Supplementary material of **HAGEN**: Homophily-Aware Graph Convolutional Recurrent Network for Crime Forecasting

Anonymous authors

In this supplementary material, we will provide more details on the experiment settings, implementations and more experiment results.

1 Experiment Settings

1.1 Data Collection and Description

We evaluated HAGEN on two real-world benchmarks collected in Chicago and Los Angeles by CrimeForecaster (Sun et al. 2021), details are as follows:

Table 1: Los Angeles City crime dataset statistics.

Category	Theft	Vehicle Theft	Burglary	Fraud	Assault	Sexual Offenses	Robbery	Vandalism
Counts	66,697	17,123	14,517	15,578	32,372	6,161	8,864	17,123

Table 2: Chicago crime dataset statistics.

Category	Theft	Criminal Damage	Narcotics	Robbery	Assault	Deceptive Practices	Burglary	Battery
Counts	56,695	28,589	21,607	9,632	16,692	14,085	13,103	48,824

- **Los Angeles City Crime Dataset**: The crime data is collected across 216 regions of Los Angeles County from 2010 to 2019, and all crimes are aggregated into 8 categories. Table 1 shows all eight crime categories and the total counts of each in our dataset.
- **Chicago Crime Dataset**: In order to compare with existing crime forecasting applications on Chicago's crime (Huang et al. 2019; Sun et al. 2021), we use the official dataset published by the **City of Chicago** (of Chicago 2021). This dataset includes all reported crime events that occurred in Chicago during 2015. Table 2 shows the eight categories of crime events and their counts.
- **Geographical Data**: Geographical data of each region is used to construct spatial graphs. The geographical information of Los Angeles regions sourced from **Census Bureau** (Bureau 2021), which only contains boundary information for 113 geographical neighborhoods within Los Angeles City. Later in experiments, we only focused on crime activities within these 113 regions whose geographical information is available. The information of 77 Chicago regions is collected from the **City of Chicago** (of Chicago 2021)
- **Demographic and Point of Interest (POI) Data**: The number of criminal activities can be influenced by external factors such as POI's. We collected the information (class, description, location) of 134 unique classes of POI's from **OpenStreetMap** (OpenStreetMap 2021). Examples of POI include police stations, fire stations, hospitals, etc.

1.2 Training Configuration and Evaluation Metrics

We evaluate the performance of HAGEN on two real-world datasets described above. In our experiments, we use the same "train-validation-test" setting as the previous work (Sun et al. 2021; Huang et al. 2019). We chronologically split the dataset as 6.5 months for training, 0.5 months for the validation, and 1 month for testing. We implemented our HAGEN with the Pytorch (Paszke et al. 2019) architecture on a computer with an Intel Xeon Silver 4210R CPU and a 32 GB RAM. To be specific, in our experiments, we set *learning rate* as 0.01, *the number of epochs* as 100 and the *batch size* as 15. For the vital hyperparameters in HAGEN, we use two stacked layers of RNNs. Within each RNN layer, we set 64 as the size of the hidden dimension. Moreover, we set the subgraph size of the sparsity operation as 50 and the saturation rate as 3. For the learning objective, we fix the trade-off parameter λ as 0.01, similar to the common practice of other regularizers. We utilize Adam optimizer (Kingma and Ba 2014) with learning rate annealing.

We utilize Micro-F1 (Grover and Leskovec 2016) and Macro-F1 (Lin et al. 2020) as general metrics to evaluate prediction performance across all crime categories, which are formulated as follows:

$$\begin{aligned}\text{Macro-F1} &= \frac{1}{C}\sum_{l=1}^{C}\frac{2\cdot TP_l}{2\cdot TP_l + FN_l + TP_l},\\ \text{Micro-F1} &= \frac{2\cdot\sum_{l=1}^{C} TP_l}{2\cdot\sum_{l=1}^{C} TP_l + \sum_{l=1}^{C} FN_l + \sum_{l=1}^{C} FP_l}.\end{aligned} \tag{1}$$

1.3 Baselines

We introduce the three types of baselines in the paper, which can be further broke down into five categories and are elaborated in detail as follows:

Time-series Forecasting (i.e., ARIMA). ARIMA (Contreras et al. 2003) is a traditional time-series prediction model for predicting future records given historical records.

Classical Machine Learning Methods (i.e., SVR, Decision Tree, and Random Forest). Epsilon-Support Vector Regression (SVR) (Chang and Lin 2011) is a supervised learning model for regression problems based on the RBF kernel function. Decision Tree (Safavian and Landgrebe 1991) is a non-parametric supervised learning model used for classification and regression by learning simple decision rules. Random Forest (Verikas, Gelzinis, and Bacauskiene 2011) is an ensemble learning method for classification by building multiple decision trees at training time and predicting the class. For the three basic machine learning models, we use historical crime records as features and the predicting crime events as labels.

Traditional Neural Network (Multi-layer Perceptron classifier and Recurrent Neural Network). Multi-layer Perceptron classifier (MLP) (Covington, Adams, and Sargin 2016) is a classical deep neural network that learns the non-linearity from the historical distributions of crime records. Long Short-term Memory (LSTM) (Gers, Schmidhuber, and Cummins 1999) and Gated Recurrent Unit (GRU) (Chung et al. 2014) are RNN variants with feedback connections. We leverage them to predict future crime occurrences based on historical crime records.

Spatial-temporal graph neural network (i.e., MTGNN and Graph WaveNet). We chose two spatio-temporal neural network methods designed for multivariate time-series forecasting for comparison. Graph WaveNet (GW) (Wu et al. 2019) is a spatial-temporal graph neural network that combines diffusion convolutions with 1D dilated convolutions. MTGNN (Wu et al. 2020) is the state-of-the-art spatial-temporal graph neural network in previous works, which integrates the adaptive graph structure, mix-hop graph convolution layers, and dilated temporal convolution layers.

Spatial-temporal learning models for crime forecasting (i.e., MiST and CrimeForecaster). MiST (Huang et al. 2019) learns both the inter-region temporal and spatial correlations. To the best of our knowledge, CrimeForecaster (CF) (Sun et al. 2021) is the most recent crime forecasting model, which is an end-to-end framework to model the dependencies between regions by geographical neighborhoods instead of grid partitions and captures both spatial and temporal dependencies using diffusion convolution layer and Gated Recurrent Network.

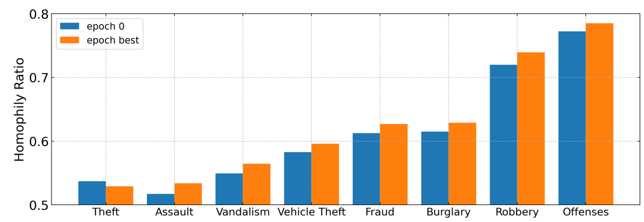
2 Experiment Results

2.1 Influence of Homophily-aware Constraint

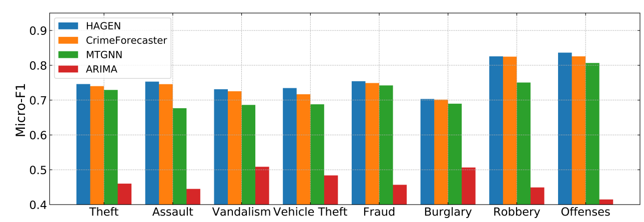
Heterophily problem significantly affects the performance of graph neural network (Zhu et al. 2020).

We can observe from Figure 1(a) that the graph learned by `HAGEN` (the orange bars) has a higher homophily level than the initial state, indicating that the homophily constraint enables our model to learn more homophilous graph structure effectively. In Figure 1(b), we show the performance comparison of each crime category for different models, where we sort crime categories by their homophily in ascending order like Figure 1(a).

According to the performance of each crime category, although previous graph-based models (e.g., CrimeForecaster, MTGNN) outperform pure sequential based model (e.g., ARIMA), which indicates the effectiveness of graph structure and spatial and temporal modeling in crime forecasting, their advantages over non-graph based models are more obvious for crime categories with high homophily ratio (e.g, offenses, robbery), and relatively faint for crime categories with low homophily ratio. On the contrary, with the homophily ratio constraint, `HAGEN` can achieve higher performance for low homophily ratio crime categories, which is elaborated in detail in supplementary materials. Such finding emphasizes the importance of incorporating the homophily constraint in crime forecasting.



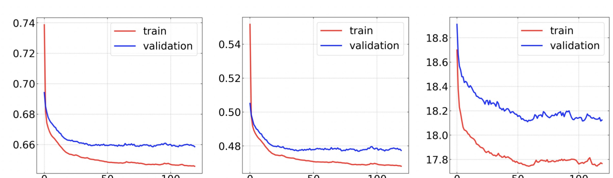
(a) Homophily ratio of different crime categories



(b) Comparison with baselines across different crime categories.

Figure 1: Influence of homophily-aware constraint.

We show changes of total loss, Cross-Entropy loss (CE Loss), and Homophily-aware Constraint loss (HR Loss) respectively for both training and validation sets in Figure 2. Both CE loss and HR loss decay with the training process.



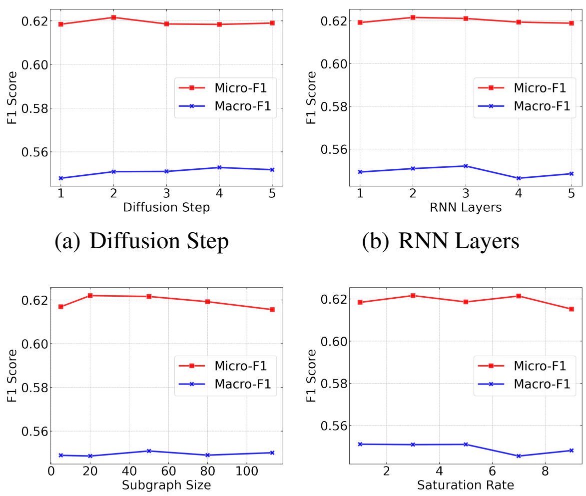
(a) Total Loss (b) CE Loss (c) HR Loss

Figure 2: Total loss, cross entropy loss and homophily constraint loss versus the number of training epoch.

2.2 Parameter Study

`HAGEN` preserves the spatial-temporal information by conducting diffusion process on the adaptive region graph. *Max*

Diffusion Step M reflects the depth of the spatial information that will be aggregated into each region node. Figure 3(a) shows that both the Micro-F1 and Macro-F1 of HAGEN achieve a performance gain when M increases from one to two as it receives more crime information from neighboring regions. However, the performance starts to drop when the max diffusion step reaches 3. The reason could be that deeper information of multi-hop connected nodes would add noise to crime forecasting. Similar patterns could also be observed in Figure 3(b), where *the number of RNN layers* reflects the depth of temporal information that the GRU network can capture. In addition, we show the performance of HAGEN under different hyperparameter settings of region graph learning, including *the subgraph size* that controls the sparsity of the region graph in Figure 3(c) and *the saturation rate* of the corresponding activation function in Figure 3(d).



(a) Diffusion Step (b) RNN Layers

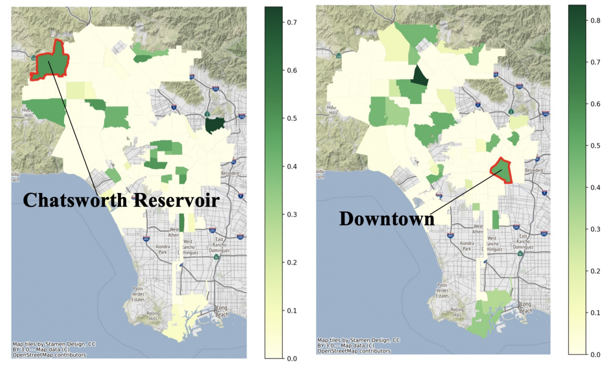
(c) Subgraph Size (d) Saturation Rate

Figure 3: Effect of parameters.

2.3 Model Explainability of Region-crime Dependency

We conduct case studies and present visualization to interpret the influence of the region-crime dependency in HAGEN.

In the region-crime dependency encoder, a matrix is learned to capture spatial and categorical correlations in crime forecasting of a specific category in a certain region. We show two cases, "Theft" and "Fraud" to grasp a better understanding of the captured correlations. Figure 4 depicts the dependency of "Theft" and "Fraud" on different regions learned from the region and crime category embedding (a darker color indicates higher dependency). It is interesting but not surprising to observe that "Theft" is dominating to Chatsworth Reservoir while "Fraud" is the dominating crime in Downtown. Intuitively, the small population in Chatsworth Reservoir creates more chances for thievery without witness, and the mass population in Downtown increases the difficulty of tracking swindlers.



(a) Theft (b) Fraud

Figure 4: Region-crime dependency of Theft and Fraud.

References

Bureau, C. 2021. Census Bureau. https://www.census.gov/.

Chang, C.-C.; and Lin, C.-J. 2011. LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3): 1–27.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Contreras, J.; Espinola, R.; Nogales, F. J.; and Conejo, A. J. 2003. ARIMA models to predict next-day electricity prices. *IEEE transactions on power systems*, 18(3): 1014–1020.

Covington, P.; Adams, J.; and Sargin, E. 2016. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, 191–198.

Gers, F. A.; Schmidhuber, J.; and Cummins, F. 1999. Learning to Forget: Continual Prediction with LSTM. In *Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale*. IET.

Grover, A.; and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 855–864.

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 Web Conference (WWW)*, 717–728.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lin, Z.; Lyu, S.; Cao, H.; Xu, F.; Wei, Y.; Samet, H.; and Li, Y. 2020. HealthWalks: Sensing Fine-grained Individual Health Condition via Mobility Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4): 1–26.

of Chicago, T. C. 2021. The City of Chicago. https://data.cityofchicago.org/Public-Safety/Crimes2015/vwwp-7yr9.

OpenStreetMap. 2021. OpenStreetMap. https://download.geofabrik.de/north-america/us/california.html.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.;

et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*.

Safavian, S. R.; and Landgrebe, D. 1991. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3): 660–674.

Sun, J.; Yue, M.; Lin, Z.; Yang, X.; Nocera, L.; Kahn, G.; and Shahabi, C. 2021. CrimeForecaster: Crime Prediction by Exploiting the Geographical Neighborhoods' Spatiotemporal Dependencies. In *The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 52–67.

Verikas, A.; Gelzinis, A.; and Bacauskiene, M. 2011. Mining data with random forests: A survey and results of new tests. *Pattern recognition*, 44(2): 330–349.

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

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In *Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19*.

Zhu, J.; Yan, Y.; Zhao, L.; Heimann, M.; Akoglu, L.; and Koutra, D. 2020. Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs. *Conference on Neural Information Processing Systems (NeurIPS)*, 33.