H}_p$ is taken as input, and the hawkes-enhanced representation $\mathbf{H}$ will be generated by the following equation.

$$\mathbf{H} = \boldsymbol{f}_t(\mathbf{H}_p), \tag{2}$$

where the representations ($\mathbf{H}$ and $\mathbf{H}_p$) are composed of a sequence of representations at different timestamps *i.e.,* $\mathbf{H} = \{\mathbf{H}^{t_j} \mid j \in [0, T)\}$, and $\mathbf{H}_p = \{\mathbf{H}_p^{t_j} \mid j \in [0, T)\}$.

Notation	Explanation
C, B, T	set of crime categories, region blocks, timestamps
c_i, b_k, t_j	i^{th} crime category, k^{th} region block, j^{th} timestamp
$\|\cdot\|$	quantity of the elements in set
$\mathbf{X}$	initial crime feature
$\boldsymbol{g}_b(\cdot)$	adopted hypergraph learning backbone
$\boldsymbol{f}_t(\cdot)$	hawkes process enhancement procedure
$\mathbf{H}_p$, $\mathbf{H}$	primitive and hawkes-enhanced representation
$\mathbf{H}^{t_j}$, $\mathbf{H}_p^{t_j}$	primitive and hawkes-enhanced representation at t_j
w	calculated hawkes weight
s	historical scope for hawkes enhancement
δ	trade-off hyperparameter for hawkes enhancement
$\mathcal{T}_{b_k,t_j}$	category set of occurred crimes in b_k at t_j
$\mathcal{T}^*_{b_k,t_j}$	correlated category set of the crime in b_k at t_j
N_{c_i,b_k,t_j}	anchor block with c_i in b_k at t_j
$\mathcal{N}^*_{c_i,b_k,t_j}$	correlated neighbor spatial set with c_i in b_k at t_j
$\mathcal{C}^t$, $\mathcal{C}^n$	type, neighbor spatial correlation
$\mathcal{S}_{tc}$, $\mathcal{S}_{nc}$	type, neighbor spatial correlation set
$\boldsymbol{p}_c(\cdot)$	correlation-based contrastive pair constructor
$\mathcal{P}_{tc}$, $\mathcal{P}_{nc}$	set of type, neighbor correlation-based contrastive pairs
β_{c_i}	calculated category coefficient
λ_{tc}, λ_{nc}	trade-off hyperparameters of type, neighbor correlation
$\mathcal{L}_{tc}$, $\mathcal{L}_{nc}$	contrastive losses of type, neighbor spatial correlations
$\mathcal{L}$, $\mathcal{L}_{task}$	overall loss, and original task loss

Table 1: Notation summary

Specifically, hawkes process modeling is performed on the representation level (the primitive representation $\mathbf{H}$) to enhance the dynamic interactions between different features via different hawkes weights (Zuo et al. 2018) within a predefined historical scope along the timeline. Eq. 3 shows the enhancement procedure at timestamp t_j.

$$\mathbf{H}^{t_j} = \mathbf{H}_p^{t_j} + \delta \cdot \sum_{i=1}^{s} w_{j-i,j} \cdot \mathbf{H}_p^{t_{j-i}}, \tag{3}$$

where δ is the trade-off hyperparameter for hawkes enhancement. s denotes the predefined size of the historical scope, and $w_{j-i,j}$ represents the hawkes weights between timestamp t_{j-i} and t_j, which can be calculated by Eq. 4.

$$w_{j-i,j} = \exp\left(-\frac{t_j - t_{j-i} + 1}{t_j - t_{j-s} + 1}\right). \tag{4}$$

Spatial Criminal Correlation Extraction

Inspired by previous works (Liang et al. 2023a; Thakoor et al. 2021), we mine the information underlying two aforementioned spatial correlations *i.e.,* type spatial correlations and neighbor spatial correlations via contrastive learning techniques. To prepare for it, we first formally define these two spatial correlations in Def. 1 and Def. 2, defined as tuples corresponding to the target samples and their correlated samples for contrastive learning. Both correlations are extracted in this procedure and further contribute to the construction of contrastive pairs in the next section.

Type Spatial Correlation Extraction Type spatial correlation describes the characteristic, *i.e.,* co-occurrence of different types of crimes in the same region block. Thus, the feature representations of different crimes will affect each other. In particular, the occurred crimes will influence the probability of occurrence of the others. Besides, the interactions among crime types will also differ in region blocks. For example, assuming two situations, *i.e.,* (a) three types of crimes co-occur in one block, and (b) one type of crime occurs, the rest types of crime are more likely to happen in (a) compared to (b). To mine such information, we first traverse all of the region blocks and then extract the type spatial correlations $\mathcal{C}^t$ satisfying the Def. 1. As a result, all of the $\mathcal{C}^t$ constitute correlation set $\mathcal{S}^{tc}$ (See an example in Fig. 1).

Definition 1 (Type Spatial Correlation) *Given crime data $\mathbf{X} \in \mathbb{R}^{|C|\times|B|\times|T|}$, type spatial correlation in k^{th} block at j^{th} timestamp is denoted as $\mathcal{C}^t_{b_k,t_j} = (\mathcal{T}_{b_k,t_j}, \mathcal{T}^*_{b_k,t_j})$, iff the correlation exists.*

$$\begin{aligned} \mathcal{T}_{b_k,t_j} &= \{c_i \mid \mathbf{X}[c_i][b_k][t_j] > 0\}, \\ \mathcal{T}^*_{b_k,t_j} &= \{c_i \mid \mathbf{X}[c_i][b_k][t_j] = 0\}, \end{aligned} \tag{5}$$

where $c_i \in C$, $b_k \in B$ and $t_j \in T$.

Neighbor Spatial Correlation Extraction Neighbor spatial correlation describes the characteristic, *i.e.,* the closer to the crime center, the more dangerous it is around the neighborhood area. Thus, the crime in the center block will affect their occurrence in neighborhood blocks. However, different types of crimes may have different impacts on their neighbors. Fig. 1 (c) indicates that Larceny is more influenced by this correlation than other crimes. To mine such information, we first locate the blocks where the crimes exist and traverse their neighbors. Then, we extract the neighbor spatial correlations $\mathcal{C}^n$ satisfying the Def. 2 and further build the correlation set $\mathcal{S}^{nc}$ (See an example in Fig. 2).

Definition 2 (Neighbor Spatial Correlation) *Given crime data $\mathbf{X} \in \mathbb{R}^{|C|\times|B|\times|T|}$, neighbor spatial correlation with i^{th} crime category in k^{th} block at j^{th} timestamp is denoted as $\mathcal{C}^n_{c_i,b_k,t_j} = (N_{c_i,b_k,t_j}, \mathcal{N}^*_{c_i,b_k,t_j})$, iff the correlation exists.*

$$\begin{aligned} N_{c_i,b_k,t_j} &= b_k \qquad \text{if } \mathbf{X}[c_i][b_k][t_j] > 0, \\ \mathcal{N}^*_{c_i,b_k,t_j} &= \{\phi(N_{c_i,b_k,t_j}) \mid \mathbf{X}[c_i][\phi(N_{c_i,b_k,t_j})][t_j] = 0\}, \end{aligned} \tag{6}$$

where $c_i \in C$, $b_k \in B$, $t_j \in T$ and $\phi(\cdot)$ denotes the operation for locating the neighborhood block of a specific block. $\phi(N) = [N \pm 1, N \pm |\text{cols.}|]$, e.g.,, $|\text{cols.}| = 4$ in Fig. 2.

Correlation-based Contrastive Learning

After extracting spatial criminal correlations, we leverage the contrastive learning paradigm to mine their hidden semantics. It will endow the selected hypergraph learning backbone for better expressive and discriminative ability. We leverage a self-supervised alignment loss, *i.e.,* MSE loss, to pull together the representations of the elements within each correlation tuple in latent space due to the positive effects on each other described in previous sections.

Contrastive Pair Construction The contrastive pairs are constructed based on the former extracted type and neighbor spatial correlations. Whichever category it is, we should first encode the elements, *i.e.,* $\mathcal{T}_{b_k,t_j}, \mathcal{T}^*_{b_k,t_j}, N_{c_i,b_k,t_j}, \mathcal{N}^*_{c_i,b_k,t_j}$, in the correlation tuples, and further construct the contrastive pairs. Specifically, we leverage the average pooling strategy

to generate the embeddings based on the hawkes-enhanced representation $\mathbf{H} \in \mathbb{R}^{|C|\times|B|\times|T|\times D}$ by Eq. 7.

$$\begin{aligned}\begin{bmatrix}\mathbf{H}_{\mathcal{T}_{b_k,t_j}}\\ \mathbf{H}_{\mathcal{T}^*_{b_k,t_j}}\\ \mathbf{H}_{N_{c_i,b_k,t_j}}\\ \mathbf{H}_{\mathcal{N}^*_{c_i,b_k,t_j}}\end{bmatrix} &= p(\begin{bmatrix}\mathcal{T}_{b_k,t_j}\\ \mathcal{T}^*_{b_k,t_j}\\ N_{c_i,b_k,t_j}\\ \mathcal{N}^*_{c_i,b_k,t_j}\end{bmatrix})\\ &= \begin{bmatrix}\frac{1}{|\mathcal{T}_{b_k,t_j}|}\sum_{c_i\in\mathcal{T}_{b_k,t_j}}\mathbf{H}[c_i][b_k][t_j]\\ \frac{1}{|\mathcal{T}^*_{b_k,t_j}|}\sum_{c_i\in\mathcal{T}^*_{b_k,t_j}}\mathbf{H}[c_i][b_k][t_j]\\ \mathbf{H}[c_i][b_k][t_j]\\ \frac{1}{|\mathcal{N}^*_{c_i,b_k,t_j}|}\sum_{b_k\in\mathcal{N}^*_{c_i,b_k,t_j}}\mathbf{H}[c_i][b_k][t_j]\end{bmatrix}\end{aligned} \tag{7}$$

Afterward, we construct two corresponding contrastive pair sets, *i.e.,* $\mathcal{P}_{tc}$ and $\mathcal{P}_{nc}$, based on the generated correlation representations following the Eq. 8.

$$\begin{aligned}\mathcal{P}_{tc} &= \{(\mathbf{H}_{\mathcal{T}_{b_k,t_j}}, \mathbf{H}_{\mathcal{T}^*_{b_k,t_j}}) \mid \mathcal{C}^t_{b_k,t_j} \in \mathcal{S}_{tc}\},\\ \mathcal{P}_{nc} &= \{(\mathbf{H}_{N_{c_i,b_k,t_j}}, \mathbf{H}_{\mathcal{N}^*_{c_i,b_k,t_j}}) \mid \mathcal{C}^n_{c_i,b_k,t_j} \in \mathcal{S}_{nc}\}.\end{aligned} \tag{8}$$

Correlation-based Training Objective As shown in Eq. 9, the overall training objective composes of two parts, *i.e.,* task loss and correlation-based contrastive loss. Our HCL is optimized by minimizing the overall loss $\mathcal{L}$.

$$\mathcal{L} = \mathcal{L}_{task} + \underbrace{\lambda_{tc}\cdot\mathcal{L}_{tc} + \lambda_{nc}\cdot\mathcal{L}_{nc}}_{\text{contrastive loss}}. \tag{9}$$

Specifically, compared to some of the other contrastive learning frameworks, HCL is more like a plug-and-play auxiliary module, which should be coupled with the task losses, *i.e.,* regression loss for the quantity prediction of different crimes and classification loss for binary crime occurrence prediction (Xia et al. 2022c; Li et al. 2022).

Besides, the correlation-based contrastive loss contains two parts, both relying on the self-supervised alignment loss, *i.e.,* MSE loss, which aim to pull together the features in the correlations for training (See Eq. 10).

$$\begin{aligned}\mathcal{L}_{tc} &= \frac{1}{|\mathcal{P}_{tc}|}\sum_{\mathcal{C}^t_{b_k,t_j}}^{\mathcal{P}_{tc}}\text{MSE}(\mathbf{H}_{\mathcal{T}_{b_k,t_j}}, \mathbf{H}_{\mathcal{T}^*_{b_k,t_j}}),\\ \mathcal{L}_{nc} &= \frac{1}{|\mathcal{P}_{nc}|}\sum_{\mathcal{C}^n_{c_i,b_k,t_j}}^{\mathcal{P}_{nc}}\beta_{c_i}\cdot\text{MSE}(\mathbf{H}_{N_{c_i,b_k,t_j}}, \mathbf{H}_{\mathcal{N}^*_{c_i,b_k,t_j}}).\end{aligned} \tag{10}$$

Here, the β_{c_i} is the category coefficient corresponding to the probability that the same type of crime occurs in the 1-hop neighbors when a crime occurs, which is calculated during the spatial criminal correlation extraction(See Fig. 1 (c) as an example). Besides, the $\text{MSE}(\mathbf{H}_a, \mathbf{H}^+_a)$ can be calculated as follows:

$$\begin{aligned}\text{MSE}(\mathbf{H}_a, \mathbf{H}^+_a) &= \left\| \frac{\mathbf{H}_a}{\|\mathbf{H}_a\|_2} - \frac{\mathbf{H}^+_a}{\|\mathbf{H}^+_{\boldsymbol{a}}\|_2}\right\|^2_2\\ &= 1 - 2\cdot\frac{\langle\mathbf{H}_a, \mathbf{H}^+_a\rangle}{\|\mathbf{H}_a\|_2\cdot\|\mathbf{H}^+_a\|_2},\end{aligned} \tag{11}$$

where $\|\cdot\|_2$ denotes the $\mathcal{L}_2$-*norm*, and $\langle\cdot\rangle$ denotes the cosine value. Concretely, the correlated samples in each contrastive pair are pulled together in the latent space in this manner, thus improving the discriminative capability of the network. Algorithm 1 presents the pseudo-code of HCL.

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

Table 2: Two benchmark datasets for crime prediction.

Complexity Analysis The overall complexity is $O(|B|\times|C|\times(|B|\times D + |T|))$. Considering the complexity of adopted hypergraph models, *e.g.,* $O(|B|^2\times|C|^2\times D)$ in ST-HSL (Li et al. 2022), There is no redundant complexity and only a small proportion of extra complexity is added by HCL, thus leading to comparable model efficiency compared to previous models.

Experiment

This section presents the evaluation and analysis of HCL. Concretely, we first introduce the experiment setup. Then, different experiments are conducted for evaluation. We further put forward the following four questions in this section:

- **Q1: Superiority.** Does HCL outperform the existing state-of-the-art existing crime prediction models?
- **Q2: Transferability.** Will the proposed strategies be able to be transferred to different hypergraph backbones?
- **Q3: Effectiveness.** Are the proposed strategies effective in leveraging criminal correlations?
- **Q4: Sensitivity.** How does the performance fluctuation of HCL with different hyper-parameters?

Experiment Setup

Dataset Two real-world crime datasets in New York City (NYC) and Chicago (CHI) are used. Their statistics are reported in Tab. 2. They are split into disjoint regions with 3km × 3km spatial units. As in previous works, We use one day as the time interval to map crime records into a series. The training and test set are split with a ratio of 7:1 and use the crimes of the last month in the training set for validation.

Implementation Detail All experiments are conducted with a single NVIDIA 3090 Ti GPU (24 GB) and intel core i9-9900K CPU. For a fair comparison, the parameter settings in HCL for the selected hypergraph backbones are the same as the original papers of them, *i.e.,* ST-HSL (Li et al. 2022) for quantity prediction of different crimes, ST-SHN (Xia et al. 2022c) for occurrence prediction. We find the best combination of the hyperparameters in a grid-search manner. Concretely, the hyperparameter λ_{tc} and λ_{tc} are both searched in $\{0.05, 0.1, 0.15, 0.2, 0.25\}$. The size of the historical scope s is searched in $\{1, 3, 5, 7, 9\}$, and the temporal influential weight δ is searched in $\{0.001, 0.01, 0.05, 0.1\}$.

Models	New York City								Chicago							
	Burglary		Larceny		Robbery		Assault		Theft		Battery		Assault		Damage	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
SVM (2011)	1.1604	0.7653	1.4979	0.6417	1.1278	0.6733	1.1928	0.6964	1.7711	0.5629	1.3493	0.6027	1.0879	0.6560	1.1313	0.5721
ARIMA (2012)	0.8999	0.6305	1.3015	0.6268	0.9558	0.5969	0.9983	0.6198	1.5965	0.5720	1.3212	0.5792	0.8691	0.6044	1.0430	0.6134
ST-ResNet (2017)	0.8680	0.5603	1.1082	0.5329	0.8717	0.5209	0.9645	0.5749	1.3931	0.5488	1.1519	0.5719	0.7679	0.4633	0.9064	0.5018
DCRNN (2018)	0.8176	0.5324	1.0732	0.5492	0.9189	0.5532	0.9692	0.5955	1.3699	0.5770	1.1583	0.5528	0.7639	0.4600	0.8764	0.4756
STGCN (2018)	0.8366	0.5404	1.0629	0.5295	0.9035	0.5441	0.9375	0.5757	1.3628	0.5359	1.1512	0.5761	0.7963	0.4810	0.9068	0.4959
DeepCrime (2018)	0.8227	0.5508	1.0618	0.5351	0.8841	0.5537	0.9222	0.5677	1.3391	0.5430	1.1290	0.5389	0.7737	0.4616	0.9096	0.4960
GWN (2019)	0.7993	0.5235	1.0493	0.5405	0.8681	0.5351	0.8866	0.5646	1.3211	0.5502	1.1331	0.5503	0.7493	0.4580	0.8584	0.4850
STDN (2019)	0.8831	0.5768	1.1442	0.5889	0.9230	0.5649	0.9498	0.5661	1.5303	0.6287	1.2076	0.5791	0.8052	0.4820	0.9169	0.4869
ST-MetaNet (2019)	0.8285	0.5369	1.0697	0.5627	0.9214	0.5766	0.9323	0.5702	1.3369	0.5369	1.1762	0.5748	0.7904	0.4753	0.8907	0.4756
STtrans (2020)	0.8617	0.5592	1.0896	0.5478	0.8839	0.5651	0.9363	0.5679	1.3404	0.5356	1.1466	0.5684	0.7671	0.4499	0.8987	0.4842
GMAN (2020)	0.8652	0.5633	1.0503	0.5340	0.9234	0.5671	0.9338	0.5803	1.3235	0.5307	1.1442	0.5560	0.7852	0.4714	0.8823	0.4838
AGCRN (2020)	0.8260	0.5397	1.0499	0.5404	0.9013	0.5383	0.9063	0.5519	1.3281	0.5304	1.1432	0.5697	0.7669	0.4612	0.8712	0.4859
MTGNN (2020)	0.8329	0.5439	1.0473	0.5330	0.8759	0.5457	0.9090	0.5714	1.3054	0.5378	1.1307	0.5597	0.7571	0.4572	0.8667	0.4859
ST-SHN (2021)	0.8012	0.5198	1.0431	0.5291	0.8717	0.5362	0.9169	0.5682	1.3292	0.5310	1.1348	0.5544	0.7758	0.4574	0.8741	0.4747
DMSTGCN (2021)	0.8376	0.5485	1.0410	0.5464	0.8597	0.5403	0.9036	0.5601	1.3292	0.5291	1.1297	0.5552	0.8058	0.4759	0.8698	0.4877
ST-HSL (2022)	0.5413	0.3431	0.9131	0.4347	0.6397	0.3689	0.6672	0.3802	1.2917	0.4887	1.0895	0.4821	0.6417	0.3802	0.8246	0.4424
HCL (Ours)	**0.5228**	**0.3146**	**0.9017**	**0.4142**	**0.6116**	**0.3391**	**0.6487**	**0.3662**	**1.2506**	**0.4502**	**1.0601**	**0.4621**	**0.6284**	**0.3712**	**0.8006**	**0.4269**

Table 3: Performance (MAE, MAPE) comparison of quantity prediction of different crimes on NYC and CHI datasets.

Evaluation Metric Following previous works (Li et al. 2022; Huang et al. 2018), (1) Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) (Huang et al. 2018) are used for quantity prediction of different crimes (the lower, the better), and (2) Macro-F1 and Micro-F1 (Geng et al. 2019) are used for occurrence prediction (the higher, the better). The mean results of five runs are reported.

Compared Baseline We compare HCL with seventeen SOTA baselines divided into five groups. **(1)** conventional machine learning models, *i.e.,* ARIMA (Pan, Demiryurek, and Shahabi 2012), SVM (Chang and Lin 2011); **(2)** CNN-based models, *i.e.,* ST-ResNet (Zhang, Zheng, and Qi 2017), UrbanFM (Liang et al. 2019); **(3)** RNN-based models, *i.e.,* STDN (Yao et al. 2019), DeepCrime (Huang et al. 2018), STtrans (Wu et al. 2020a); **(4)** GNN-based models, *i.e.,* DCRNN (Li et al. 2017), STGCN (Yu, Yin, and Zhu 2017), GWN (Wu et al. 2019), AGCRN (Bai et al. 2020), MTGNN (Wu et al. 2020b), GMAN (Zheng et al. 2020), ST-MetaNet (Pan et al. 2019), DMSTGCN (Han et al. 2021); **(5)** Hypergraph Learning-based models, *i.e.,* ST-SHN (Xia et al. 2022c), ST-HSL (Li et al. 2022). We rerun ST-SHN and ST-HSL, as they are selected as backbones. The other results are copied from the original papers. More details for each compared baseline can be found in their references.

Main Performance (RQ1 & RQ2)

Tab. 3 and Tab. 4 show the main performances of HCL on two crime prediction tasks. Based on them, the **Q1** and **Q2** about the superiority and transferability can be answered.

Superiority Analysis (RQ1) HCL outperforms other models on all the metrics. Specifically, for crime quantity prediction, HCL can boost best MAE and MAPE performances on average by 2.96% and 6.20% on NYC and improve MAE and MAPE performance by 2.72% and 4.74% on CHI. In particular, HCL shows great superiority in Burglary predictions in NYC, *i.e.,* improving MAE and MAPE performances by 3.42% and 8.31%. Similar to it, the MAE and MAPE performance of Theft prediction in CHI can also be improved by 3.18% and 7.88%. Meanwhile, as for crime occurrence prediction, our method can also achieve the best performance. Concretely, the values of Micro-F1 and Macro-F1 are raised by 3.1% and 2.9% compared to the previous best performances in NYC. Based on the above performances, we can find the performance of HCL in NYC is better than it on CHI.

Models	New York City		Chicago	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1
SVM (2011)	0.4982	0.5049	0.6089	0.6068
ARIMA (2012)	0.4591	0.4629	0.4591	0.4629
ST-ResNet (2017)	0.5461	0.5497	0.6268	0.6339
DCRNN (2018)	0.5715	0.5773	0.6517	0.6507
STGCN (2018)	0.5722	0.5771	0.6752	0.6755
DeepCrime (2018)	0.5717	0.5820	0.6653	0.6717
STDN (2019)	0.5316	0.5355	0.6279	0.6261
UrbanFM (2019)	0.5631	0.5695	0.6464	0.6420
ST-MetaNet (2019)	0.5301	0.5316	0.6748	0.6765
STtrans (2020)	0.5767	0.5792	0.6501	0.6498
GMAN (2020)	0.5517	0.5570	0.6723	0.6759
ST-SHN (2021)	0.6111	0.6126	0.6868	0.6901
HCL (Ours)	**0.6301**	**0.6304**	**0.6955**	**0.6973**

Table 4: Performance (Micro-F1, Macro-F1) comparison of crime occurrence prediction on NYC and CHI datasets.

Transferability Analysis (RQ2) HCL with two different backbones, *i.e.,* ST-HSL and ST-SHN, are evaluated on two subtasks, *i.e.,* quantity prediction of crimes and crime occurrence prediction, respectively. Tracking all the records shows the superiority and effectiveness of HCL when adopting different backbones. Specifically, our HCL makes an average of 2.84% MAE and 5.47% MAPE improvements compared to the first backbone and boosts F1 performance by 2.08% compared to the second backbone.

As shown in the above analysis, HCL outperforms previous typical crime prediction baselines, which shows its superiority. Moreover, HCL can also achieve promising performances when applied to different crime prediction tasks

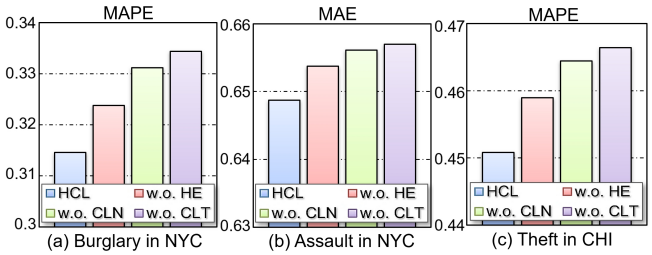

Figure 3: Ablation study, where "w.o. HE" denotes removing hawkes enhancement. "w.o. CLT" and "w.o. CLN" denote removing contrastive learning techniques for type and neighbor spatial correlations separately.

and integrated with different hypergraph backbones, demonstrating its great transfer ability and generalizability of HCL.

Ablation Study (RQ3)

We conduct the ablation study to answer the **Q3**, *i.e.,* the effectiveness of three leveraged correlations, *i.e.,* type spatial correlations, neighbor spatial correlations and hawkes temporal correlations. Fig. 3 shows the performance of the ablation studies. Concretely, the results show that the information effectively mined from all these three correlations can benefit the prediction performance. In general, two spatial correlations are more effective than the temporal correlation in our HCL. Besides, the type spatial correlations are more useful than the neighbor spatial correlations in most situations. It is reasonable since the type spatial correlations are more common characteristics compared to neighbor spatial correlations. Moreover, there are slightly different categories of crimes on how effective these correlations are. Specifically, the spatial correlations are more meaningful in Burglary and Theft prediction than Assault prediction. These performance variances of different correlations are mainly because of the characteristic differences for different crimes. Theft and Burglary are more similar, thus leading to similar performances with different correlations.

In conclusion, Both of the spatial correlations, *i.e.,* type spatial correlations and neighbor spatial correlations, and temporal correlations, *i.e.,* hawkes temporal correlations, are effectively mined by the designed strategies, which improves the capacities of the models for crime prediction.

Hyperparameter Analysis (RQ4)

We investigate the influence of the four extra introduced parameters, *i.e.,* type spatial correlation tradeoff weight λ_{tc}, neighbor spatial correlation tradeoff weight λ_{nc}, hawkes enhancement tradeoff weight δ, and historical scope size s, in this section to answer **Q4**. We report the MAE performance on two crime prediction tasks, *i.e.,* Robbery prediction in NYC and Damage prediction in CHI. The performances of different combinations of hyperparameters are present in Fig. 4. According to the figure, we can get the following three observations. (1) Referring to the absolute value fluctuation, the performances of the crime prediction are insensitive to different values within a specific scope, *e.g.,*

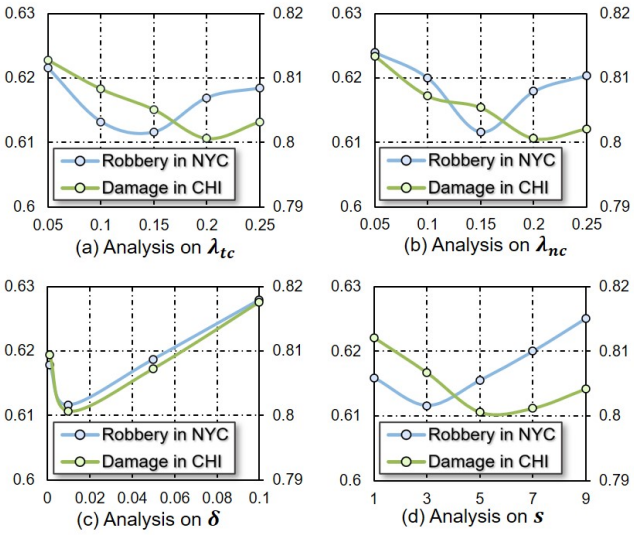

Figure 4: Hyperparameter analysis. The X-axes represent the scope of hyper-parameters. Both two Y-axes in each subfigure represent MAE, where the left axis is for "Robbery" prediction and the right one is for "Damage" prediction.

$[0.1, 0.15]$ for λ_{tc} for Robbery prediction in NYC, $[0.2, 0.25]$ for λ_{tc} for Damage prediction in CHI. (2) The trends of performance fluctuation of each parameter are similar in different datasets but may have different best values. For example, HCL achieve the best performance for robbery prediction in NYC when $s = 3$, while for Damage prediction in CHI when $s = 5$. (3) The influence of the Hawkes enhancement weights is more stable than the other two correlations. When δ is around 0.01, HCL usually reaches the best performance.

In summary, the hyperparameter discussion reveals how the performance fluctuates of HCL with different hyperparameters. Regarding to the results, we set $\lambda_{tc} = 0.15$, $\lambda_{nc} = 0.15$, $\delta = 0.01$, $s = 3$ for crime prediction in NYC, and $\lambda_{tc} = 0.2$, $\lambda_{nc} = 0.2$, $\delta = 0.01$, $s = 5$ for crime prediction in CHI.

Conclusion

In this paper, we first point out three omitted crime correlations, *i.e.,* type spatial correlations, neighbor spatial correlations, and hawkes temporal correlations, which will benefit the crime prediction. Besides, we design **H**awkes-enhanced Spatial-Temporal Hypergraph **C**ontrastive **L**earning framework (HCL), which adopts contrastive learning techniques, and hawkes process modeling to mine the information underlying the spatial and temporal correlation, respectively. In the future, we plan to develop hyperparameter-efficient models by reducing the hyperparameters more adaptively and optimising the extra time consumption brought by HCL.

Acknowledgments

This work is supported by the National Key R&D Program of China (No.2021YFB3100700) and the National Natural Science Foundation of China (no. 62325604, 62276271).

References

Angelidou, M. 2014. Smart city policies: A spatial approach. *Cities*, 41: S3–S11.

Bai, L.; Yao, L.; Li, C.; Wang, X.; and Wang, C. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *NeurIPs*, 33: 17804–17815.

Chang, C.-C.; and Lin, C.-J. 2011. LIBSVM: a library for support vector machines. *TIST*, 2(3): 1–27.

Chen, M.; Liu, T.; Wang, C.; Huang, D.; and Lai, J. 2022a. Adaptively-weighted Integral Space for Fast Multiview Clustering. In *MM*.

Chen, M.; Wang, C.; Huang, D.; Lai, J.; and Yu, P. S. 2022b. Efficient Orthogonal Multi-view Subspace Clustering. In *SIGKDD*.

Ermolov, A.; Siarohin, A.; Sangineto, E.; and Sebe, N. 2021. Whitening for self-supervised representation learning. In *ICML*, 3015–3024. PMLR.

Feng, J.; Lin, Z.; Xia, T.; Sun, F.; Guo, D.; and Li, Y. 2020. A Sequential Convolution Network for Population Flow Prediction with Explicitly Correlation Modelling. In *IJCAI*, 1331–1337.

Geng, X.; Li, Y.; Wang, L.; Zhang, L.; Yang, Q.; Ye, J.; and Liu, Y. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *AAAI*, volume 33, 3656–3663.

Han, L.; Du, B.; Sun, L.; Fu, Y.; Lv, Y.; and Xiong, H. 2021. Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting. In *SIGKDD*, 547–555.

Hawkes, A. G. 1971. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3): 438–443.

Huang, C.; Zhang, C.; Zhao, J.; Wu, X.; Yin, D.; and Chawla, N. 2019. Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting. In *The world wide web conference*, 717–728.

Huang, C.; Zhang, J.; Zheng, Y.; and Chawla, N. V. 2018. DeepCrime: Attentive hierarchical recurrent networks for crime prediction. In *CIKM*, 1423–1432.

Ji, C.; Li, J.; Peng, H.; Wu, J.; Fu, X.; Sun, Q.; and Yu, P. S. 2023a. Unbiased and Efficient Self-Supervised Incremental Contrastive Learning. In *WSDM*, 922–930.

Ji, C.; Zhao, T.; Sun, Q.; Fu, X.; and Li, J. 2023b. Higher-Order Memory Guided Temporal Random Walk for Dynamic Heterogeneous Network Embedding. *Pattern Recognition*, 109766.

Lee, N.; Lee, J.; and Park, C. 2022. Augmentation-free self-supervised learning on graphs. In *AAAI*, 7372–7380.

Li, L.; Zhang, J.; Wang, S.; Liu, X.; Li, K.; and Li, K. 2023. Multi-View Bipartite Graph Clustering With Coupled Noisy Feature Filter. *T-KDE*, 35(12): 1–13.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2017. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*.

Li, Z.; Huang, C.; Xia, L.; Xu, Y.; and Pei, J. 2022. Spatial-temporal hypergraph self-supervised learning for crime prediction. In *ICDE*.

Liang, K.; Liu, Y.; Zhou, S.; Tu, W.; Wen, Y.; Yang, X.; Dong, X.; and Liu, X. 2023a. Knowledge Graph Contrastive Learning Based on Relation-Symmetrical Structure. *T-KDE*.

Liang, K.; Meng, L.; Liu, M.; Liu, Y.; Tu, W.; Wang, S.; Zhou, S.; and Liu, X. 2023b. Learn from relational correlations and periodic events for temporal knowledge graph reasoning. In *SIGIR*, 1559–1568.

Liang, K.; Zhou, S.; Liu, Y.; Meng, L.; Liu, M.; and Liu, X. 2023c. Structure Guided Multi-modal Pre-trained Transformer for Knowledge Graph Reasoning. *arXiv preprint arXiv:2307.03591*.

Liang, Y.; Ouyang, K.; Jing, L.; Ruan, S.; Liu, Y.; Zhang, J.; Rosenblum, D. S.; and Zheng, Y. 2019. Urbanfm: Inferring fine-grained urban flows. In *SIGKDD*, 3132–3142.

Liu, M.; and Liu, Y. 2021. Inductive representation learning in temporal networks via mining neighborhood and community influences. In *SIGIR*, 2202–2206.

Liu, M.; Wu, J.; and Liu, Y. 2022. Embedding global and local influences for dynamic graphs. In *CIKM*, 4249–4253.

Liu, Y.; Liang, K.; Xia, J.; Yang, X.; Zhou, S.; Liu, M.; Liu, X.; and Li, S. Z. 2023a. Reinforcement Graph Clustering with Unknown Cluster Number. In *MM*, 3528–3537.

Liu, Y.; Liang, K.; Xia, J.; Zhou, S.; Yang, X.; ; Liu, X.; and Li, Z. S. 2023b. Dink-Net: Neural Clustering on Large Graphs. In *ICML*.

Liu, Y.; Tu, W.; Zhou, S.; Liu, X.; Song, L.; Yang, X.; and Zhu, E. 2022a. Deep Graph Clustering via Dual Correlation Reduction. In *AAAI*.

Liu, Y.; Xia, J.; Zhou, S.; Wang, S.; Guo, X.; Yang, X.; Liang, K.; Tu, W.; Li, Z. S.; and Liu, X. 2022b. A Survey of Deep Graph Clustering: Taxonomy, Challenge, and Application. *arXiv preprint arXiv:2211.12875*.

Liu, Y.; Zhang, R.; Guo, J.; de Rijke, M.; Chen, W.; Fan, Y.; and Cheng, X. 2023c. Black-Box Adversarial Attacks against Dense Retrieval Models: A Multi-View Contrastive Learning Method. *CIKM*.

Liu, Y.; Zhang, R.; Guo, J.; de Rijke, M.; Chen, W.; Fan, Y.; and Cheng, X. 2023d. Topic-Oriented Adversarial Attacks against Black-Box Neural Ranking Models. *SIGIR*.

Luo, X.; Tian, Z.; Zhang, T.; Yu, B.; Tang, Y. Y.; and Jia, J. 2023. PFENet++: Boosting Few-Shot Semantic Segmentation With the Noise-Filtered Context-Aware Prior Mask. *T-PAMI*.

Pan, B.; Demiryurek, U.; and Shahabi, C. 2012. Utilizing real-world transportation data for accurate traffic prediction. In *ICDM*, 595–604. IEEE.

Pan, Z.; Liang, Y.; Wang, W.; Yu, Y.; Zheng, Y.; and Zhang, J. 2019. Urban traffic prediction from spatio-temporal data using deep meta learning. In *SIGKDD*, 1720–1730.

Peng, B.; Tian, Z.; Wu, X.; Wang, C.; Liu, S.; Su, J.; and Jia, J. 2023. Hierarchical Dense Correlation Distillation for Few-Shot Segmentation.

Shao, Z.; Wang, F.; Xu, Y.; Wei, W.; Yu, C.; Zhang, Z.; Yao, D.; Jin, G.; Cao, X.; Cong, G.; et al. 2023. Exploring

Progress in Multivariate Time Series Forecasting: Comprehensive Benchmarking and Heterogeneity Analysis. *arXiv preprint arXiv:2310.06119*.

Shao, Z.; Zhang, Z.; Wei, W.; Wang, F.; Xu, Y.; Cao, X.; and Jensen, C. S. 2022. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. *arXiv preprint arXiv:2206.09112*.

Su, K.; Li, J.; and Fu, H. 2011. Smart city and the applications. In *ICECC*, 1028–1031. IEEE.

Tang, J.; Xia, L.; Hu, J.; and Huang, C. 2023. Spatio-Temporal Meta Contrastive Learning. In *CIKM*.

Tang, J.; Xia, L.; and Huang, C. 2023. Explainable Spatio-Temporal Graph Neural Networks. In *CIKM*, 2432–2441.

Thakoor, S.; Tallec, C.; Azar, M. G.; Azabou, M.; Dyer, E. L.; Munos, R.; Veličković, P.; and Valko, M. 2021. Large-scale representation learning on graphs via bootstrapping. *arXiv preprint arXiv:2102.06514*.

Tian, Z.; Zhao, H.; Shu, M.; Yang, Z.; Li, R.; and Jia, J. 2020. Prior Guided Feature Enrichment Network for Few-Shot Segmentation. *T-PAMI*.

Wang, H.; Kifer, D.; Graif, C.; and Li, Z. 2016. Crime rate inference with big data. In *SIGKDD*.

Wang, X.; Ma, Y.; Wang, Y.; Jin, W.; Wang, X.; Tang, J.; Jia, C.; and Yu, J. 2020. Traffic flow prediction via spatial temporal graph neural network. In *WWW*, 1082–1092.

Wei, W.; Huang, C.; Xia, L.; Xu, Y.; Zhao, J.; and Yin, D. 2022. Contrastive meta learning with behavior multiplicity for recommendation. In *WSDM*, 1120–1128.

Wei, W.; Huang, C.; Xia, L.; and Zhang, C. 2023. Multi-Modal Self-Supervised Learning for Recommendation. In *WWW*, 790–800.

Wei, W.; Xia, L.; and Huang, C. 2023. Multi-Relational Contrastive Learning for Recommendation. In *RecSys*, 338–349.

Wortley, R.; and Townsley, M. 2016. *Environmental criminology and crime analysis*. Taylor & Francis.

Wu, H.; Li, N.; Zhang, J.; Chen, S.; Ng, M. K.; and Long, J. 2024a. Collaborative contrastive learning for hypergraph node classification. *Pattern Recognition*, 146: 109995.

Wu, H.; Yan, Y.; and Ng, M. K.-P. 2023. Hypergraph collaborative network on vertices and hyperedges. *T-PAMI*.

Wu, H.; Yip, A.; Long, J.; Zhang, J.; and Ng, M. K. 2024b. Simplicial Complex Neural Networks. *T-PAMI*, 46(1): 561–575.

Wu, X.; Huang, C.; Zhang, C.; and Chawla, N. V. 2020a. Hierarchically structured transformer networks for fine-grained spatial event forecasting. In *WWW*, 2320–2330.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; and Zhang, C. 2020b. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *SIGKDD*, 753–763.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*.

Xia, J.; Wu, L.; Chen, J.; Hu, B.; and Li, S. Z. 2022a. Simgrace: A simple framework for graph contrastive learning without data augmentation. In *WWW*, 1070–1079.

Xia, J.; Wu, L.; Wang, G.; and Li, S. Z. 2022b. ProGCL: Rethinking Hard Negative Mining in Graph Contrastive Learning. In *ICML*. PMLR.

Xia, L.; Huang, C.; Huang, C.; Lin, K.; Yu, T.; and Kao, B. 2023. Automated Self-Supervised Learning for Recommendation. In *Proceedings of the ACM Web Conference 2023*, 992–1002.

Xia, L.; Huang, C.; Xu, Y.; Dai, P.; Bo, L.; Zhang, X.; and Chen, T. 2022c. Spatial-temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning. *arXiv preprint arXiv:2201.02435*.

Yang, Y.; Guan, Z.; Li, J.; Zhao, W.; Cui, J.; and Wang, Q. 2023a. Interpretable and Efficient Heterogeneous Graph Convolutional Network. *IEEE TKDE*, 35(2): 1637–1650.

Yang, Y.; Guan, Z.; Wang, Z.; Zhao, W.; Xu, C.; Lu, W.; and Huang, J. 2022. Self-supervised Heterogeneous Graph Pre-training Based on Structural Clustering. In *NeurIPS*, volume 35, 16962–16974.

Yang, Y.; Guan, Z.; Zhao, W.; Lu, W.; and Zong, B. 2023b. Graph Substructure Assembling Network with Soft Sequence and Context Attention. *IEEE TKDE*, 35(5): 4894–4907.

Yang, Y.; Yang, J.; Bao, R.; Zhan, D.; Zhu, H.; Gao, X.; Xiong, H.; and Yang, J. 2023c. Corporate Relative Valuation Using Heterogeneous Multi-Modal Graph Neural Network. *T-KDE*.

Yao, H.; Tang, X.; Wei, H.; Zheng, G.; and Li, Z. 2019. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In *AAAI*, volume 33, 5668–5675.

Yu, B.; Yin, H.; and Zhu, Z. 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*.

Yu, C.; Wang, F.; Shao, Z.; Sun, T.; Wu, L.; and Xu, Y. 2023a. DSformer: a double sampling transformer for multivariate time series long-term prediction. In *CIKM*.

Yu, C.; Yan, G.; Yu, C.; and Mi, X. 2023b. Attention mechanism is useful in spatio-temporal wind speed prediction: Evidence from China. *Applied Soft Computing*, 148: 110864.

Yu, C.; Yan, G.; Yu, C.; Yu, Z.; and Xiwei, M. 2023c. A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks. *Energy*, 263: 126034.

Zhang, J.; Zheng, Y.; and Qi, D. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, volume 31.

Zhang, Q.; Huang, C.; Xia, L.; Wang, Z.; Li, Z.; and Yiu, S. 2023. Automated Spatio-Temporal Graph Contrastive Learning. In *WWW*, 295–305.

Zhao, T.; Zhang, X.; and Wang, S. 2022. Exploring edge disentanglement for node classification. In *WWW*, 1028–1036.

Zheng, C.; Fan, X.; Wang, C.; and Qi, J. 2020. Gman: A graph multi-attention network for traffic prediction. In *AAAI*, volume 34, 1234–1241.

Zuo, Y.; Liu, G.; Lin, H.; Guo, J.; Hu, X.; and Wu, J. 2018. Embedding temporal network via neighborhood formation. In *SIGKDD*, 2857–2866.