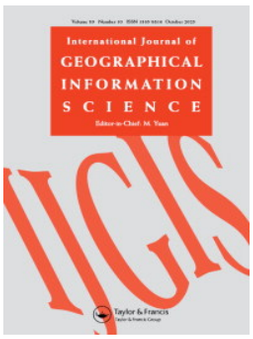

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Predicting human-activity intensity in urban areas with a prior-enhanced probabilistic-deterministic model

Cheng Zhong, Sheng Wu, Peixiao Wang, Hengcai Zhang, Shifen Cheng & Feng Lu

To cite this article: Cheng Zhong, Sheng Wu, Peixiao Wang, Hengcai Zhang, Shifen Cheng & Feng Lu (23 Sep 2025): Predicting human-activity intensity in urban areas with a prior-enhanced probabilistic-deterministic model, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2025.2562250

To link to this article: https://doi.org/10.1080/13658816.2025.2562250

Published online: 23 Sep 2025.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2025.2562250

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Predicting human-activity intensity in urban areas with a prior-enhanced probabilistic-deterministic model

Cheng Zhong[a], Sheng Wu[a], Peixiao Wang[b,c], Hengcai Zhang[b,c], Shifen Cheng[b,c] and Feng Lu[b,c]

[a]The Academy of Digital China (Fujian), Fuzhou University, Fuzhou, China; [b]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, China; [c]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

Although numerous models have been proposed to predict the intensity of human activities in urban areas, two major issues hamper the performance of existing models: (1) fail to incorporate appropriate prior knowledge instrumental for improving accuracy and interpretability; (2) fail to integrate probabilistic and deterministic predictions to achieve complementary strengths, namely uncertainty quantification and high predictive accuracy. To address these challenges, we proposed a prior-enhanced dual-mode spatiotemporal graph neural network (PED-STGNN) to support both probabilistic and deterministic predictions. Specifically, we introduced a hypergraph node-to-vector (hypernode2vec) method to capture the multivariate functional similarity prior derived from complex and multivariate relations between urban regions. This functional similarity characterizes urban systems more precisely than existing methods relying on first-order pairwise relations. It improves accuracy and interpretability while enabling spatial modeling of higher-order multivariate relations beyond first-order pairwise relations. We also designed a plug-and-play probabilistic prediction module that enables switches between probabilistic and deterministic modes. Experiments based on the human activity intensity in Fuzhou, China, demonstrated the advantages in accuracy, interpretability and multi-scenario applicability.

ARTICLE HISTORY
Received 28 March 2025
Accepted 10 September 2025

KEYWORDS
Prior knowledge; dual-mode; human activity intensity; probabilistic prediction; deterministic prediction

1. Introduction

Human activity intensity, defined as the population density within urban regions over specific time periods, exhibits clear diurnal and weekly variation across urban zones with distinct functions (Wang *et al.* 2022, Yang *et al.* 2023, Wang and Zhu 2024). These spatiotemporal patterns influence daily life, from the crowded subway trains during morning rush hours to the relative calm of parks on weekday afternoons. For urban residents, accurate prediction of human activity intensity supports intelligent route planning and optimizes recreational scheduling (Chen *et al.* 2023, Cai *et al.* 2024). It also

CONTACT Peixiao Wang wpx@lreis.ac.cn

enhances urban development, commercial strategies and the safety management of potential high-density areas, which contributes to smart city advancement (Wang *et al.* 2019, jia *et al.* 2021).

Predicting human activity intensity is a classic spatiotemporal prediction task (Li *et al.* 2022, Wang and Zhu 2024). The methods are generally divided into two categories: knowledge-driven and data-driven methods (Cheng *et al.* 2018, Ren *et al.* 2020, Li *et al.* 2021, 2022). Among these methods, deep learning, a core subset of data-driven methods, has become dominant because of end-to-end learning and its effective modeling of complex spatiotemporal dependencies (Xing *et al.* 2020, Li *et al.* 2021, 2022). Within this domain, dynamic graph neural networks (GNNs) have received increasing attention for their flexibility in spatial modeling and their ability to capture time-varying spatial relationships (Liu *et al.* 2023, Wang *et al.* 2023, Sun *et al.* 2024, Wang *et al.* 2025a). Although numerous dynamic GNN models have been developed, they face two main challenges. First, the performance of GNNs depends heavily on how adjacency matrices are constructed and the construction suffers from a significant limitation. Fully data-driven approaches often struggle to accurately identify nodes with high spatial correlations, which reduces both predictive accuracy and model interpretability (Liu *et al.* 2023, Wang and Zhu 2024). In contrast, traditional priors (e.g. distance thresholds and topological constraints) can construct adjacency matrices and strengthen local focus but inevitably constrain the modeling of hidden dependencies (Yu *et al.* 2018, Li *et al.* 2021). Points of interest (POIs) capture the semantics of urban functions and offer valuable priors for refining the sparsification of adjacency matrices (Zhang *et al.* 2023). However, urban regions exhibit multivariate relations. Most current frameworks fail to incorporate the heterogeneous geographic knowledge in POIs into the construction of adjacency matrices. Second, most existing models are deterministic and struggle to quantify prediction uncertainty, which limits decision-makers' ability to evaluate risks associated with extreme values (Kendall and Gal 2017). While some probabilistic models can estimate uncertainty, they often suffer from accuracy loss. Few models can switch seamlessly between probabilistic and deterministic modes to preserve high accuracy and quantify uncertainty. In practice, adaptive systems are essential, as they allow users to choose prediction modes according to specific requirements. However, creating a unified framework that supports both modes without degrading accuracy remains a technical challenge.

To overcome these challenges, we proposed a novel prior-enhanced dual-mode spatiotemporal graph neural network (PED-STGNN), which incorporates multivariate functional similarity embedded in POIs as a prior to enhance the accuracy and interpretability. It also unifies probabilistic and deterministic prediction for the urban intensity of human activities. The main contributions of this study are as follows:

1. Drawing on hypergraph theory, we developed a hypergraph node-to-vector (hypernode2vec) method to capture multivariate functional similarity from POIs, which extends spatial modeling from first-order pairwise relations to high-order multivariate relations. We employed this functional similarity as a prior to enhance both prediction accuracy and model interpretability.

2. We designed a plug-and-play probabilistic prediction module that enables most existing deterministic models to switch flexibly between probabilistic and deterministic modes. This adaptability allows users to select appropriate prediction modes for specific scenarios.
3. We validated the proposed PED-STGNN by using the intensity of human activities derived from mobile location data in Fuzhou, Fujian Province, China. The results demonstrate that the multivariate functional similarity prior improves both accuracy and interpretability and that the plug-and-play module enables seamless switches between deterministic and probabilistic prediction modes.

2. Literature review

Graph-based methods have become mainstream for predicting the urban intensity of human activities. Central to these methods is the construction of adjacency matrices. In both graph-based and other spatiotemporal models, the choice of prediction mode – deterministic or probabilistic – remains a key research area. This section reviews the development of graph-based spatiotemporal prediction models and prediction modes. It focuses on their methodological evolution and current challenges.

2.1. Graph-based spatiotemporal prediction models

Graph-based spatiotemporal prediction models use graph structures to represent spatial or functional relationships between urban regions (Wang 2023, Wang *et al.* 2024, Zeghina *et al.* 2024, Wang *et al.* 2025b). Classic models in this domain include GNNs and variants, such as spatiotemporal graph convolutional networks (ST-GCN) (Yu *et al.* 2018), attention-based spatiotemporal graph convolutional networks (ASTGCN) (Guo *et al.* 2019), temporal graph convolutional networks (T-GCN) (Zhao *et al.* 2020), spatiotemporal hypergraph convolutional networks (STHGCN) (Wang *et al.* 2021) and attention-based spatiotemporal graph convolutional recurrent networks (ASTGCRN) (Liu *et al.* 2023). The adjacency matrix is central to graph-based spatiotemporal prediction, as its construction directly influences model performance. Early approaches typically relied on data-driven techniques to construct adjacency matrices. However, such techniques suffer from reduced accuracy due to sampling bias and limited interpretability due to black-box nature (Lan *et al.* 2022, Zhao *et al.* 2023, Kong *et al.* 2024, Wang and Zhu 2024). Subsequently, prior-based methods addressed these limitations by incorporating spatial constraints, such as physical distances or topological relations between geographic entities (Li *et al.* 2021, Guan *et al.* 2024, Chen *et al.* 2025). However, distance-based approaches inherently constrain the modeling of hidden dependencies, as they rely on Euclidean proximity rather than semantic or functional similarity. To overcome this limitation, some studies explored alternative priors. Zhang *et al.* (2023) employed Pearson correlation coefficients between POI sequences to construct adjacency matrices. Wang *et al.* (2023) proposed Traffic BERT to capture implicit semantic associations from traffic sequences. Geng *et al.* (2019) constructed the adjacency matrix by integrating spatial proximity, road-network connectivity and functional

similarity measured by POI sequence distance. Although prior-based methods, particularly those utilizing POIs, have demonstrated promising results, the full potential of POI-based priors remains underexplored. Urban systems are inherently complex, interactive and often involve multivariate relations. The adjacency matrices derived from simple priors struggle to characterize such relations. Therefore, accurately extracting multivariate relations as priors is essential for improving both the accuracy and interpretability of GNN-based models. However, achieving this remains a significant technical challenge.

2.2. Prediction modes in spatiotemporal prediction models

Spatiotemporal prediction modes can be primarily divided into deterministic and probabilistic ones. Deterministic prediction, which generates single-point estimates, has been widely used in predicting the urban intensity of human activities, traffic flow, air quality and movement trajectories (Wang, *et al.* 2025c, Hong *et al.* 2023, Lun *et al.* 2025, Wang *et al.* 2024, 2023, Zhao *et al.* 2025). However, its inability to quantify uncertainty limits identification of regions at risk of extreme values. In contrast, probabilistic prediction offers greater flexibility. By producing both point estimates and uncertainty intervals, these models provide a deeper understanding of potential extreme values (Ryu *et al.* 2019, Abdar *et al.* 2021, Gawlikowski *et al.* 2023, Wen *et al.* 2023). This strength has led to its widespread adoption in weather forecasting and groundwater estimation (Liu *et al.* 2020, Zhu *et al.* 2025). However, the flexibility often comes at the cost of reduced accuracy and lower efficiency. Deterministic models can mitigate these shortcomings by offering high precision and efficiency. Therefore, a dual-mode prediction framework is needed to let users choose prediction modes based on specific requirements. However, most existing models remain single-modal and provide either deterministic or probabilistic predictions exclusively. Developing an architecture that can transition seamlessly between the two modes remains a major technical challenge.

2.3. Challenges and solutions

Two key challenges persist in this domain. First, existing graph-based models struggle to capture multivariate relations between geographic entities in urban systems. They fail to incorporate this relational knowledge as priors when constructing adjacency matrices. Second, few studies have proposed a unified framework that combines deterministic and probabilistic prediction. It is difficult for users to leverage their respective strengths and adapt to task-specific demands.

To address these challenges, we proposed a novel PED-STGNN, which incorporates a hypernode2vec method to extract multivariate functional similarity from POIs and to use this prior for constructing adjacency matrices. In addition, we introduced a plug-and-play probabilistic prediction module that enables seamless switches between the two prediction modes within a unified architecture.

3. Methodology

3.1. Problem definition and the framework of PED-STGNN

We represented the study area as a graph and employed a graph-based approach for spatiotemporal modeling, which leverages the flexibility of GNNs to capture complex spatial dependencies. As shown in Figure 1(a), we modeled the study area as a graph structure $G =< V, E, \boldsymbol{A} >$, where $V = \{v_i\}_{i=1}^{N}$ denotes the set of N graph nodes (i.e. urban regions), E represents the set of edges between nodes, and $\boldsymbol{A} \in \mathrm{R}^{N \times N}$ is the adjacency matrix ($\boldsymbol{A}_{i,j}$ indicates the connection between nodes v_i and v_j). Figure 1(b) illustrates the input and output of the model. $\boldsymbol{x}^t \in \mathrm{R}^{N \times 1}$ denotes the human activity intensity across all the graph nodes at time t. We aimed to learn a mapping from historical node features to future values, defined as $\hat{\boldsymbol{X}} = \mathrm{f}(\boldsymbol{X})$, where f denotes the prediction model. $\boldsymbol{X} = \{\boldsymbol{x}^{t-Q+1}, \boldsymbol{x}^{t-Q+2}, \ldots, \boldsymbol{x}^t\} \in \mathrm{R}^{N \times Q}$ represents the input data. $\hat{\boldsymbol{X}} = \{\hat{\boldsymbol{x}}^{t+1}, \hat{\boldsymbol{x}}^{t+2}, \ldots, \hat{\boldsymbol{x}}^{t+L}\} \in \mathrm{R}^{N \times L}$ denotes the predicted output. Q is the length of the historical window and L is the prediction horizon.

Figure 2 presents the overall framework of the proposed PED-STGNN. It consists of three primary modules: a multivariate functional similarity prior learning module based on hypernode2vec, a prior-enhanced spatiotemporal dependency learning module and a plug-and-play probabilistic prediction module. The first module extracts the multivariate functional similarity prior between urban regions by using a hypergraph-based structure. The second module integrates the prior into the GNN architecture to improve both prediction accuracy and interpretability. The third module enables PED-STGNN to switch seamlessly between probabilistic and deterministic prediction modes.

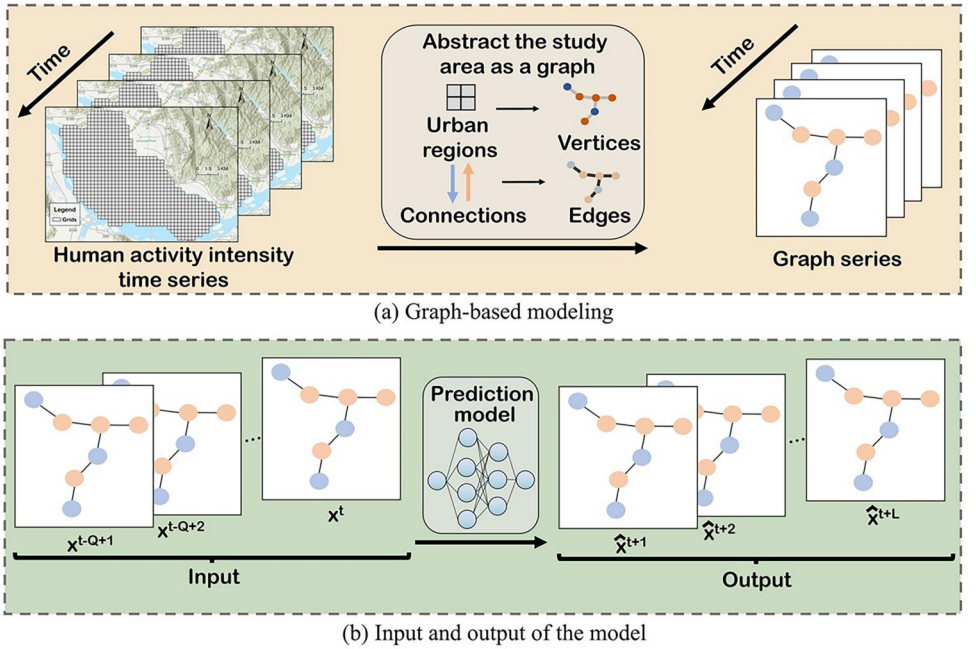

Figure 1. Relevant definitions.

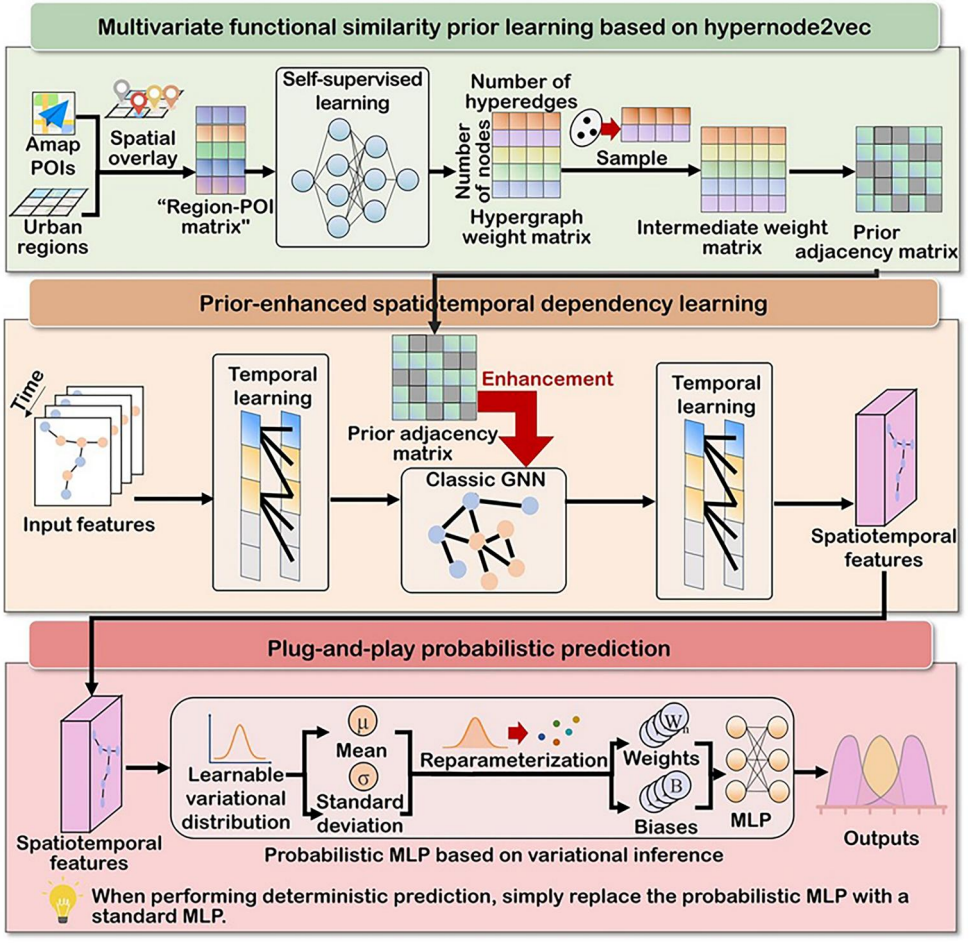

Figure 2. Illustrations of PED-STGNN.

To maintain computational efficiency, we precomputed the multivariate functional similarity prior offline, which eliminated additional cost during training and inference.

3.2. Multivariate functional similarity prior learning based on hypernode2vec

Temporal patterns of residents' activities are closely related to urban functions. Characterizing functional similarity and incorporating it as sparse constraints on the parameters of deep learning models can significantly improve both accuracy and interpretability. POIs are widely used to explicitly represent urban functions (Yao *et al.* 2017, Niu and Silva 2021). Previous studies used the distances between POI sequences to quantify functional similarity and to construct corresponding graph structures (Geng *et al.* 2019). However, relying solely on sequence distance introduces bias and reduces precision. First, the limited variety of POI types hinders the accurate representation of complex and diverse urban functions. Second, regions with mixed functions often experience temporal shifts in the dominant function, such as office use during the day and entertainment use at night. Sequence-based approaches oversimplify

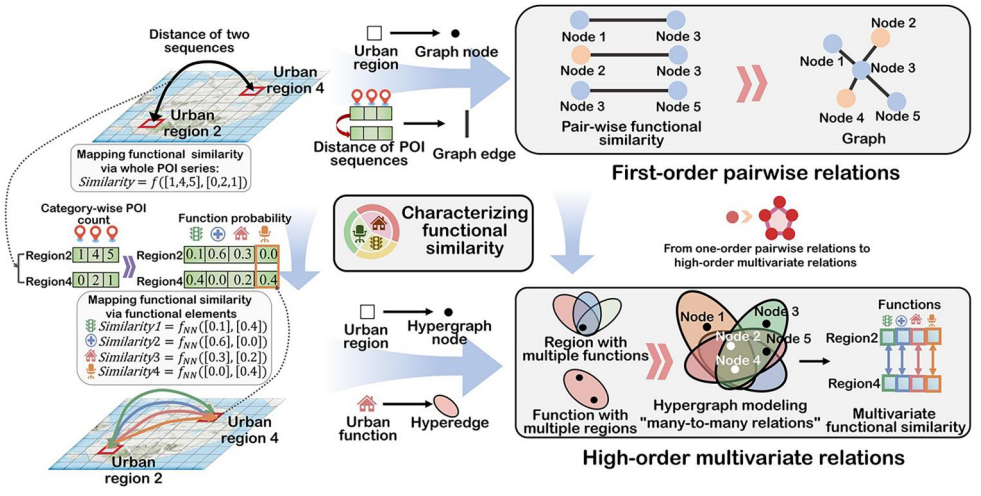

Figure 3. Rationale for using hypergraphs to model multivariate functional similarity.

these time-sensitive patterns and may overlook the temporal importance of specific functional roles. To overcome these limitations, we proposed a novel framework called hypernode2vec. This framework captures functional similarity between urban regions by first learning a hypergraph weight matrix in a self-supervised manner and then constructing a prior adjacency matrix.

Figure 3 illustrates the rationale for using hypergraphs to model multivariate functional similarity. The complex relations between urban regions and functions resemble hypergraph structures: an urban region with multiple functions corresponds to a hypergraph node connected to several hyperedges, while a function spanning multiple regions corresponds to a hyperedge connecting multiple nodes. Based on this analogy, we constructed a hypergraph $G_H =< V_H, E_H, \boldsymbol{W}_{Region-Func} >$ to represent the multivariate relations between urban regions and functions. In this structure, $V_H = \{v_{H:i}\}_{i=1}^{N}$ denotes the set of N hypergraph nodes (i.e. N urban regions); $E_H = \{e_{H:j}\}_{j=1}^{M}$ denotes the set of M hyperedges (i.e. M urban functions) and $\boldsymbol{W}_{Region-Func} \in \mathrm{R}^{N \times M}$ is the hypergraph weight matrix encoding the connections between nodes and hyperedges. The connections represent the relations between regions and their functions. It is important that the hypergraph was used only to support the subsequent GNN modeling and was not an independent prediction framework.

As described earlier, the hypergraph weight matrix encodes multivariate relations between urban regions and diverse functions. To learn this matrix and accurately capture these relations, we adopted a self-supervised learning framework. Figure 4 illustrates the workflow of learning the hypergraph weight matrix. First, we initialized a matrix of shape $N \times M$ and computed the initial hypergraph weight matrix $\boldsymbol{W}_{Region-Func}$ by using a one-dimensional convolution and a softmax function. Second, we constructed training samples $\tilde{\boldsymbol{M}}_{Region-POI}$ by perturbing the original 'Region-POI matrix', which we overlaid reclassified POIs on urban regions to generate. The perturbations included random row masking, row substitution and row retention. The perturbed matrix served as the input for self-supervised learning. Finally, we minimized the

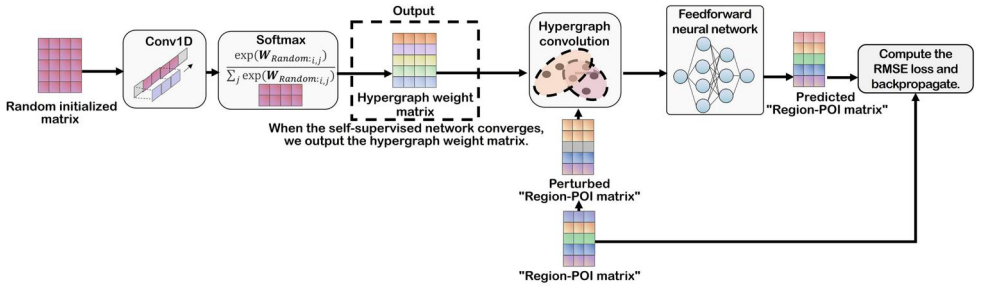

Figure 4. Workflow of learning the hypergraph weight matrix.

Table 1. Reclassification of Amap POIs.

Original first-level categories	Reclassification categories
Scenic spots	Recreation category
Science, education and cultural services	Science, education and cultural service category
Offices and corporations	Work category
Government agencies and social organizations	
Automobile sales	
Business-oriented residences	Residence category
Passage facilities	Transportation category
Transportation facilities services	
Automotive repair	Basic living support facility category
Automotive services	
Public facility	
Motorcycle services	
Lifestyle services	
Financial and insurance services	
Accommodation services	Consumption category
Sports and leisure services	
Shopping services	
Catering services	
Healthcare services	Healthcare service category

difference between the predicted and the original 'Region-POI matrix' to derive the learned $\boldsymbol{W}_{Region-Func}$. Equation (1) summarizes the process:

$$\begin{cases} \boldsymbol{W}_{Region-Func} = \text{softmax}(\text{Conv1D}(\boldsymbol{W}_{Random})) \\ \hat{\boldsymbol{M}}_{Region-POI} = \text{FNN}(\text{HGCN}(\tilde{\boldsymbol{M}}_{Region-POI}, \boldsymbol{W}_{Region-Func})) \end{cases} \tag{1}$$

where $\boldsymbol{W}_{Random} \in R^{N \times M}$ is the randomly initialized matrix; Conv1D denotes a one-dimensional convolution with a kernel size of 3; $\tilde{\boldsymbol{M}}_{Region-POI} \in R^{N \times P}$ denotes the perturbed 'Region-POI matrix', where P is the number of reclassified Amap POI categories in Table 1; HGCN represents the hypergraph convolution operation; FNN denotes the feedforward neural network, $\hat{\boldsymbol{M}}_{Region-POI}$ is the predicted 'Region-POI matrix'.

Although we derived the matrix $\boldsymbol{W}_{Region-Func}$, it cannot be directly used to augment GNN because its shape differs from that of a standard adjacency matrix. To address this issue, we generated a prior adjacency matrix compatible with the GNN input based on $\boldsymbol{W}_{Region-Func}$. Figure 5 illustrates the workflow. Inspired by node2vec (Grover and Leskovec 2016), which samples node sequences based on adjacency matrices, we probabilistically sampled hypergraph node sequences from the hypergraph weight matrix. We then used the skip-gram method to obtain the hypergraph node embeddings and computed the pairwise similarity between the nodes. For each row in the

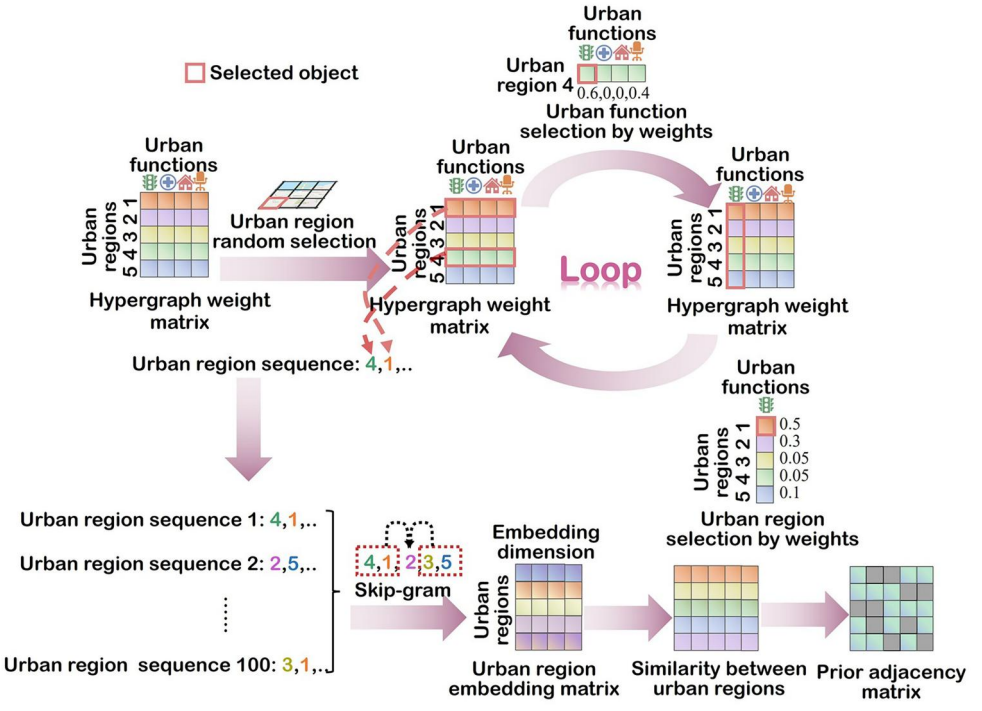

Figure 5. Generation process of the prior adjacency matrix.

resulting similarity matrix, we assigned 1 to the top-k values and 0 to the others. Consequently, we obtained the prior adjacency matrix $\boldsymbol{A}_{Prior}$, where the ratio k/N is λ_1.

3.3. Prior-enhanced spatiotemporal dependency learning

Priors help mitigate the reduced accuracy caused by sampling bias and the poor interpretability due to the black-box nature of data-driven models (Wang and Zhu 2024, Yang and Wu 2025). By incorporating causality or semantic constraints into an explicit modeling framework, priors improve both accuracy and interpretability. To integrate these benefits, we designed a prior-enhanced spatiotemporal dependency learning module, which improves the performance and interpretability of conventional GNN architectures.

Figure 6 presents the workflow of the prior-enhanced spatiotemporal dependency learning module. The module includes two temporal convolution layers and one prior-enhanced GNN layer. In the GNN layer, we used the prior adjacency matrix $\boldsymbol{A}_{Prior}$ to enhance the conventional GNN. The forward propagation process of this module is described in Equation (2):

$$\begin{cases} \tilde{\boldsymbol{H}} = \mathrm{TempConv}(\boldsymbol{X}) \\ \tilde{\boldsymbol{h}}_k^{Prior} = \mathrm{ReLU}\left(\sum_{l \in \{l \mid \boldsymbol{A}_{Prior}[k,l]=1\}} \alpha_{kl} \times \tilde{\boldsymbol{h}}_l \cdot \boldsymbol{W}_{Prior}\right) \in \tilde{\boldsymbol{H}}^{Prior} \\ \tilde{\boldsymbol{h}}_m^{Classic} = \mathrm{ReLU}\left(\sum_{n \in \{n \mid \boldsymbol{A}_{Classic}[m,n]=1\}} \alpha_{mn} \times \tilde{\boldsymbol{h}}_n \cdot \boldsymbol{W}_{Classic}\right) \in \tilde{\boldsymbol{H}}^{Classic} \\ \boldsymbol{H}^{st} = \mathrm{ReLU}(\mathrm{Concat}\left(\tilde{\boldsymbol{H}}^{Prior}, \tilde{\boldsymbol{H}}^{Classic}\right) \cdot \boldsymbol{W}_{st}) \\ \boldsymbol{H}^{out} = \mathrm{TempConv}(\boldsymbol{H}^{st}) \end{cases} \tag{2}$$

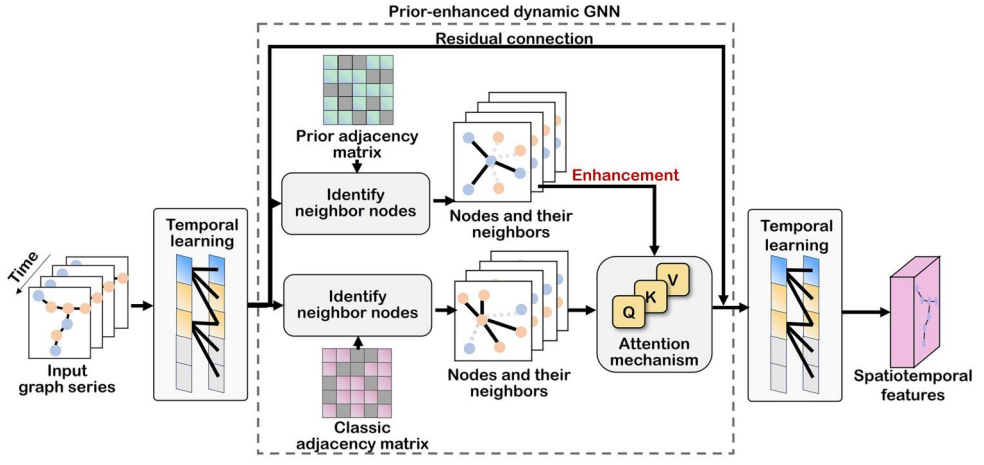

Figure 6. Workflow of prior-enhanced spatiotemporal dependency learning.

where $\boldsymbol{X} \in \mathrm{R}^{N\times Q}$ denotes the input human activity intensity across N urban regions over Q historical timesteps; $\boldsymbol{H}^{out}$ is the output of the prior-enhanced spatiotemporal dependency learning module; $\boldsymbol{H}^{st}$ is the intermediate representation of spatiotemporal features; TempConv denotes the temporal convolution operation. The matrix $\boldsymbol{A}_{Prior}$ introduced in Section 3.2 guided the forward propagation process. Specifically, $l \in \left\{l|\boldsymbol{A}_{Prior}[k,l]=1\right\}$ indicates the process was constrained by the prior implicitly represented in $\boldsymbol{A}_{Prior}$. $\boldsymbol{A}_{Classic}$ is the conventional adjacency matrix, which was constructed by using Wasserstein distance between region-wise activity sequences (Lan *et al.* 2022). Similarly to $\boldsymbol{A}_{Prior}$, we assigned 1 to the top-k' smallest values and 0 to the others. with the ratio k/N defined as λ_2; α_{kl} and α_{mn} are the attention scores; $\boldsymbol{W}_{Prior} \in \mathrm{R}^{h\times h}$, $\boldsymbol{W}_{Classic} \in \mathrm{R}^{h\times h}$ and $\boldsymbol{W}_{st} \in \mathrm{R}^{2h\times h}$ denote learnable weights matrices with h being the hidden dimension; $\tilde{\boldsymbol{h}}_l \in \tilde{\boldsymbol{H}}$ and $\tilde{\boldsymbol{h}}_n \in \tilde{\boldsymbol{H}}$ denote the hidden features of urban regions l and n, respectively; $\tilde{\boldsymbol{h}}_k^{Prior} \in \tilde{\boldsymbol{H}}^{Prior}$ denotes the output features of the kth region *via* the prior adjacency matrix and $\tilde{\boldsymbol{h}}_m^{Classic} \in \tilde{\boldsymbol{H}}^{Classic}$ is the output features of the mth region *via* the classic adjacency matrix; $\mathrm{Concat}\left(\tilde{\boldsymbol{H}}^{Prior}, \tilde{\boldsymbol{H}}^{Classic}\right)$ denotes the concatenation operation, i.e. the process of enhancing the classic GNN by using prior knowledge. Residual connections were incorporated into the GNN to prevent problems such as gradient explosion or vanishing.

3.4. Plug-and-play probabilistic prediction

Unlike deterministic prediction, probabilistic prediction provides uncertainty estimates to support the identification of regions susceptible to extreme values (Liu *et al.* 2020, Gawlikowski *et al.* 2023, Gao *et al.* 2024). However, not all application scenarios require uncertainty quantification. In scenarios where uncertainty is unnecessary, probabilistic prediction may reduce both computational efficiency and predictive accuracy. To support various application scenarios, we developed a plug-and-play probabilistic

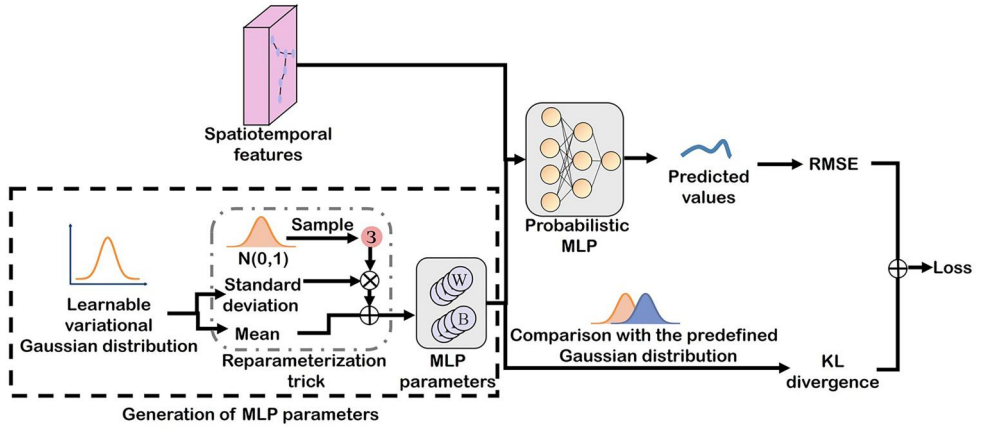

Figure 7. Workflow of plug-and-play probabilistic prediction module.

prediction module, which enables users to switch between deterministic and probabilistic outputs according to specific requirements.

Inspired by Bayesian neural networks based on variational inference, we learned the distributions of multilayer perceptron (MLP) parameters to enable probabilistic prediction (Abdar *et al.* 2021). Figure 7 illustrates the workflow of the plug-and-play probabilistic prediction module. We employed the reparameterization trick to sample parameters from the learnable variational Gaussian distribution $\mathrm{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}})$, where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\sigma}}$ denote the learnable mean and standard deviation, respectively. We fed $\boldsymbol{H}^{out}$ in Section 3.3 and the sampled parameters into a probabilistic MLP to obtain the predicted values. Equation (3) describes this process:

$$\begin{cases} \boldsymbol{s}^{Prob} = \hat{\boldsymbol{\mu}} + \varepsilon \times \hat{\boldsymbol{\sigma}}, \ \ \varepsilon \sim \mathrm{N}(0,1) \\ \hat{\boldsymbol{X}} = \mathrm{Prob\_MLP}(\boldsymbol{H}^{out}; \boldsymbol{s}^{Prob}) \end{cases} \tag{3}$$

where Prob_MLP denotes the probabilistic MLP, with its parameters modeled as distributions instead of deterministic values; $\boldsymbol{s}^{Prob}$ represents all the sampled parameters; ε is a random variable drawn from a standard Gaussian distribution; $\hat{\boldsymbol{X}}$ denotes the predicted values. Drawing on the bootstrap method (DiCiccio and Efron 1996), we estimated the predicted values by averaging across repeated resamples and quantified uncertainty by using 95% confidence intervals. When single-value parameters (instead of $\boldsymbol{s}^{Prob}$) are used in Prob_MLP, the module simplifies to a standard MLP for deterministic prediction. This design enables dual-mode prediction.

For probabilistic prediction, we first computed the root mean square error (RMSE) between predicted and true values. We then constructed a loss function by combining RMSE with the Kullback-Leibler (KL) divergence between the learnable variational distribution and a predefined posterior distribution, $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$ based on Liu *et al.* (2020). Equation (4) describes the details:

$$\begin{cases} KL = \log\boldsymbol{\sigma} - \log\hat{\boldsymbol{\sigma}} + \dfrac{\hat{\boldsymbol{\sigma}}^2 + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^2}{2 \times \boldsymbol{\sigma}^2} - \dfrac{1}{2}) \\ Loss = \sqrt{\dfrac{\sum_i (\hat{x}_i - x_i)^2}{n}} + \beta \times KL \end{cases} \tag{4}$$

where *KL* denotes the KL divergence; $\hat{x}_i$ denotes the *i*th predicted value; x_i denotes the *i*th true value; β is the regularization coefficient. For deterministic prediction, we adopted RMSE as the loss function.

3.5. Model evaluation

3.5.1. Evaluation metrics

We assessed model accuracy by using three metrics: RMSE, mean absolute error (MAE) and mean absolute percentage error (MAPE). The equations show the details:

$$RMSE = \sqrt{\frac{\sum_i (\hat{x}_i - x_i)^2}{n}} \tag{5}$$

$$MAE = \frac{\sum_i |\hat{x}_i - x_i|}{n} \tag{6}$$

$$MAPE = \frac{\sum_i \left|\frac{\hat{x}_i - x_i}{x_i}\right|}{n} \times 100\% \tag{7}$$

where $\hat{x}_i$ denotes the *i*th predicted value; x_i is the *i*th true value; n is the total number of samples.

3.5.2. Baseline models

We selected several baseline models including both deterministic and probabilistic ones for comparison with our proposed model.

1. STGCN (Yu *et al.* 2018): STGCN is a deterministic prediction model that integrates graph convolutional network (GCN) and gated convolutional neural network (CNN). GCN is used to learn spatial dependencies and gated CNN is used to capture temporal dependencies. This model is widely used for spatiotemporal prediction tasks.
2. ASTGCN (Guo *et al.* 2019): ASTGCN is a deterministic model that incorporates GCN and CNN. The model is further enhanced with an attention mechanism to selectively weight input features. It effectively captures key spatiotemporal dependencies in the data.
3. T-GCN (Zhao *et al.* 2020): T-GCN is a deterministic model that combines GCN with a gated recurrent unit (GRU). It uses GCN to learn spatial dependencies and uses GRU to extract temporal features.
4. STHGCN (Wang *et al.* 2021): STHGCN is a deterministic prediction model that constructs a hypergraph from historical traffic data and metro station topologies. It uses hypergraph convolution for spatial learning and GRU for temporal learning.
5. ASTGCRN (Liu *et al.* 2023): ASTGCRN is a deterministic model based on the graph convolutional recurrent network (GCRN) framework. It incorporates an attention mechanism to improve the representation of both temporal and spatial features.
6. SVGP (Titsias 2009): Sparse variational Gaussian process (SVGP) is a probabilistic model that approximates Gaussian processes by using sparse variational inference and inducing points. This approach significantly reduces computational cost while maintaining probabilistic interpretability.

7. STNN (Liu *et al.* 2020): STNN is a probabilistic deep learning model for spatiotemporal prediction. It employs a multi-layer structure that combines the convolutional gated recurrent unit (ConvGRU) with CNN. This model provides accurate probabilistic predictions of wind speed.

4. Experiment

In this section, we present the study area and the experiments we performed. We conducted a series of experiments to explore the following research questions:

RQ1: Does the prior learned through hypernode2vec improve predictive accuracy?
RQ2: Does the prior learned through hypernode2vec enhance interpretability?
RQ3: Does the plug-and-play probabilistic prediction module function as intended?
RQ4: How does PED-STGNN perform compared with existing models?

4.1. Study area, data preprocessing and experimental settings

As shown in Figure 8, the study area comprises the Gulou, Taijiang and Cangshan Districts of Fuzhou City, Fujian Province, China. These urban districts are characterized by large populations and high human activity intensity. Consequently, they are well-suited for evaluating spatiotemporal prediction models. We partitioned the area into 885 regular grid cells, with each cell measuring 500 m $\times$ 500 m. Human activity intensity was derived from anonymized mobile location data and used for model training, validation and testing. As summarized in Table 2, the dataset includes anonymized user IDs, geolocation coordinates (latitude and longitude), and timestamp information. Raw mobile location data are prone to noise and inconsistencies because of phenomena such as the ping-pong and drift effects (Nalin *et al.* 2024, Song *et al.* 2024). The ping-pong effect refers to spurious location switches between adjacent base stations caused by signal fluctuations. Drift describes gradual deviations from true positions caused by signal interference (e.g. from buildings), hardware or software limitations and cumulative environmental errors in positioning systems. We preprocessed the raw data through deduplication, outlier removal and filtering of ping-pong and drift effects. Then, we aggregated the number of individuals within each region at 15-min intervals as human activity intensity. To ensure temporal representativeness across weekdays and weekends, we designated data from 1 March to 21 March 2023 as the training set, 22 March to 25 March 2023 as the validation set and 26 March to 28 March 2023 as the test set. We conducted hyperparameter tuning to optimize model performance. We set the number of input historical timesteps as 9, the hidden dimension as 512, and the prior, classic sparsification thresholds λ_1, λ_2 (Sections 3.2 and 3.3) 0.02 and 0.03, respectively. Based on iterative experimentations, we set the regularization coefficient β (Section 3.4) as 0.01, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ of the predefined posterior distribution for PED-STGNN (Section 3.4) as 0 and 0.1 and set the initial learning rate as 0.0001.

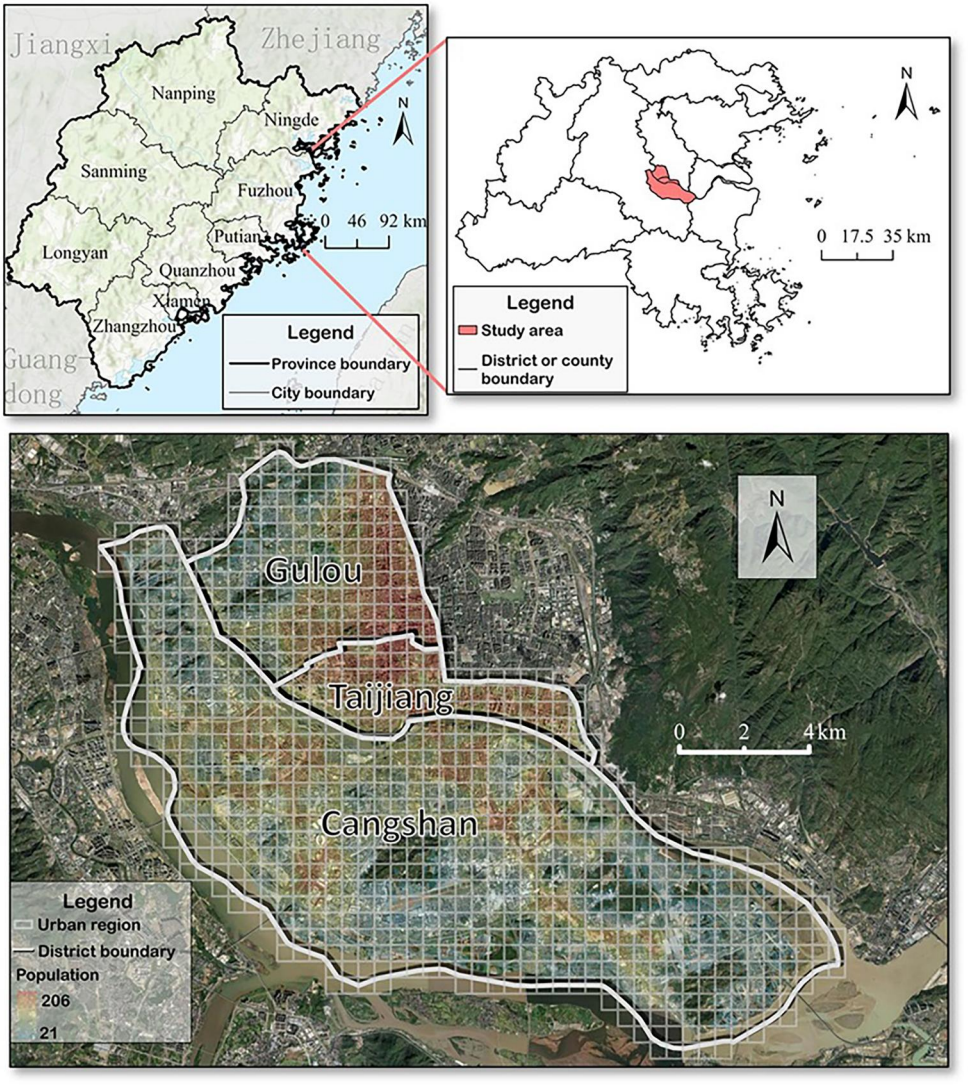

Figure 8. Study area.

Table 2. Examples of mobile location data.

Anonymized IDs	latitude/°E	longitude/°N	Start time
03244040-5cac-4ffc-9594-6073b5f589e7	118.168526	24.489397	2023/3/1 19:23
04974c4b-4dd7-44c1-9991-2ee3c892b8b3	121.251564	30.172546	2023/3/1 13:52
…	…	…	…
1219f065-f87c-3562-bb4a-5067d58338aa	118.99641	25.430447	2023/3/1 20:22

4.2. Comparison with baseline models (RQ1)

To validate the accuracy of PED-STGNN, we first conducted a quantitative comparison under the deterministic mode. As shown in Table 3, ST-GCN achieved the best performance among all baselines, with an RMSE of 20.68 and an MAE of 11.73. Compared with ST-GCN, PED-STGNN reduced RMSE and MAE by approximately 10.3% and 11.6%,

Table 3. Comparison with baseline models (deterministic mode).

Model	RMSE	MAE	MAPE (%)
ST-GCN	20.68	11.73	23.73
T-GCN	22.14	11.76	19.06
ASTGCRN	22.99	12.40	20.45
STHGCN	24.22	14.87	19.58
ASTGCN	22.00	12.20	26.87
PED-STGNN (deterministic)	**18.54**	**10.37**	**17.32**

Note: The bold values indicate the best performance.

Table 4. Comparison with probabilistic baseline models.

Model	RMSE	MAE	MAPE (%)
ST-GCN (with the plug-and-play probabilistic prediction module)	22.04	12.53	24.15
T-GCN (with the plug-and-play probabilistic prediction module)	25.76	12.99	19.68
ASTGCRN (with the plug-and-play probabilistic prediction module)	22.88	12.70	22.68
STHGCN (with the plug-and-play probabilistic prediction module)	25.15	13.05	19.71
ASTGCN (with the plug-and-play probabilistic prediction module)	23.86	12.45	22.25
STNN	25.41	**10.41**	46.64
SVGP	27.73	15.34	41.54
PGD-STGNN (with the plug-and-play probabilistic prediction module)	**19.99**	10.98	**16.92**

Note: The bold values indicate the best performance.

respectively. While T-GCN showed the lowest MAPE among baseline models, PED-STGNN still outperformed it by approximately 9.1%. These findings underscore the benefits of incorporating the multivariate functional similarity prior. Unlike baseline models that rely on data-driven adjacency learning or simple priors, PED-STGNN incorporates multivariate functional similarity between urban regions, which improves prediction accuracy. These results highlight the importance of appropriate priors for enhancing GNN performance in complex urban environments (Zhang *et al.* 2024, Wang *et al.* 2025d).

We then compared the probabilistic PED-STGNN with two sets of models: probabilistic models derived from deterministic ones through the plug-and-play module and specialized probabilistic prediction models. As shown in Table 4, PED-STGNN demonstrated strong overall performance across all evaluation metrics. Specifically, it achieved the best RMSE and MAPE scores and outperformed the best baselines by approximately 9.3% and 14.2%, respectively. Although STNN – a deep learning model designed for probabilistic prediction – achieved the lowest MAE, PED-STGNN ranked a close second. It increased MAE by approximately 5.5%. Importantly, Tables 3 and 4 reveal that converting deterministic models to probabilistic ones can increase RMSE by up to 3.62 units. However, this moderate increase is acceptable given the substantial benefit of uncertainty quantification. Moreover, since human activity intensity typically ranges from hundreds to thousands, this minor loss of accuracy has limited impact on decision-making.

4.3. Sensitivity analysis of hyperparameters

To examine the robustness and the influence of the key hyperparameters on accuracy, we conducted a sensitivity analysis on λ_1 (from Section 3.2), λ_2 (from Section 3.3), the hidden dimension, and the number of input historical timesteps. First, we varied λ_1 and λ_2 from 0.01 to 0.04 and used grid search to explore their effects. As shown in

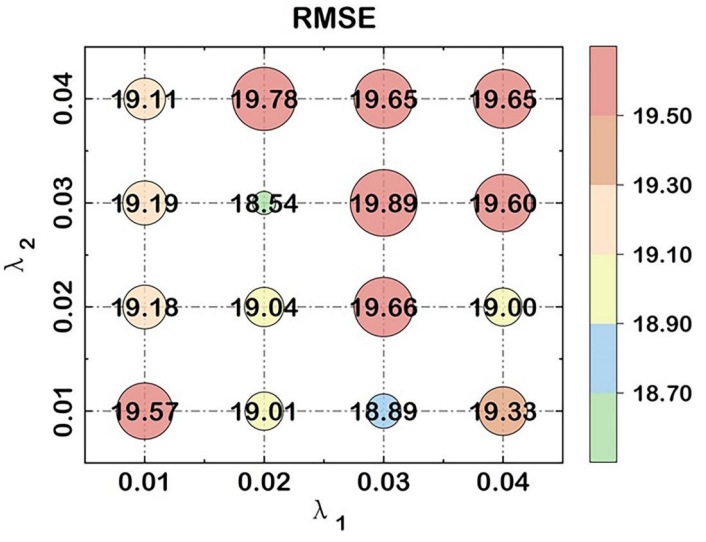

Figure 9. Sensitivity analysis of λ_1 and λ_2.

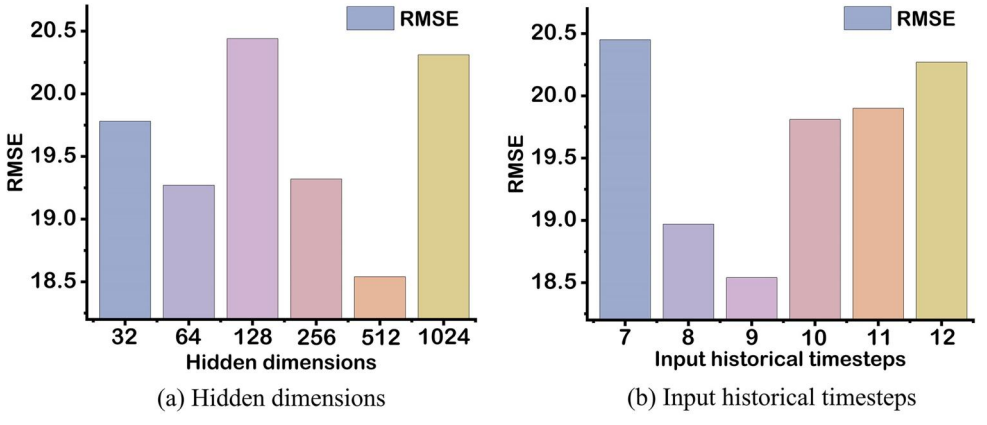

Figure 10. Sensitivity analysis of hidden dimension and input timesteps.

Figure 9, the model demonstrated strong robustness to these parameters. RMSE fluctuated by only about 1.24. Generally, as λ_1 and λ_2 increased, RMSE first declined and then rose and reached its global minimum (18.54) when λ_1 and λ_2 were set to 0.02 and 0.03, respectively. We then analyzed the sensitivity to hidden dimensions and tested values of 32, 64, 128, 256, 512, and 1024. As illustrated in Figure 10(a), the model was less robust to this parameter, with RMSE varying by about 1.9. The lowest RMSE occurred at the hidden dimension of 512. Lastly, we assessed the sensitivity to the number of input historical timesteps by setting it between 7 and 12. As shown in Figure 10(b), RMSE again varied by about 1.9, with the lowest RMSE occurring at the timestep of 9. These findings suggest that PED-STGNN is relatively robust to hyperparameter changes, as the total RMSE fluctuated by no more than 2. The results also provide practical guidance for hyperparameter selection.

Table 5. The results of ablation experiments.

Ablation method	RMSE	MAE	MAPE (%)
w/o Prior enhancement and classic adjacency relationship	21.81	12.02	20.02
w/o Prior enhancement adjacency relationship	21.72	11.94	19.22
w/o Classes adjacency relationship	21.70	11.94	18.94
PGD-STGNN (deterministic)	**18.54**	**10.37**	**17.32**

Note: The bold values indicate the best performance.

4.4. Ablation experiments (RQ1)

To assess the individual contributions of our proposed modules, we performed a series of ablation experiments in the deterministic mode. As shown in Table 5, removing the prior-enhanced adjacency relationship resulted in an approximate 17.2% increase in RMSE, while excluding the classic adjacency relationship led to a 17.0% increase. Eliminating both adjacency structures forced the model to use a dense adjacency matrix and caused RMSE, MAE and MAPE to increase by 17.6%, 15.9% and 15.6%, respectively. These findings confirm that incorporating both prior and classic constraints significantly enhances prediction accuracy in spatiotemporal GNN models.

4.5. Interpretability analysis of prior enhancement (RQ2)

To investigate whether hypernode2vec captured multivariate functional similarity, we randomly selected four regions with multiple urban functions. We extracted adjacency matrices across all timesteps in the test dataset and analyzed the corresponding rows for the selected regions. Over 99% of the rows contained only one nonzero entry, which indicates that the model concentrates on a minimal set of urban regions. For the rows corresponding to the selected regions, we identified the columns with the adjacency values of 1, aggregated the indices of these columns across all timesteps and recorded their frequencies. The regions corresponding to the two most frequent column indices were then labeled as closely connected to the selected regions. To determine the dominant urban functions of the regions, we referred to Baidu map, Amap and Amap POIs. Some of the closely connected region pairs and their dominant functions were subsequently visualized. As shown in Figure 11, despite functional heterogeneity and spatial separation, the region pairs demonstrated high multivariate functional similarity. These pairs spanned a broad area of Fuzhou and included both frequent functions, such as residence and office, and less common ones, such as healthcare and recreation (e.g. parks and scenic spots). These findings indicate that hypernode2vec effectively captures functional similarity without spatial bias.

4.6. Qualitative analysis of prediction results (RQ3)

We conducted a qualitative analysis to evaluate the validity of the probabilistic predictions. From a temporal perspective, we randomly selected one region and compared the deterministic predictions with the probabilistic ones across all timesteps in the test dataset. As shown in Figure 12(a), deterministic prediction produced single-point estimates, while Figure 12(b) illustrates that probabilistic prediction provided both point estimates and prediction intervals. The uncertainty, quantified by these intervals,

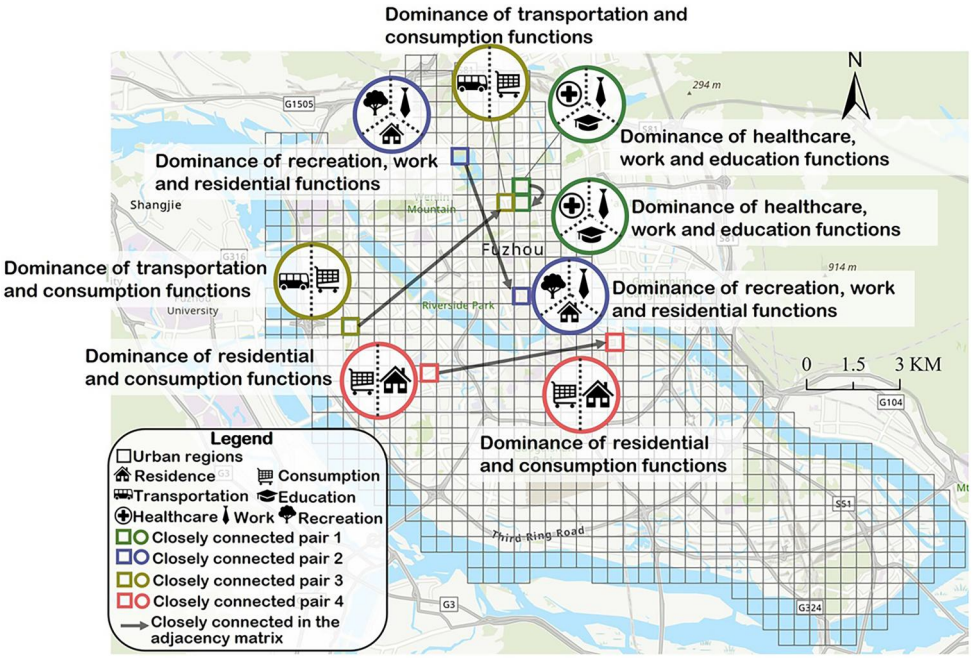

Figure 11. Visualization of regions closely connected in the adjacency matrix.

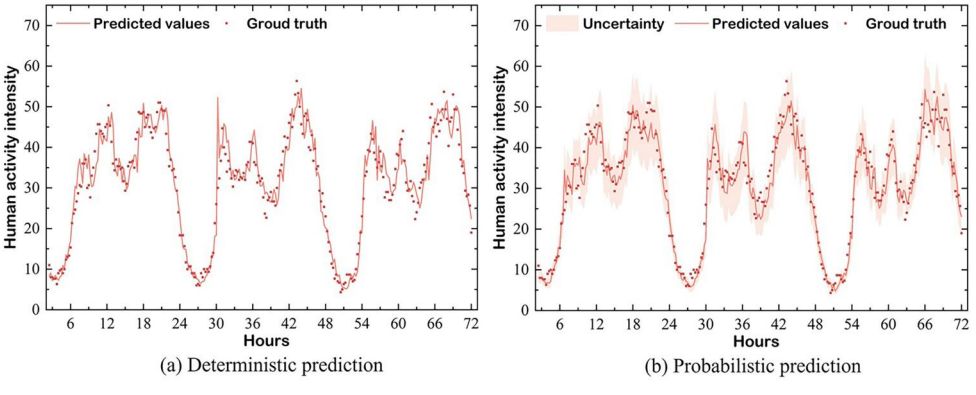

Figure 12. Comparative visualization of probabilistic and deterministic predictions.

exhibited clear diurnal periodicity. The intervals were broader between 7:00 and 22:00 (corresponding to peak hours of urban activities) and narrower during nighttime. Such a temporal pattern aligns with well-established human mobility rhythms (Guo *et al.* 2018, Yang *et al.* 2025). In practice, regions with high uncertainty are more vulnerable to extreme events and require proactive risk management such as preparing for sudden crowd surges.

From a spatial perspective, we visualized the distributions of prediction uncertainty, actual human activity intensity and the risk of extreme values (the ratio of uncertainty to human activity intensity) at 2:30, 10:30 and 18:30 in the test dataset. As illustrated in Figure 13(d–f), uncertainty was higher in the northern regions than in the southern

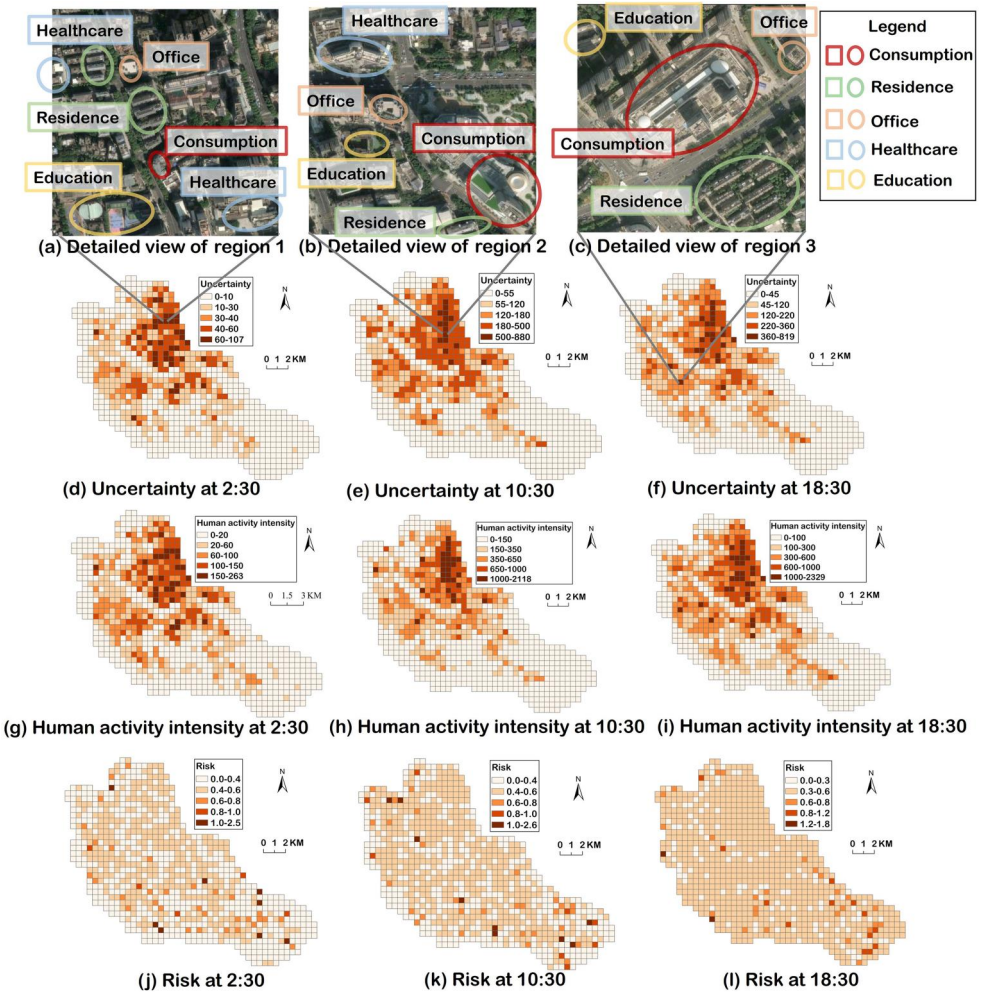

Figure 13. Qualitative analysis of prediction results.

ones and much higher during daytime. These patterns align with both the population distribution shown in Figure 8 and the daily activity patterns of urban residents. For each time, we randomly selected one region exhibiting high uncertainty and then identified its primary urban functions by integrating remote sensing imagery, Baidu Map and Amap. As shown in Figure 13(a–c), these regions demonstrated diverse urban functions. The region in Figure 13(a) contained two major healthcare institutions, Fuzhou Chinese Medicine Hospital and Fujian Provincial Hospital. These institutions contributed to increased nighttime uncertainty at 2:30. The selected regions in Figure 13(b,c) contained commercial zones, Shangri-La Commercial Center and Wanda Plaza, respectively. The high density of consumption-related entities contributed to the sharp fluctuations in human activity intensity and led to high uncertainty. Although uncertainty levels in these regions were relatively high, they remained acceptable relative to the actual human activity intensity shown in Figure 13(g–i). They did not compromise the reliability of the model's predictions. Importantly, for risk prevention, regions

prone to extreme values should be identified by jointly considering both uncertainty and observed human activity intensity (Cheong *et al.* 2020, Zhu *et al.* 2024). To this end, we computed the ratio of uncertainty to observed intensity as a proxy for relative risk. As shown in Figure 13(j–l), this analysis reveals distinct patterns of risk across urban regions, which highlights the spatial heterogeneity of extreme-value susceptibility.

5. Discussion

This section addresses RQ4. In this section, we summarize the improvements of PED-STGNN over conventional spatiotemporal prediction models and discuss its broader real-world applications. Most existing models struggle to capture multivariate relations between urban geographical entities, which leads to both accuracy loss and interpretability degradation. Additionally, the absence of dual-mode functionality restricts their flexibility and applicability in diverse real-world scenarios.

PED-STGNN addresses these limitations and represents a substantive advancement in the domain of spatiotemporal prediction by offering three key technological advantages. First, PED-STGNN improves prediction accuracy by approximately 10.3% in RMSE compared with baselines including ASTGCRN (Liu *et al.* 2023), ST-GCN (Yu *et al.* 2018), ASTGCN (Guo *et al.* 2019), STHGCN (Wang *et al.* 2021) and T-GCN (Zhao *et al.* 2020). The accuracy gain is primarily attributed to the incorporation of appropriate prior knowledge *via* the hypernode2vec framework. The framework captures semantic associations between regions with mixed functions and informs the construction of adjacency matrices. It characterizes higher-level urban semantics compared with the models that rely on observed data or simple priors and accurately represents latent functional similarity, which enhances both generalizability and robustness. Second, the incorporation of the prior markedly improves interpretability. As discussed in Section 4.5, PED-STGNN identified the linkages between the urban regions with shared but complex functions. This not only demonstrates the hypernode2vec's ability to capture complex semantics but also provides a deeper understanding of the PED-STGNN's prediction logic. Compared with black-box models, the interpretability allows users to better understand the rationale for the model's outputs, which facilitates informed decision-making, improves user trust and enables the targeted refinement of the model design (Guo *et al.* 2022, Zhang *et al.* 2021, Kong *et al.* 2024). Third, the proposed plug-and-play probabilistic prediction module enables seamless switches between deterministic and probabilistic modes without significant accuracy loss. For deterministic models, this module allows the output of uncertainty in the form of confidence intervals, which has substantial practical significance (Liu *et al.* 2020, Lan *et al.* 2022, Gao *et al.* 2024, Wang and Zhu 2024). For example, urban administrators may adopt probabilistic prediction in risk-sensitive scenarios, such as the prevention of crowd surge. In this scenario, the probabilistic mode supports the identification of regions where human activity intensity may exceed safety thresholds, which enables early intervention. In scenarios where uncertainty is unnecessary, the deterministic mode is preferred for its efficiency and accuracy. The plug-and-play design of this

component enhances the flexibility and real-world applicability of existing prediction models.

Despite these advantages, this study has some limitations: (1) when mobile phones are powered off or disconnected from Wi-Fi, the data may be unsampled or undersampled, which introduces sampling bias; (2) we did not incorporate external influences such as weather conditions or emergency events that may affect human activity, which limits the real-time density estimation during disruptive events; (3) because of dataset constraints, we were unable to assess model's robustness across different seasons or geographical regions. To address these limitations, future research will focus on three areas: (1) integrating mobile location data with complementary sources, such as public transport smart-card records and bike-sharing demand to mitigate sampling bias; (2) incorporating meteorological and emergency data from official sources to enhance the performance during emergency events; (3) expanding the dataset both spatially and temporally to evaluate robustness in different seasons and regions.

6. Conclusion

In this study, we proposed a novel PED-STGNN that integrates multivariate functional similarity as a prior to enhance both predictive accuracy and interpretability. The model further incorporates a plug-and-play probabilistic prediction module that enables users to switch between deterministic and probabilistic modes. Experiments based on mobile location data from Fuzhou demonstrated the model's advantages in accuracy, interpretability and operational adaptability. Specifically, prior enhancement significantly increases predictive accuracy. Compared with the state-of-the-art baselines (STGCN and T-GCN), the model achieved a 10.3% reduction in RMSE, an 11.6% decrease in MAE, and a 9.1% improvement in MAPE. The prior also enhances the model's capacity to capture spatial dependencies among urban regions with similar functions, which provides more interpretable rationales for predictions. The plug-and-play probabilistic prediction module effectively transforms deterministic models into probabilistic ones while preserving relatively high precision. Compared with the best probabilistic baselines, PED-STGNN achieved a 9.3% reduction in RMSE and a 14.2% reduction in MAPE. In terms of MAE, it ranked second and achieved a 5.5% increase compared with STNN, a deep learning model specialized in probabilistic prediction. Overall, these advancements support the development of GeoAI by enhancing predictive accuracy and mitigating the black-box nature of deep learning models. They also expand the practical scope of geospatial decision-making.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This project was supported by National Natural Science Foundation of China under grant 42401524 and 42571546.

Notes on contributors

Cheng Zhong is a M.S. candidate at the Academy of Digital China (Fujian), Fuzhou University. His research interests include spatial modeling and geospatial big data analysis. He contributed to the conceptualization, methodology, formal analysis, visualization, original draft and review of this article.

Sheng Wu is a Professor at the Academy of Digital China (Fujian), Fuzhou University. His research interests include spatiotemporal data mining and geospatial big data analysis. He contributed to conceptualization, methodology, formal analysis, visualization and review of this article.

Peixiao Wang is an Assistant Professor at State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests include spatiotemporal data mining and spatiotemporal prediction, with a particular focus on the prediction of transportation systems. He contributed to conceptualization, methodology, formal analysis, visualization and review of this article.

Hengcai Zhang is an Associate Professor of State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. His interests focus on spatial-temporal data mining and 3D-Computing. He contributed to methodology and review of this article.

Shifen Cheng is an Associate Professor of the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests include spatiotemporal data mining, urban computing and intelligent transportation. He contributed to methodology and review of this article.

Feng Lu is a Professor at the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests cover trajectory data mining, computational transportation science and location-based services. He contributed to methodology and review of this article.

ORCID

Peixiao Wang http://orcid.org/0000-0002-1209-6340
Hengcai Zhang http://orcid.org/0000-0002-5004-9609

Data and codes availability statement

The data and codes supporting this study are available at https://doi.org/10.6084/m9.figshare.28683131.

References

Abdar, M., *et al.*, 2021. A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Information Fusion*, 76, 243–297.

Cai, J., *et al.*, 2024. Quantifying spatial interaction centrality in urban population mobility: a mobility feature- and network topology-based locational measure. *Sustainable Cities and Society*, 114, 105769.

Chen, J., *et al.*, 2025. Spatio-temporal graph neural networks for missing data completion in traffic prediction. *International Journal of Geographical Information Science*, 39 (5), 1057–1075.

Chen, Y., *et al.*, 2023. Exploring the spatiotemporal patterns and correlates of urban vitality: Temporal and spatial heterogeneity. *Sustainable Cities and Society*, 91, 104440.

Cheng, S., *et al.*, 2018. Short-term traffic forecasting: an adaptive ST-KNN model that considers spatial heterogeneity. *Computers, Environment and Urban Systems*, 71, 186–198.

Cheong, L., *et al.*, 2020. Evaluating the impact of visualization of risk upon emergency route-planning. *International Journal of Geographical Information Science*, 34 (5), 1022–1050.

DiCiccio, T.J. and Efron, B., 1996. Bootstrap confidence intervals. *Statistical Science*, 11 (3), 189–228.

Gao, X., *et al.*, 2024. Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction. *Accident; Analysis and Prevention*, 208, 107801.

Gawlikowski, J., *et al.*, 2023. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 56 (S1), 1513–1589.

Geng, X., *et al.*, 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33 (01), 3656–3663.

Grover, A. and Leskovec, J., 2016. node2vec: Scalable feature learning for networks. *In*: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. Presented at the KDD '16: The 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 13 August San Francisco, CA. New York, NY: ACM, 855–864.

Guan, Q., *et al.*, 2024. Predicting short-term $PM_{2.5}$ concentrations at fine temporal resolutions using a multi-branch temporal graph convolutional neural network. *International Journal of Geographical Information Science*, 38 (4), 778–801.

Guo, B., *et al.*, 2018. Using rush hour and daytime exposure indicators to estimate the short-term mortality effects of air pollution: a case study in the Sichuan Basin, China. *Environmental Pollution (Barking, Essex: 1987)*, 242 (Pt B), 1291–1298.

Guo, H., *et al.*, 2022. ASTCN: an attentive spatial–temporal convolutional network for flow prediction. *IEEE Internet of Things Journal*, 9 (5), 3215–3225.

Guo, S., *et al.*, 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33 (01), 922–929.

Hong, Y., *et al.*, 2023. Context-aware multi-head self-attentional neural network model for next location prediction. *Transportation Research Part C: Emerging Technologies*, 156, 104315.

Jia, C., *et al.*, 2021. Evaluation of urban vibrancy and its relationship with the economic landscape: a case study of Beijing. *ISPRS International Journal of Geo-Information*, 10 (2), 72.

Kendall, A. and Gal, Y., 2017. What uncertainties do we need in Bayesian deep learning for computer vision? *In*: *Proceedings of the 31st international conference on neural information processing systems*. Red Hook, NY: Curran Associates Inc., 5580–5590.

Kong, J., *et al.*, 2024. ADCT-Net: Adaptive traffic forecasting neural network via dual-graphic cross-fused transformer. *Information Fusion*, 103, 102122.

Lan, S., *et al.*, 2022. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. *In: International conference on machine learning*, 17–23 July Baltimore, Maryland. PMLR, 11906–11917.

Li, M., *et al.*, 2021. Prediction of human activity intensity using the interactions in physical and social spaces through graph convolutional networks. *International Journal of Geographical Information Science*, 35 (12), 2489–2516.

Li, M., *et al.*, 2022. Fine-grained crowd distribution forecasting with multi-order spatial interactions using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 144, 103908.

Liu, H., *et al.*, 2023. Attention-based spatial-temporal graph convolutional recurrent networks for traffic forecasting. *In: International conference on advanced data mining and applications*, 21–23 August Shenyang, China. Springer, 630–645.

Liu, Y., *et al.*, 2020. Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model. *Applied Energy*, 260, 114259.

Lun, M., *et al.*, 2025. Predicting the next location of urban individuals via a representation-enhanced multi-view learning network. *ISPRS International Journal of Geo-Information*, 14 (8), 302.

Nalin, A., *et al.*, 2024. Assessing veracity of big data: an in-depth evaluation process from the comparison of Mobile phone traces and groundtruth data in traffic monitoring. *Journal of Transport Geography*, 118, 103930.

Niu, H. and Silva, E.A., 2021. Delineating urban functional use from points of interest data with neural network embedding: a case study in Greater London. *Computers, Environment and Urban Systems*, 88, 101651.

Ren, Y., *et al.*, 2020. A hybrid integrated deep learning model for the prediction of citywide spatio-temporal flow volumes. *International Journal of Geographical Information Science*, 34 (4), 802–823.

Ryu, S., Kwon, Y., and Kim, W.Y., 2019. A Bayesian graph convolutional network for reliable prediction of molecular properties with uncertainty quantification. *Chemical Science*, 10 (36), 8438–8446.

Song, X., *et al.*, 2024. Inferring the accurate locations of noise records in mobile phone location data. *Transactions in GIS*, 28 (8), 2668–2686.

Sun, J., *et al.*, 2024. NWSTAN: a lightweight dynamic spatial–temporal attention network for traffic prediction. Neural *c*omputing and *a*pplications. Berlin, Germany: Springer.

Titsias, M., 2009. Variational learning of inducing variables in sparse Gaussian processes. *In*: *Artificial intelligence and statistics*, 16–18 April Clearwater Beach, Florida, USA. PMLR, 567–574.

Wang, H., *et al.*, 2019. Early warning of burst passenger flow in public transportation system. *Transportation Research Part C: Emerging Technologies*, 105, 580–598.

Wang, J., *et al.*, 2021. Metro passenger flow prediction via dynamic hypergraph convolution networks. *IEEE Transactions on Intelligent Transportation Systems*, 22 (12), 7891–7903.

Wang, P., *et al.*, 2025a. Lightweighting the prediction process of urban states with parameter sharing and dilated operations. *International Journal of Digital Earth*, 18 (1), 2468414.

Wang, P., *et al.*, 2025b. Quickly forecasting the future state of urban sensors by the missing-data-tolerant deep learning approach. *Sustainable Cities and Society*, 118, 106044.

Wang, P., *et al.*, 2024. A lightweight spatiotemporal graph dilated convolutional network for urban sensor state prediction. *Sustainable Cities and Society*, 101, 105105.

Wang, P., *et al.*, 2025c. Efficient inference of large-scale air quality using a lightweight ensemble predictor. *International Journal of Geographical Information Science*, 39 (4), 900–924.

Wang, P., *et al.*, 2023. Urban traffic flow prediction: a dynamic temporal graph network considering missing values. *International Journal of Geographical Information Science*, 37 (4), 885–912.

Wang, X., *et al.*, 2022. Investigating the spatiotemporal pattern of urban vibrancy and its determinants: spatial big data analyses in Beijing, China. *Land Use Policy*, 119, 106162.

Wang, Y., 2023. Advances in spatiotemporal graph neural network prediction research. *International Journal of Digital Earth*, 16 (1), 2034–2066.

Wang, Y. and Zhu, D., 2024. A hypergraph-based hybrid graph convolutional network for intracity human activity intensity prediction and geographic relationship interpretation. *Information Fusion*, 104, 102149.

Wang, Z., *et al.*, 2025d. Capturing spatial heterogeneity of population-level human mobility via a prior-guided graph neural network. *International Journal of Geographical Information Science*, 1–28.

Wen, H., *et al.*, 2023. Diffstg: Probabilistic spatio-temporal graph forecasting with denoising diffusion models. *In:* P*roceedings of the 31st ACM international conference on advances in geographic information systems*. New York, NY: ACM, 1–12.

Xing, X., *et al.*, 2020. Mapping human activity volumes through remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5652–5668.

Yang, H., *et al.*, 2025. Exploring human mobility: a time-informed approach to pattern mining and sequence similarity. *International Journal of Geographical Information Science*, 39 (3), 627–651.

Yang, S. and Wu, Q., 2025. SDSINet: a spatiotemporal dual-scale interaction network for traffic prediction. *Applied Soft Computing*, 173, 112892.

Yang, W., *et al.*, 2023. A two-level random forest model for predicting the population distributions of urban functional zones: a case study in Changsha, China. *Sustainable Cities and Society*, 88, 104297.

Yao, Y., *et al.*, 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, 31 (4), 825–848.

Yu, B., Yin, H., and Zhu, Z., 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. *In*: *Proceedings of the 27th international joint conference on artificial intelligence*. Stockholm, Sweden: AAAI Press, 3634–3640.

Zeghina, A., *et al.*, 2024. Deep learning on spatiotemporal graphs: a systematic review, methodological landscape, and research opportunities. *Neurocomputing*, 594, 127861.

Zhang, T., *et al.*, 2024. Prior-Guided gated convolutional networks for rainstorm forecasting. *Journal of Hydrology*, 633, 130962.

Zhang, X., *et al.*, 2021. Traffic flow forecasting with spatial-temporal graph diffusion network. *In: Proceedings of the AAAI conference on artificial intelligence*. Washington DC: AAAI, 15008–15015.

Zhang, Y., *et al.*, 2023. An urban crowd flow model integrating geographic characteristics. *Scientific Reports*, 13 (1), 1695.

Zhao, L., *et al.*, 2020. T-GCN: a temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21 (9), 3848–3858.

Zhao, T., *et al.*, 2023. Developing a multiview spatiotemporal model based on deep graph neural networks to predict the travel demand by bus. *International Journal of Geographical Information Science*, 37 (7), 1555–1581.

Zhao, Y., *et al.*, 2025. Predicting origin-destination flows by considering heterogeneous mobility patterns. *Sustainable Cities and Society*, 118, 106015.

Zhu, F., *et al.*, 2025. A robust Bayesian Multi-Machine learning ensemble framework for probabilistic groundwater level forecasting. *Journal of Hydrology*, 650, 132567.

Zhu, J., *et al.*, 2024. A flood knowledge-constrained large language model interactable with GIS: enhancing public risk perception of floods. *International Journal of Geographical Information Science*, 38 (4), 603–625.