2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

SPATIAL-TEMPORAL INDOOR CROWD INTERPOLATION AND PREDICTION WITH GRAPH NEURAL NETWORKS

Cong Huang[1], Jack C.P. Cheng[1], Mingkai Li[2], Zhaoji Wu[1], Fangli Hou[1], and Chengyao Peng[3]
[1]The Hong Kong University of Science and Technology, Hong Kong, China
[2]National University of Singapore, Singapore
[3]The Hong Kong Polytech University, Hong Kong, China

Abstract

Crowd analysis is crucial for smart city applications, such as space management, safety, and well-being. While studies address spatial, temporal, and spatial-temporal dimensions, spatial-temporal data granularity is often constrained by the high cost and maintenance of high-resolution sensor deployment. Existing approaches focus heavily on prediction and rely on predefined graph structures, overlooking dynamic spatial-temporal dependencies and underperforming in interpolation tasks. To address this, we propose a hybrid model, the Spatial-Temporal Attention-based Interpolation and Prediction Graph Convolutional Gated Recurrent Unit (STAIP-GCGRU), which integrates interpolation and prediction tasks to deliver robust spatial-temporal predictions, even with incomplete graph structures.

Introduction

Crowd monitoring, analysis, and prediction leverage data and algorithms to monitor and predict the behaviors and actions of a group of people. By examining the crowd patterns, trends, and periodicity, researchers can accurately predict various crowd profiles precisely. Indoor scenario requires higher privacy levels compared with outdoor environments. Consequently, direct monitoring via Closed Circuit Television (CCTV) is often restricted, prompting the development of indirect measurement via different surrogate approaches. For instance, He and Chan (2016) employ WiFi access points to estimate the indoor positions of occupants, while Schauer et al. (2014) utilize Bluetooth to measure indoor occupancy levels, offering higher-resolution occupant detection with lower energy consumption.

Predictive crowd management provides valuable opportunities for informed decision-making and anomaly detection, enabling managers to effectively perform tasks such as space management, facility maintenance, indoor environment analysis, and occupant-centric ventilation control. Recognizing its significance, researchers from diverse fields have extensively studied crowd prediction. Traditional models often rely on statistical approaches. For example, Yoshida et al. (2018) proposed a new Hidden Markov Model (HMM) that takes advantage of monitored environmental parameters to predict occupancy levels. Similarly, Ding et al. (2021) employed a Gaussian distribution model to analyze occupant patterns, offering enhanced energy-saving strategies tailored to different types of buildings. These studies underscore the potential of predictive models to optimize resource utilization and improve occupant comfort.

The development of machine learning techniques provides more advanced models in crowd prediction, Sun et al. (2015) develop one hybrid model combined Support Vector Machine (SVM) with Wavelet decomposition, achieving superior performance in short-term crowd flow forecasting in a subway station. Xu and Qiu (2016) adopt Random Forest (RF) model in image feature selection and random projection for crowd density estimation, compared with traditional regression models, the efficiency and accuracy have been improved. Wu et al. (2018) utilize eXtreme Gradient Boost (XGBoost) and compare with other Gradient Boosting Tree (GBT) models and RF models, XGBoost outperforms in the campus crowd flow prediction tasks. However, machine learning approaches face significant limitations, particularly in their reliance on extensive feature engineering, which can be time-consuming and require domain expertise. Additionally, these models often struggle with scalability, making them less effective when applied to large-scale datasets. Furthermore, given the high dimensionality and complexity of spatial-temporal data, traditional machine learning models exhibit limited capacity in capturing spatial dependencies and temporal dynamics, reducing their effectiveness in tasks that require the integration of both spatial and temporal information. These challenges highlight the need for more advanced methods capable of handling the intricacies of spatial-temporal data seamlessly.

Deep learning models, which utilize neural network architectures, are considered a subfield of machine learning. Unlike traditional Artificial Neural Network (ANN) models, deep learning incorporates multi-layer structures, enabling it to capture more complex patterns and relationships within diverse datasets. Sudo et al. (2017) propose a column-structured Deep Neural Network (DNN) model that outperforms different baseline machine learning models in predicting the occupant density. Follow-up studies try to extract various dimensional features from the dataset, with particular emphasis on the spatial and tem-

poral dimensions of crowd data. Recurrent Neural Network (RNN) and its variants, tailored to deal with sequences and time series data, have been widely adopted in this context. Poon et al. (2022) structured the occupant dataset with two-dimensional data and successfully predicted the occupant for the long time gap using Long-Short Term Memory (LSTM) model. The Gated Recurrent Unit (GRU) model is often regarded as a simplified version of the LSTM model, with a less complicated cell structure. He et al. (2019) adopted a GRU model in the crowd flow prediction, compared with traditional machine learning models, the loss in the prediction has greatly reduced. It is also obvious that, when exploiting the temporal patterns, the prediction performance has been greatly enhanced compared with the models that do not integrate temporal feature extraction modules. With advancements in sensing technologies and the growing demand for spatial big data, spatial modeling has increasingly integrated temporal components, leading to the development of spatial-temporal modeling approaches. These methods effectively capture spatial dependencies and temporal dynamics, enabling more accurate predictions and insights for applications like urban planning, environmental monitoring, and transportation management. At the early stage, Convolutional Neural Networks (CNNs), While predominantly known for image processing, the ability of CNNs for 2-Dimensional data processing has also been generalized and adapted to extract spatial features. Combined with RNN structures, ConvRNNs and their variants have been adopted in spatial-temporal prediction. Jiang et al. (2019) developed a ConvLSTM-based encoder-decoder structure and predicted crowd density during events, compared with spatial models, this approach shows superior performance. The graph-based machine learning technology and Graph Neural Networks (GNNs) further allowed the capture of dependencies within structured topology data. The scalability of GNN models has been greatly enhanced compared with CNN-based models. Similarly, hybrid models have been developed. Cheng et al. (2022) developed a hybrid model based on Graph Convolutional Gated Recurrent Unit (GCGRU) and predicted the campus crowd density considering holiday and non-holiday features, compared with other spatial and temporal baseline models, the enhancement in prediction accuracy shows robustness in different time gaps. However, many GNN studies suffer from limitations such as ad hoc graph construction and over-reliance on spatial distance for defining topology.

To tackle this problem, transformer models, which excel at capturing long-range dependencies and contextual relationships in data, are widely utilized in the spatial-temporal data, especially for large-scale geospatial data. Jiang et al. (2023) proposed one Propagation Delay-Aware Dynamic Long-range Transformer (PDFormer) for traffic prediction to tackle the propagation delay in the traffic network, which significantly enhances the performance by utilizing the attention mechanism to capture remote dependencies. Wang et al. (2024) propose a hybrid modem called spatial-temporal graph transformer (STGformer), which addresses the large-scale complex spatial-temporal interactions in a hybrid model structure, combining the capability of GCN and transformer. The STGformer outperforms aforementioned PDFormer in different traffic prediction tasks, validating the effectiveness of the unique message-passing mechanism designed in the graph-based models.

The major limitations of the studies mentioned before can be summarized as follows: 1) Existing graph-based spatial-temporal prediction methods heavily rely on the predefined graph structure and ignore the spatial-temporal dynamics, which hinders the performance of GNN in the sense of interpolation and prediction;

2) Spatial-temporal interpolation studies have been relatively under-explored by researchers, despite the noteworthy challenge posed by spatial sparsity in spatial-temporal data. Therefore, in this paper, we propose a spatial-temporal graph prediction and interpolation framework. Leveraging the information-rich Building Information Modeling (BIM) and geo-referenced IoT sensors for crowd density interpolation and prediction, the data is analyzed within spatial-temporal dimensions using a GNN-based framework;

3) Although some studies have utilized attention-based models like transformer, which is good at dealing with sequential data, there is a lack of local awareness of spatial information.

To address these challenges, we propose a hybrid model called the Spatial-Temporal Attention-based Interpolation and Prediction Graph Convolutional Gated Recurrent Unit(STAIP-GCGRU), which integrates Graph Attention Networks (GAT) to capture spatial-temporal dependencies and a Graph Convolutional Gated Recurrent Unit (GCGRU) to validate and update the learned representations. In the following sections, the detailed information about our proposed methodology is shown in Section 2. A case study is presented to demonstrate the effectiveness of the proposed methodology in Section 3. Finally, conclusions are provided in Section 4.

Methodology

The proposed framework for the spatial-temporal prediction and interpolation of crowd density is illustrated in Figure 1. Crowd density data are retrieved from the IoT sensor portal via an API. The raw data, originally in JSON format, are processed during dataset construction. The sensor data are organized based on sensor location ID and timestamp, then converted into CSV format for further use. As for spatial graph construction, BIM models function as effective tools in various building management applications. When integrated with IoT technologies, BIM models facilitate the geo-referencing of sensor locations and provide essential indoor topological information for spatial graph construction.

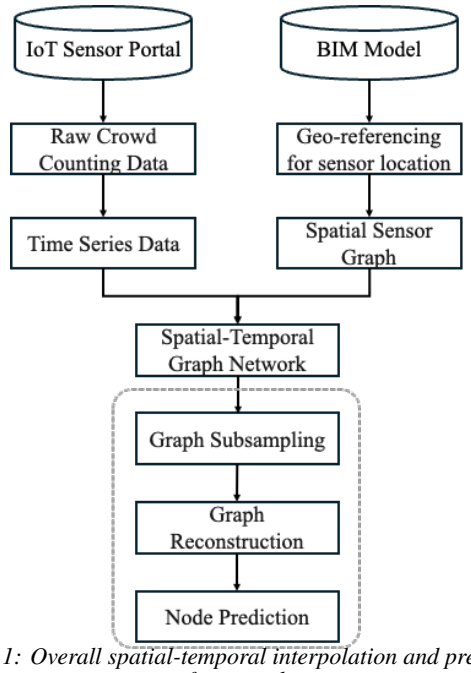

Figure 1: Overall spatial-temporal interpolation and prediction framework

Spatial-Temporal Graph Formulation

After the integration of the crowd density time series and the spatial sensor graph structure, the initialized spatial-temporal graph network, noted as G_0 is formulated: $G_0 = G(V,E)$, where V denotes different nodes, $V = \{V_1, V_2, ..., V_n\}, V \in R^{n \cdot t}$, where $V_i = \{x_1^i, x_2^i, ..., x_t^i\}$ is the time series data measured by sensors, n is the total number of nodes, and t is the corresponding time steps. E denotes the set of edges that define the connectivity of different pairs of nodes, with the edge weight calculated using inverse distance weighting, as shown in Equation (1), d_{ij} is the distance between the sensors, for simplicity, the α is equal to 2 according to the common practice values. The initialized graph G_0. In our study, interpolation and prediction tasks are performed. Rather than keeping the original graph topology, masking, and sub-graph sampling are required. The node masking can be defined as M, for simplicity, binary node masking is utilized. The masked node can be represented as V_M, and it can be calculated as shown in Equation (2), where $\odot$ is the element-wise multiplication. The binary masking first hides the crowd counting data at the masked sensor locations, the masked time series is noted as X_M.

$$Ew_{ij} = \frac{1}{d_{ij}^{\alpha}} \tag{1}$$

$$V_M = V \odot M \tag{2}$$

After masking, the sub-graph structure will be defined based on the residual node connections. Dynamic graph technology is applied. This approach enables the extraction of more complex node interactions while avoiding the creation of additional separate graphs, which can hinder GNN model training due to the importance of the message-passing mechanism in GNNs. In the following GNN model training tasks, our GNN model can be formulated as shown in the Equation (3). In the Graph representation learning tasks, we aim to learn a function to predict masked node states X_M and future node states Y, based on the reconstructed graph topology.

$$Y, X_M = f(X, G_0, M) \tag{3}$$

Proposed Spatial-Temporal GNN model

Our proposed STAIP-GCGRU is shown in Figure 2, this model performs graph interpolation tasks and node states prediction in the single model. The spatial graph is initially defined with proximity relationships in its structure. Next, graph sub-sampling with node masking is performed to construct a sparse graph signal, which often results in the incomplete graph topology. To address this, sequential similarity analysis is firstly conducted for initial reconstruction of the graph topology, avoiding the formulation of different subgraphs, as described in Equation (4).

$$\text{Cosine similarity} = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \tag{4}$$

where $\mathbf{a}$ and $\mathbf{b}$ represent the feature vectors of two different nodes, and $\|\mathbf{a}\|$ and $\|\mathbf{b}\|$ are the Euclidean norms. The cosine similarity between node pairs is assigned as a new edge attribute.

Since the sequential similarity is not enough to embed complex spatial-temporal information, we employ a subsequent Graph Attention Network (GAT) module, introduced by Veličković et al. (2017), is then applied to compute a normalized attention coefficient. Unlike the traditional Graph Convolutional Network (GCN) model, which utilizes a simple mean pooling approach for neighborhood aggregation, GAT dynamically assigns learnable attention weights to each neighbor, allowing the model to focus on more relevant neighbors during aggregation. By restricting aggregation to first-hop neighborhoods and weighting neighbors adaptively, GAT mitigates the over-smoothing issues common in GCN-based approaches, where equal treatment of neighbors often leads to loss of discriminative features during training.

After applying the GAT module, the GCGRU model, proposed by Seo et al. (2018), is utilized to validate the graph structure and predict both future states and masked node states within the graph representation learning framework. The GNN-RNN architecture has demonstrated its effectiveness in capturing spatial information and handling temporal or sequential prediction tasks. Compared with the spatial model including GCN and GAT, this structure ensures a comprehensive understanding on the temporal information in the model. Although some studies tried on the dynamic topologies, such as Temporal Graph Networks TGN (Rossi et al., 2020), the main focus of the model is on temporal dynamics. In our proposed model, the space and time information is integrated in the GNN-RNN structure, the GCGRU modules. The cell structure can be ex-

plained as shown in Figure 2, firstly, a graph convolutional module is applied to encode the input x_t with the graph structure. This is followed by integrating the graph convolutional modules with the GRU cell structure. Similar to the original GRU, the GCGRU model also employs gate control mechanisms to handle the sequential data. Finally, a Multi-Layer Perceptron (MLP) is applied before the output layer, leveraging the neural network structure to extract information from the hidden states produced by the graph. It also performs node-wise projection to address issues with inconsistent dimensions.

In addition, it is also worth noting that, there are two scenarios to consider in graph sub-sampling and reconstruction. In the first scenario, node filtering does not result in disconnected sub-graphs, ensuring that all sampled nodes remain connected. In the second scenario, subsampling leads to the formation of multiple sub-graphs, potentially including isolated single nodes. Since the message-passing mechanism is the core of graph representation learning, it is not particularly effective for learning the underlying graph structure itself. This limitation is why we incorporate cosine similarity to measure relationships between different nodes, ensuring a more meaningful dynamic graph construction.

Evaluation Metrics

The findings of this study are evaluated using node prediction and interpolation tasks in graph representation learning. Our analysis predicts the states of time-series nodes. To assess prediction precision, we use metrics commonly employed in regression analysis, including Mean Absolute Error (MAE) and R-square (R^2). Considering two different graph representation learning tasks conducted in the proposed model, interpolation loss for the masked nodes and prediction loss are calculated separately first, as defined in Equation (5) and Equation (6). Finally, two types of loss are integrated as shown in Equation (7). In the equations, y_i represents the actual value, $\bar{y}$ is the mean of the exact values, $\hat{y}_i$ is the predicted value, $|M|$ and $|S|$ denotes the number of interpolated nodes and the number of nodes in prediction, and α is the hyper-parameter to balance the importance of the two different tasks. Additionally, the R^2 value is added to evaluate the goodness of model fitting and capture the model's relative performance, as shown in Equation (8), similarly, the R^2 is calculated separately in the two tasks and aggregated using α value.

$$L_{\text{prediction}} = \frac{1}{|S|} \sum_{i \in S} |\hat{y}_i - y_i| \tag{5}$$

$$L_{\text{interpolation}} = \frac{1}{|M|} \sum_{i \in M} |\hat{y}_i - y_i| \tag{6}$$

$$L_{\text{total}} = \alpha \cdot L_{\text{prediction}} + (1 - \alpha) \cdot L_{\text{interpolation}} \tag{7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{8}$$

Case Study

Dataset pre-processing

A case study is conducted in the HKUST campus. Some representative locations are selected for the model validation, including the library (Lib g, Lib 1, Lib lg1, Lib lg3, and Lib lg4), the atrium, the canteens (Cateen 1 and Cateen 7), and the academic building area. The selected locations are high-traffic areas in the campus and the crowd movement is frequent. These areas represent diverse spatial configurations and usage patterns, offering a rich and varied time series for testing the effectiveness of the proposed STAIP-GCGRU framework. Additionally, these areas are critical to campus operations, making them particularly relevant for real-world applications such as crowd management and facility optimization. In total, 9 sensing locations are selected to provide a comprehensive representation of the campus environment while maintaining practical coverage for data collection. In general, the workflow of the spatial graph generation is shown in Figure 3. The HKUST Digital Twin project provides the campus BIM models with informative semantics. From the campus BIM model, multiple floorplans can be exported, and integrated with the geo-referenced IoT sensors, with the marked sensor locations in the BIM model and the topological connectivity information, the spatial sensor graph is built. The initialized spatial graph represents the sensors and their corresponding locations in the graph embedding space, and accordingly, the node features are different crowd density time series recorded by sensors. The sensor data can be requested from the HKUST API portal (QU, 2021), the resolution of the crowd density data is 6-min. A 2-month dataset is constructed, considering the sensor data quality and completeness, crowd-counting data in 2018 October and November are selected, after the dataset processing, There are 14640 timestamps in the dataset. After the request of spatial-temporal crowd sensor data, crowd time series are integrated as node features, combined with the spatial graph topology, and the initialized spatial-temporal graph is built. Our proposed STAIP-GCGRU aims to provide robust forecasting and interpolation for the spatial-temporal data. Thus, a random masking is assigned and the sub-graph is generated as the input of the model. The dataset is divided into training and validation sets, with a 4:1 ratio for training to validation. Accordingly, there are 11702 timestamps for the training set and 2918 timestamps for the testing set.

Model training and graph representation learning

The GNN models training is conducted on a device with Intel i7-11800 CPU and laptop version GPU NVIDIA GeForce RTX 3070. The prediction time is around 23 seconds an epoch. For comparison with other baseline models, the GCN model and raw GCGRU model have been selected as baseline models. And each model is trained for 100 epochs. The sub-sampling of the graph dataset, the masked node states are filtered out from the

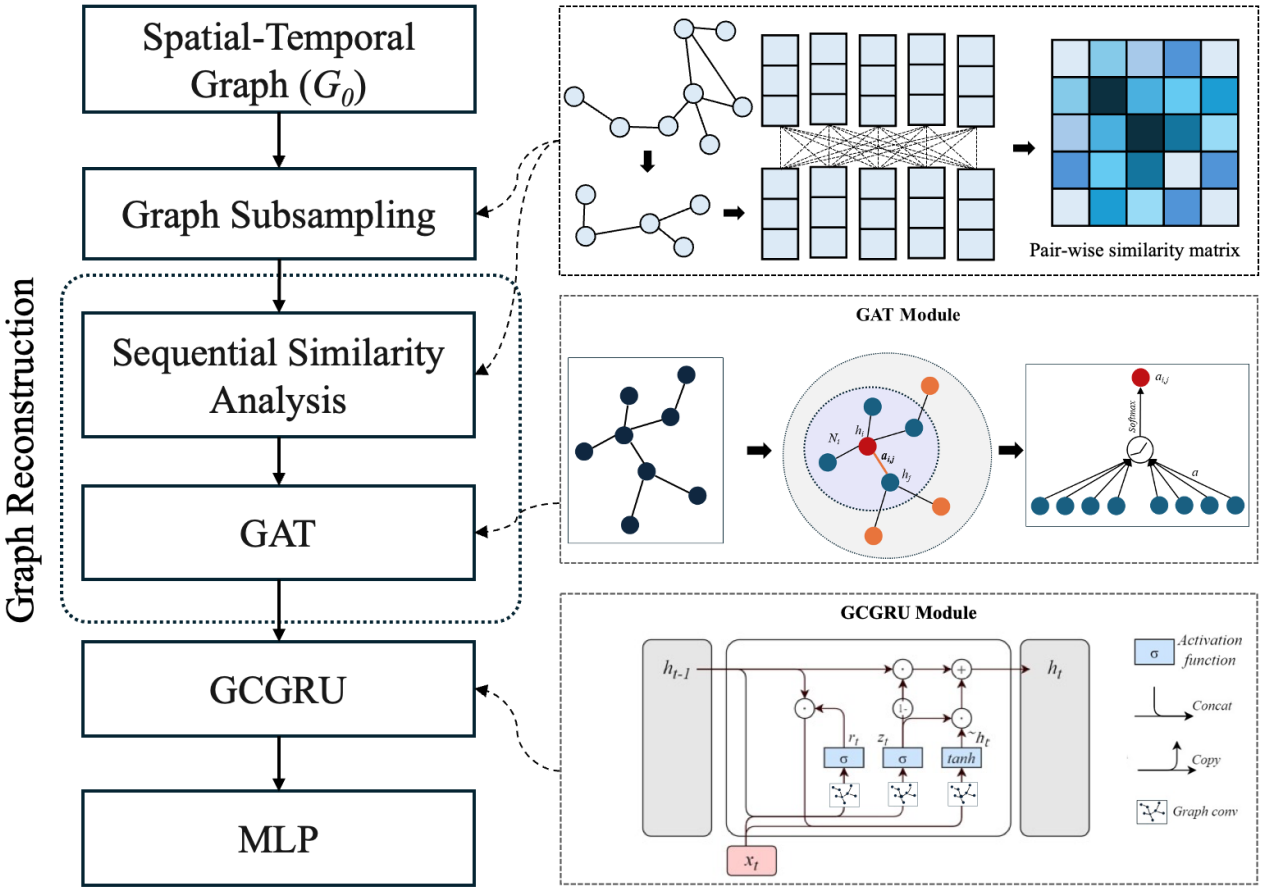

Figure 2: The proposed STAIP-GCGRU model structure

input, and utilized as one output, noted as X_M for graph reconstruction, and the future node states Y are also predicted. To validate the effectiveness of the proposed model, the graph structure 'damage' situation, which formulates several sub-graphs even single nodes, is considered in the case study to enhance the difficulty of the GNN training. The loss curve comparison of GCGRU and our proposed STAIP-GCGRU is shown in Figure 4. Although the same learning rate is used throughout the training process, the loss graph ofSTAIP-GCGRU shows that the training is not very stable compared with GCGRU training. This instability arises due to the adaptive changes in the edge connections, which affect the message-passing mechanism adaption and validation.

Quantitative model training results are shown in Table 1 and Table 2, showing the MAE loss and R^2, respectively. The mask rate is set from 22% to 44% to validate the model's robustness when dealing with different conditions. As released from tables, the model's MAE loss shows similar in the 22% and 33% missing rate conditions and suffers a much larger loss when dealing with 44% missing nodes. Compared with raw GCGRU, when handling 22% and 33% masked graphs, our proposed model can enhance the prediction accuracy by above 10%, and lower the MAE loss by 12 people. When the mask rate increases to 44% and more single nodes in the graph appear, our proposed model also shows enhanced robustness, with an enhanced prediction accuracy by 20% and the MAE loss decreased by 25 people. It is also noticed from the results that it is relatively easier for a simple model (i.e., GCN model) to converge within the same epoch number, the complex GCGRU graph model fails to deal with incomplete graph structure and performs worse than the basic GCN model.

Table 1: MAE loss comparison at different node masking rate

Models	22% Missing	33% Missing	44% Missing
GCN	17.18	23.02	58.45
GCGRU	26.43	28.14	72.12
STAIP-GCGRU	14.48	14.50	46.50

The predicted and interpolated curves are shown in the figures below. In this comparison, we take the 33% masking condition as an example, in this task, nodes with ID 3, 4, and 7 are masked. The performance of GCN and STAIP-GCGRU is compared. It is obvious that taking the prediction tasks, as shown in Figure 5, the performance of the different models is similar and accurate, the major dif-

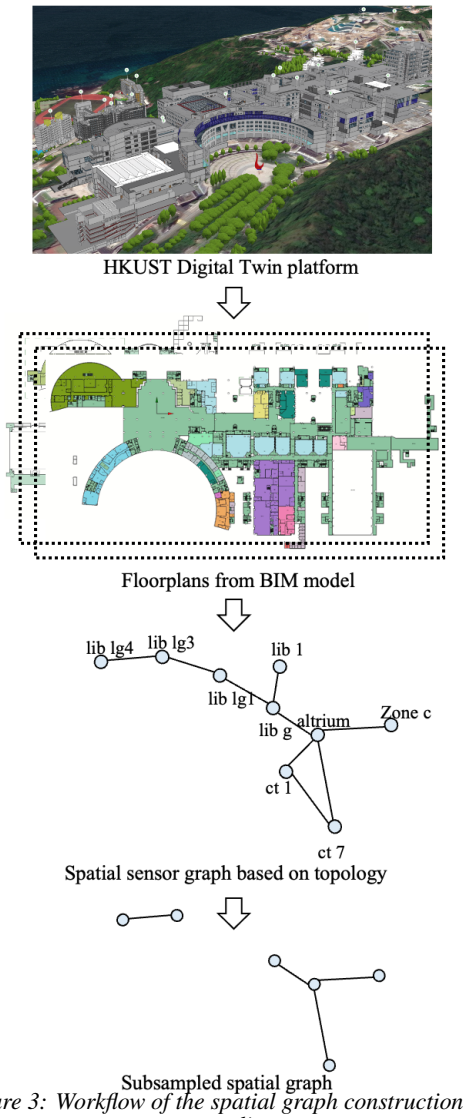

Figure 3: Workflow of the spatial graph construction and sampling

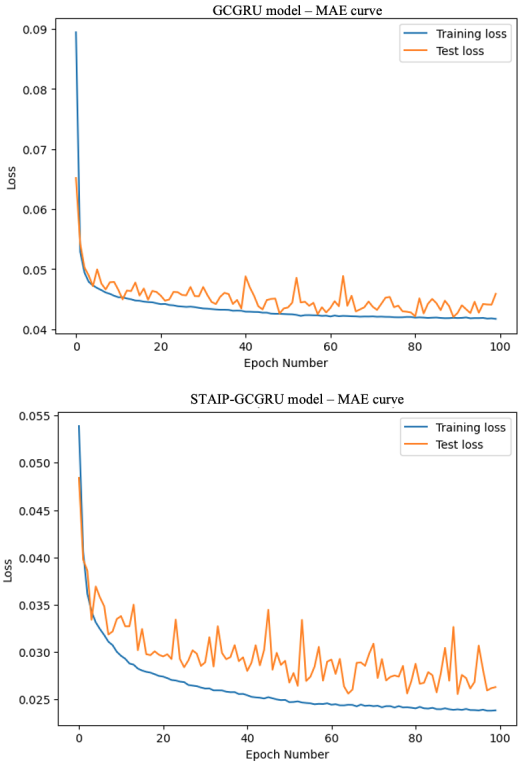

Figure 4: MAE loss during the GNN model training

Table 2: R^2 loss comparison at different node masking rate

Models	22% Missing	33% Missing	44% Missing
GCN	94.34%	93.16%	60.12%
GCGRU	85.41%	84.98%	53.46%
STAIP-GCGRU	97.67%	94.99%	73.70%

ference is the interpolation task, as shown in Figure 6, our proposed STAIP-GCGRU shows superior performance when dealing with the long-missing/masked sequence with its well defined dynamic GNN-RNN structure.

Conclusions

This paper introduces a novel integrated framework, the STAIP-GCGRU model, designed for spatial-temporal prediction and interpolation tasks. The proposed framework effectively combines graph reconstruction and node prediction tasks within a single model. Its architecture includes sub-graph sampling, similarity-based reconnection, and GAT modules for neighborhood information aggregation, followed by a GCGRU-MLP structure for node prediction as output. The capability of the designed model structure is validated through the designed tasks. Another distinct contribution of this study is the approach to graph construction. Unlike traditional static graph methods in spatial-temporal studies, which primarily rely on physical proximity, our framework initializes the graph based on spatial connectivity and adaptively updates node connections using similarity and the graph attention mechanism. This is particularly crucial in interpolation tasks, where sub-sampling can lead to disconnections, potentially increasing loss. By addressing this issue, our approach significantly enhances prediction accuracy. It enables effective mapping and prediction of spatial-temporal data, improving accuracy while compensating for missing values and spatial sparsity in a data-driven manner. In real-world applications such as building automation and management, a dynamic graph framework offers greater flexibility by adapting to changing connectivity patterns. This is especially beneficial in scenarios where node relationships evolve over time. The advantage of our approach is evident when comparing the performance of the raw GCGRU model with our proposed STAIP-GCGRU. The model's accuracy improves significantly by 20%, highlighting the impact of incorporating dynamic updates. An incomplete graph structure can lead

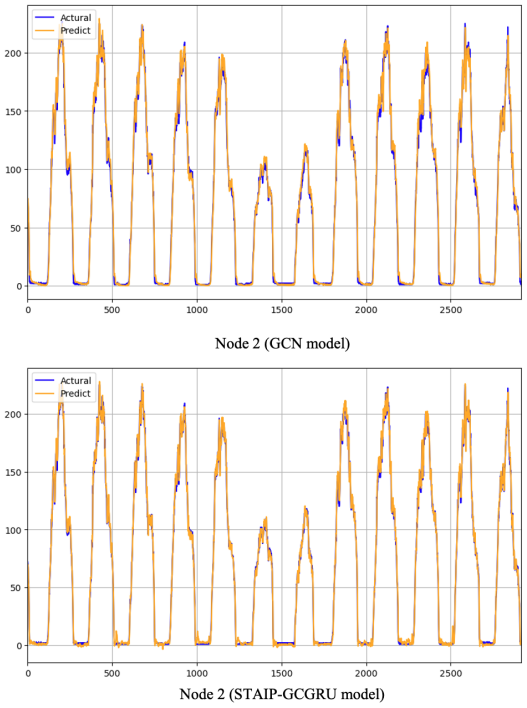

Figure 5: Predicted time series at Lib g (Node 2)

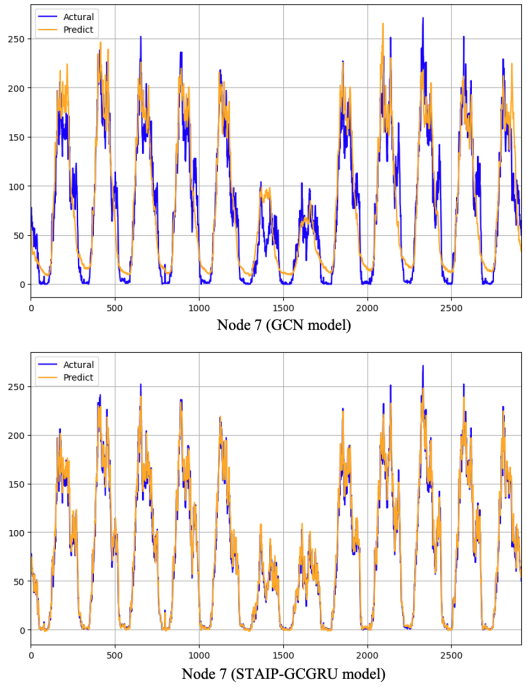

Figure 6: Interpolated time series at canteen 1 (Node 7)

to substantial loss during GNN model training, sometimes causing GCGRU to underperform compared to a basic GCN model (i.e., An average 8% accuracy loss) However, by integrating dynamic updating mechanisms, including similarity-based connections and the GAT module, our approach effectively mitigates this issue and enhances overall model performance.

In the future, research should be expanded to incorporate a wider range of crowd profiles, including crowd density, behavior, trajectories, and other relevant features. Besides, due to the limit of data acquisition, this study only conducts on a 9 nodes dataset, in the follow-up studies, more dataset considering more nodes would be included to validate the effectiveness of this model structure. Additionally, to improve the model's robustness, incorporating calendar features could be valuable, particularly for long-term predictions, as crowd dynamics are often influenced by various calendar-related factors.

References

Cheng, J. C., Ho Poon, K., and Kok-Yiu Wong, P. (2022). Long-time gap crowd prediction with a two-stage optimized spatiotemporal hybrid-gcgru. Advanced Engineering Informatics, 54:101727.

Ding, Y., Chen, W., Wei, S., and Yang, F. (2021). An occupancy prediction model for campus buildings based on the diversity of occupancy patterns. Sustainable Cities and Society, 64:102533.

He, J., Wang, J., and Luo, Y. (2019). Deep architectures for crowd flow prediction. In Proceedings of the 2019 2nd International Conference on Data Science and Information Technology, DSIT 2019, page 236–241, New York, NY, USA. Association for Computing Machinery.

He, S. and Chan, S.-H. G. (2016). Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons. IEEE Communications Surveys & Tutorials, 18(1):466–490.

Jiang, J., Han, C., Zhao, W. X., and Wang, J. (2023). Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In AAAI. AAAI Press.

Jiang, R., Song, X., Huang, D., Song, X., Xia, T., Cai, Z., Wang, Z., Kim, K.-S., and Shibasaki, R. (2019). Deepurbanevent: A system for predicting citywide crowd dynamics at big events. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, page 2114–2122, New York, NY, USA. Association for Computing Machinery.

Poon, K. H., Wong, P. K.-Y., and Cheng, J. C. (2022). Long-time gap crowd prediction using time series deep learning models with two-dimensional single attribute inputs. Advanced Engineering Informatics, 51:101482.

QU, H. (2021). Pulse | a data-driven smart campus. Accessed: Oct. 10, 2024.

Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., and Bronstein, M. (2020). Temporal graph networks for deep learning on dynamic graphs. In ICML 2020 Workshop on Graph Representation Learning.

Schauer, L., Werner, M., and Marcus, P. (2014). Estimating crowd densities and pedestrian flows using wi-fi and bluetooth. In Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, MOBIQUITOUS '14, page 171–177, Brussels, BEL. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Seo, Y., Defferrard, M., Vandergheynst, P., and Bresson, X. (2018). Structured sequence modeling with graph convolutional recurrent networks. In Neural Information Processing, volume 11301 of Lecture Notes in Computer Science, pages 362–373. Springer.

Sudo, A., Teng, T.-H., Lau, H. C., and Sekimoto, Y. (2017). Predicting indoor crowd density using column-structured deep neural network. In Proceedings of the 1st ACM SIGSPATIAL Workshop on Prediction of Human Mobility, PredictGIS'17, New York, NY, USA. Association for Computing Machinery.

Sun, Y., Leng, B., and Guan, W. (2015). A novel wavelet-svm short-time passenger flow prediction in beijing subway system. Neurocomputing, 166:109–121.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.

Wang, H., Chen, J., Pan, T., Dong, Z., Zhang, L., Jiang, R., and Song, X. (2024). Stgformer: Efficient spatiotemporal graph transformer for traffic forecasting.

Wu, C., Yin, T., Ge, S., and Yu, K. (2018). Ensemble learning for crowd flows prediction on campus. In Smart Computing and Communication: Second International Conference, SmartCom 2017, Shenzhen, China, December 10-12, 2017, Proceedings 2, pages 103–113. Springer.

Xu, B. and Qiu, G. (2016). Crowd density estimation based on rich features and random projection forest. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1–8.

Yoshida, M., Kleisarchaki, S., Gtirgen, L., and Nishi, H. (2018). Indoor occupancy estimation via location-aware hmm: An iot approach. In 2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), pages 14–19.