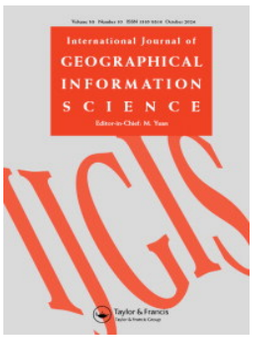

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Next location prediction using heterogeneous graph-based fusion network with physical and social awareness

Sijia He, Wenying Du, Yan Zhang, Lai Chen, Zeqiang Chen & Nengcheng Chen

To cite this article: Sijia He, Wenying Du, Yan Zhang, Lai Chen, Zeqiang Chen & Nengcheng Chen (2024) Next location prediction using heterogeneous graph-based fusion network with physical and social awareness, International Journal of Geographical Information Science, 38:10, 1965-1990, DOI: 10.1080/13658816.2024.2375725

To link to this article: https://doi.org/10.1080/13658816.2024.2375725

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2024, VOL. 38, NO. 10, 1965–1990
https://doi.org/10.1080/13658816.2024.2375725

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Next location prediction using heterogeneous graph-based fusion network with physical and social awareness

Sijia He[a], Wenying Du[a], Yan Zhang[b], Lai Chen[a], Zeqiang Chen[a] and Nengcheng Chen[a]

[a]National Engineering Research Center for Geographic Information System, China University of Geosciences (Wuhan), Wuhan, China; [b]Institute of Space and Earth Information Science, Chinese University of Hong Kong, Hong Kong, China

ABSTRACT

Location prediction based on social media information is highly valuable in human mobility research and has multiple real-life applications. However, existing research methods often ignore social influences, largely ignoring implicit information regarding interactions between users and geographical locations. Additionally, they generally employ single modeling structures, which restricts the effective integration of complex spatiotemporal characteristics and factors influencing user mobility. In this context, we propose a novel network with physical and social awareness that expresses both physical and social influences of user mobility from a global perspective based on a heterogeneous graph constructed using users and spatial locations as nodes and relationships between them as edges. This graph enables the model to leverage information from connected nodes and edges to infer missing or unobserved data. The model predicts future locations of users by effectively integrating the temporal and spatial features of user trajectory series. The proposed model is validated using three social media datasets. The experimental results demonstrate that the proposed method outperforms the state-of-the-art baseline models. This indicates the importance of considering complex interactions between users and locations, as well as the various influences of physical and social spaces.

ARTICLE HISTORY
Received 5 June 2023
Accepted 30 June 2024

KEYWORDS
Next location prediction; social media data; heterogeneous graph network; spatial–temporal information fusion

1. Introduction

From population migration across cities to commuters in large metropolitan areas, human mobility occurs on a daily basis, is a fundamental property of human behavior, and wields an enormous impact on human societies (Barbosa *et al.* 2018, Alessandretti *et al.* 2020). One of the main problems in human mobility research is forecasting future destinations of individuals based on their historical mobility data (Barbosa *et al.* 2018). This has long been a topic of great interest in a wide range of fields because of its relevance to numerous real-world applications. These include anticipating changes in population distributions in specific regions to aid urban planning, business site

CONTACT Nengcheng Chen chennengcheng@cug.edu.cn; Wenying Du duwenying@cug.edu.cn

selection, and tourism resource development (Wang *et al.* 2019, Song *et al.* 2010a); leveraging geotagged epidemic information shared on social media for virus transmission prediction (Leung *et al.* 2021, Ilin *et al.* 2021); forecasting population location information for timely emergency responses during disasters; and risk assessment. Moreover, by analyzing the high-level and regular temporal and spatial patterns inherent in human mobility data (González *et al.* 2008, Song *et al.* 2010b, Jia *et al.* 2020), personal location prediction technology is rapidly gaining significance in several advanced location-based services (LBSs) and applications, such as next-point of interest (POI) recommendations, targeted advertising, and location-based content delivery (Noulas *et al.* 2012, Li *et al.* 2022).

With the widespread popularity of mobile devices and rapid development in location technology, very high amounts of individual-level tracking data, eg, mobile phone signaling data, global positioning system (GPS) log data, and geotagged social media data, are collected over the Internet and by application service providers (Miller and Goodchild 2015). This provides significant opportunities for further research on human mobility. Social media data such as location-based social network (LBSN) service check-in data are a form of volunteered geographic information (VGI) data (Sui *et al.* 2012) that offer the advantages of openness, high data volume, and rich contextual and semantic information associated with users and geographical locations. Thus, their analysis enables the study of mobility in a broad context (Jurdak *et al.* 2015, Yin *et al.* 2022). For instance, authors have utilized real-time information on disaster responses contained in social media data to measure the impact of disasters on population dynamics (Martín *et al.* 2020). However, not all valuable information available on social media includes geographical tags, eg, only approximately 1.5% of the 775,000 tweets about hurricanes collected by Kryvasheyeu *et al.* (2016) contained location tags. Research on location prediction methods based on social media data has the potential to assist researchers in deducing plausible locations for significant tweets based on users' historical geotagged tweets. This, in turn, enhances the utility of social media data and enables emergency management agencies to investigate the impact of disasters on human society accurately.

Predicting future locations of users requires analysis of spatial and temporal patterns that characterize user habits based on trajectory data. A considerable amount of literature on personal location prediction has been published by capturing and analyzing these patterns. For example, recurrent neural networks (RNNs) have been used to capture temporal patterns in human trajectories for prediction (Liu *et al.* 2016, Kong and Wu 2018). However, certain limitations remain. First, user activity recorded on LBSNs is typically sparse in terms of temporal and spatial aspects (Hawelka *et al.* 2014, Sloan and Morgan 2015, Afyouni *et al.* 2022). Temporal sparsity refers to the fact that collected check-in data typically represent only a small proportion of user activity, with users checking in only a few times over long periods. Geotagged user activity records are represented as temporally arranged location sequences, with the time intervals between locations being expressed in minutes, hours, or days (Tu *et al.* 2017). Spatial sparsity refers to the tendency of users to check in at a limited number of locations (Huang *et al.* 2019a). Previously conducted studies have typically alleviated data sparsity by either incorporating additional information or establishing features related to human mobility, such as time and space (Yao *et al.* 2023, Yang *et al.* 2023).

Second, the lack of consideration of social influences hinders the assessment of complex influences of physical and social spaces from a global perspective. These complex contextual factors include individual preferences, geographic influences, temporal contexts, and social influences (Barbosa *et al.* 2018, Luca *et al.* 2021, Wang *et al.* 2022a, Wang *et al.* 2023a). Individual preferences involve the degree of favorability that individuals have for different locations (Lim *et al.* 2020). Geographical influences represent a pivotal consideration; however, they are highly complex and often challenging to accurately approximate using linear calculations (Wang *et al.* 2022b). For example, individuals may be more interested in locations near their current positions to avoid travelling long distances. However, inherent transition patterns may exist between locations. If a user is currently at a restaurant and already had lunch, it is quite plausible for them to visit an entertainment site rather than another restaurant, even if another restaurant is geographically closer. Temporal contexts refer to the tendency of users to prefer different locations at different times, representing high variability in temporal patterns of locations (Han *et al.* 2020, Wang *et al.* 2021a). For example, people may visit a restaurant during the midday on weekdays or a forested park on weekends. Social influences are related to the intersection between human mobility and social networks that extend beyond the obvious connection caused by the social nature of human beings (Barbosa *et al.* 2018). For instance, the movement of two individuals who often interact socially, such as friends, family, or colleagues, can be reasonably expected to not be independent (Carrasco and Miller 2006). People may visit locations frequented by their friends. Social influence plays a significant role in mobility prediction (Takens 2006, Lim *et al.* 2020, Wang *et al.* 2021a).

However, most existing studies have ignored social influences and the complexity of geographical influences (only spatial distance has been considered). Moreover, such complex features cannot be analyzed solely using sequence-based modeling, particularly with regard to users, locations, and other complex relationships. Human mobility typically involves complex interactions between users and locations. These interactions form complex networks that can help model the complex influences of both physical and social spaces. Finally, besides capturing various factors influencing user mobility, it is important to consider the different contributions of relationships or features that impact on next location prediction. Effectively integrating the captured information to enhance prediction accuracy remains challenging.

In this context, we propose a novel method that leverages physical and social awareness based on a heterogeneous graph (PS-HGAN) for next location prediction. By elaborately defining a data schema, we construct a heterogeneous graph (HG) network that considers user preferences, geographic distances, time period similarities, transition frequencies, and social space influences. Users and spatial venues are considered as nodes, and the different relationships between them are considered as edges. The multidimensional relationships within the heterogeneous graph enable denser representations of available data. Additionally, through indirect connections (such as users sharing similar preferences or locations with closer distances), the model can infer information not directly observed in the data, thus compensating for sparse data. Using the flexible propagation mechanism of HG attention networks (Lv *et al.* 2021), nodes in HGs can be easily mapped into a low-dimensional complex

space, and incorporating edge-type attention allows the influences of different relationships to be accounted for. Next, we employ a long short-term memory (LSTM)-based method to model the temporal features of users. Finally, we fuse the temporal features with the spatial features learned from HGs using an attention mechanism, enabling the model to better understand the underlying patterns and dynamics of user mobility, thereby reducing the impact of data sparsity and improving prediction performance. In summary, our study makes the following contributions:

- We employ an HG to establish relationships between global locations and users, providing a holistic perspective for modeling human mobility features in both physical and social spaces.
- We propose a novel PS-HGAN framework for predicting the next locations of individuals. The integrated spatial features learned from the HG and the time-series features learned from the sequence model are effectively integrated using a fusion module. Additionally, by leveraging the advantages of the attention technique, this model considers the distinct influences of different types of relationships as well as spatial and temporal features.
- Experiments are conducted on three different datasets across the LBSN, and the results demonstrate that our approach outperforms the baseline and state-of-the-art methods.

2. Related works

2.1. Objective of the next location prediction task

The next location prediction task aims to predict future destinations of individuals based on their historical mobility records. Next POI recommendation (Agrawal *et al.* 2021, Rahmani *et al.* 2022, Ma *et al.* 2024), which is a variant of the next location prediction task, is the area most relevant to this study. It involves recommending specific POIs to users that they are most likely to visit based on their historical data. However, owing to spatial and temporal sparsity of trajectory data records in LBSNs, predicting specific locations to be visited by users may result in low accuracy. Therefore, the POI recommendation task tends to calculate the likelihood of a large number of possible POIs that a user may visit and ranks them in descending order of visiting likelihood for user selection (Han *et al.* 2020). However, in other tasks requiring location prediction, such as regional planning, selecting business sites, infectious disease transmission prediction, and urban flood risk measurement, approximate location prediction is sufficient for coarse-scale surveys. Thus, Bao *et al.* (2021) proposed a model combining a convolutional neural network (CNN) and a bidirectional LSTM architecture, inspired by the idea of converting trajectories into finite number of locations to achieve higher prediction accuracy based on GPS trajectory data. The proposed system employs spatial clustering to convert POI predictions into coarse-scale clustered region predictions. This approach provides insights into related applications that require coarse-scale location prediction results. However, spatial clustering-based approaches may fragment semantically meaningful spatial units with the same functional attributes. To avoid this, we adopt the traffic analysis zone (TAZ) (Yao *et al.* 2017), which is the basic unit

of urban research formed by partitioning city areas using the urban transportation network as the basis. for next location prediction.

2.2. Sequential next location prediction

Early studies on next location prediction typically used conventional probabilistic, pattern-based, or collaborative filtering approaches, such as Markov chain (MC)-based approaches (Du *et al.* 2018, Gambs *et al.* 2012, Mathew *et al.* 2012), pattern mining algorithms (Monreale *et al.* 2009) and matrix factorization (MF) (Koren 2008). For example, Guo *et al.* (2010) extracted frequent patterns of user motion from trajectory data and constructed a pattern tree to predict next locations. Factorizing personalized Markov chains (FPMCs) (Rendle *et al.* 2010) integrate the advantages of both MF (Koren 2008) and MCs to recommend next locations by considering localized region constraints. To consider higher number of previously visited locations, Gambs *et al.* (2012) extended the MC model to a mobility Markov chain (MMC) model, achieving 70–95% accuracy with the extended n-MMC model. Mathew *et al.* (2012) clustered access records into meaningful locations and trained hidden Markov models (HMMs) for each, where every state represents a probabilistic distribution rather than a single location. Based on the GeoLife project dataset, the prediction accuracy of HMM becomes 13.85% when considering an area of approximately 1280 square meters.

These methods explore the temporal features of user trajectories and perform well when handling small amounts of data. However, they often require extensive feature engineering and are time-consuming processes. Additionally, traditional approaches cannot capture long-range temporal and spatial dependencies.

With the rapid development of artificial intelligence technology, deep learning-based approaches have become widely used, as shown in Table 1. RNN-based methods (Liu *et al.* 2016) have exhibited excellent mobility prediction performance by considering user trajectories as successive sequences and capturing the dynamic

Table 1. Deep learning approaches and their base modules for next location prediction.

Problem	References	Model name	Base modules
Next Location Prediction	Liu *et al.* (2016)	STRNN	RNN
	Kong and Wu (2018)	HST-LSTM	LSTM
	Wang *et al.* (2021c)	STKG	Knowledge Graph
	Lin *et al.* (2021)	CTLE	Transformer
Next POI Recommendation	Huang *et al.* (2019b)	ATST-LSTM	LSTM, Attention
	Zhao *et al.* (2022)	STGN	LSTM
	Luo *et al.* (2021)	STAN	Attention
	Yu *et al.* (2020)	CatDM	LSTM
	Sun *et al.* (2020)	LSTPM	LSTM
	Lian *et al.* (2020)	GeoSAN	Feed-Forward Neural Network (FFN)
	Wu *et al.* (2022)	PLSPL	LSTM
	Lim *et al.* (2020)	STP-UDGAT	Graph Attention Networks (GAT)
	Bao *et al.* (2021)	BiLSTM-CNN	Graph learning, LSTM, (Convolutional Neural Networks) CNN
	Wang *et al.* (2021b)	ASGNN	GNN, Attention
	Wang *et al.* (2021a)	CTHGAT	GNN, LSTM
	Wang *et al.* (2022b)	GSTN	Graph learning, LSTM, Attention
	Kim *et al.* (2021)	DynaPosGNN	Graph learning, Multilayer Perceptron (MLP)

preferences and temporal regularity of human movements on this basis. To address the effects of both temporal and geographic factors while acquiring sequence features effectively, researchers have concentrated on integrating various types of spatial and temporal information into the classical RNN architecture, which can partially alleviate data sparsity issues. For example, Hierarchical spatial–temporal LSTM (HST-LSTM) (Kong and Wu 2018) accepts spatial–temporal factors as the input parameter of the LSTM gate mechanism and extends the model to a hierarchical one to improve prediction performance.

Some studies have focused on the long- and short-term interests of sequential patterns. For example, the spatiotemporal gated network (STGN) (Zhao *et al.* 2022) adds two pairs of time gates and distance gates to LSTM for long-term interest and short-term interest modeling, respectively. Sun *et al.* (2020) developed long- and short-term preference modeling (LSTPM). Yang *et al.* (2023) proposed a user preference transfer and drift network (UPTDNet) for cross-city next POI recommendation.

Some studies have leveraged attention mechanisms as not all historical records are relevant to the records to be visited. For instance, Huang *et al.* (2019b) introduced the attention-based spatiotemporal LSTM (ATST-LSTM). Luo *et al.* (2021) developed the spatiotemporal attention network (STAN). Li *et al.* (2020) attempted two hierarchical attention-based encoder–decoder approaches to capture potential long- and short-term dependencies in individual longitudinal trajectories, including daily local time patterns and weekly global time travel patterns. Yao *et al.* (2023) proposed a geographical embedding and multilayer attention BiLSTM (GEMA-BiLSTM) model that combines location and spatiotemporal information to extract the semantics of human mobility. Other studies have considered the semantic features of visited locations. For instance, the category-aware deep model (CatDM) (Yu *et al.* 2020) uses two different LSTM encoders to mine user preferences, one of which is used for POI categories.

In summary, most existing studies use time-series models to predict user trajectories based on temporal features and improve prediction performance by combining relevant backgrounds, including frequent patterns (Guo *et al.* 2010), spatiotemporal contexts (Liu *et al.* 2016), semantic features (Wang *et al.* 2021a), and long- and short-term user preferences (Sun *et al.* 2020). This illustrates the importance of considering multiple influencing factors during location prediction. The additional information provides models with more background knowledge, enabling better understanding of the underlying motivations behind user mobility in sparse data scenarios (Yao *et al.* 2023, Yang *et al.* 2023). However, these methods ignore the complexity of nonlinear geographical effects (merely considering the influence of geographical distance) and usually disregard social spatial information, which has been demonstrated to affect user mobility significantly. Further, methods based solely on sequential analysis struggle to handle and integrate these complex features, and suffer from limitations induced by local perspectives, in which only the information in the current sequence affects the prediction results.

2.3. Graph-based next location prediction

Human mobility flow generated by movement in the physical world involves interactions between people and geographic entities. These interactions naturally constructing a spatial

association graph (Zhang *et al.* 2022). Hence, several approaches have modeled user trajectories as spatial correlation graphs in order to obtain global spatial features. Researchers have mainly explored the following five types of graph structures for location prediction. (1) Simple isomorphic graphs: spatial distance, temporal distance, and transitions between locations are relationships that are often used to construct relational graphs, such as those in the spatial–temporal preference user dimensional graph attention (STP-UDGAT) network (Lim *et al.* 2020), BiLSTM-CNN (Bao *et al.* 2021), and the graph-enhanced spatial–temporal network (GSTN) (Wang *et al.* 2022b). Unlike the BiLSTM-CNN and GSTN methods, STP-UDGAT considers social relationships among users. This consideration enables the transfer of different user preferences, resulting in a collaborative filtering effect that aids the prediction for the current user based on the preferences of similar users. (2) Sequential graphs: sequential graphs model users' access sequences as a continuous graph with input–output relationships. The attentive sequential model based on a graph neural network (ASGNN) (Wang *et al.* 2021b) uses this type of graph structure to learn the users' long- and short-term preferences and adds an attention mechanism for next position recommendation. (3) Dynamic graph: Yin *et al.* (2023) constructed a sequence-based dynamic neighbor graph (SDNG) to determine similarity neighborhoods and developed a multistep dependency prediction (MSDP) model that explicitly leverages information from previous states. (4) HG: Wang *et al.* (2021a) first utilized HG modeling of user movement records for location prediction, which integrates complex relationships between locations and users in a way that captures implicit information present in human mobility networks more comprehensively. In addition, semantic category information of the locations is added using LSTM and parameter sharing. (5) Knowledge graph: the knowledge graph-based approach provides a new way to combine spatiotemporal trajectory records with knowledge obtained from multiple sources. For example, the urban spatio-temporal knowledge graph established by Wang *et al.* (2021c), which mainly contains spatio-temporal movement relations and affiliation relations, combines physical, social, and semantic spaces. Other recent studies on hypergraphs have also demonstrated their effectiveness (Yan *et al.* 2023). For example, Wang *et al.* (2023b) proposed an enhanced encoder–decoder network (EEDN) to exploit rich latent features between users, POIs, and interactions for POI recommendation and to derive more robust cold start-aware user representations.

These graph structures exhibit outstanding advantages in modeling complex interaction relationships. However, although CTHGAT (Wang *et al.* 2021a) considers different relationships in both physical and social spaces, these relationships contribute differently to users' access behaviors. Our proposed approach uses an attention mechanism to focus on more relevant relationships and combines the features learned from the graph with those learned from the LSTM-based model using an attention mechanism. This approach enables the model to better comprehend the underlying patterns and dynamics of user mobility, thereby reducing the impact of data sparsity and improving prediction performance.

3. Methodology

In this section, we provide a detailed description of the proposed PS-HGAN model. A flowchart of the model is presented in Figure 1, and the different components is introduced in Sections 3.1–3.4. First, because users' historical check-in data consist of

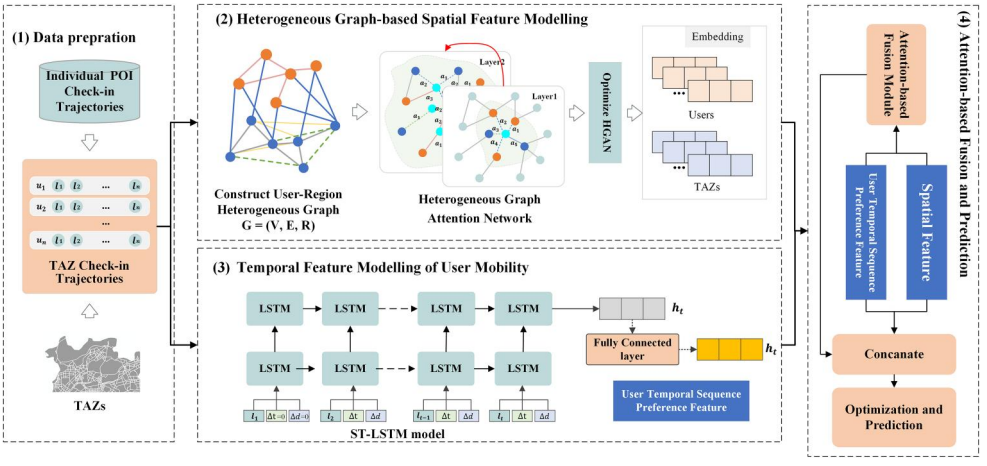

Figure 1. Workflow of the proposed methodology.

a series of POI points with geographical coordinates (longitudes and latitudes), we partition the city area into a finite number of TAZs using traffic roads and map the original POI check-in sequences to corresponding TAZ check-in sequences as the input for the model. Next, we construct a user–region interaction HG that integrates five relationships using users' historical TAZ check-in sequences and then updates node features with a HG attention network. Third, we adopt an LSTM-based sequential model that considers time and distance intervals to model temporal features of users' TAZ check-in sequences. Finally, we fuse the spatial and temporal features learned in the previous steps using an attention layer and then concatenate all the features to obtain the prediction result.

3.1. Data preprocessing and definitions

We use the POI check-in sequence of users recorded on social media as the input for the model. By utilizing transportation networks to partition urban areas into a finite number of TAZs, we map each POI at which a user has checked in to a TAZ. The problem of predicting the next POI is then transformed into the prediction of meaningful TAZs containing the next POI. Because of the typically uneven spatial distribution of social media points, with most POIs concentrated in the city center, using a regular grid division method would result in an imbalance in the number of POIs in each area. However, as the spatial distribution of traffic networks is denser in urban centers, TAZs (also referred to as 'regions' in this paper) defined using the transportation network are smaller in the city center, which helps balance the number of POIs. The final inputs to the model consist of a sequence of TAZs and their corresponding timestamps. Based on the aforementioned data preparation, this paper introduces the following basic definitions:

Definition 1. (Check-in Sequence): Let $\mathcal{U} = \{u_1, u_2, ..., u_M\}$ be a set of users and $\mathcal{L} = \{l_1, l_2, ..., l_N\}$ be a set of regions. Then, the check-in sequence of a user $u \in \mathcal{U}$ can be represented by $H^u = \{(t_i, l_j) | i = 1, 2, ..., n; j = 1, 2, ..., N\}$, where M and N denote the

total numbers of users and regions, respectively, (t_i, l_j) indicates that u visited l_j at timestamp t_i, and n denotes the total number of check-ins.

Definition 2. (Problem Formulation): Given $\mathcal{U}$, $\mathcal{L}$, and the set of check-in sequences $H = \{H^{u_1}, H^{u_2}, ..., H^{u_M}\}$, the target of the prediction task is to estimate the next TAZ $l^* \in \mathcal{L}$ of each user at a given future moment t^*.

3.2. HG-based spatial feature modeling

3.2.1. Constructing a user–region HG

To fuse physical and social information, we establish a user–region HG using check-in trajectories. This HG is defined as a graph $G = (V, \mathcal{E}, \mathcal{R})$, where V denotes the set of nodes comprising both users and regions, $\mathcal{E}$ denotes the set of edges or relations, and $\mathcal{R}$ represents the set of relation types between all nodes. The graph incorporates the following five relation types, and the fixed threshold is explained in Section 4.2.

1. **Region–Region Distance Relation:** This relation is based on the first law of geography, which suggests that nearby things are more closely related than distant things. Thus, we assume that if the distance (denoted by d) between the centers of mass of any two regions l_i and l_j is less than the distance threshold $d_{\max}$, these two regions are considered to be adjacent (Figure 2a).
2. **Region–Region Transition Relation:** The regional transition relation assumes that there is a close connection between pairs of regions that are always continuously visited within a certain time frame. In particular, if regions l_i and l_j are continuously visited by users within three days (those visited over longer time intervals

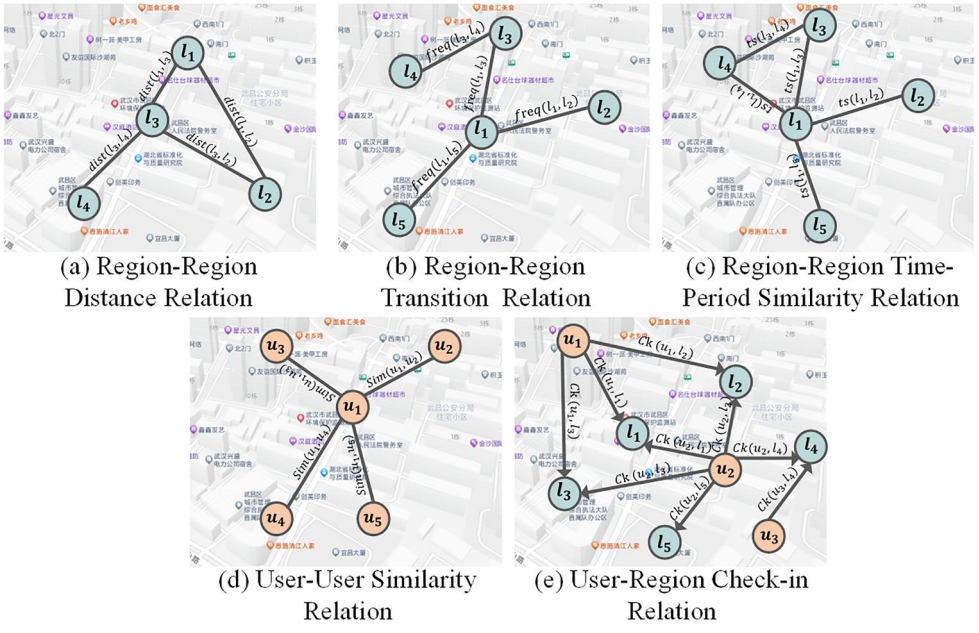

Figure 2. Illustration of five types of relationships in an HG.

are considered to have lower correlations), these two regions are considered to be adjacent (Figure 2b).

3. **Region–Region Time-period Similarity Relation:** We propose that two locations may have similar properties if they are consistently visited during the same time periods. To capture this, we divide the day into 24 time slots and the week into weekdays and non-working days, resulting in 48 time slots in aggregate. We then compute the Pearson similarity coefficient (Cohen *et al.* 2009) *ts* using the logarithm of the number of times the regions are visited during the same time slot. If the similarity between any two regions l_i and l_j, $ts > ts_{\min}$ (where $ts_{\min}$ is the similarity threshold), an undirected edge is established between them (Figure 2c). A region's population may affect the number of visits to nearby regions.
4. **User–User Similarity Relation:** Human mobility may be related to social relationships because humans are inherently social beings. We introduce this similarity relation by calculating the cosine similarity (Lim *et al.* 2020) between users based on the number of times they have visited the same location. For any two users u_i and u_j, if their similarity $s > s_{\min}$ (where $s_{\min}$ is the similarity threshold), an undirected edge is established between them (Figure 2d).
5. **User–Region Check-in Relation:** This relation stipulates that if user u_i has visited region l_j at some time, a directed edge $u_i \rightarrow l_j$ is established between them (Figure 2e).

3.2.2. HG attention network

The aforementioned HG contains two types of nodes—users and regions—connected via five different types of edges. To obtain the features of users and regions, we must embed the features of the HG from a high-dimensional space into a low-dimensional space. Several graph models, such as graph convolutional networks (GCNs) (Kipf and Welling 2016), GAT networks (Veličković *et al.* 2017), and relational data with GCN (RGCNs) (Schlichtkrull *et al.* 2018), have been proposed for this purpose. GAT focuses on assigning different attention coefficients to neighboring nodes during the propagation of node features, thereby emphasizing more important information. Simple-HGNN, proposed by Lv *et al.* (2021), extends GAT by incorporating edge-type information into the attention calculation, enabling more efficient processing of information in HG, and is applied to process the HG in this paper. This approach considers multiple relationships simultaneously and accounts for their different impacts.

Simple-HGNN assumes that each node in the graph tends to have features similar to those of its neighboring nodes. Thus, the features of the central node can be updated using an objective function motivated by the features of the neighboring nodes. Because not all neighboring node features are equally important, Simple-HGNN adds independent representation vectors for each type of edge and calculates the weights of different edge types. An attention mechanism a is added to calculate the attention coefficient e_{ij} for the neighboring node j with respect to the central node i :

$$e_{ij} = a\left(\mathrm{W}d_i, \mathrm{W}d_j, \mathrm{W}_r \mathrm{r}_{\varphi(<\mathrm{i,j}>)}\right) = \sigma(A^T[\mathrm{W}d_i || \mathrm{W}d_j || \mathrm{W}_r \mathrm{r}_{\varphi(<\mathrm{i,j}>)}]) \quad (1)$$

where d_i and d_j denote the node feature vectors; A^T, W, and W_r denote learnable weights; $\mathrm{r}_{\varphi(<\mathrm{i,j}>)}$ represents the representation vector for each edge type $\varphi_{(<\mathrm{i,j}>)}$ for considering different edge types; $||$ denotes the concatenation operation; and σ

represents the activation function LeakyReLU. The Softmax function is used to ensure that coefficients are easily comparable across different nodes:

$$\begin{aligned}\alpha_{ij} &= \mathrm{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k\in\mathcal{N}_i}\exp(e_{ik})} \\ &= \frac{\exp\left(\mathrm{LeakyReLU}\left(A^T\left[\mathrm{W}d_i||\mathrm{W}d_j||\mathrm{W}_r\mathrm{r}_{\varphi(<\mathrm{i,j}>)}\right]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\mathrm{LeakyReLU}\left(A^T\left[\mathrm{W}d_i||\mathrm{W}d_k||\mathrm{W}_r\mathrm{r}_{\varphi(<\mathrm{i,k}>)}\right]\right)\right)}\end{aligned} \tag{2}$$

Subsequently, the feature of each node i can be updated using those of all of its neighboring nodes:

$$d_i = \sigma\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}Wd_j\right), \tag{3}$$

where σ denotes the activation function. To enhance the expressiveness of the model, we adopt a two-layer network and use multiheaded attention, where K independent attention mechanisms are used to aggregate neighboring nodes. The resulting K features are subsequently concatenated and passed through a nonlinear activation function to obtain the final representation:

$$d_i^{(l)} = \sigma\left(\|_{k=1}^{K} d_{ik}^{(l)} + W_{k0}^{(l)}d_i^{(l-1)}\right), \tag{4}$$

where l denotes the network layer, $W_{k0}^{(l)}d_i^{(l-1)}$ represents the residual connection part used to address the limitation of GNN with respect to depth, W_{k0} denotes the weight matrix of the corresponding linear transformation, and $\|$ denotes the concatenation operation.

Unsupervised loss is employed to optimize the model, encouraging nodes with shared edges in the graph to have similar embeddings:

$$L_1 = -\log\sigma\left(d_i^T d_j\right) - \sum_{\mathcal{Z}\in NEG(i)}\log\sigma\left(-d_i^T d\mathcal{Z}\right), \tag{5}$$

where $d_i^T d_j$ and $d_i^T d_{\mathcal{Z}}$ represent the similarity between the two nodes measured using the dot product and $NEG(i)$ represents the negative-sampling neighbor node set of node i. To control memory usage and computational overhead, a neighbor-sampling strategy is used to obtain a fixed number of neighbors corresponding to each node.

3.3. Temporal feature modeling for user mobility

To capture the temporal characteristics of user check-in behaviors, we focus on leveraging the advantages of RNNs to capture the temporal characteristics of sequences. The ST-LSTM (Kong and Wu 2018) method, an LSTM-based approach that addresses the issue of gradient vanishing and captures long-term preferences, is employed to model temporal dependencies. We introduce an embedding layer into the model to encode all regions in the set $\mathcal{L}$. Subsequently, the check-in sequence for each user can be represented by $(r_{t_1}, r_{t_1}, \ldots, r_{t_n})$, where r_{t_i} is a d-dimensional embedding vector. The embedding layer vectors are initialized randomly and their weights are updated during network training.

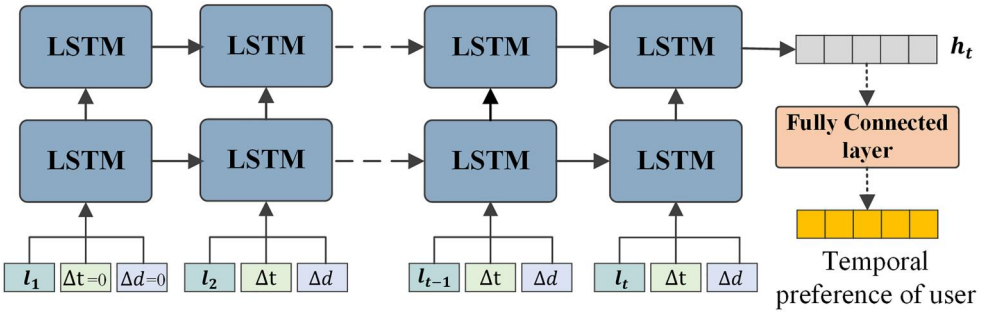

Figure 3. The structure of the ST-LSTM model.

We adopt ST-LSTM to capture temporal dependencies and dynamic changes within the sequence because local temporal and spatial context information comprises an important factor affecting the next step. ST-LSTM includes two components for simulating spatial and temporal distances between consecutive locations. Figure 3 illustrates the structure of the ST-LSTM model. Given an input r at a time stamp t, the output h_t of the hidden layer of the model can be calculated as follows:

$$\begin{cases} \hat{c}_t = \tanh(W_{rc} r_t + W_{hc} h_{t-1} + b_c) \\ i_t = \sigma(W_{ri} r_t + W_{hi} h_{t-1} + W_{qi} q_{t-1} + W_{si} s_{t-1} + b_i) \\ f_t = \sigma\left(W_{rf} r_t + W_{hf} h_{t-1} + W_{qf} q_{t-1} + W_{sf} s_{t-1} + b_f\right) \\ o_t = \sigma(W_{ro} r_t + W_{ho} h_{t-1} + W_{qo} q_{t-1} + W_{so} s_{t-1} + b_o) \\ c_t = f_t \cdot \mathrm{c}_{t-1} + \mathrm{i}_t \cdot \hat{\mathrm{c}}_t \\ h_t = \varphi(o_t \cdot \tanh(\mathrm{c}_t)) \end{cases} \tag{6}$$

where $i, f, o,$ and c_t denote the input gate, forggate, output gate, and cell state of the LSTM network, respectively; $\hat{c}_t$ denotes the new candidate state vector at step t; q is the temporal factor vector; and s denotes the spatial distance factor vector. q_{t-1} and s_{t-1} represent the representation vectors of time intervals Δt and spatial intervals Δd between r_t and r_{t-1}, respectively, where $q_0 = 0$ and $s_0 = 0$. W_{kz}, where $k = r, h, q, s$ and $z = i, f, o,$ denote linear transition matrices, with corresponding biases, $b_{x=c,i,f,o}$. σ denotes the sigmoid function, and φ denotes the fully connected layer. By partitioning the time and spatial intervals into discrete slots and only encoding their upper and lower bounds, q_{t-1} and s_{t-1} can be calculated as follows:

$$\begin{cases} q_{t-1} = \dfrac{Q_{u(\Delta t)}[u(\Delta t) - \Delta t] + Q_{l(\Delta t)}[\Delta t - l(\Delta t)]}{u(\Delta t) - l(\Delta t)} \\ s_{t-1} = \dfrac{S_{u(\Delta d)}[u(\Delta d) - \Delta d] + S_{l(\Delta d)}[\Delta d - l(\Delta d)]}{u(\Delta d) - l(\Delta d)} \end{cases} \tag{7}$$

where $u(\Delta t)$ and $l(\Delta t)$ denote the upper and lower bounds of the time interval Δt, respectively, and $u(\Delta d)$ and $l(\Delta d)$ denote the upper and lower bound values of the spatial distance interval Δd, respectively. Further, $Q \in R^{N_q * d}$ denotes the temporal factor matrix, $S \in R^{N_s * d}$ denotes the spatial factor matrix, N_q and N_q denote the total number of time and spatial slots, and d denotes the vector dimension. Finally, we use the final layer output h_t as the temporal sequence preference of each user at time t.

3.4 Attention-based fusion module and prediction layer

3.4.1 Attention-based Fusion Module

To better integrate the features learned by the two aforementioned models, we employ the spatio-temporal attention aggregation method proposed by Wang *et al.* (2022b). In particular, we capture the dynamic user-personalized preferences d_A for each visited region as follows:

$$d_A = \sum_{i=1}^{t} \alpha_i (W_1 d_i), \alpha_i = \sum_{i=1}^{t} softmax\left(\frac{(W_2 h_t)(W_3 d_i)^T}{\sqrt{k}}\right)(W_1 d_i) \tag{8}$$

where d_i represents the region embeddings learned from the HG, α_i denotes the weight coefficient, h_t denotes the output of the LSTM-based model, $W_1, W_2,$ and W_3 denote trainable weight matrices, k denotes the embedding dimension, and *Softmax* is an activation function.

3.4.2 Prediction and optimization

After incorporating the aforementioned modules, we obtain embeddings d_i of users and regions that represent global physical and social space information, as well as temporal sequential features h_t^u and dynamic user-personalized preferences ${d_A}^u$. Then, the probability $\hat{y}_{u,j}^{t+1}$ of user u visiting region l_j at a given time $t+1$ can be calculated by integrating the influence of these features on destination l_j as follows:

$$\hat{y}_{u,j}^{t+1} = \sigma(\varphi(l_{HG} || r_j^T h_t^u || d_j^T {d_A}^u)) \tag{9}$$

where $l_{HG} = d_u^T d_j + \frac{1}{|H^u|}\sum_{l_k \in L^u} \left({d_k}^T d_j\right)$, H^u denotes the check-in sequence of the user, $|H^u|$ denotes the sequence length, r_j denotes the embedding vector of the targret region j, as defined in Section 3.3, φ denotes a fully connected nonlinear layer, and σ represents the sigmoid function.

The model is optimized using binary cross-entropy (BCE) loss based on negative sampling, which can be formulated as follows:

$$L_2 = \sum_{u \in \mathcal{U}} \sum_{k \in \{j\} \cup NEG\{j\}} (y_{u,k}^{t+1} \log\left(\hat{y}_{u,k}^{t+1}\right) + (1 - y_{u,k}^{t+1}) \log(1 - \hat{y}_{u,k}^{t+1})) \tag{10}$$

where $y_{u,k}^{t+1}$ denotes the probability of the true region being visited, which is equal to 1 if $k = j$ and 0 otherwise; and $NEG\{j\}$ denotes the negative sampling obtained by randomly replacing the real regions. Finally, the final loss function is taken to be $L = L_1 + L_2 + \lambda ||\theta||_2$, where θ represents the parameters used for prediction and $\lambda ||\theta||_2$ represents the L2 regularization used to prevent overfitting.

The training process of the PS-HGAN model is presented in Algorithm 1. To mitigate information loss, we employ a single-stage approach to train both the feature modeling module and the final prediction component simultaneously. During each training iteration, the node features of the graph are updated by utilizing node pairs from the HG as input, and then the corresponding node embeddings are integrated into the prediction part. The optimization process utilizes the Adam optimizer to minimize both the L1 and L2 loss functions by adjusting all the trainable parameters.

Algorithm 1. Training process of PS-HGAN

Input: adjacent node pairs (set of edges) $\mathcal{E}$, heterogeneous graph $G = (V, \mathcal{E}, \mathcal{R})$, check-in sequences $H = \{H^{u_1}, H^{u_2}, ..., H^{u_M}\}$.

Output: learned PS-HGAN model

1: Initialize parameters θ
2: **while** *epoch* $\leq$ *epoch_size* **do:**
3: **while** *batch* $<$ *train_size*//*batch_size*+1 **do:**
4: Sample from heterogeneous graph G
5: Optimize loss function L_1 and update node embedding vectors
6: **end**
7: **while** *batch* $<$ *train_size*//*batch_size*+1 **do:**
8: calculate the probability $\hat{y}_{u,j}^{t+1}$ base on Eq. (9)
9: Optimize loss function L_2 and update θ
10: **end**
11: **end**

4. Results and discussion

4.1. Data preparation

In our experiment, real data from three popular LBSN platforms, Sina Weibo, Gowalla, and Foursquare, are used to evaluate the performance of the proposed model. The spatial coverages of the three datasets correspond to the administrative districts of three large cities with rich and diverse human activities: Shenzhen in China, New York City in the United States, and Tokyo in Japan.

These datasets contain the check-in records of users, including user ID, timestamp, POI ID, and exact latitude and longitude of each POI. To ensure the reliability of the experiments, the original POI check-in data are preprocessed to eliminate invalid and duplicate data. The datasets are filtered and only users with more than ten check-in records in each of the three datasets are retained. The processed check-in records of each user are sorted by timestamp—the first 70% are used as training data and the final 30% as testing data. The resulting datasets are presented in Table 2.

The spatial distributions and data statistics of the three datasets are depicted in Figure 4. For clarity, some POIs have been diluted. The figure indicates that POIs tend to form distinct spatial aggregation centers, with those closer to the centers exhibiting higher check-in counts (Figure 4a–c). The data distribution of the three social media datasets follows a long-tailed distribution (Figure 4d), with a small number of POIs

Table 2. Statistics of datasets.

Dataset	Records	Users	Regions	POIs	Time	Source
Shenzhen	161147	9427	1885	16696	January 2017 to December 2018	Sina Weibo[a]
New York	212231	6197	1224	20702	April 2009 to July 2011	Gowalla[b]
Tokyo	481492	7785	2104	61455	April 2012 to September 2013	Foursquare[c]

[a]https://weibo.com/.
[b]http://snap.stanford.edu/data/loc-gowalla.html.
[c]https://sites.google.com/site/yangdingqi/home/foursquare-dataset.

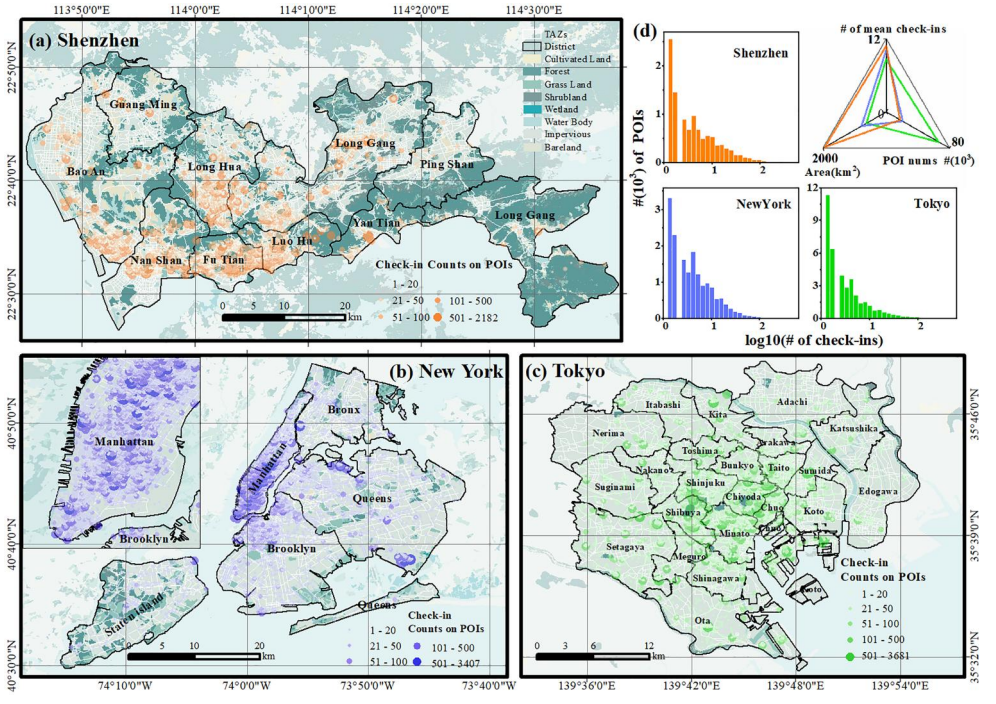

Figure 4. Spatial distributions of POI check-in data for the (a) Shenzhen, (b) New York, and (c) Tokyo datasets. (d) Histograms of the logarithmic check-in counts for the three datasets, where the *x* axis represents the logarithmic check-in numbers of POI base 10, denoted log10 (# of check-ins), and the *y* axis represents the number of POIs with the corresponding number of check-ins (similar to Figure 5). The check-in values obtained are heavy-tailed, where most POIs exhibit low check-in activity, and only a few POIs exhibit very high check-in counts. To visualize the data distribution more clearly, we compute the logarithm of the check-in activity relative to base 10. (For simplicity, we henceforth omit the map north arrows and scale bars.).

exhibiting high check-in counts, and the majority of POIs corresponding to relatively few check-ins. The Shenzhen datasdemonstrates the largest spatial extent and sparsest distribution of POIs, whereas the Tokyo datasshares similar spatial coverage but exhibits a higher concentration and denser distribution of POIs. In comparison, the New York datasexhibits a relatively small spatial extent and fewer POIs.

Figure 5 illustrates the temporal distribution of the three datasets, revealing significant temporal patterns that underscore the crucial role of integrating temporal patterns into our models. Further, the average time sparsities (which measures the proportion of days without check-ins compared to the total number of days between the first and last check-ins) for the Shenzhen, New York, and Tokyo datasets are 97.7%, 98.1%, and 92.4%, respectively, indicating a high level of sparsity within the datasets. The average check-in sequence lengths for users in the Shenzhen, New York, and Tokyo datasets are 21, 44, and 83, respectively.

We utilize road network data from the Open StreMap website along with administrative division data to partition Shenzhen, New York, and Tokyo into 1885, 1224, and 2104 TAZ units, respectively. The spatial distribution of check-in counts in the TAZs is

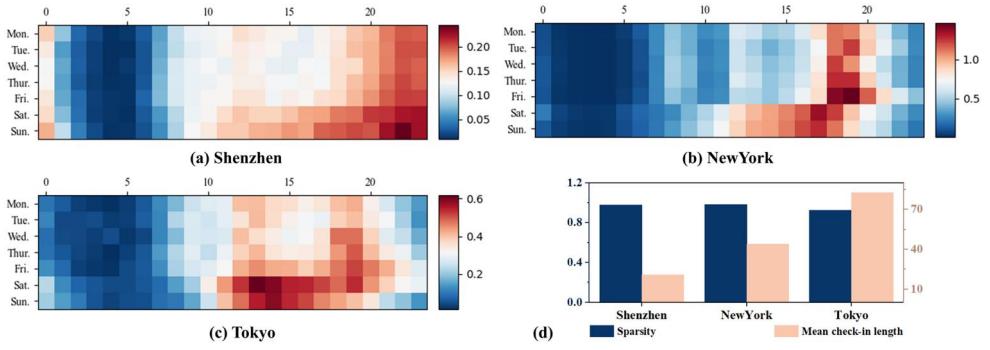

Figure 5. TAZ check-in data spatial distribution of (a) Shenzhen, (b) New York, and (c) Tokyo datasets. (d) Histograms of the logarithmic check-in counts of three datasets.

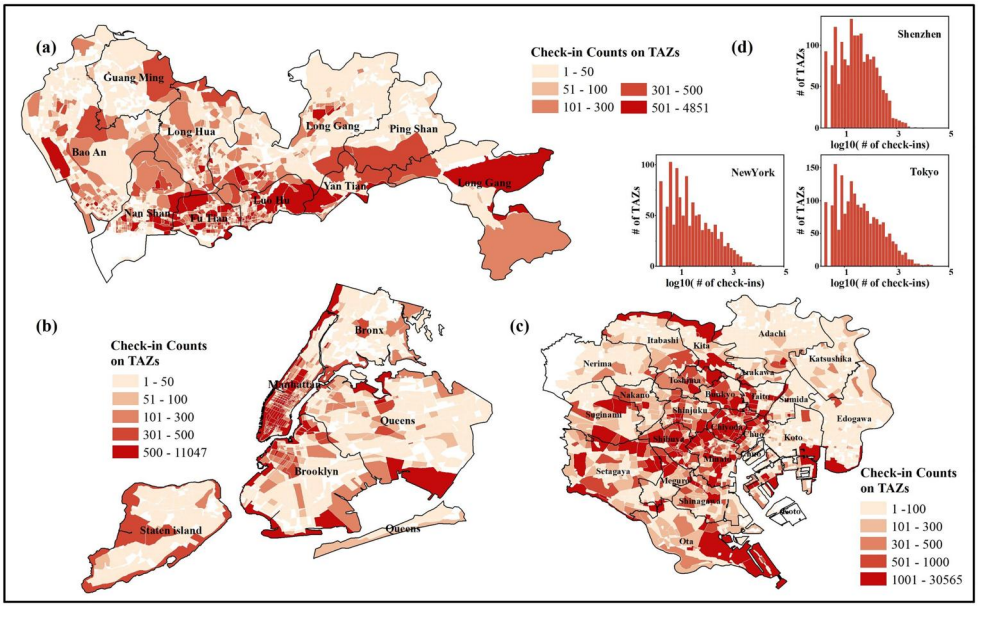

Figure 6. Data time distribution of (a) Shenzhen, (b) New York, and (c) Tokyo datasets; (d) data sparsity and average user check-in sequence lengths for the three datasets.

depicted in Figure 6. As the distance from the aggregation center with a high number of check-ins decreases, the density of the TAZ increases, and its area decreases.

4.2. Evaluation metrics and parameter setting

We utilize two of the most widely used evaluation metrics, *Accuracy@k* and normalized discounted cumulative gain (*NDCG@k*), to evaluatenext location prediction performance. *Accuracy@k* is 1 if the targregion appears in the top-*k* ranking lists and 0 otherwise. *NDCG@k* measures the order of the ranking lists. In this study, we s $k = \{1, 5, 10\}$, and these evaluation metrics are calculated as the average values over all test instances. The functions for *NDCG@k* are defined as follows:

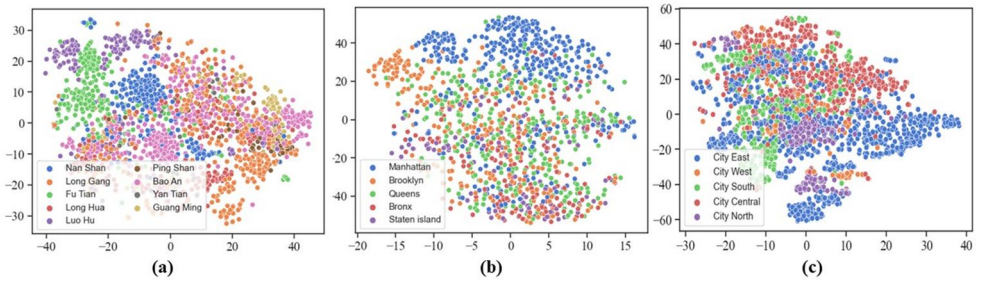

Figure 7. The dimensionality reduction of graph node embedding of TAZs in (a) Shenzhen, (b) New York, and (c) Tokyo datasets. In the Tokyo dataset, the 23 wards are mapped into five commonly used wards—City East (Edogawa, Katsushika, Koto, Sumida, Taito, Arakawa, Adachi), City West (Nakano, Suginami), City South (Setagaya, Meguro. Shinagawa, Ota), City North (Nerima, Itabashi, Kita, Toshima) and City Central (Shibuya, Minato, Chuo, Chiyoda, Shinjukn, Bunkyo).

$$NDCG@k = \sum_{i=1}^{n} \frac{1}{\log_2(Rank_i + 1)}, Rank_i \leq k \tag{11}$$

where $Rank_i$ denotes the position of the targlocation in the ranking list.

As our experiments involve up to 9427 users and 2104 TAZs, the dimension, d, of the node embedding is set to 128. To simplify this process, we initialize a feature vector for each node. We implement ST-LSTM based on 2-layer LSTM networks. The embedding dimension is also set to 128, and the numbers of time and spatial slots are set to 10. When extracting spatial information using the HG attention network, we use five sample neighbors, two network layers, and three attention heads. After evaluating different learning rates $\{0.01, 0.001, 0.0001, \text{and } 0.00001\}$ and l2 regularization coefficients λ $\{0.001, 0.0001, 0.00001, \text{and } 0.000001\}$, we select 0.0001, 0.00001, and 0.00001 as the final learning rate parameters and 0.000001 as the final λ parameter for the New York, Shenzhen and Tokyo datasets, respectively. After evaluating different $d_{\max}$ values in $\{1, 2, 3, 4, 5\}$ and $s_{\min}$ and $ts_{\min}$ in $\{0.5, 0.6, 0.7, 0.8\}$, we select 0.2, 0.7 and 0.7 as the final parameters.

4.3. Analysis of graph-based spatial features

To demonstrate the spatial features learned by our model, we utilize t-SNE to map the high-dimensional vectors of the region node embedding from our model-trained HG attention network onto a two-dimensional plane, while maintaining their connections. Figure 7 depicts the results, with different colors representing the administrative regions of the respective cities. The results reveal that TAZ vectors within the same administrative area, such as Nanshan and Futian in Shenzhen, Manhattan in New York, and City East in Tokyo, are typically clustered closely. These results indicate that these TAZs are more similar from the perspective of human activities than other pairs. Additionally, there are instances where TAZs from different administrative regions appear similar to each other in the vector space, suggesting a strong connection between human activities in these regions despite their spatial separation. The proposed model effectively captures these spatial features.

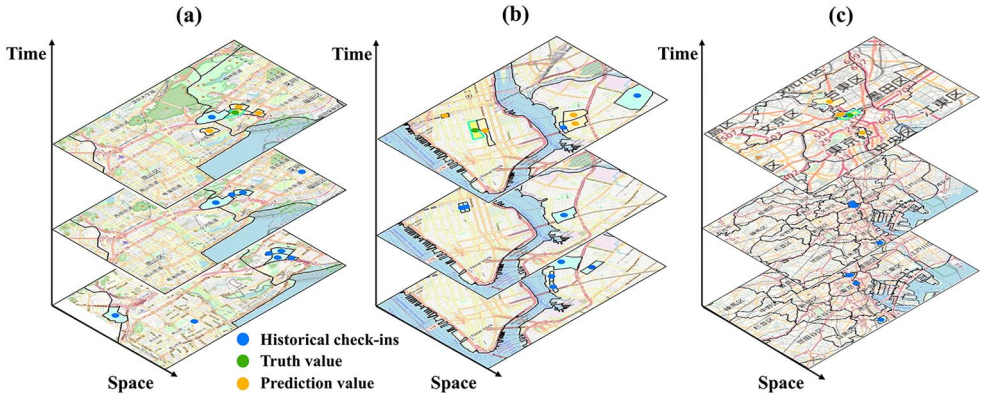

Figure 8. Historical check-ins, next truth check-in values and top-5 model prediction results for users in the (a) Shenzhen, (b) New York, and (c) Tokyo datasets.

4.4. Prediction results

To evaluate the prediction performance of the model, we randomly select three users from the test sets of the three datasets and visualize their historical check-in records and the top five prediction results—these are depicted in Figure 8. The different map layers represent consecutive repetitive check-ins in a user's history trajectory, and the check-in records before and after the repetitive check-ins are located on different maps. The results demonstrate that the proposed model can determine the exact areas that users will visit in the future. The upper map depicts the last location visited by the user and the top-5 predicted results. Although the top-1 prediction results do not always correspond to real user visits, the top-5 prediction results are always distributed around the true value. The top-5 nontrue prediction results also approximate the true values, which establishes the effectiveness of the model in incorporating spatial relationships.

4.5. Comparison with baselines

To evaluate the effectiveness of the proposed PS-HGAN model, we compare its next location prediction performance withthose of seven state-of-the-art baselines: 1. RNN-based methods, these include LSTM (Hochreiter and Schmidhuber 1997), ST-RNN (Liu *et al.* 2016) and ST-LSTM (Kong and Wu 2018), where ST-RNN and ST-LSTM incorporate spatiotemporal information sequentially; 2. CTLE (Lin *et al.* 2021), which considers context-specific information in sequences based on the transformer model; 3. BiLSTM-CNN (Bao *et al.* 2021), which constructs a graph of temporal distance relationships for spatially clustered regions to acquire global information; 4. ASGNN (Wang *et al.* 2021b), which models user check-in sequences as sequential graphs and combines graph neural networks and personalized hierarchical attention networks to capture users' long- and short-term preferences; and 5. STGCN (Han *et al.* 2020), which constructs multirelational graphs for users, POIs, and regions where the POIs are located. STGCN performs location prediction by utilizing different relations and scoring functions to model user-region and user-POI periodic patterns. It relies solely on a graph model

Table 3. Performance comparison of different approaches in terms of Accuracy@K (Acc@K) and NDCG@K.

Shenzhen									
Model	LSTM	ST-RNN	ST-LSTM	CTLE	BiLSTM-CNN	ASGNN	STGCN-LSTM	PS-HGAN	Improve
Acc@1	0.221	0.2243	0.2298	0.3378	0.396	**0.4256**	0.0235	0.4227	−0.68%
Acc@5	0.3329	0.332	0.3434	0.4338	0.4749	0.5111	0.1093	**0.5359**	4.85%
Acc@10	0.3876	0.3987	0.4012	0.4755	0.5068	0.5297	0.1708	**0.5761**	8.76%
NDCG@1	0.221	0.2243	0.2298	0.3378	0.396	**0.4256**	0.0235	0.4227	−0.68%
NDCG@5	0.2792	0.2801	0.2895	0.3899	0.4386	0.4734	0.0659	**0.4857**	2.60%
NDCG@10	0.2967	0.2988	0.3081	0.4033	0.4488	0.4795	0.0857	**0.4986**	3.98%
New York									
Model	LSTM	ST-RNN	ST-LSTM	CTLE	BiLSTM-CNN	ASGNN	STGCN-LSTM	PS-HGAN	Improve
Acc@1	0.0356	0.0382	0.0382	0.0382	0.0384	0.0384	0.0384	**0.0401**	4.43%
Acc@5	0.2575	0.2485	0.2422	0.2535	0.2828	0.212	0.2385	**0.2926**	3.47%
Acc@10	0.409	0.41	0.4135	0.4265	0.425	0.4147	0.4269	**0.4322**	1.24%
NDCG@1	0.0356	0.0382	0.0382	0.0382	0.0384	0.0384	0.0384	**0.0401**	4.43%
NDCG@5	0.1342	0.1267	0.1273	0.1347	0.144	0.1511	0.1273	**0.1517**	0.40%
NDCG@10	0.1819	0.1803	0.1808	0.1889	0.1921	0.2159	0.1857	**0.2172**	0.60%
Tokyo									
Model	LSTM	ST-RNN	ST-LSTM	CTLE	BiLSTM-CNN	ASGNN	STGCN-LSTM	PS-HGAN	Improve
Acc@1	0.0751	0.075	0.0759	0.0653	0.0748	0.1335	0.0502	**0.1474**	10.41%
Acc@5	0.2047	0.2103	0.2114	0.2007	0.2488	0.2565	0.1979	**0.3199**	24.72%
Acc@10	0.3217	0.3233	0.3221	0.3234	0.3305	0.306	0.3229	**0.4097**	23.96%
NDCG@1	0.0751	0.075	0.0759	0.0653	0.0748	0.1335	0.0502	**0.1474**	10.41%
NDCG@5	0.1402	0.1443	0.1451	0.1365	0.1637	0.1984	0.1246	**0.2369**	19.41%
NDCG@10	0.1781	0.1811	0.1809	0.1761	0.1901	0.2144	0.1644	**0.2660**	24.07%

Note. Bold scores are the best results for each metric, while the second-best scores are underlined. Improvement indicates the relative improvements of our model over the best baselines.

and focuses on the POI prediction problem rather than on sequence-based POI prediction. Therefore, we combine it with LSTM (forming STGCN-LSTM) for model performance comparison. In particular, the graph node representation obtained by the STGCN model is integrated as the input of the LSTM model. We utilize the grid search method in a neural network intelligence (NNI) automatic machine learning toolkit to perform optimal parameter search for these models. The maximum number of experiments for each pair of baseline methods and datasets is set to 50, and the maximum training epoch is set to 200. To prevent overfitting, we implement an early stopping strategy during the experiments—training is terminated when the validation loss value exhibits an increasing trend for five consecutive rounds.

Table 3 lists the performances of the proposed PS-HGAN model on the three datasets. Compared to existing models, PS-HGAN exhibits relatively good improvements in the two metrics on all three datasets. Among the existing models, the ASGNN model performs the best, followed by the BiLSTM-CNN model. These two models are better equipped to capture spatial information than the LSTM, ST-LSTM, and CTLE models. However, our model outperforms all other models, with each metric exhibiting improvement by more than 10% on the Tokyo dataset, with the largest improvement in the accuracy @5 metric of 24.72%. Clearly, the LSTM and ST-LSTM methods based solely on RNN models and the CTLE method based on the transformer are unable to capture the global relationships between locations as well as the proposed model, and consequently exhibit poorer prediction results. PS-HGAN achieves these significant

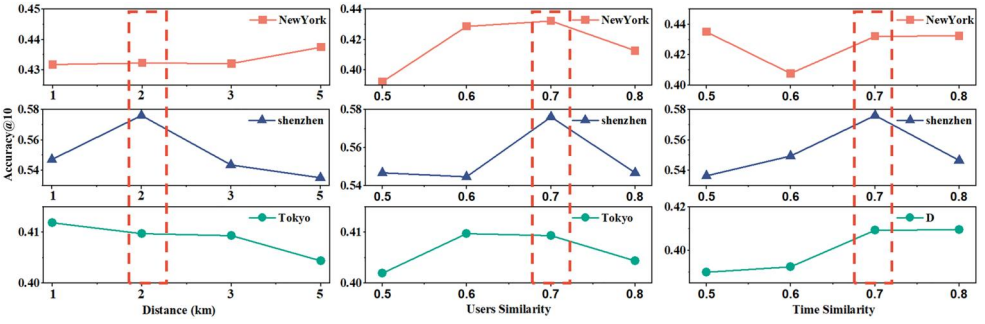

Figure 9. Effect of different hyperparameters on model performance.

improvements by utilizing a HG structure to synthesize complex relationships of physical and social spaces. This allows the model to observe similar users and locations from a global perspective, thereby achieving a collaborative filtering-like effect. Moreover, the advantages of LSTM networks are combined to capture sequential information. These improvements enable our model to achieve superior performance compared to the others.

4.6. Influence of hyperparameters

Model parameters are critical to the performance of the proposed model. To investigate their impact, we focus on several key hyperparameters, the HG: distance threshold d_{max}, user similarity threshold s_{min}, and the time period similarity threshold ts_{min}, while constructing the HG. We systematically vary d_{max} from 1 km to 5 km, and s_{min} and ts_{min} from 0.5 to 0.8. These ranges are selected because similarities below 0.5 might introduce excessive noise and increase the size of the HG.

Figure 9 illustrates the performance of the proposed model in terms of Accuracy@10 corresponding to different threshold parameters. The results indicate a consistent decrease in performance on all three datasets as the distance threshold increases. This observation aligns with the first law of geography—larger distance thresholds introduce more noise and hinder the model's ability to capture adjacency relationships between nodes effectively. In addition, this leads to an expanded HG, subsequently degrading model performance. After considering all factors, we select 2 km as the distance threshold.

Regarding the user similarity threshold, s_{min} and time period similarity threshold, ts_{min}, we observe an initial significant improvement in model performance as the thresholds are increased, followed by a gradual decline. A substantial number of edges connecting adjacent nodes are pruned, resulting in fewer similar nodes. Consequently, the model struggles to extract meaningful information from sparse graphs. Setting s_{min} and ts_{min} to 0.7 yields relatively optimal performance. Hence, we set the threshold to 0.7.

4.7. Effect of different components

Ablation experiments are performed to evaluate the contribution of each component to model performance. The results of the ablation experiments are presented in

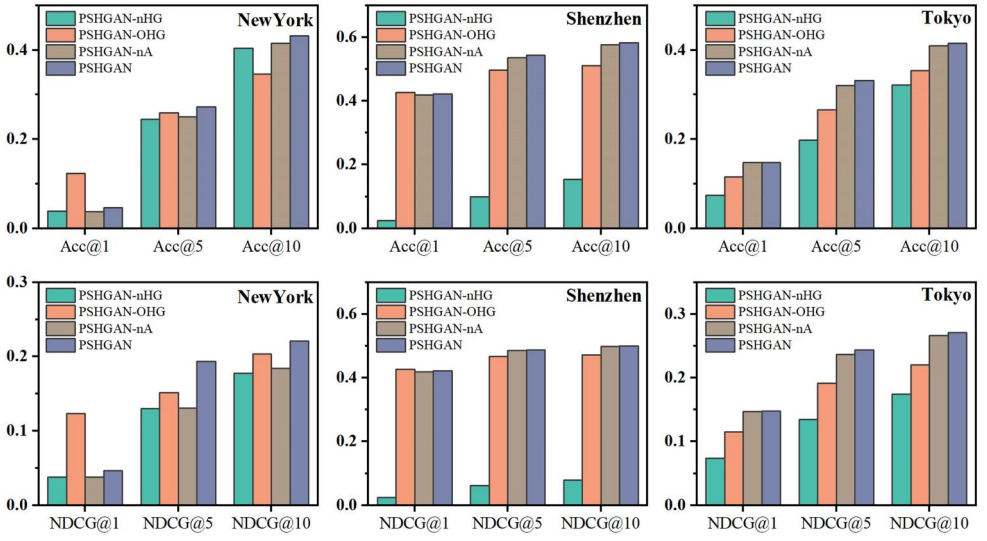

Figure 10. Effectiveness of using only time-series information (PSHGAN-nHG), only spatial information (PSHGAN-OHG), fusing both spatial and temporal information without using an attention network (PSHGAN-nA) and a complete model network (PSHGAN).

Figure 10, and the following observations are made. First, PSHGAN-nHG, which excludes the HG attention network and considers only temporal sequential features, exhibits the poorest performance in terms of most metrics across all datasets. This indicates that the complex spatial information captured by HGAT from the HG provides valuable insights into users' next locations, thereby enhancing the model's predictive capabilities. The largest performance gap between PSHGAN-nHG and PSHGAN is observed on the Shenzhen dataset, possibly because of its increased sparsity. The graph-based model captures information from sparse data more effectively. PSHGAN-OHG, which excludes the LSTM module responsible for capturing sequential information, achieves the second-worst performance on all datasets. This highlights the importance of considering the temporal dependencies in user check-ins. Third, PSHGAN consistently achieves the best performance across all datasets, whereas PSHGAN-nA, which excludes the attention fusion mechanism, achieves the second-best performance on the two datasets, with only a slight difference in accuracy compared to PSHGAN. This suggests that combining complex spatial information with temporal sequential information using graphs and recurrent neural networks yields impressive results. The performance of PSHGAN on the New York dataset further demonstrates that attention mechanisms can enhance the integration of these two types of information, leading to improved model performance in certain scenarios.

5. Conclusions

Predicting the next location of humans is a crucial issue in human mobility research, with significant value in practical applications, such as location services. Previous studies have revealed the vital role of capturing spatiotemporal features from sparse data

concerning user mobility and various influencing factors from both the physical and social domains to understand human mobility patterns and improve predictive performance. However, in these studies, social influences were often disregarded, and implicit information on the interactions between users and geographical locations needs to be explored further. Moreover, the limitations of a single modeling structure restrict the effective integration of complex spatiotemporal information and multiple influencing factors. This constrains the model's ability to extract implicit information from user trajectories, thereby affecting its predictive performance.

To address this issue, we propose a novel method with physical and social awareness based on a heterogeneous graph, PS-HGAN. The primary focus of this model is to establish a complex interaction network between users and locations through a HG network, thereby capturing the spatial dependencies of various high-order geographical influences, social influences, and user preferences. In addition, we use attention mechanisms to fuse the spatial features obtained from graph networks with the temporal features obtained from sequence models.

Extensive experiments conducted on three datasets demonstrate the effectiveness of the proposed approach. Compared to experiments incorporating interaction networks, ablation experiments considering only temporal sequence features perform worse in terms of most metrics, highlighting the importance of considering complex interactions between users and locations and various influences from physical and social spaces. This may be because sequence-based methods typically address data sparsity by integrating additional information such as time and space features (Yang *et al.* 2023, Yao *et al.* 2023), whereas the modeling approach of HGs can represent various relationships between users and locations, resulting in denser representations of the available data (Wang *et al.* 2021a, Dai *et al.* 2022). Further, the indirect relationships in the graph enable the model to infer missing or unobserved data, thereby alleviating data sparsity and enhancing robustness. By considering these complex relationships, we provide a model with more information regarding the factors that influence mobility most significantly.

Experiments with no attention mechanism and direct integration of temporal and spatial features yielded the second-best performance, indicating the significance of integrating these spatiotemporal features with complex influencing factors. Compared with existing models (Kong and Wu 2018, Sun *et al.* 2020, Wang *et al.* 2021b), our approach can integrate multidimensional information related to user mobility from a global perspective, including temporal, spatial, social, geographical, and complex implicit information from physical and social interactions. The addition of attention mechanisms allowed the model to focus on the most important features, thereby providing additional benefits for predictions. Overall, our model achieved a significant improvement of up to 24.72% in terms of the Accuracy@10 metric compared with the best available model, demonstrating its ability to better understand user mobility patterns and enhance predictive efficiency.

In future works, we intend to investigate novel fusion methods that can effectively combine spatial and temporal features and explore other types of contextual information, such as POI categories, land-use functions, and user profile information and other contextual data, to enhance the performance of our location prediction model.

Acknowledgements

We would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions.

Disclosure statement

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

Funding

This work was supported by grants from the National Key Research and Development Program of China (No. 2021YFF0704400), the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (No. CUG230601), the National Nature Science Foundation of China Program (No. 42201438), and the Special Fund of Hubei Luojia Laboratory (No. 220100034).

Notes on contributors

Sijia He is currently a PhD student in the National Engineering Research Center of Geographic Information System at China University of Geosciences (Wuhan). Her research interests focus on flood monitoring methods and knowledge graph construction. She contributed to the conceptualization of the research idea, data analysis, methodology, code development and original draft preparation.

Wenying Du is an Associate Professor in the National Engineering Research Center of Geographic Information System at China University of Geosciences (Wuhan). Her research interests focus on flood monitoring methods and knowledge graph construction. She contributed to review, editing and formal analysis.

Yan Zhang is a Postdoctoral Researcher of Institute of Space and Earth Information Science at the Chinese University of Hong Kong. His research focus on collaborative sensing and geographic knowledge services. He contributed to the conceptualization of the research idea and technical support.

Lai Chen is currently a PhD student in the National Engineering Research Center of Geographic Information System at China University of Geosciences (Wuhan). His research focus on spatiotemporal process mining and prediction. He contributed to data collection and collation.

Zeqiang Chen is a Professor in the National Engineering Research Center of Geographic Information System at China University of Geosciences (Wuhan). His research interests include sensor Web, spatiotemporal big data intelligence and its application in smart watersheds. He contributed to review and editing.

Nengcheng Chen is a Professor in the National Engineering Research Center of Geographic Information System at China University of Geosciences (Wuhan). His research interests focus on Earth observation sensor Web, spatiotemporal big data, Web GIS and smart city. He contributed to the conceptualization of the research idea, supervision and project administration.

Data and codes availability statement

The data and codes that support the findings of this study are available on 'figshare.com' with the identifier https://doi.org/10.6084/m9.figshare.23285594

References

Afyouni, I., Al Aghbari, Z., and Razack, R.A., 2022. Multi-feature, multi-modal, and multi-source social event detection: a comprehensive survey. *Information Fusion*, 79, 279–308.

Agrawal, S., Roy, D., and Mitra, M., 2021. Tag embedding based personalized point of interest recommendation system. *Information Processing and Management*, 58 (6), 102690.

Alessandretti, L., Aslak, U., and Lehmann, S., 2020. The scales of human mobility. *Nature*, 587 (7834), 402–407.

Bao, Y., *et al.*, 2021. A BiLSTM-CNN model for predicting users' next locations based on geo-tagged social media. *International Journal of Geographical Information Science*, 35 (4), 639–660.

Barbosa, H., *et al.*, 2018. Human mobility: models and applications. *Physics Reports*, 734, 1–74.

Carrasco, J.A., and Miller, E.J., 2006. Exploring the propensity to perform social activities: a social network approach. *Transportation*, 33 (5), 463–480.

Cohen, I., *et al.*, 2009. Pearson correlation coefficient. *Noise Reduction in Speech Processing*, (5), 1–4.

Dai, S., *et al.*, 2022. Spatio-temporal representation learning with social tie for personalized POI recommendation. *Data Science and Engineering*, 7 (1), 44–56.

Du, Y., *et al.*, 2018. A geographical location prediction method based on continuous time series Markov model. *PloS One*, 13 (11), e0207063.

Gambs, S., Killijian, M.-O., and del Prado Cortez, M.N., 2012. Next place prediction using mobility Markov chains. *In: Proceedings of the first workshop on measurement, privacy, and mobility*, 1–6.

González, M.C., Hidalgo, C.A., and Barabási, A.-L., 2008. Understanding individual human mobility patterns. *Nature*, 453 (7196), 779–782.

Guo, L., *et al.*, 2010. Uncertain path prediction of moving objects on road networks. *Journal of Computer Research Development*, 47 (1), 104–112.

Han, H., *et al.*, 2020. STGCN: a spatial–temporal aware graph learning method for POI recommendation. In: *2020 IEEE International Conference on Data Mining (ICDM)*, 1052–1057.

Hawelka, B., *et al.*, 2014. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41 (3), 260–271.

Hochreiter, S., and Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, 9 (8), 1735–1780.

Huang, F., *et al.*, 2019a. STPR: a personalized next point-of-interest recommendation model with spatio-temporal effects based on purpose ranking. *IEEE Transactions on Emerging Topics in Computing*, (99), 1–1.

Huang, L., *et al.*, 2019b. An attention-based spatiotemporal lSTM network for next poi recommendation. *IEEE Transactions on Services Computing*, 14 (6), 1585–1597.

Ilin, C., *et al.*, 2021. Public mobility data enables COVID-19 forecasting and management at local and global scales. *Scientific Reports*, 11 (1), 13531.

Jia, J.S., *et al.*, 2020. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, 582 (7812), 389–394.

Jurdak, R., *et al.*, 2015. Understanding human mobility from Twitter. *PloS One*, 10 (7), e0131469.

Kim, J., *et al.*, 2021. DynaPosGNN: dynamic-positional GNN for next POI recommendation. In: *2021 International Conference on Data Mining Workshops (ICDMW)*, 36–44.

Kipf, T.N., and Welling, M., 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:.02907*.

Kong, D., and Wu, F., 2018. HST-LSTM: a hierarchical spatial–temporal long-short term memory network for location prediction. *In: IJCAI*, 2341–2347.

Koren, Y., 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining*, 426–434.

Kryvasheyeu, Y., *et al.*, 2016. Rapid assessment of disaster damage using social media activity. *Science Advances*, 2 (3), e1500779.

Leung, K., Wu, J.T., and Leung, G.M., 2021. Real-time tracking and prediction of COVID-19 infection using digital proxies of population mobility and mixing. *Nature Communications*, 12 (1), 1501.

Li, F., *et al.*, 2020. A hierarchical temporal attention-based LSTM encoder–decoder model for individual mobility prediction. *Neurocomputing*, 403, 153–166.

Li, G., *et al.*, 2022. Potential destination discovery for low predictability individuals based on knowledge graph. *Transportation Research Part C: Emerging Technologies*, 145, 103928.

Lian, D., *et al.*, 2020. Geography-aware sequential location recommendation. In: *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery and data mining*, 2009–2019.

Lim, N., *et al.*, 2020. STP-UDGAT: spatial–temporal-preference user dimensional graph attention network for next POI recommendation. In: *Proceedings of the 29th ACM international conference on information and knowledge management*, 845–854.

Lin, Y., *et al.*, 2021. Pre-training context and time aware location embeddings from spatial–temporal trajectories for user next location prediction. *In: Proceedings of the AAAI conference on artificial intelligence*, 4241–4248.

Liu, Q., *et al.*, 2016. Predicting the next location: a recurrent model with spatial and temporal contexts. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30, 1.

Luca, M., *et al.*, 2021. A survey on deep learning for human mobility. *ACM Computing Surveys*, 55 (1), 1–44.

Luo, Y., Liu, Q., and Liu, Z., 2021. STAN: spatio-temporal attention network for next location recommendation. In: *Proceedings of the web conference 2021*, 2177–2185.

Lv, Q., *et al.*, 2021. Are we really making much progress? Revisiting, benchmarking and refining heterogeneous graph neural networks. In: *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery and data mining*, 1150–1160.

Ma, G., *et al.*, 2024. Successive POI recommendation via brain-inspired spatiotemporal aware representation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38 (1), 574–582.

Martín, Y., *et al.*, 2020. Using geotagged tweets to track population movements to and from Puerto Rico after Hurricane Maria. *Population and Environment*, 42 (1), 4–27.

Mathew, W., Raposo, R., and Martins, B., 2012. Predicting future locations with hidden Markov models. *In: Proceedings of the 2012 ACM conference on ubiquitous computing*, 911–918.

Miller, H.J., and Goodchild, M.F., 2015. Data-driven geography. *GeoJournal*, 80 (4), 449–461.

Monreale, A., *et al.*, 2009. Wherenext: a location predictor on trajectory pattern mining. In: *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*, 637–646.

Noulas, A., *et al.*, 2012. Mining user mobility features for next place prediction in location-based services. In: *2012 IEEE 12th international conference on data mining*, 1038–1043.

Rahmani, H.A., *et al.*, 2022. A systematic analysis on the impact of contextual information on point-of-interest recommendation. arXiv preprint arXiv:.08150.

Rendle, S., Freudenthaler, C., and Schmidt-Thieme, L., 2010. Factorizing personalized Markov chains for next-bask recommendation. In: *Proceedings of the 19th international conference on world wide web*, 811–820.

Schlichtkrull, M., *et al.*, 2018. 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*, 593–607.

Sloan, L., and Morgan, J., 2015. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. *PloS One*, 10 (11), e0142209.

Song, C., *et al.*, 2010a. Modelling the scaling properties of human mobility. *Nature Physics*, 6 (10), 818–823.

Song, C., *et al.*, 2010b. Limits of predictability in human mobility. *Science (New York, NY)*, 327 (5968), 1018–1021.

Sui, D., Elwood, S., and Goodchild, M., 2012. *Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice*. Berlin: Springer Science & Business Media.

Sun, K., *et al.*, 2020. Where to go next: modeling long-and short-term user preferences for point-of-interest recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34 (01), 214–221.

Takens, F., 2006. Detecting strange attractors in turbulence. *In: Dynamical systems and turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*, 366–381.

Tu, W., *et al.*, 2017. Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science*, 31 (12), 2331–2358.

Veličković, P., *et al.*, 2017. Graph attention networks. arXiv preprint arXiv:.10903.

Wang, C., *et al.*, 2021a. CTHGAT: category-aware and time-aware next point-of-interest via heterogeneous graph attention network. In: *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2420–2426.

Wang, D., *et al.*, 2021b. Attentive sequential model based on graph neural network for next poi recommendation. *World Wide Web*, 24 (6), 2161–2184.

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, H., *et al.*, 2021c. Spatio-temporal urban knowledge graph enabled mobility prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5 (4), 1–24.

Wang, P., *et al.*, 2022a. A multi-view bidirectional spatiotemporal graph network for urban traffic flow imputation. *International Journal of Geographical Information Science*, 36 (6), 1231–1257.

Wang, P., *et al.*, 2023a. 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.*, 2023b., EEDN: enhanced encoder–decoder network with local and global context learning for poi recommendation. In: *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*, 383–392.

Wang, Z., *et al.*, 2022b. Graph-enhanced spatial–temporal network for next POI recommendation. *ACM Transactions on Knowledge Discovery from Data*, 16 (6), 1–21.

Wu, Y., *et al.*, 2022. Personalized long-and short-term preference learning for next POI recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 34 (4), 1944–1957.

Yan, X., *et al.*, 2023. Spatio-temporal hypergraph learning for next POI recommendation. In: *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*, 403–412.

Yang, T., *et al.*, 2023. UPTDNet: a user preference transfer and drift network for cross-city next POI recommendation. *International Journal of Intelligent Systems*, 2023 (1), 1–17.

Yao, Y., *et al.*, 2023. Predicting mobile users' next location using the semantically enriched geo-embedding model and the multilayer attention mechanism. *Computers, Environment Urban Systems*, 104, 102009.

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.

Yin, F., *et al.*, 2023. Next POI recommendation with dynamic graph and explicit dependency. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37 (4), 4827–4834.

Yin, J., Gao, Y., and Chi, G., 2022. An evaluation of geo-located Twitter data for measuring human migration. *International Journal of Geographical Information Science: IJGIS*, 36 (9), 1830–1852.

Yu, F., *et al.*, 2020. A category-aware deep model for successive POI recommendation on sparse check-in data. In: *Proceedings of the web conference 2020*, 1264–1274.

Zhang, Y., *et al.*, 2022. City2vec: urban knowledge discovery based on population mobile network. *Sustainable Cities and Society*, 85, 104000.

Zhao, P., *et al.*, 2022. Where to go next: a spatio-temporal gated network for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 34 (5), 2512–2524.