Graph Neural Network-based Tourism Demand Forecasting in Multivariate Time Series

Qun Yang[1], **Weijun Li**[2] and **Wencai Du**[3]

[1]Department of Computer Science, The University of Auckland, Auckland, New Zealand
[2]Faculty of International Tourism Management, City University of Macau, Macao SAR, China
[3]Institute for Data Engineering and Science, University of Saint Joseph, Macao SAR, China
qyan327@aucklanduni.ac.nz, vivian_lwjun@hotmail.com, george.du@usj.edu.mo

Abstract

Precise tourism demand forecasting is essential for efficient and effective tourism management. In recent studies, deep learning-based methods have dominated tourism demand forecasting. However, these models cannot explain the spatial effects, such as spatial spillover and spatial heterogeneity. Spatial dependence and regional interactions have significant influences on tourism. To this end, we propose a graph neural network-based model by explicitly integrating spatial effects. We learn the graph of relationships between attractions by introducing embedding vectors for each attraction. We also design a graph attention mechanism to compute the weights of neighboring nodes to highlight spatial effects. We first collect a real-world dataset of tourist volumes in Beijing. Comprehensive experiments are conducted to validate and compare the proposed method against baselines. Results show the forecasting accuracy of the proposed model outperforms baselines by a great margin. In addition, the learned graph structure and graph attention mechanism increase the interpretability of our model, which can be used as a recommendation tool for tourists.

1 Introduction

The tourism industry has played a major role in the global economy, with tourist volumes recovering 57% of pre-pandemic levels in the first seven months of 2022 [UNWTO, 2022]. Forecasting tourism demand for different tourist attractions is critical for efficient and effective tourism management, which can guide governments and destination management organizations to formulate policies and make strategic plans. However, accurate tourism demand forecasting is challenging due to the volatility of determining factors and external interventions.

Popular models for predicting tourism demand include conventional statistical and deep learning-based models. Statistical models, such as the error correction model, vector autoregressive model, autoregressive moving average (ARIMA), and seasonal ARIMA, rely on historical patterns to forecast future time series. In addition, the statistical models often explicitly consider the causality between explanatory variables (such as income and consumer price index) and tourism demand. Although conventional statistical models have achieved desirable forecasting performance, their rigorous assumptions of stationary and distribution result in difficulty when forecasting nonlinear time series, such as tourism demand data. Deep learning-based models have recently enabled improvements in tourism demand forecasting with multivariate time series. For example, the long short-term memory (LSTM) model [Bi *et al.*, 2020] is a promising application in tourism demand forecasting. More recently, a variant of the LSTM model by integrating the attention mechanism improves forecasting performance [Bi *et al.*, 2021]. However, these models do not explicitly capture dependencies between tourist attractions (i.e., spatial effect), resulting in difficulty in modeling tourism demand data using underlying inter-relationships. This limits their ability to improve the forecasting accuracy from such relationships.

The spatial effect (i.e., domain knowledge in tourism) includes spillover effect and spatial heterogeneity [Balli *et al.*, 2015; Li *et al.*, 2016; Yang and Fik, 2014], which can be explicitly considered in forecasting models to improve accuracy. The spillover effect is typically referred to as multi-destination tourism. That is, tourists tend to visit a few tourist attractions within one destination. Studies have confirmed significant spillover effects in certain regions [Balli *et al.*, 2015; Cao *et al.*, 2017; Assaf *et al.*, 2019]. On the other hand, because of the uniqueness of each destination or attraction, the spatial heterogeneity is related to spatial differences, which can be reflected in variables such as history and culture [Zhang *et al.*, 2011; Lin *et al.*, 2019]. Geographic heterogeneity has been identified in senior tourists' travel patterns [Losada *et al.*, 2019].

Therefore, a novel graph neural network-based model with spatial effects for enhanced tourism demand forecasting is proposed in this study. Recently, graph neural network models have shown success in modeling graph-structured data. The spillover effect can be represented by graph edges (i.e., relationships between tourist attractions) in our proposed model. We also introduce the embedding vector for each tourist attraction to model the spatial heterogeneity. Therefore, the domain knowledge in tourism is explicitly integrated into our model. The inner product between embedding vectors (i.e., cosine similarity) quantifies the differences between tourist attractions. Moreover, we incorporate a graph attention mechanism to compute the weights of related tourist attractions.

This study makes contributions to tourism demand forecasting in the following aspects:

- We develop a novel model based on the graph neural network with the attention mechanism, which learns a graph of relationships between tourist attractions to explicitly consider the domain knowledge in tourism (such as spillover effect and spatial heterogeneity).
- We evaluate the proposed model using a dataset of tourist volumes in Beijing. Experimental results show that our model forecasts tourism demand more accurately than baselines.
- Our model is an explainable model through embedding vectors (i.e., spatial heterogeneity) and learned graph structure (i.e., spillover effect). We show that the fusion of domain knowledge in tourism effectively improves forecasting performance.

2 Related Work

Time Series Modeling. Deep learning-based techniques have shown promising applications in modeling nonlinear high-dimensional time series. For instance, the convolutional neural network-based model [Munir *et al.*, 2018], long short-term memory model [Filonov *et al.*, 2016; Park *et al.*, 2018], autoencoder model [Yang *et al.*, 2021], and sequence to sequence model [Yang and Shen, 2022], have achieved desirable performance in time series forecasting tasks. More recently, the success of Transformer architecture has inspired time series tasks to use the self-attention mechanism. For example, Informer [Zhou *et al.*, 2021] and Autoformer [Wu *et al.*, 2021] outperform the classical methods in long-term time series forecasting tasks. However, these models cannot capture the inter-series dependencies, ignoring the spatial effect in forecasting tourism demand.

Graph Neural Networks. The graph neural network (GNN) has emerged as a promising method for modeling graph-structured data. GNN updates the representation of each node according to its neighbors, which builds the relationships between nodes. The graph convolution network (GCN) [Kipf and Welling, 2016] aggregates representations of each node's one-hop neighbors to fuse information. Based on GCN, the graph attention network [Veličković *et al.*, 2017] assigns weight coefficients for different neighbors according to the attention function when operating aggregation. Related variants have shown advances in various time series tasks. For instance, GNN-based models achieve desirable performance in traffic forecasting tasks [Yu *et al.*, 2017; Chen *et al.*, 2020]. Extensive applications in recommendation systems [Lim *et al.*, 2020; Schlichtkrull *et al.*, 2018] and relevant applications [Wang *et al.*, 2020] highlight the validity of GNN in terms of modeling large-scale multivariate time series. In addition, GNN shows a competitive application in anomaly detection [Deng and Hooi, 2021]. GNN models the relationship between sensor data, thereby improving anomaly detection accuracy compared with non-GNN models. However, the aforementioned GNN models share parameters to learn the representation of each node without considering spatial heterogeneity (i.e., different nodes are assumed to have different attributions). Furthermore, GNN updates nodes according to the graph structure, which is initially unknown in our tourism demand forecasting task. Therefore, we first need to learn the graph structure from time series data.

3 Proposed Method

3.1 Problem Definition

In this work, the training data consists of tourist volumes of N tourist attractions (i.e., multivariate time series) over T_{train} timestamps. The tourist arrival volume data is denoted as $\mathbf{X}_{\text{train}} = \left[\mathbf{x}_{\text{train}}^{(1)}, \ldots, \mathbf{x}_{\text{train}}^{(T_{\text{train}})}\right]$. At each timestamp t, $\mathbf{x}_{\text{train}}^{(t)} \in \mathbb{R}^N$ represents the number of tourists of N tourist attractions. We aim to predict tourist volumes in the testing data from the same N tourist attractions but across separate T_{test} timestamps. The testing data is represented as $\mathbf{X}_{\text{test}} = \left[\mathbf{x}_{\text{test}}^{(1)}, \ldots, \mathbf{x}_{\text{test}}^{(T_{\text{test}})}\right]$.

3.2 Model Overview

Our proposed model learns relationships between tourist attractions as a graph and then generates predictions using the historical data from the learned graph structure. As shown in Figure 1, the proposed model consists of three components:

- **Tourist Attraction Embedding** introduces embedding vectors to represent the unique attributions of each tourist attraction (i.e., spatial heterogeneity) for learning the graph structure;
- **Graph Structure Learning** builds a graph structure based on embedding vectors to capture dependencies between tourist attractions, thereby explicitly considering spatial effects;
- **Graph Attention-based forecasting** computes the weight coefficients of each node's neighbors to forecast the future tourist volume of each attraction, which distinguishes the importance of related attractions in the learned graph structure.

3.3 Tourist Attraction Embedding

Different tourist attractions can have very different characteristics, which can be crucial to affect the tourist volumes. For example, Beijing Olympic Green Park is a forest green park, while the Palace Museum is a historical monument. Then, it is plausible that these two tourist attractions would appeal to visitors to Beijing with various preferences. Therefore, we introduce a trainable **embedding vector** $\mathbf{v}_i \in \mathbb{R}^d$ ($i \in \{1, 2, \ldots, N\}$) for the i^{th} tourist attraction to explicitly consider its unique characteristics. Furthermore, the implicit factor underlying the tourist volumes can be captured.

The similarity between embeddings quantifies the similarity of tourist attractions. Therefore, tourist attractions with similar embedding vectors are assumed to have a high tendency to be associated with one another. In our method, the

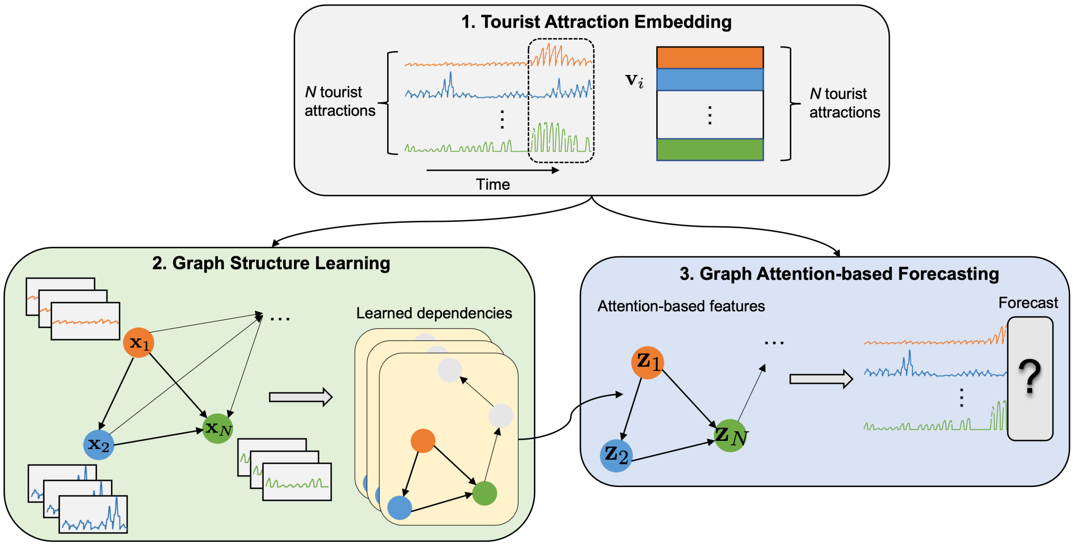


Figure 1: Model overview.

embedding plays two roles: (1) constructing the graph structure based on the characteristics of tourist attractions and (2) computing the weight coefficients of each node's neighbors to distinguish different types of tourist attractions in our attention mechanism.

3.4 Graph Structure Learning

The first step of our model is to capture the dependencies between tourist attractions because the graph structure is unknown. That is, we aim to learn a **directed graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where $\mathcal{V} = \{v_1, \ldots, v_n\}$ represents a set of nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represents a set of edges, and $A \in \mathbb{R}^{n \times n}$ represents a weighted adjacency matrix. In the context of tourism demand forecasting, a node $v_i \in \mathcal{V}$ represents a tourist attraction, while an edge e_{ij} represents the dependency between node i and node j. An edge from one tourist attraction to another indicates that this attraction contributes to the modeling of its neighbors. We use the directed graph because relationships between tourist attractions are not necessarily mutual.

We assume that the graph structure is related to attraction attributes. That is, tourist attractions can be classified into different types, where tourist attractions belonging to the same type are supposed to have interaction. Therefore, we sample these dependencies from a set of **candidate dependencies** $\mathcal{C}_i \subseteq \{1, 2, \ldots, N\} \backslash \{i\}$ for each tourist attraction i. The candidate dependencies of tourist attraction i are all tourist attractions other than itself.

The normalized inner product between node i's embedding vector and the embedding vector of its candidate node $j \in \mathcal{C}_i$ measures the similarity between the two nodes, which is used to determine the dependencies of tourist attraction i among these candidate nodes:

$$e_{ji} = \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{\|\mathbf{v}_i\| \|\mathbf{v}_j\|} \tag{1}$$

$$A_{ji} = \mathbb{1}\{j \in \mathsf{TopK}(\{e_{pi} : p \in \mathcal{C}_i\})\} \tag{2}$$

In summary, we learn a graph structure by grouping the nodes and their similar neighbors in terms of embeddings. We first compute e_{ji}, the normalized inner product, which measures the similarity between node i's embedding vector and the embedding vector of its candidate nodes. Then, we choose the top k nodes from $\mathcal{C}_i$ as node i's neighbors. Here, TopK represents the corresponding indices of the top-k normalized inner products. The k determines the sparsity level of the learned graph, a hyperparameter for graph structure learning. The constant k for all attractions creates a regular graph.

3.5 Graph Attention-based Forecasting

Following graph structure learning, we design a graph attention mechanism for tourism demand forecasting. At timestamp t, we define the input $\mathbf{x}^{(t)} \in \mathbb{R}^{N \times w}$ with a sliding lookback window of size w across the whole time series:

$$\mathbf{x}^{(t)} := \left[\mathbf{x}^{(t-w)}, \ldots, \mathbf{x}^{(t-1)}\right] \tag{3}$$

Given a lookback window w, the output of our model is the tourist volumes at the next s timestamps (i.e., prediction horizon of size s), $\hat{\mathbf{x}}^{(t)} = g(\mathbf{x}^{(t)})$, where $\hat{\mathbf{x}}^{(t)} \in \mathbb{R}^{N \times s}$, and $g(\cdot)$ denotes the graph-based prediction function, and $\hat{\mathbf{x}}^{(t)}$ predicts the next s timestamps of $\mathbf{x}^{(t)}$.

Temporal Feature Extractor. To distinguish the importance of neighboring nodes, we design a temporal feature extractor according to the graph attention mechanism to aggregate the node's representation. Compared with the usual graph attention-based models, our model explicitly characterizes different types of tourist attractions by introducing embedding vectors. To this end, we fuse node i's temporal feature $\mathbf{z}_i$ as follows:

$$\mathbf{z}_i^{(t)} = \mathsf{LeakyReLU}\left(a_{i,i}\mathbf{W}\mathbf{x}_i^{(t)} + \sum_{j \in \mathcal{N}(i)} a_{i,j}\mathbf{W}\mathbf{x}_j^{(t)}\right) \tag{4}$$

where $\mathbf{x}_i^{(t)} \in \mathbb{R}^w$ is node i's input time series within a look-back window, $\mathcal{N}(i)$ denotes a set of node i's neighbors according to the learned graph structure, $\mathbf{W} \in \mathbb{R}^{d \times w}$ is the weight matrix of a linear layer to extract temporal features. The attention coefficients $a_{i,j}$ are computed as:

$$\mathbf{h}_i^{(t)} = \mathbf{W}\mathbf{x}_i^{(t)} \oplus \mathbf{v}_i \tag{5}$$

$$\tilde{h}(i,j) = \mathsf{LeakyReLU}\left(\mathbf{att}^\mathsf{T}\left(\mathbf{h}_i^{(t)} \oplus \mathbf{h}_j^{(t)}\right)\right) \tag{6}$$

$$a_{i,j} = \frac{\exp\left(\tilde{h}(i,j)\right)}{\sum_{p \in \mathcal{N}(i) \cup \{i\}} \exp\left(\tilde{h}(i,p)\right)} \tag{7}$$

where $\mathbf{h}_i^{(t)}$ concates (denoted as $\oplus$) the transformed temporal feature $\mathbf{W}\mathbf{x}_i^{(t)}$ and the corresponding tourist attraction embedding vector $\mathbf{v}_i$, and $\mathbf{att}$ is a vector of attention coefficients. The attention coefficients are activated by the LeakyReLU function and then normalized by the softmax function.

Output Layer Following the graph attention-based feature extractor, we aggregate representation $\mathbf{z}_i^{(t)}$ with its tourist attraction embedding $\mathbf{v}_i$ across all nodes in the graph. The output is the predicted tourist volume at the next s timestamps, i.e., $\hat{\mathbf{x}}_i^{(t)}$:

$$\hat{\mathbf{x}}_i^{(t)} = f\left(\mathbf{v}_i \circ \mathbf{z}_i^{(t)}\right) \tag{8}$$

where $\circ$ denotes element-wise multiplication, and f denotes a stacked fully-connected linear layers.

We use the *Mean Squred Error* between the observed data $\mathbf{x}_i^{(t)}$ and the predicted data $\hat{\mathbf{x}}_i^{(t)}$ across all sliding windows for N attractions as the loss function:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=w+1}^{T_{\text{train}}}\left\|\hat{\mathbf{x}}_i^{(t)} - \mathbf{x}_i^{(t)}\right\|_2^2 \tag{9}$$

4 Experiments

We first collect a real-world dataset of tourist volumes in Beijing and then conduct experiments to evaluate our GNN-based model in this section.

4.1 Dataset

Beijing Tourism Official Website (http:www.visitbeijing.com.cn/) provides real-time tourist volumes of 134 attractions. The frequency of updating tourist volumes for all attractions is 15 min. The data from February 1st, 2023, to March 31st, 2023, is collected as the dataset in this study. Since most attractions are open from 9:00 to 21:00 every day, we select the data during this period as the dataset in this study. The 2007 samples from February 1st, 2023, to March 15th, 2023, are divided into the training dataset, while another 798 samples from March 16th, 2023, to March 31st, 2023, are used to test the model.

Figure 2 illustrates the tourist volumes of four famous attractions in Beijing from March 1st, 2023, to March 31st, 2023. We find that tourist volumes of four attractions exhibit different patterns. First, the peak tourist volumes vary significantly for different attractions. The peak tourist volume for Olympic Green Park is approximately 6.0×10^4, whereas the peak for Badaling Great Wall is about 0.04×10^4. This difference reflects the heat of attractions. Second, some attractions have more peaks in tourist volumes than others, such as Palace Museum and Summer Palace, which exhibit an obvious periodical pattern. Lastly, we observe that several surges occur at certain timestamps, such as Badaling Great Wall on March 24th, 2023, and Summer Palace on March 19th, 2023. To visualize the distribution of tourists in all attractions, we plot the heatmap of tourist volumes on February 4th, 2023, shown in Figure 3 as an example. We demonstrate that the geographical location of attraction significantly affects tourist volumes, with most tourists concentrating at the center of Beijing. This distinct distribution of tourists in different attractions highlights the necessity of using the GNN-based model because the relationship between attractions is a key factor affecting tourist volumes.

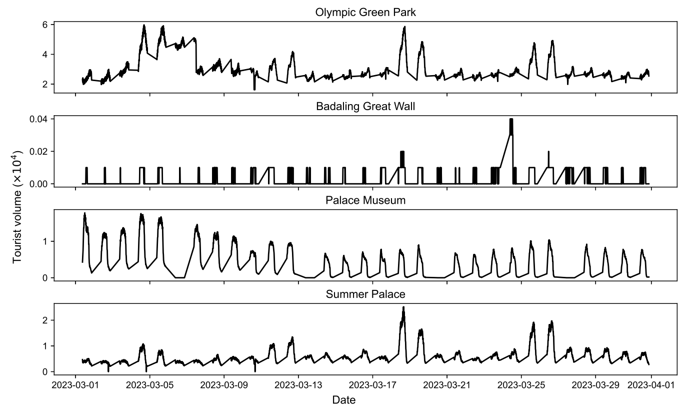


Figure 2: Tourist volumes of four famous attractions in Beijing from March 1st, 2023, to March 31st, 2023.

4.2 Baselines

We compare our GNN-based model against five popular tourism demand forecasting models, including:

- **ARIMA**: the autoregressive integrated moving average model [Box *et al.*, 2015], which is a statistical model using past values of a given time series.
- **LSTM**: the long short-term memory model [Hochreiter and Schmidhuber, 1997], which remembers values over arbitrary time intervals.
- **LSTM-AM**: a variant of LSTM by adding an attention mechanism (AM) [Law *et al.*, 2019], which endows the model with explainability by highlighting important features in time series.
- **CTS-LSTM**: a variant of LSTM by considering the correlated time series (CTS) [Wan *et al.*, 2020], which enables the spatial correlation of time series in the model.
- **CTS-LSTM-AM**: a variant of CTS-LSTM by using AM [Zheng *et al.*, 2021], which explicitly considers the

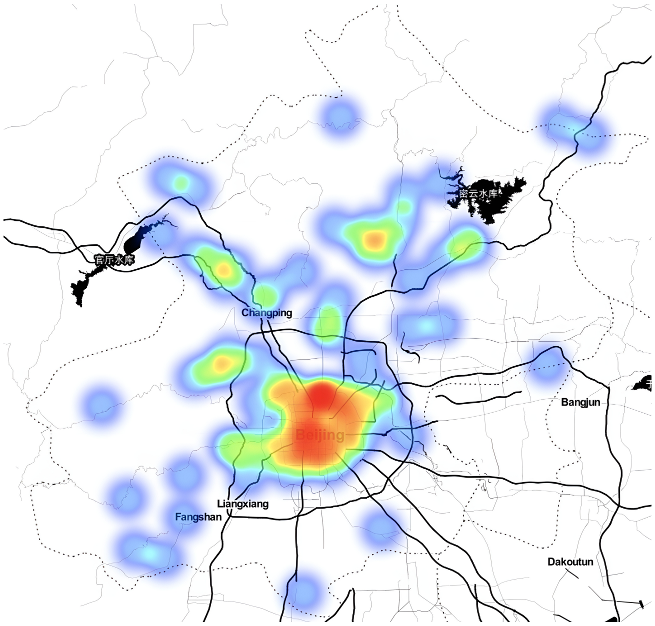


Figure 3: Heatmap of tourist volumes for all attractions on February 4th, 2023. Here, the red region indicates a high number of tourists, while the blue region indicates a less number of tourists.

spatial correlation and attention mechanism in the meantime.

4.3 Experimental Setup

We implement our proposed model in PyTorch [Paszke *et al.*, 2019] version 2.0 with CUDA 11.8 and PyTorch Geometric Library [Fey and Lenssen, 2019]. We run codes on a server with Intel(R) Core(TM) CPU i7-13700KF @ 3.40 GHz and a single NVIDIA RTX 3060 graphics card. The models are optimized using the Adam for up to 500 epochs. We set the learning rate as 5×10^{-3} and $(\beta_1, \beta_2) = (0.9, 0.99)$. We set the lookback window size w and the prediction horizon s as 15 and 10, respectively. We use 64-dimensional embedding vectors to represent different attractions and transform the initial series into 64-dimensional features.

4.4 Forecasting accuracy

The *Root Mean Squared Error* (RMSE) and *Mean Absolute Error* (MAE) of our proposed model and baselines are summarized in Table 1. Results show the RMSE and MAE of our model outperform baselines by a great margin. Both our model and CTS-LSTM-AM integrate the spatial correlation and attention mechanism. However, our model achieves a smaller MAE than CTS-LSTM-AM. This finding indicates that the key component (such as tourist attraction embedding) contributes to the forecasting performance.

Figure 4 illustrates the predicted tourist volumes of four famous attractions in Beijing. Olympic Green Park's predicted tourist arrival volume aligns well with the ground truth. Although our model overestimates/underestimates peak tourist volumes of the Palace Museum/Summer Palace, our model successfully captures the trend in time series. For Badaling Great Wall, the deviation between the predicted tourist arrival volume and ground truth is significantly larger than another three attractions. This finding highlights our model's limitation in predicting the pulse-like time series.

Table 1: Forecasting accuracy in terms of RMSE and MAE of our proposed model and baselines

	Metric	
Model	RMSE	MAE
ARIMA	0.608	0.424
LSTM	0.153	0.122
LSTM-AM	0.110	0.100
CTS-LSTM	0.075	0.058
CTS-LSTM-AM	0.071	0.053
Ours	**0.063**	**0.035**

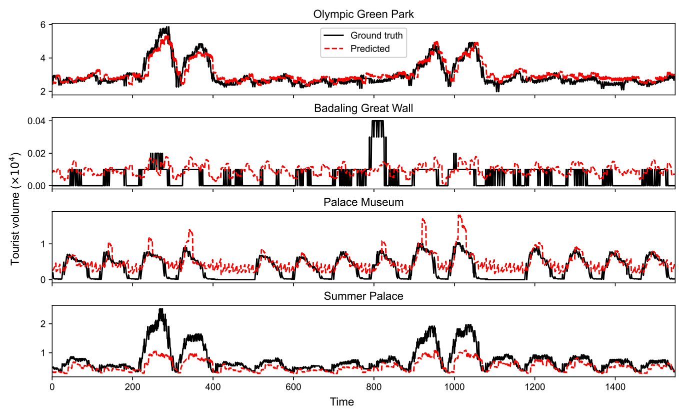


Figure 4: Predicted tourist volumes of four famous attractions in Beijing.

4.5 Ablation study

We gradually disable components to examine whether the forecasting performance degrades to investigate the importance of each component in the GNN-based model. First, we replace the learned graph with a complete graph (i.e., $\mathcal{N}(i) = \mathcal{C}_i$) to evaluate its necessity. Second, we remove attraction embedding vectors from the attention mechanism, i.e., $\mathbf{h}_i^{(t)} = \mathbf{W}\mathbf{x}_i^{(t)}$ in Eq. (5). Lastly, we assign equal weight coefficients to each node's neighbors instead of the attention mechanism.

Table 2: Forecasting accuracy in terms of RMSE and MAE of our proposed model and its variants

Model	RMSE	MAE
Ours		
-TOPK	0.108	0.062
-EMB	0.131	0.064
-ATT	0.244	0.131

We summarize the results of the ablation study in Table 2 and demonstrate the following aspects:

- The forecasting accuracy degrades after substituting the learned graph structure with a complete graph, indicating that the sub-graph sampling is effective in enhancing forecasting performance.

- Ignoring attraction embedding vectors in the attention mechanism degrades the forecasting accuracy, implying that modeling spatial heterogeneity improves forecasting performance.
- Disabling the attention mechanism most degrades the forecasting performance. Because tourist attractions differ significantly in characteristics, the model cannot distinguish the importance of each node's neighbors if treating them equally. This validates the necessity of using the graph attention mechanism.

The ablation study shows that each component in our GNN-based model (i.e., graph structure learning, tourist attraction embedding, and graph attention mechanism) enables better forecasting performance than baselines.

5 Conclusions

A novel GNN-based model for tourism demand forecasting is proposed in this study. The domain knowledge in tourism, such as spillover effect and spatial heterogeneity, is explicitly considered in the model. We learn the graph of relationships between attractions by introducing embedding vectors for each attraction. We also design a graph attention mechanism to compute the importance of neighboring nodes to highlight spatial effects. We conduct extensive experiments to validate and compare our GNN-based model against baselines. Results show the RMSE and MAE of our model outperform baselines by a great margin. The ablation study verifies the effectiveness of each component in our model. In addition, the learned graph structure and graph attention mechanism increase the interpretability of our model, which can be used as a recommendation tool for tourists.

Acknowledgments

This work was supported by the China Scholarship Council under Grant 201806710072.

References

[Assaf *et al.*, 2019] A George Assaf, Gang Li, Haiyan Song, and Mike G Tsionas. Modeling and forecasting regional tourism demand using the bayesian global vector autoregressive (bgvar) model. *Journal of Travel Research*, 58(3):383–397, 2019.

[Balli *et al.*, 2015] Faruk Balli, Joshua Curry, and Hatice Ozer Balli. Inter-regional spillover effects in new zealand international tourism demand. *Tourism Geographies*, 17(2):262–278, 2015.

[Bi *et al.*, 2020] Jian-Wu Bi, Yang Liu, and Hui Li. Daily tourism volume forecasting for tourist attractions. *Annals of Tourism Research*, 83:102923, 2020.

[Bi *et al.*, 2021] Jian-Wu Bi, Hui Li, and Zhi-Ping Fan. Tourism demand forecasting with time series imaging: A deep learning model. *Annals of tourism Research*, 90:103255, 2021.

[Box *et al.*, 2015] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[Cao *et al.*, 2017] Zheng Cao, Gang Li, and Haiyan Song. Modelling the interdependence of tourism demand: The global vector autoregressive approach. *Annals of Tourism Research*, 67:1–13, 2017.

[Chen *et al.*, 2020] Weiqi Chen, Ling Chen, Yu Xie, Wei Cao, Yusong Gao, and Xiaojie Feng. Multi-range attentive bicomponent graph convolutional network for traffic forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3529–3536, 2020.

[Deng and Hooi, 2021] Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4027–4035, 2021.

[Fey and Lenssen, 2019] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

[Filonov *et al.*, 2016] Pavel Filonov, Andrey Lavrentyev, and Artem Vorontsov. Multivariate industrial time series with cyber-attack simulation: Fault detection using an lstm-based predictive data model. *arXiv preprint arXiv:1612.06676*, 2016.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Law *et al.*, 2019] Rob Law, Gang Li, Davis Ka Chio Fong, and Xin Han. Tourism demand forecasting: A deep learning approach. *Annals of tourism research*, 75:410–423, 2019.

[Li *et al.*, 2016] Hengyun Li, Jason Li Chen, Gang Li, and Carey Goh. Tourism and regional income inequality: Evidence from china. *Annals of Tourism Research*, 58:81–99, 2016.

[Lim *et al.*, 2020] Nicholas Lim, Bryan Hooi, See-Kiong Ng, Xueou Wang, Yong Liang Goh, Renrong Weng, and Jagannadan Varadarajan. Stp-udgat: Spatial-temporal-preference user dimensional graph attention network for next poi recommendation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 845–854, 2020.

[Lin *et al.*, 2019] Vera Shanshan Lin, Yang Yang, and Gang Li. Where can tourism-led growth and economy-driven tourism growth occur? *Journal of Travel Research*, 58(5):760–773, 2019.

[Losada *et al.*, 2019] Nieves Losada, Elisa Alen, Tomas R Cotos-Yanez, and Trinidad Dominguez. Spatial heterogeneity in spain for senior travel behavior. *Tourism Management*, 70:444–452, 2019.

[Munir *et al.*, 2018] Mohsin Munir, Shoaib Ahmed Siddiqui, Andreas Dengel, and Sheraz Ahmed. Deepant: A deep learning approach for unsupervised anomaly detection in time series. *Ieee Access*, 7:1991–2005, 2018.

[Park *et al.*, 2018] Daehyung Park, Yuuna Hoshi, and Charles C Kemp. A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. *IEEE Robotics and Automation Letters*, 3(3):1544–1551, 2018.

[Paszke *et al.*, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[Schlichtkrull *et al.*, 2018] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*, pages 593–607. Springer, 2018.

[UNWTO, 2022] UNWTO. World tourism barometer. https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2022-01/UNWTO_Barom22_01_January_Excerpt.pdf?lXWPdwSpA0FBYClC_xMHrLYarzBnARrV, 2022. Accessed: (17 May 2023).

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[Veličković *et al.*, 2017] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[Wan *et al.*, 2020] Huaiyu Wan, Shengnan Guo, Kang Yin, Xiaohui Liang, and Youfang Lin. Cts-lstm: Lstm-based neural networks for correlatedtime series prediction. *Knowledge-Based Systems*, 191:105239, 2020.

[Wang *et al.*, 2020] Yiwei Wang, Wei Wang, Yujun Ca, Bryan Hooi, and Beng Chin Ooi. Detecting implementation bugs in graph convolutional network based node classifiers. In *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)*, pages 313–324. IEEE, 2020.

[Wu *et al.*, 2021] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34:22419–22430, 2021.

[Yang and Fik, 2014] Yang Yang and Timothy Fik. Spatial effects in regional tourism growth. *Annals of Tourism Research*, 46:144–162, 2014.

[Yang and Shen, 2022] Qun Yang and Dejian Shen. Learning damage representations with sequence-to-sequence models. *Sensors*, 22(2):452, 2022.

[Yang *et al.*, 2021] Qun Yang, Dejian Shen, Wencai Du, and Weijun Li. A deep learning-based framework for damage detection with time series. *IEEE Access*, 9:66570–66586, 2021.

[Yu *et al.*, 2017] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

[Zhang *et al.*, 2011] Yan Zhang, Jian-Hua Xu, and Pei-Jun Zhuang. The spatial relationship of tourist distribution in chinese cities. *Tourism Geographies*, 13(1):75–90, 2011.

[Zheng *et al.*, 2021] Weimin Zheng, Liyao Huang, and Zhibin Lin. Multi-attraction, hourly tourism demand forecasting. *Annals of Tourism Research*, 90:103271, 2021.

[Zhou *et al.*, 2021] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

A Explainability

Tourist Attraction Embeddings. To explain our model from a perspective of learning tourist attraction embeddings, we visualize original 64-dimensional embeddings by transforming them to 2-dimensional embeddings using t-SNE [Van der Maaten and Hinton, 2008]. As shown in Figure 5, the distance between nodes (i.e., tourist attractions) represents the similarity between attractions' attributes, which may be helpful in classifying tourist attractions in Beijing. The learned attraction embeddings exhibit a local clustering pattern in a 2-dimensional space, indicating our model's validity in reflecting the local attractions' attribute similarity. We observe a group of attractions forming 4 local clusters, shown in the dashed circled region. In detail, green nodes (such as Palace Museum and Summer Palace) represent famous historical monuments, blue nodes (such as Olympic Green Park) represent forest green parks, yellow nodes represent mountaineering scenic areas, and red nodes represent other types of attractions. Our model effectively distinguishes attractions according to attributes, which validates the effectiveness of tourist attraction embeddings in tourist demand forecasting.

Learned Graph Structure. Figure 6 shows the learned graph structures of two attractions, with each having 20 edges (i.e., $k = 20$). We find that the learned graph structure is related to the geographical location of the attraction. If there are many other attractions around an attraction, such as the Palace Museum, the majority of edges are connected between neighboring nodes (see Figures 6a). Most connections are between distant attractions if an attraction is far from other attractions, such as Badaling Great Wall shown in Figure 6b. The learned graph structure can be used as a recommendation tool for tourists, which baselines cannot achieve (i.e., non-GNN models).

Attention weights. In our model, we fuse information (i.e., attraction embeddings and transformed temporal features) to the target node according to the learned attention weights.

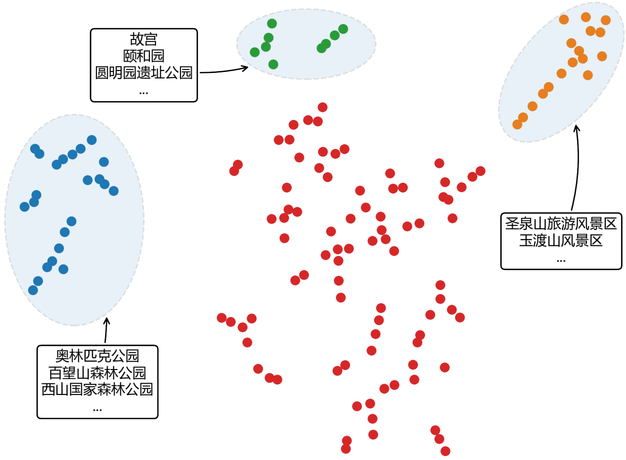
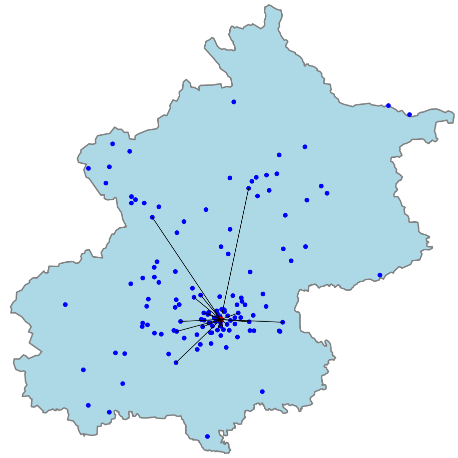
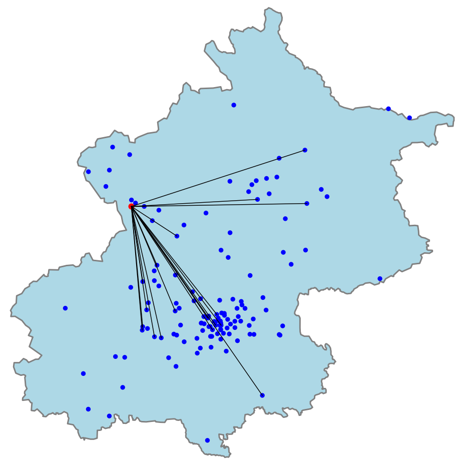


Figure 5: A visualization of the attraction embeddings in the 2-dimensional space using t-SNE. Here, green nodes represent famous historical monuments, blue nodes represent forest green parks, yellow nodes represent mountaineering scenic areas, and red nodes represent other types of attractions.

The attention weights quantify the importance of related nodes to the target node. Figure 7 shows the attention weights of each attraction and its related attractions. The attention weight matrix is sparse because each attraction has 20 links, with the attention weights of the rest attraction being zero. The sparse attention weight matrix effectively reduces the computational cost. In addition, the ablation study has shown that a complete graph decreases the model performance in terms of forecasting accuracy. Figure 8 illustrates the attention weights of two attractions in Beijing. The attention weights reflect the contribution of each node's connected neighbors when aggregating its information. The distribution of attention weights is also related to the geographical location of the attraction. For attractions that many other attractions are around (such as Palace Museum), neighboring attractions contribute more to tourism demand forecasting. For Badaling Great Wall, which is far from other attractions, distant attractions contribute more than neighboring attractions. The attention weights explain the spillover phenomenon in tourism demand forecasting. Therefore, our model can recommend hot attractions to tourists by ranking the attention weights.



(a) Palace Museum



(b) Badaling Great Wall

Figure 6: Learned graph structure.

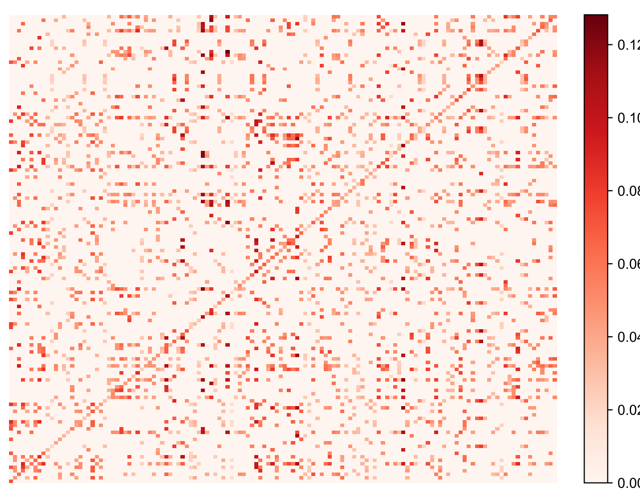
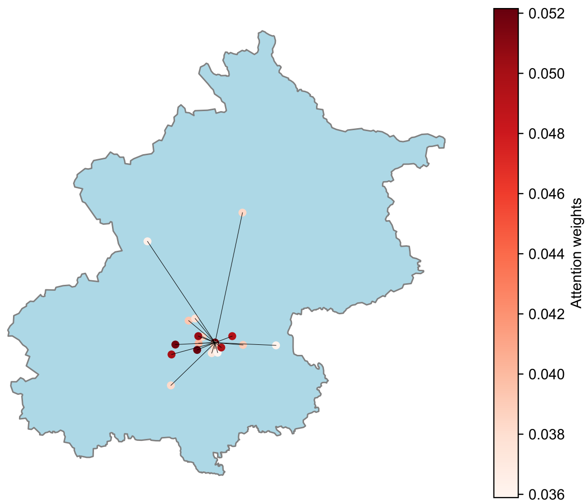
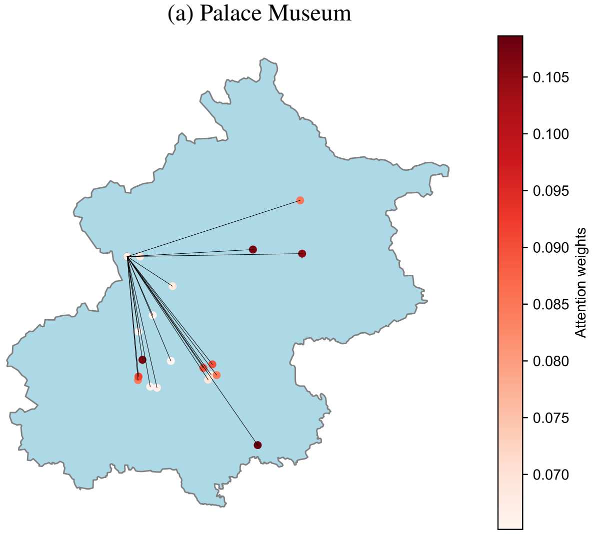


Figure 7: Attention weight matrix of the learned graph.



(a) Palace Museum



(b) Badaling Great Wall

Figure 8: Attention weights of attractions