Multi-View Spatial-Temporal Enhanced Hypergraph Network for Next POI Recommendation

Yantong Lai[1,2], Yijun Su[3,4] ✉, Lingwei Wei[1,2], Gaode Chen[1,2], Tianci Wang[1,2], and Daren Zha[1]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3] JD iCity, JD Technology, Beijing, China
[4] JD Intelligent Cities Research, Beijing, China
{laiyantong,weilingwei,chengaode,wangtianci,zhadaren}@iie.ac.cn
suyijun.ucas@gmail.com

Abstract. Next point-of-interest (POI) recommendation has been a prominent and trending task to provide next suitable POI suggestions for users. Current state-of-the-art studies have achieved considerable performances by modeling user-POI interactions or transition patterns via graph- and sequential-based methods. However, most of them still could not well address two major challenges: 1) Ignoring important spatial-temporal correlations during aggregation within user-POI interactions; 2) Insufficiently uncovering complex high-order collaborative signals across users to overcome sparsity issue. To tackle these challenges, we propose a novel method Multi-View Spatial-Temporal Enhanced Hypergraph Network (MSTHN) for next POI recommendation, which jointly learns representations from local and global views. In the local view, we design a spatial-temporal enhanced graph neural network based on user-POI interactions, to aggregate and propagate spatial-temporal correlations in an asymmetric way. In the global view, we propose a stable interactive hypergraph neural network with two-step propagation scheme to capture complex high-order collaborative signals. Furthermore, a user temporal preference augmentation strategy is employed to enhance the representations from both views. Extensive experiments on three real-world datasets validate the superiority of our proposal over the state-of-the-arts. To facilitate future research, we release the codes at https://github.com/icmpnorequest/DASFAA2023_MSTHN.

Keywords: Next POI recommendation · Spatial-temporal graph · Hypergraph neural network

1 Introduction

Location-based social networks (LBSNs) have provided open platforms for users to share their experience at different point-of-interests (POIs), such as restaurants and shopping malls. Therefore, POI recommender systems have been widely

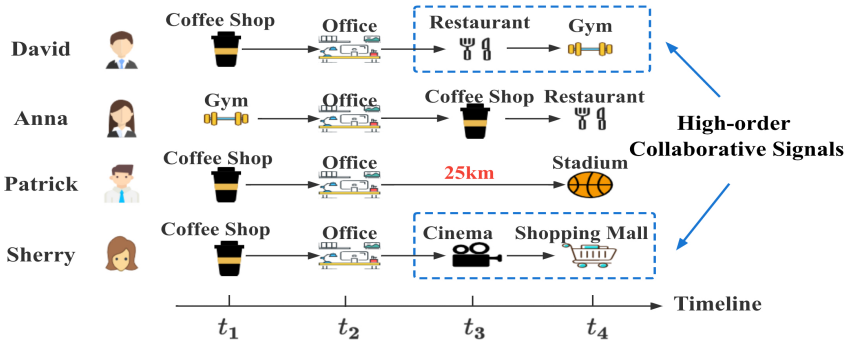

Fig. 1. A motivating example of our proposed framework

utilized to help users and service providers for exploration and targeted advertising, respectively. Among various POI recommendation tasks, next POI recommendation is arguably a prominent and trending one. Different from conventional POI recommendation focusing on user's general long-term preference, next POI recommendation considers user's recent spatial-temporal contexts [21] and long- and short-term preferences for next suitable location suggestions [16,17,31,10].

Prior next POI recommendation approaches are mainly based on sequential methods, ranging from Markov chain [2] to recurrent neural networks (RNNs) [3]. These methods treat it as a general sequence prediction task, and ignore important spatial-temporal information. Subsequently, researchers extend various of RNNs [17,31] by incorporating geographical distance, time intervals or spatio-temporal gates. However, RNN-based methods are limited to short-term contiguous visits. Inspired by the great success of self-attention mechanism [18] in natural language processing field, researchers [12,14] employ it to capture long-term dependencies and correlations between non-consecutive POIs. Nevertheless, they only focus on intra-sequence learning but fail to explore beyond sequence information. Recently, graph-based methods [10,4,13,8,20,15] leverage graph neural networks (GNNs) to refine latent representations of POIs from a global view. Despite their success in next POI recommendation, there still exist some limitations to be better explored.

1) First, ignoring spatial-temporal correlations during aggregation within a user-POI interaction graph. In next POI recommendation, previous GNN-based studies [10,4,13,8,20,15] mainly utilize GNNs to enrich representations from a global view. However, they either ignore or could not directly model spatial-temporal correlations during aggregation and propagation with GNNs. Take David and Anna in Figure 1 for example, they visit the same POIs (coffee shop, office, restaurant and gym) but in different sequential order. If only aggregating their interacted POIs, the embeddings of them would be the same. However, latent representations of David and Anna should be different in fact, due to different sequential order. Additionally, since each user has her/his ac-

ceptance on distance, spatial influences should also be taken into account, as illustrated in Patrick's trajectory. Thus, how to mine and fuse spatial-temporal correlations within a user-POI interaction graph is well deserved to be explored in next POI recommendation.

2) Second, insufficient to uncover high-order collaborative signals. Some researchers [10] in next POI recommendation try to capture collaborative signals [19] by sampling one-hop POI neighbors randomly. Unfortunately, they overlook the high-order connectivity among POIs. As shown in Figure 1, restaurant and gym are high-order neighbors of coffee shop in the trajectory of David. While in Sherry's, its high-order neighbors are cinema and shopping mall. Thus, restaurant, gym, cinema and shopping mall are potentially related, and there might exist implicit high-order collaborative signals among them. If uncovering such signals, it would help alleviate the data sparsity issue.

To this end, we propose a novel framework **M**ulti-View **S**patial-**T**emporal Enhanced **H**ypergraph **N**etwork (MSTHN) for next POI recommendation. To capture spatial-temporal correlations within user-POI interactions, we first design a local spatial-temporal enhanced graph neural network, which aggregates and propagates in an asymmetric way. Then, we construct a global interactive hypergraph to sufficiently uncover high-order collaborative signals with a designed two-step propagation scheme. Subsequently, in contrast to simple concatenation, we utilize a user temporal preference augmentation strategy to enhance the representations from both local and global views. Empirical results show that our MSTHN consistently outperforms state-of-the-art methods, e.g., average relative improvement of 36.20% over LightGCN, 20.36% over SGRec, 17.10% over STAN and 11.09% over DHCN in terms of Recall@10.

We summarize our main contributions as follows:

- To the best of our knowledge, this is the first attempt at multi-view spatial-temporal hypergraph network in next POI recommendation, which captures spatial-temporal correlations and high-order collaborative signals from local and global views.
- We propose a novel local spatial-temporal enhanced graph neural network to jointly model complex user-POI interactions, POI-POI sequential relations and non-adjacent POI-POI geographical relations, which aggregates and propagates spatial-temporal correlations in an asymmetric way.
- We design an interactive hypergraph to depict global interaction dependencies, which empowers to distill high-order collaborative signals effectively.
- Extensive experiments on three public available datasets validate the effectiveness of our proposed MSTHN over various state-of-the-art methods for next POI recommendation.

2 Related Work

2.1 Next POI Recommendation.

Next POI recommendation aims to suggest next suitable location for users based on their recent spatial-temporal context and visiting behaviours. Early studies in

next POI recommendation are mainly based on sequential methods, ranging from Markov chain [2] to recent RNN and its variants [3,31,17]. Limited to short-term contiguous visits in these methods, more recent studies [14,12] solve by utilizing self-attention mechanism [18] to model spatial-temporal information in an explicit or implicit way. However, the above studies only rely on each user's trajectory and overlook the potential collaborative signals among users. Since graph structure is naturally suitable to represent data in LBSN, some researchers have started to leverage graph-based techniques for next POI recommendation, ranging from graph [23] and hypergraph embeddings [24,25] to more recent GNNs [10,4,13,8,20,15]. However, most GNNs-based works [10,4,13,8,20] do not consider important spatial-temporal information within user-POI interactions. Rao et al. [15] noticed the importance of spatial-temporal and chronological information for next POI recommendation, but they still could not directly model such information in GNNs during aggregation and propagation. To tackle the challenge, we propose a local spatial-temporal enhanced graph neural network to capture spatial-temporal correlations during aggregation and propagation within a user-POI interaction graph.

2.2 Hypergraph Neural Network-based Recommendation.

Due to the extension structure and the ability in modeling complex high-order dependencies, hypergraph neural network [5,1] has been recently developed in various recommendation tasks, such as session recommendation [22,11], social recommendation [29,6] and group recommendation [30]. Inspired by these works, we design a learnable interactive hypergraph neural network to uncover global high-order collaborative signals across users.

3 Problem Formulation

Let $\mathcal{U} = \{u_1, u_2, ..., u_{|\mathcal{U}|}\}$ and $\mathcal{L} = \{l_1, l_2, ..., l_{|\mathcal{L}|}\}$ be a set of users and POIs, respectively. Each POI $l \in \mathcal{L}$ has unique geographical coordinates (*longitude, latitude*) tuple, i.e., (*lon, lat*). For each user $u \in \mathcal{U}$, we split her/his trajectory sequence into several sessions by specific time interval (i.e., 1 day) and obtain a trajectory sequence $S^u = \{S_1^u, S_2^u, ..., S_n^u\}$, where n denotes the number of sessions. Each session is denoted as $S_i^u = \{(l_j^u, t_{l_j^u}) | j = 1, 2, ...\}$, where each tuple $(l_j^u, t_{l_j^u})$ indicates user u visited POI l_j^u at timestamp $t_{l_j^u}$.

Given a target user u and her/his trajectory sequence S^u, the goal of next POI recommendation is to recommend top-K POIs that u may visit in the next timestamp.

4 Methodology

In this section, we present our proposed framework **M**ulti-View **S**patial-**T**emporal Enhanced **H**ypergraph **N**etwork (MSTHN) in detail. As illustrated in Figure 2,

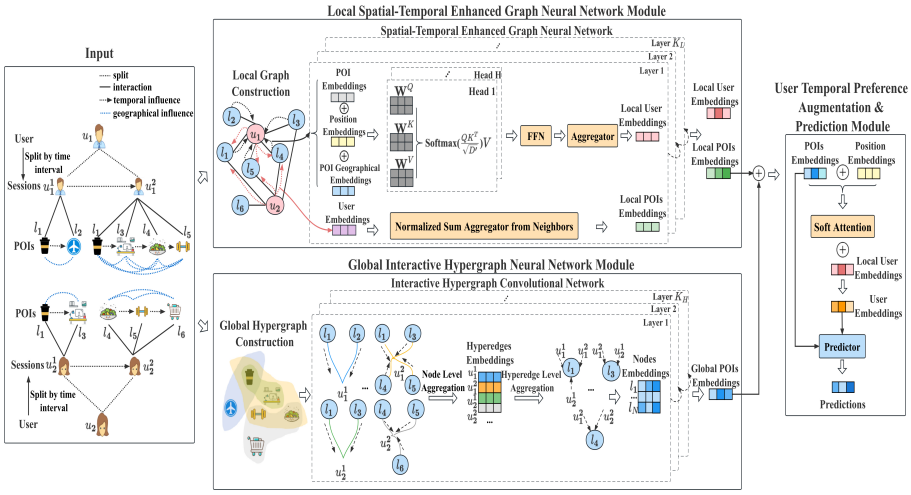

Fig. 2. The framework of our proposed MSTHN. It mainly contains three modules: 1) *Local spatial-temporal enhanced graph neural network module* to capture spatial-temporal correlations; 2) *Global interactive hypergraph neural network module* to uncover high-order collaborative signals; 3) *User temporal preference augmentation module* to augment user preference for prediction.

our MSTHN mainly consists of: 1) Local spatial-temporal enhanced graph neural network module captures spatial-temporal correlations within a user-POI interaction graph in the local view; 2) Global interactive hypergraph neural network module uncovers high-order collaborative signals with a two-step propagation scheme in the global view; 3) User temporal preference augmentation module fuses POIs latent representations from both local and global views and augments long- and short-term user temporal preference; 4) Prediction and optimization module predicts visiting probability from the learned POIs and users latent representations.

4.1 Local Spatial-Temporal Enhanced Graph Neural Network Module

The local spatial-temporal enhanced graph neural network module aims to capture spatial-temporal correlations during aggregation and propagation within user-POI interactions in the local view.

Local Spatial-Temporal Enhanced Graph Construction. To represent user-POI interactions and spatial-temporal correlations among interacted POIs, we firstly construct the local spatial-temporal enhanced graph $\mathcal{G}_L = (\mathcal{V}_L, \mathcal{E}_L)$ (Figure 2). In the local graph $\mathcal{G}_L$, nodes $\mathcal{V}_L$ are users and POIs, and edges $\mathcal{E}_L$ consist of user-POI interactions, POI-POI sequential relations and non-adjacent POI-POI geographical relations.

Spatial-Temporal Message Embedding. To leverage important spatial-temporal information within local spatial-temporal enhanced graph $\mathcal{G}_L$, we firstly sort the interacted POIs of user u chronologically and the sorted POIs set is denoted as $T_u = \{l_1^u, l_2^u, ..., l_m^u\}$, where m is the sequence length of user u. Then, through look-up table, we obtain the initial embeddings for each POI in the sorted set $\mathbf{E}^u = \{\mathbf{e}_1^u, \mathbf{e}_2^u, ..., \mathbf{e}_m^u\}$, where $\mathbf{e}_i^u \in \mathbb{R}^d$ and the embedding dimension is d. Since the interacted POIs are in temporal sequential dependencies, we employ positional encoding [18], which has been proved effective in sequence modeling, to represent the sequential relationship among POIs. The position embeddings of the sorted POI set is $\mathbf{P}^u = \{\mathbf{p}_1^u, \mathbf{p}_2^u, ..., \mathbf{p}_m^u\}$, where $\mathbf{p}_i^u \in \mathbb{R}^d$.

As described in Figure 1, each user has different spatial acceptance on choosing POIs. Thus, the geographical influence among interacted POIs should be taken into account. To achieve this goal, we construct a geographical adjacent matrix $\mathbf{A}_{geo} \in \mathbb{R}^{m \times m}$ to reflect the edge constraints among interacted POIs. For each (l_i^u, l_j^u) pair in the sorted set T_u, the geographical influence a_{ij} is defined as:

$$a_{ij} = \exp(-dist(d_i, d_j)^2) \tag{1}$$

here we choose Haversine distance as $dist(\cdot, \cdot)$ and d_i denotes the geographical coordinates of POI l_i^u. Additionally, we use Δ_d as the distance threshold, if $dist(d_i, d_j) > \Delta_d$, we set $a_{ij} = 0$. For simplicity, we modify Gaussian kernel function to represent the geographical influence between two POIs, which depicts the inverse correlation between geographical influence and distance, and controls the constraint ranging from 0 to 1. To capture the non-linear geographical influence among interacted POIs, we employ the graph convolutional network [9] as follows:

$$\mathbf{V}^u = \mathbf{A}_{geo}\mathbf{E}^u\mathbf{W}_{geo} + \mathbf{b}_{geo} \tag{2}$$

where $\mathbf{W}_{geo} \in \mathbb{R}^{d \times d}$ represents a transition matrix and $\mathbf{b}_{geo} \in \mathbb{R}^d$ is a bias vector.

Subsequently, we obtain the spatial-temporal message embeddings $\mathbf{Z}^u = \mathbf{E}^u + \mathbf{P}^u + \mathbf{V}^u$ by performing element-wise addition on initial embeddings, position embeddings and geographical embeddings, where $\mathbf{Z}^u \in \mathbb{R}^{m \times d}$.

Spatial-Temporal Graph Aggregation Layer. To aggregate important spatial-temporal message collected from interacted POI neighbors in graph $\mathcal{G}_L$, we design a novel spatial-temporal graph aggregation layer that models temporal dependency and non-linear geographical influence among POIs in the local view.

We utilize self-attention [18], an effective mechanism in sequence modeling, to capture sequential dependency and assign different weights to each POI within the interactions. Given the spatial-temporal message embeddings $\mathbf{Z}^u \in \mathbb{R}^{m \times d}$, the spatial-temporal graph aggregation layer firstly performs multi-head scaled dot-product attention operation to get spatial-temporal aware representations $\mathbf{h}^{T_u} \in \mathbb{R}^{m \times d}$ as follows:

$$\mathbf{h}_i^{T_u} = \text{softmax}\left(\frac{(\mathbf{Z}^u\mathbf{W}_Q)(\mathbf{Z}^u\mathbf{W}_K)^T}{\sqrt{D'}}\right)(\mathbf{Z}^u\mathbf{W}_V) \tag{3}$$

$$\mathbf{h}^{T_u} = \text{FFN}([\mathbf{h}_1^{T_u}; \mathbf{h}_2^{T_u}; ...; \mathbf{h}_H^{T_u}]) \tag{4}$$

where H denotes the number of heads in multi-head attention, $[\cdot;\cdot]$ represents the concatenation operation and $D' = \sqrt{d/H}$. Here, $\mathbf{W}_Q$, $\mathbf{W}_K$ and $\mathbf{W}_V \in \mathbb{R}^{d\times D'}$ are shared weight transformations. Additionally, feed-forward network could be represented as $\mathrm{FFN}(\mathbf{x}) = \mathbf{x}\mathbf{W}_0 + \mathbf{b}_0$, where $\mathbf{W}_0 \in \mathbb{R}^{D'\times d}$ and $\mathbf{b}_0 \in \mathbb{R}^d$ are trainable parameters.

Then, we apply mean pooling to obtain local central user representation $\mathbf{x}_L^u = \frac{1}{m+1}\sum_{i=1}^{m} \mathbf{h}_i^{T_u}$, where $\mathbf{x}_L^u \in \mathbb{R}^d$. It aggregates spatial-temporal information from one-hop neighbors in the local view and updates corresponding local node embeddings.

The user-item interaction matrix is $\mathbf{R} \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{L}|}$ and we define the adjacency matrix $\mathbf{A} \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{L}|)\times(|\mathcal{U}|+|\mathcal{L}|)}$ of local spatial-temporal enhanced graph $\mathcal{G}_L$ as:

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{R} \\ \mathbf{R}^T & \mathbf{0} \end{pmatrix} \tag{5}$$

Inspired by LightGCN [7], we also omit non-linear transformation and stack several spatial-temporal graph aggregation layers for propagation to update nodes embeddings:

$$\mathbf{X}_L^{(k+1)} = (\mathbf{D}_L^{-\frac{1}{2}}\mathbf{A}\mathbf{D}_L^{-\frac{1}{2}})\mathbf{X}_L^{(k)} \tag{6}$$

where $\mathbf{D}_L \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{L}|)\times(|\mathcal{U}|+|\mathcal{L}|)}$ is a diagonal matrix and the 0-layer embedding matrix $\mathbf{X}_L^{(0)} \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{L}|)\times d}$ contains initial users embeddings and POIs embeddings. After propagating K_L layers, the final local nodes representations $\mathbf{X}_L \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{L}|)\times d}$ are generated by aggregator (i.e., mean-pooling or sum-pooling).

Different from STAM [27], our proposed spatial-temporal graph aggregation layer models both temporal sequential dependency and non-linear geographical influence among POIs jointly in the local view. Since a user may visit the same POI several times, if taking the chronologically interacted users into account, it would lead to a sub-optimal performance. That is, we only perform spatial-temporal graph aggregation operation in an asymmetric way for local central user node. Detailed empirical analysis would be introduced in section 5.5.

4.2 Global Interactive Hypergraph Neural Network Module

The global interactive hypergraph neural network module aims to uncover high-order collaborative signals effectively with a two-step propagation scheme in the global view.

Global Interactive Hypergraph Construction. Motivated by the strength of hypergraph for unifying nodes beyond pairwise relations, we construct an interactive hypergraph $\mathcal{G}_H = (\mathcal{V}_H, \mathcal{E}_H)$ to uncover high-order collaborative signals across sessions. In the hypergraph $\mathcal{G}_H$, we represent each user's session in her/his trajectory sequence as an hyperedge and the interacted POIs within the session consist of nodes in the hyperedge (Figure 2). Incidence matrix $\mathbf{H} \in \mathbb{R}^{|\mathcal{L}|\times|S|}$ is introduced to describe the topology structure of hypergraph, with entries defined

as:

$$h(v,e) = \begin{cases} 1, & \text{if } e \text{ connects } v, \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

For each node $v \in \mathcal{V}_H$, its degree is defined as $d(v) = \sum_{e \in \mathcal{E}_H} W_e h(v,e)$, calculating the occurrence of node v in all hyperedges. W_e is an assigned positive weight and all the weights formulate a diagonal matrix $\mathbf{W} \in \mathbb{R}^{|S| \times |S|}$. For each hyperedge $e \in \mathcal{E}_H$, its degree is $d(e) = \sum_{v \in \mathcal{V}_H} h(v,e)$. All the node degree and hyperedge degree form diagonal node degree matrix $\mathbf{D}_H$ and diagonal hyperedge degree matrix $\mathbf{B}$ respectively.

Hypergraph Convolutional Network. After the construction of hypergraph $\mathcal{G}_H$, we develop a hypergraph convolutional network with two-step information propagation scheme to capture high-order POI-level relations iteratively. In the node-hyperedge-node propagation scheme, hyperedges serve as mediums for nodes aggregation within the hyperedge and propagation across hyperedges (Figure 2). Particularly, we design our hypergraph convolutional network as follows:

$$\mathbf{X}_H^{(k+1)} = \mathbf{D}_H^{-\frac{1}{2}} \mathbf{H} \mathbf{W} \mathbf{B}^{-1} \mathbf{H}^{\mathrm{T}} \mathbf{D}_H^{-\frac{1}{2}} \mathbf{X}_H^{(k)} \tag{8}$$

where $\mathbf{X}_H^{(k)} \in \mathbb{R}^{|\mathcal{L}| \times d}$ represents the embeddings of POIs, encoded from the $k-$th hypergraph convolutional network layer. In the first node to hyperedge propagation stage, we use multiplication $\mathbf{H}^T \mathbf{X}_H^{(k)}$ to denote the aggregation process, for $\mathbf{H}^T$ reflects the hyperedge-node relation. After aggregating nodes representations within each hyperedge, we then premultiply $\mathbf{H}$ to aggregate information from hyperedges to nodes. Since incidence matrix $\mathbf{H}$ represents the node-hyperedge relation, the second hyperedge to node propagation stage aims to leverage global information beyond current hyperedge to enrich nodes representations.

Distinct to spectral hypergraph convolutional HGNN [5], we omit nonlinear activation function for simplification. Unlike the simplified row normalization in DHCN [22], we keep the same row normalization as HGNN since it is more stable in propagation than the simplified one $\mathbf{D}^{-1}\mathbf{H}\mathbf{W}\mathbf{B}^{-1}\mathbf{H}^{\mathrm{T}}$ in DHCN. According to [5], the symmetric hypergraph Laplacian matrix $\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{H}\mathbf{W}\mathbf{B}^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-\frac{1}{2}}$ is a positive semi-definite matrix, where $\mathbf{I} \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{L}|}$. Therefore, the eigenvalue of $\mathbf{D}^{-\frac{1}{2}}\mathbf{H}\mathbf{W}\mathbf{B}^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-\frac{1}{2}}$ is no larger than 1, solving the instability problem in propagation.

After propagating K_H hypergraph convolutional layers, we average the POIs representations obtained at each layer and output the final global POIs representations $\mathbf{X}_H \in \mathbb{R}^{|\mathcal{L}| \times d}$.

4.3 User Temporal Preference Augmentation Module

The user temporal preference augmentation module aims to fuse the learned representations from both local and global views and augment temporal-aware user preference.

In next POI recommendation, the final decision heavily depends on user's recent preference. Instead of simply aggregating or concatenating the interacted

POIs representations, inspired by [22], we integrate the reversed position embeddings for user temporal preference augmentation. After learning nodes representations from both views, we could obtain the embeddings of all POIs by element-wise addition, e.g., $\mathbf{X}^L = \mathbf{X}_L^L + \mathbf{X}_H$, where $\mathbf{X}_L^L \in \mathbb{R}^{|\mathcal{L}| \times d}$ and $\mathbf{X}_H \in \mathbb{R}^{|\mathcal{L}| \times d}$ denote POIs embeddings in the local and global view, respectively. The i-th POI temporal augmented embedding $\mathbf{x}_i^{u^*}$ in user u's sorted sequence T^u is defined as following:

$$\mathbf{x}_i^{u^*} = \tanh(\mathbf{W}_1[\mathbf{x}_i^u; \mathbf{p}_{m+1-i}] + \mathbf{b}_1) \tag{9}$$

where $\mathbf{W}_1 \in \mathbb{R}^{d \times 2d}$ and $\mathbf{b}_1 \in \mathbb{R}^d$ are trainable parameters. $\mathbf{x}_i^u \in \mathbb{R}^d$ could be indexed from POIs embeddings $\mathbf{X}^L$. Moreover, $\mathbf{p}_{m+1-i} \in \mathbb{R}^d$ denotes reverse position embedding.

Thus, with soft-attention mechanism, we could get temporal preference augmented embedding $\mathbf{x}_T^u \in \mathbb{R}^d$ of user u by assigning different attention weights:

$$\mathbf{x}_T^u = \sum_{i=1}^{m} \alpha_i \mathbf{x}_i^{u^*} \tag{10}$$

$$\alpha_i = \mathbf{q}^T \sigma(\mathbf{W}_2 \mathbf{x}^{u^*} + \mathbf{W}_3 \mathbf{x}_i^{u^*} + \mathbf{b}_2) \tag{11}$$

where $\mathbf{q} \in \mathbb{R}^d$, $\mathbf{W}_2, \mathbf{W}_3 \in \mathbb{R}^{d \times d}$ and $\mathbf{b}_2 \in \mathbb{R}^d$ are trainable attention parameters. $\mathbf{x}^{u^*} \in \mathbb{R}^d$ is aggregated by performing mean-pooling on all the interacted POIs embeddings of user u. σ denotes sigmoid activation function here.

4.4 Prediction and Optimization Module

Having obtained user u's local representation $\mathbf{x}_L^u \in \mathbb{R}^d$ and local-global aware temporal-augmented user representation $\mathbf{x}_T^u \in \mathbb{R}^d$, we apply element-wise addition to get the final user representation as $\mathbf{u} = \mathbf{x}_L^u + \mathbf{x}_T^u, \mathbf{u} \in \mathbb{R}^d$. After that, we compute the score by doing inner product between the final user representation $\mathbf{u}$ and target POI representation $\mathbf{x}^l \in \mathbb{R}^d$:

$$\hat{\mathbf{y}}_l^u = \text{softmax}(\mathbf{u}^T \mathbf{x}^l) \tag{12}$$

We formulate the learning objective as a cross-entropy loss function, which has been largely used in next POI recommendation:

$$\mathcal{J} = -\sum_{u \in \mathcal{U}} \sum_{i \in T^u} \sum_{j=1}^{|\mathcal{L}|} \mathbf{y}_{i,j}^u \log(\hat{\mathbf{y}}_{i,j}^u) + \lambda \|\Theta\|_2 \tag{13}$$

where $\mathbf{y}_{i,j}^u$ is an indicator that is equal to 1 if l_j is the ground truth and 0 otherwise. $\|\Theta\|_2$ represents the $L2$ regularization of all parameters for preventing over-fitting under the control of λ.

5 Experiments

In this section, we present our empirical results to evaluate the effectiveness of our MSTHN.

Table 1. Dataset statistics

	#Users	#POIs	#Check-ins	#Sessions	Sparsity
NYC	834	3,835	44,686	8,841	98.61%
TKY	2,173	7,038	308,566	41,307	97.82%
Gowalla	5,802	40,868	301,080	75,733	99.87%

5.1 Experimental Setting

Datasets We conduct experiments on three public LBSN datasets: Foursquare-NYC (NYC for abbreviation), Foursquare-TKY (TKY) [26] and Gowalla[28]. NYC and TKY were collected from Apr. 2012 to Feb. 2013 in New York City and Tokyo, respectively, while Gowalla contains check-ins from Feb. 2009 to Oct. 2010. Following [17], we first eliminate unpopular POIs that are visited by less than 10 users and 5 users for Gowalla and Foursquare, respectively. Then, we split each user's complete check-ins into sessions within 1 day and remove those which includes fewer than 3 records. Furthermore, inactive users with less than 5 sessions for Gowalla and 3 sessions for Foursquare are filtered out. According to [17], the first 80% sessions of each user are used for training and the rest for testing. The statistics of pre-processed datasets are shown in Table 1.

Evaluation Metrics Following previous works in next POI recommendation, we adopt two widely used evaluation metrics: Recall@K and Normalized Discounted Cumulative Gain (NDCG@K). Specifically, Recall@K measures the rate of the label within top-K recommendations and NDCG@K reflects the quality of ranking lists. In this paper, we repeat experiments on each metric for 10 times and report the averaged Recall@K and NDCG@K with the popular K $\in \{5, 10\}$.

Baselines We compare our MSTHN with following representative methods for next POI recommendation, including 1) statistical-based method UserPop; 2) RNN-based methods GRU, STGN and LSTPM; 3) self-attention-based method STAN; 4) GNN-based methods LightGCN and SGRec and 5) hypergrpah neural network-based method DHCN:

- **UserPop**: It ranks the most popular POIs according to each user's visiting frequency.
- **GRU** [3]: A popular variant of RNN, which controls the information flow with two gates.
- **STGN** [31]: A state-of-the-art LSTM-based model, which introduces spatial and temporal gates for users' long- and short-term preferences.
- **LSTPM** [17]: A state-of-the-art LSTM-based model, which captures long- and short-term preferences with a non-local network and geo-dilated LSTM.
- **STAN** [14]: A state-of-the-art method based on self-attention mechanism, which explicitly models spatial-temporal influences within a user's check-in sequence.

- **LightGCN** [7]: A state-of-the-art simplified GNN-based collaborative filtering framework, which omits the non-linear activation and feature transformation during propagation.
- **SGRec** [10]: A state-of-the-art GNN-based method, which proposes Seq2Graph augmentation and captures collaborative signals among one-hop neighbors. For fairness comparison, we remove the POI categorical information that other methods do not use.
- **DHCN** [22]: A state-of-the-art hypergraph neural network-based method for session recommendation, which could be applied for next POI recommendation.

Parameter Settings Our experiments are conducted with PyTorch 1.9.1 on a 32 GB Tesla V100 GPU. For baselines, we firstly preserve the settings as provided in original papers and fine-tune each model's hyperparameters on three datasets. For our MSTHN, we adopt Adam as optimizer with a learning rate of 1e-3, weight decay of 1e-5 and dropout rate of 0.3. We apply the same dimension size $d = 128$ for user and POI embeddings and set batch size as 100. In each batch, we pad sessions which do not meet the maximum session length in batch. Furthermore, we empirically choose 2.5km (for NYC and TKY) and 100km (for Gowalla) as distance threshold and use 1 layer spatial-temporal graph aggregation layer in all datasets. The number of stable hypergraph convolutional layer and head of self-attention is chosen from $\{1, 2, 3, 4\}$ and $\{1, 2, 4, 8, 16\}$, respectively.

5.2 Performance Comparison

The results of all the methods are reported in Table 2. For the results, we have the following observations.

Our proposed MSTHN achieves the best results on all datasets. On NYC dataset, our MSTHN improves the performance over the best baseline by 8.84%-15.58%. Additionally, MSTHN outperforms the best results by 5.12%-13.69% on TKY dataset and 7.12%-8.98% on Gowalla dataset. We contribute the improvements to the following aspects: 1) Capturing important spatial-temporal

Table 2. Performances comparison on three datasets. The best and the second best performances are bolded and underlined, respectively. The improvements are calculated between the best and the second best scores.

Method	NYC				TKY				Gowalla			
	Rec@5	Rec@10	NDCG@5	NDCG@10	Rec@5	Rec@10	NDCG@5	NDCG@10	Rec@5	Rec@10	NDCG@5	NDCG@10
UserPop	0.2866	0.3297	0.2283	0.2423	0.2229	0.2668	0.1718	0.1861	0.0982	0.1489	0.0907	0.1336
GRU	0.236	0.2471	0.2252	0.2279	0.1549	0.1734	0.1371	0.1436	0.1282	0.1606	0.1102	0.1225
STGN	0.2371	0.2594	0.2261	0.2307	0.2112	0.2587	0.1482	0.1589	0.1600	0.2041	0.1191	0.1333
LSTPM	0.2495	0.2668	0.2425	0.2483	0.2203	0.2703	0.1556	0.1734	0.2021	0.2510	0.1523	0.1681
STAN	0.3523	0.3827	0.3025	0.3137	0.2621	0.3317	0.2074	0.2189	0.2449	0.2878	0.1837	0.1942
LightGCN	0.3221	0.3488	0.2958	0.3042	0.2213	0.2594	0.1977	0.2098	0.2356	0.2590	0.1801	0.1915
SGRec	0.3451	0.3723	0.3052	0.3178	0.2537	0.3213	0.2221	0.2447	0.2395	0.2813	0.1862	0.2002
DHCN	0.3745	0.3966	0.3126	0.3203	0.3172	0.3454	0.2442	0.2543	0.2653	0.3124	0.2038	0.2191
MSTHN	**0.4076**	**0.4398**	**0.3612**	**0.3702**	**0.3378**	**0.3927**	**0.2567**	**0.2721**	**0.2842**	**0.3396**	**0.2221**	**0.2365**
%Improv.	8.84%	10.89%	15.55%	15.58%	6.49%	13.69%	5.12%	7.00%	7.12%	8.71%	8.98%	7.94%

correlations within user-POI interactions in the local view. 2) Uncovering complex high-order collaborative signals in the global view.

Capturing spatial-temporal correlations is important for next POI recommendation. Methods which leverage spatial-temporal information explicitly or implicitly perform better than that do not use. For example, our MSTHN reaches up to 15.58% on NDCG@10 on NYC dataset than DHCN. On sparser Gowalla dataset, our MSTHN still outperforms DHCN on both metrics. Leveraging temporal information, SGRec surpasses LightGCN by 23.86% in terms of Recall@10 on TKY dataset. Beneficial from well-designed spatial-temporal gates, LSTPM and STGN also outperforms GRU on three datasets, especially on more sparser Gowalla dataset.

Uncovering high-order collaborative signals is effective and significant to improve quality of recommendation. From Table 2, hypergraph neural network-based methods (our MSTHN and DHCN) perform better than other baselines. For example, on TKY dataset, our MSTHN improves Recall@10 by 22.22% and 51.39% against SGRec and LighGCN. The major reason for LightGCN is over-smoothing effect that makes nodes representation indistinguishable with deeper layer. Since SGRec performs one-hop neighbors sampling randomly, it would cause information losses. STAN performs better than SGRec on Recall metrics on three datasets, but worse on NDCG. Thus, uncovering high-order collaborative signals contributes more on exploring potential POIs but lacks of exploiting sequential dependency.

5.3 Ablation Study

Next we investigate the underlining mechanism of our MSTHN with three ablated models: 1) MSTHN$_{w/o\ local}$ that removes local spatial-temporal enhanced graph neural network module; 2) MSTHN$_{w/o\ global}$ that removes global interactive hypergraph neural network module; 3) MSTHN$_{w/o\ temporal}$ that removes user temporal preference augmentation module; 4) Local$_{w/o\ spatial}$ that removes spatial information in the local view; 5) Local$_{w/o\ temporal}$ that removes temporal information in the local view. From Table 3, we have the following observations:

Table 3. Ablation study on MSTHN w.r.t. Recall@10 and NDCG@10.

Method	NYC		TKY		Gowalla	
	Recall@10	NDCG@10	Recall@10	NDCG@10	Recall@10	NDCG@10
MSTHN$_{w/o\ local}$	0.4275	0.3620	0.3734	0.2606	0.3125	0.2207
MSTHN$_{w/o\ global}$	0.3812	0.3033	0.3105	0.2388	0.2832	0.1917
MSTHN$_{w/o\ temporal}$	0.4013	0.3105	0.3562	0.2541	0.3098	0.2126
Local$_{w/o\ spatial}$	0.4333	0.3651	0.3791	0.2595	0.3174	0.2187
Local$_{w/o\ temporal}$	0.4387	0.3698	0.3874	0.2632	0.3259	0.2241
MSTHN	0.4398	0.3702	0.3927	0.2721	0.3396	0.2365

First, when removing the local view module, MSTHN$_{w/o\ local}$ decreases slightly compared with the other two variants due to the losses of spatial-temporal correlations. The local view affects the correctness of recommendation more than the quality. Specifically, the average decline rates on three datasets are 5.57% on Recall@10 and 4.61% on NDCG@10. Second, when removing the global view module, MSTHN$_{w/o\ global}$ drops clearly. It strongly indicates the importance of global hypergraph network, for it could represent beyond pairwise relations and model distant POIs. Additionally, it proves the significance of high-order collaborative signals for next POI recommendation. Third, when removing the user temporal preference augmentation module, the variant MSTHN$_{w/o\ temporal}$ is less competitive than the complete MSTHN. It implies the effectiveness of user temporal preference augmentation for next POI recommendation. Fourth, spatial-temporal information is essential in the local view and spatial information contributes more to performances.

5.4 Hyperparameter Analysis

We further qualitatively analyze the impacts of layer number and head number in MSTHN.

Impact of Layer Number. To investigate the impact of stacking hypergraph convolutional layers, we conduct experiments with number of layer in $\{1, 2, 3, 4\}$. As illustrated in Figure 3, our MSTHN achieves the best performances by stacking 3 layers on NYC dataset, 4 layers on TKY dataset, and 2 layers on Gowalla dataset. The results prove that our MSTHN could uncover and distill high-order collaborative signals effectively, especially on denser dataset (i.e., TKY). The possible cause of dropping would be the over-smoothing issue.

Impact of Heads Number. To explore the impact of choosing number of heads in spatial-temporal aggregation layer, we search from set $\{1, 2, 4, 8, 16\}$. From Figure 4, our MSTHN is insensitive to the number of heads on both Recall and NDCG metrics, and obtains the best performances with 8 heads on three datasets. With number of heads increasing from 8 to 16, Recall@10 and NDCG@10 on three datasets drop. The possible cause would be the over-fitting in capturing spatial-temporal correlations within user-POI interactions in the local view.

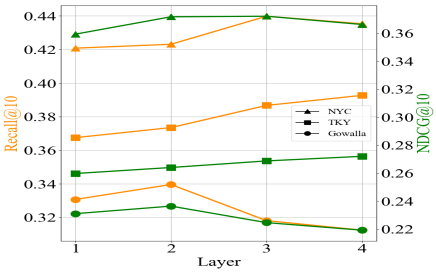

Fig. 3. Impact of Layer Number

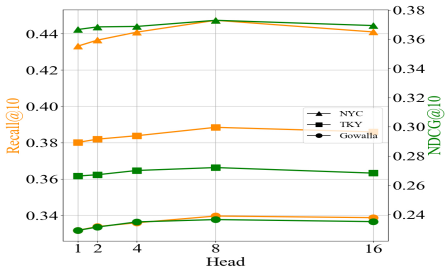

Fig. 4. Impact of Head Number

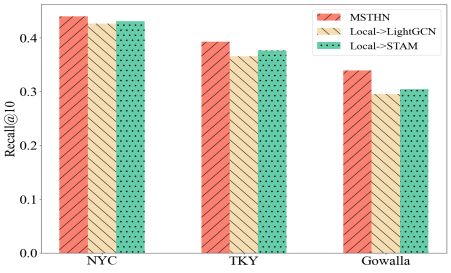

Fig. 5. Effect of Local View on Recall@10

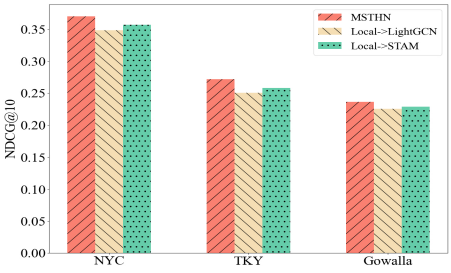

Fig. 6. Effect of Local View on NDCG@10

5.5 Further Study

To explore the effect of our proposed local view, we maintain other parts of MSTHN and replace local view with STAM [27] and LightGCN [7]. From Figure 5-6, on both Recall@10 and NDCG@10, our MSTHN outperforms these variants and the variant with STAM performs better than that with LightGCN. It proves the effectiveness of capturing spatial-temporal correlations within user-POI interactions. STAM utilizes users sequential dependency to update representations of POIs (e.g., if a POI has been visited by user $u_1 \rightarrow u_2 \rightarrow u_1$, STAM would take $u_1 \rightarrow u_2$ as input), which ignores repeated visiting patterns and another existing sequential dependency (i.e., $u_2 \rightarrow u_1$), and leads to a sub-optimal performance. Thus, our proposed local spatial-temporal enhanced graph neural network could well address the limitation by learning in an asymmetric way. Moreover, the results against variant with STAM also indicate the significance of non-adjacent POI-POI geographical relations for next POI recommendation.

6 Conclusion

In this paper, we propose a novel **M**ulti-View **S**patial-**T**emporal Enhanced **H**ypergraph **N**etwork (MSTHN) for next POI recommendation, which jointly learns representations from both local and global views. Through the spatial-temporal enhanced graph neural network and interactive hypergraph neural network, MSTHN could capture important spatial-temporal correlations within user-POI interactions and high-order collaborative signals across users. Experimental results on three datasets demonstrate the effectiveness of our MSTHN.

References

1. Bai, S., Zhang, F., Torr, P.H.: Hypergraph convolution and hypergraph attention. Pattern Recognition **110**, 107637 (2021)
2. Cheng, C., Yang, H., Lyu, M.R., King, I.: Where you like to go next: Successive point-of-interest recommendation. In: Twenty-Third international joint conference on Artificial Intelligence (2013)

3. Cho, K., van Merrienboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: EMNLP. pp. 1724–1734. ACL (2014)
4. Dang, W., Wang, H., Pan, S., Zhang, P., Zhou, C., Chen, X., Wang, J.: Predicting human mobility via graph convolutional dual-attentive networks. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. pp. 192–200 (2022)
5. Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y.: Hypergraph neural networks. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 3558–3565 (2019)
6. Han, J., Tao, Q., Tang, Y., Xia, Y.: Dh-hgcn: Dual homogeneity hypergraph convolutional network for multiple social recommendations. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2190–2194 (2022)
7. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: Lightgcn: Simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. pp. 639–648 (2020)
8. Huang, Z., Ma, J., Dong, Y., Foutz, N.Z., Li, J.: Empowering next poi recommendation with multi-relational modeling. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2034–2038 (2022)
9. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (2017)
10. Li, Y., Chen, T., Luo, Y., Yin, H., Huang, Z.: Discovering collaborative signals for next poi recommendation with iterative seq2graph augmentation. In: Proceedings of the 30th IJCAI. pp. 1491–1497 (2021)
11. Li, Y., Gao, C., Luo, H., Jin, D., Li, Y.: Enhancing hypergraph neural networks with intent disentanglement for session-based recommendation. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1997–2002 (2022)
12. Lian, D., Wu, Y., Ge, Y., Xie, X., Chen, E.: Geography-aware sequential location recommendation. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2009–2019 (2020)
13. Lim, N., Hooi, B., Ng, S.K., Goh, Y.L., Weng, R., Tan, R.: Hierarchical multi-task graph recurrent network for next poi recommendation. In: Proceedings of the 45th international ACM SIGIR conference on Research and development in Information Retrieval (2022)
14. Luo, Y., Liu, Q., Liu, Z.: Stan: Spatio-temporal attention network for next location recommendation. In: Proceedings of the Web Conference 2021. pp. 2177–2185 (2021)
15. Rao, X., Chen, L., Liu, Y., Shang, S., Yao, B., Han, P.: Graph-flashback network for next location recommendation. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 1463–1471 (2022)
16. Su, Y., Li, X., Tang, W., Xiang, J., He, Y.: Next check-in location prediction via footprints and friendship on location-based social networks. In: 2018 19th IEEE International Conference on Mobile Data Management (MDM). pp. 251–256. IEEE (2018)
17. Sun, K., Qian, T., Chen, T., Liang, Y., Nguyen, Q.V.H., Yin, H.: Where to go next: Modeling long-and short-term user preferences for point-of-interest recommenda-

tion. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 214–221 (2020)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
19. Wang, X., He, X., Wang, M., Feng, F., Chua, T.S.: Neural graph collaborative filtering. In: Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval. pp. 165–174 (2019)
20. Wang, Z., Zhu, Y., Liu, H., Wang, C.: Learning graph-based disentangled representations for next poi recommendation. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1154–1163 (2022)
21. Wang, Z., Zhu, Y., Zhang, Q., Liu, H., Wang, C., Liu, T.: Graph-enhanced spatial-temporal network for next poi recommendation. ACM Transactions on Knowledge Discovery from Data (TKDD) **16**(6), 1–21 (2022)
22. Xia, X., Yin, H., Yu, J., Wang, Q., Cui, L., Zhang, X.: Self-supervised hypergraph convolutional networks for session-based recommendation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 4503–4511 (2021)
23. Xie, M., Yin, H., Wang, H., Xu, F., Chen, W., Wang, S.: Learning graph-based poi embedding for location-based recommendation. In: Proceedings of the 25th ACM international on conference on information and knowledge management. pp. 15–24 (2016)
24. Yang, D., Qu, B., Yang, J., Cudre-Mauroux, P.: Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In: The world wide web conference. pp. 2147–2157 (2019)
25. Yang, D., Qu, B., Yang, J., Cudré-Mauroux, P.: Lbsn2vec++: Heterogeneous hypergraph embedding for location-based social networks. IEEE Transactions on Knowledge and Data Engineering (2020)
26. Yang, D., Zhang, D., Zheng, V.W., Yu, Z.: Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. IEEE Transactions on Systems, Man, and Cybernetics: Systems **45**(1), 129–142 (2014)
27. Yang, Z., Ding, M., Xu, B., Yang, H., Tang, J.: Stam: A spatiotemporal aggregation method for graph neural network-based recommendation. In: Proceedings of the ACM Web Conference 2022. pp. 3217–3228 (2022)
28. Yin, H., Cui, B., Chen, L., Hu, Z., Zhang, C.: Modeling location-based user rating profiles for personalized recommendation. ACM Transactions on Knowledge Discovery from Data (TKDD) **9**(3), 1–41 (2015)
29. Yu, J., Yin, H., Li, J., Wang, Q., Hung, N.Q.V., Zhang, X.: Self-supervised multi-channel hypergraph convolutional network for social recommendation. In: Proceedings of the Web Conference 2021. pp. 413–424 (2021)
30. Zhang, J., Gao, M., Yu, J., Guo, L., Li, J., Yin, H.: Double-scale self-supervised hypergraph learning for group recommendation. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management. pp. 2557–2567 (2021)
31. Zhao, P., Luo, A., Liu, Y., Zhuang, F., Xu, J., Li, Z., Sheng, V.S., Zhou, X.: Where to go next: A spatio-temporal gated network for next poi recommendation. IEEE Transactions on Knowledge and Data Engineering (2020)