DST-GN: Predicting Human Mobility via Disentangled Spatio-Temporal and Category Graph Networks

Jianqun Cui[⋆†], **Jianing Zhang**[⋆], **Min Wang**[⋆*], **& Yanan Chang**[⋆†*]
[⋆]School of Computer Science, Central China Normal University, Wuhan 430079, China
[†]Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning, Wuhan 430079, China
{*cswm@ccnu.edu.cn, *ynchang@mail.ccnu.edu.cn}
*Corresponding Authors

Abstract

Predicting human mobility, specifically next Point-of-Interest (POI) suggestions, requires modeling how individuals integrate spatial, temporal and category cues during decision-making. However, existing computational models often conflate these heterogeneous signals or struggle with data sparsity. We propose the **D**isentangled **S**patio-**T**emporal and Category **G**raph **N**etwork (DST-GN), a framework that reconstructs cognitive maps through representation learning. DST-GN models spatial, temporal and category contexts as independent graph views, employing Graph Attention Networks (GAT) to learn view-specific embeddings. To ensure coherence across these disentangled pathways, we introduce a cross-view contrastive learning mechanism that aligns signals into a unified semantic space. Furthermore, a global Collective Flow Enhancer is integrated to mitigate data sparsity by leveraging population-level heuristics. Experimental results on three real-world datasets show that our model outperforms state-of-the-art techniques, achieving average improvements of 22.84% in HR@10 and 18.55% in NDCG@10.

Keywords: Human Mobility; Next POI recommendation; User preference; Contrastive learning

Introduction

The ubiquity of mobile devices and Location-Based Social Networks (LBSNs) has yielded massive human mobility data, offering a transformative lens for understanding urban dynamics. Within this context, next-POI recommendation is not merely a data mining task but also serves as a computational proxy for modeling human decision-making processes. From a cognitive perspective, user trajectories can be viewed as observable outcomes of internal cognitive maps, in which individuals integrate spatial, temporal, and category cues when navigating urban environments. Despite recent advancements, the efficacy of existing trajectory prediction models remains significantly hindered by two persistent challenges (Acharya & Mohbey, 2025; Zhang et al., 2025).

First, human mobility is governed by an intricate "where–when–what" triad, encompassing spatial distance, temporal sequence and semantic category, which calls for the explicit disentanglement of heterogeneous behavioral signals (Z. Huang et al., 2024; Sun et al., 2024). However, many existing models suffer from signal entanglement, in which these distinct factors are compressed into monolithic embeddings. Such representations tend to produce context-agnostic recommendations and consequently exhibit limited generalization across diverse user routines (Lin et al., 2021; Xie & Chen, 2023). Second, the widespread issue of data sparsity—commonly manifested as the cold-start problem for new users or POIs—substantially constrains the capacity of a model to learn robust and personalized preference representations (Sanner et al., 2023; Yuan & Hernandez, 2023).

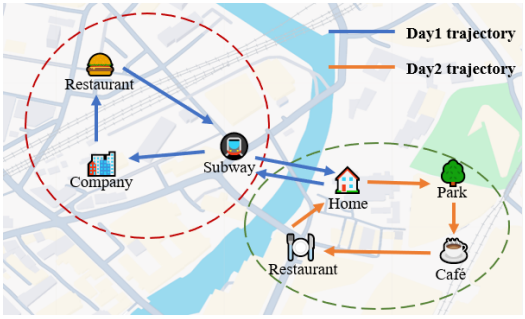


Figure 1: Distinct mobility patterns of a user, contrasting a trajectory (Day1) in a work zone with a trajectory (Day2) in a living zone.

To address these challenges, we design DST-GN, a unified framework that leverages Graph Neural Networks to disentangle and reintegrate the multifaceted drivers of human mobility. As illustrated in Figure 1, traditional models fall short by producing a single, conflated representation of a POI visit. For instance, a restaurant visited after work (Day1) is encoded without separating its identity from its specific workday context (e.g., its proximity to the office, the lunchtime visit). This leads to flawed, context-agnostic recommendations, such as suggesting the same work-lunch spot for a weekend outing. In contrast, DST-GN is designed to disentangle these contextual signals to provide more accurate, context-aware suggestions. DST-GN first constructs three distinct graphs to model spatial proximity, temporal dynamics and category correlations, and then employs view-specific GAT (Velickovic et al., 2017) to learn specialized representations from each view. To ensure semantic consistency across views, we introduce a cross-view contrastive learning mechanism that aligns the learned embeddings into a shared semantic space. Beyond personalized, disentangled preferences, our framework integrates global collective-flow information learned from all users. Finally, these rich local (personalized) and global (collective) representations are fused via a context-aware attention layer, where the user embedding serves as a query to yield highly accurate, personalized POI recommendations. In summary, the

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

main contributions of this work are summarized as follows:

- We propose a novel multi-view graph learning framework that explicitly separates spatial, temporal and category signals. By modeling these factors independently before integration, we achieve a more granular and interpretable representation of user behavior.
- We design a fusion strategy that integrates personalized short-term preferences with global collective flow patterns. This allows the model to dynamically balance individual habits with population-level trends, addressing the data sparsity issue effectively.
- We conduct extensive evaluations on three real-world POI recommendation datasets. The results demonstrate that DST-GN significantly outperforms existing baselines, validating the effectiveness of disentangled modeling in capturing the complexities of human mobility.

Related Work

Spatio-Temporal Modeling for Next POI Recommendation

Accurately capturing the multi-faceted nature of user behavior remains a focal point in mobility research. However, existing approaches often struggle to maintain the distinct properties of contextual signals. For example, Song et al. (2025) fused temporal and spatial features directly, a strategy that fails to account for the independent cognitive weights humans assign to distance versus time. Sequence-based models, such as those by T. Huang et al. (2024) and Feng et al. (2024), often prioritize temporal transition patterns at the expense of static semantic affinities, leading to a loss of location-specific context. While He et al. (2024) explored personalized modeling, the absence of explicit disentanglement allows transient contextual noise—such as a specific workday lunch hour—to overshadow the intrinsic category properties of a POI, resulting in suboptimal recommendations for non-routine scenarios. Our work addresses this by constructing independent graph views, ensuring that each mobility signal is modeled without interference before final integration.

Data Sparsity in Next POI Recommendation

Data sparsity remains a fundamental challenge in next POI recommendation, most notably manifesting as the cold-start problem, where users exhibit limited historical interactions. Existing studies generally address this issue through two dominant paradigms. The first line of work attempts to alleviate sparsity by incorporating external metadata, such as social network relations to model inter-user dependencies or rich user profiles to infer latent preferences. However, despite their potential benefits (Cai et al., 2023; Li et al., 2025), such auxiliary information is often noisy, privacy-sensitive, or unavailable in real-world applications. More recently, self-supervised and contrastive learning have been introduced to enhance representation robustness under sparse data regimes. Representative methods such as CLSPRec (Duan et al., 2023), DCLS (Liu et al., 2024), and COSTA (Lei et al., 2025), exploit feature-level augmentations and cross-view alignment strategies to extract more discriminative signals from limited user trajectories. Although these approaches have demonstrated notable effectiveness, they typically lack an explicit mechanism to connect an individual's sparse behavioral history with global mobility patterns shared across the population. In contrast to these existing methods, our proposed DST-GN does not rely on external metadata. Instead, it aims to distill generalized navigational flows from collective interaction data, leveraging collective intelligence as a cognitive heuristic. By capturing common mobility regularities emerging from the crowd, DST-GN is able to provide robust and informative recommendations even in scenarios where personal historical data is insufficient.

Method

The definitions of the main terms used in our research problem are as follows. Let $L = \{l_1, l_2, \cdots, l_M\}$ be a set of POIs; $U = \{u_1, u_2, \cdots, u_N\}$ denote a set of users; $C = \{c_1, c_2, \cdots, c_P\}$ be a set of categories of POIs, where M, N and P are respectively the total number of POIs, users and categories. User behavior is not static; it changes with the temporal context. To model distinct behavioral patterns, such as the increased tendency for novelty-seeking on weekends compared to weekdays, we explicitly categorize the temporal context into two states: weekdays (denoted as d_0) and weekends (denoted as d_1). Therefore, a check-in record $r = (u, l, c, g, t, d_x)$ signifies that user u visited the point of interest (POI) l at the time t. The POI l is defined by its category c and geocoded by g (longitude and latitude). The proposed framework is illustrated in Figure. 2.

User-Level Multi-View Graph Encoding

To capture structured dependencies across different aspects of user check-in behavior, we construct three multi-view graphs: a spatial graph ($\mathcal{G}_s$), a temporal graph ($\mathcal{G}_t$) and a category graph ($\mathcal{G}_c$). Each view is designed to encode a distinct relational structure underlying user mobility patterns. For each user, these graphs are instantiated by extracting POI transitions from the user's historical trajectories, while the graph construction rules are shared across users.

Spatial Graph ($\mathcal{G}_s$): The spatial graph $\mathcal{G}_s = (\mathcal{V}, \mathcal{E}_s, \mathbf{W}_s)$ is constructed to model the geographical influence on user mobility, where $\mathcal{V}$ denotes the set of POI nodes. An edge $(i, j) \in \mathcal{E}_s$ is established between two POIs if they are visited consecutively within a trajectory. The edge weight is defined by spatial similarity, computed from the great-circle distance via an exponential decay function:

$$W_s(i, j) = \exp\big(-\text{Haversine}(\Delta S_{ij})\big). \tag{1}$$

Here, $W_s(i, j)$ represents the weight between POIs i and j, and Haversine$(\cdot)$ computes the great-circle distance given their geographic coordinates, with ΔS_{ij} denoting the corresponding spatial distance.

Temporal Graph ($\mathcal{G}_t$): The temporal graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t, \mathbf{W}_t)$ is designed to capture temporal dependencies in user mobility behavior. The node set $\mathcal{V}$ and edge set $\mathcal{E}_t$ are constructed analogously to those of the spatial graph, while edge weights are determined by the time interval between consecutive visits. Specifically, we define the temporal weight by jointly modeling daily periodicity and temporal decay:

$$W_t(i,j) = \frac{1}{2}\left[1 + \cos\left(\frac{2\pi\Delta T_{ij}}{1440}\right)\right] \cdot \exp\left(-\frac{\Delta T_{ij}}{1440}\right). \quad (2)$$

Here, ΔT_{ij} denotes the time interval (in minutes) between check-ins at POIs i and j. The cosine term models the 24-hour(1440 minutes) periodicity of human activity, while the exponential term assigns higher weights to temporally proximate transitions.

Category Graph ($\mathcal{G}_c$): The category graph $\mathcal{G}_c = (\mathcal{V}, \mathcal{E}_c, \mathbf{W}_c)$ captures high-level semantic relations among POIs based on their functional categories. In this graph, each node corresponds to a POI, and a directed edge $(C_i, C_j) \in \mathcal{E}_c$ is created if users transition from a POI belonging to category C_i to one belonging to category C_j. The edge weight $W_c(C_i, C_j)$ is defined as the normalized transition frequency, reflecting the probability of moving from category C_i to category C_j.

After constructing these three graphs, we apply a separate GAT to learn view-specific embeddings for each trajectory: $\mathbf{z}_s, \mathbf{z}_t, \mathbf{z}_c \in \mathbb{R}^d$. Using dedicated GATs for each graph allows the model to specialize in capturing the unique relational patterns inherent in each view. Then, we use an attention mechanism Vaswani et al., 2017 to fuse its spatial, temporal and category representations into a unified embedding $\mathbf{H}_{\text{fused}}$, allowing the model to adaptively weight these factors for context-aware recommendations:

$$\mathbf{H}_{\text{fused}} = \text{Attention}\left(\mathbf{z}_s, \mathbf{z}_t, \mathbf{z}_c\right) \quad (3)$$

Cross-View Contrastive Learning

While the disentangled graph encoders capture specialized features, human navigation is fundamentally a coherent process where spatial, temporal and category signals must converge into a unified understanding. To encourage this consistency across heterogeneous contexts, we introduce a multi-view contrastive learning objective. This mechanism ensures that different "views" of the same mobility event remain semantically aligned in the latent space, thereby filtering out view-specific noise.

Each embedding from the three encoders is fed into a shared non-linear projection head f_θ (implemented as a multi-layer perceptron) to map them into a shared hypersphere suitable for similarity measurement:

$$\mathbf{z} = \frac{f_\theta(\mathbf{x})}{|f_\theta(\mathbf{x})|_2} \quad (4)$$

Positive pairs: Following the principle of semantic coincidence, embeddings derived from the same trajectory segment but across different graph types (spatial $\mathbf{z}_s$, temporal $\mathbf{z}_t$, and category $\mathbf{z}_c$) are treated as positive pairs. This encourages the model to recognize that a lunchtime visit (temporal) to a mall (category) near the office (spatial) represents the same underlying behavioral intent. Their similarity is computed as the average pairwise cosine similarity:

$$s_{\text{pos}} = \frac{1}{3}\left(\mathbf{z}_s \cdot \mathbf{z}_c + \mathbf{z}_s \cdot \mathbf{z}_t + \mathbf{z}_c \cdot \mathbf{z}_t\right) \quad (5)$$

Negative pairs: To enhance the discriminative power of the representations, we adopt a robust negative sampling strategy. We include embeddings from different time windows (cross-day) and different users (intra-day). These diverse negatives prevent the model from learning trivial solutions and force it to capture the unique "signature" of a specific user-context interaction.The contrastive objective is operationalized using the InfoNCE loss, which maximizes the mutual information between the positive views relative to the negatives:

$$\mathcal{L}_{\text{cl}} = -\frac{1}{N}\sum i = 1^N \log \frac{\exp(s_{\text{pos}}^{(i)}/\tau)}{\exp(s_{\text{pos}}^{(i)}/\tau) + \sum_j \exp(s_{\text{neg}}^{(j)}/\tau)} \quad (6)$$

where τ is a temperature hyperparameter that modulates the concentration of the distribution.

User-Aware Spatio-Temporal Transformer

While the previous graph layers capture structural relationships, human mobility is inherently sequential and governed by dynamic daily routines. To capture these fine-grained behavioral patterns, we design a User-Aware Spatio-Temporal Transformer that processes daily POI sequences through a context-sensitive attentional lens.

Spatio-Temporal Transformer To model user fine-grained daily behavioral patterns, we represent each day's POI visit sequence using a simple embedding layer. This daily sequence embedding and the fusion vector of the three previously obtained graphs $\mathbf{H}_{\text{fused}}$ are passed into a bias-aware Transformer encoder. To capture spatio-temporal correlations, we incorporate learnable bias terms into the attention mechanism. Formally, the attention weight between two positions i and j in the sequence is computed as:

$$w_{ij} = \text{softmax}_j\left(\frac{Q_i K_j^\top}{\sqrt{d}} + b_{ij}^{(t)} + b_{ij}^{(s)}\right) \quad (7)$$

where $b_{ij}^{(t)} = f_t(\Delta T_{ij})$ and $b_{ij}^{(s)} = f_s(\Delta S_{ij})$ are learnable biases that encode the temporal interval and spatial distance between the i-th and j-th POIs, respectively. By incorporating both sequence content and spatio-temporal biases, the model effectively attends to contextually relevant past behaviors when modeling user preferences. The resulting vector after this transformer is $\mathbf{M}_{\mathbf{u}}^{\mathbf{i}}$.

User-Aware Attention We introduce a user-aware attention mechanism to capture individual preferences, allowing different users to focus on different aspects of the same POI sequence. Therefore, given the user embedding $\mathbf{u_i}$ and the

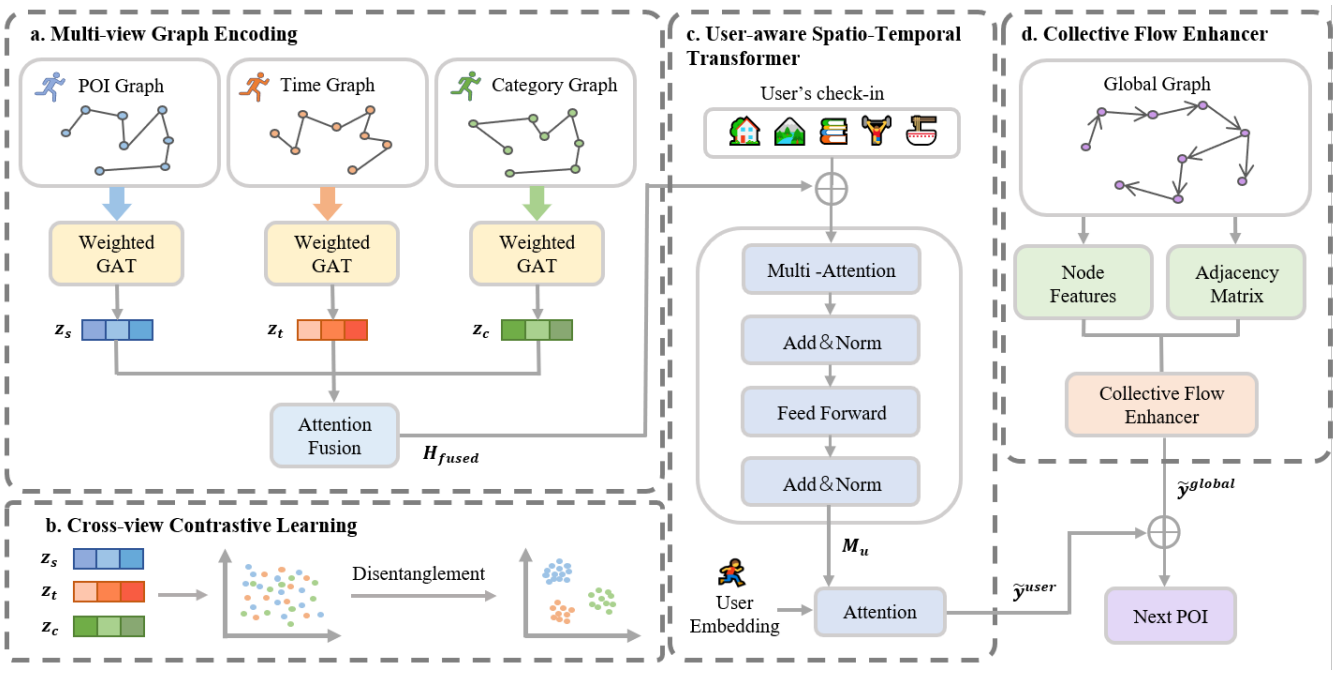


Figure 2: Overall framework of the proposed model DST-GN, which integrates multi-view graph encoding, cross-view contrastive learning, user-aware spatio-temporal Transformer and collective flow enhancer to capture user preferences for POI recommendation.

output representations $\mathbf{M_u^i}$ from the Transformer encoder, the details are as follows (Vaswani et al., 2017):

$$\mathbf{N_u^i} = Attention\left(\mathbf{u_i}, \mathbf{M_u^i}, \mathbf{M_u^i}\right) \tag{8}$$

$\mathbf{N_u^i}$ is the personalized representation vector. This vector is passed through a multi-layer perceptron to produce a user-specific prediction $\tilde{y}_i^{user}$.

Collective Flow Enhancer

To capture global transition patterns among POIs, we construct a graph from all users' historical trajectories (Yang et al., 2022). Each node represents a POI and an edge connects two POIs if they appear consecutively in any trajectory, with the weight counting the number of such transitions. The global graph is represented as (X, A), where X is the node feature matrix and A is the adjacency matrix.

We compute unnormalized attention scores between all node pairs based on their feature representations and graph structure. Given an input feature matrix $\mathbf{X} \in \mathbb{R}^{N\times d}$, we first project it into a new space using a learnable weight matrix $\mathbf{W} \in \mathbb{R}^{d\times d'}$:

$$\mathbf{I} = \mathbf{XW} \tag{9}$$

A bilinear attention formulation compute attention scores between node pairs. The learnable parameter vector $a \in \mathbb{R}^{2d'}$ is split into two parts, $a_1 \in \mathbb{R}^{d'}$ and $a_2 \in \mathbb{R}^{d'}$. For each node pair (i, j) , we apply LeakyReLU (Maas et al., 2013) as the nonlinear activation function to compute the attention coefficient between nodes, formulated as:

$$e_{ij} = \text{LeakyReLU}\left(a_1^\top \mathbf{I}_i + a_2^\top \mathbf{I}_j\right) \tag{10}$$

We add the adjacency matrix to the identity matrix E and rescale attention scores to highlight stronger structural connections:

$$\tilde{y}_i^{global} = e_{ij} \cdot (A_{ij} + E) \tag{11}$$

Loss Function

To simulate the dual-process nature of human navigational decision-making, we integrate predictions from two complementary information sources. This fusion represents the cognitive interplay between an individual's idiosyncratic history and the normative behavioral patterns of the population.

(1) A user-specific prediction derived from the personalized representation $\tilde{y}_i^{user}$, (2) a global prediction obtained from the global graph $\tilde{y}_i^{global}$. The final prediction combines both components:

$$\tilde{y}_i = \tilde{y}_i^{user} + \tilde{y}_i^{global} \tag{12}$$

This fusion enables the model to combine individual preferences with population-level patterns. The softmax function was applied to obtain the predicted probability distribution over candidate POIs:

$$p_i = \text{softmax}(\tilde{y}_i) \tag{13}$$

Finally, the model is trained by minimizing the cross-entropy loss with respect to the ground truth y_i:

$$\mathcal{L}_{poi} = -\sum_{i=1}^{Q} y_i \log(p_i) \tag{14}$$

Table 1: Experimental results on the three datasets. In each column, the best result is shown in **bold** and the second-best is underlined.

Model	PHO				SIN				NYC			
	HR@5	HR@10	NDCG@5	NDCG@10	HR@5	HR@10	NDCG@5	NDCG@10	HR@5	HR@10	NDCG@5	NDCG@10
CTLE (Lin et al., 2021)	0.2632	0.3605	0.1995	0.2068	0.2041	0.2784	0.1313	0.1556	0.2421	0.3205	0.1513	0.1841
STAR-HiT (Xie & Chen, 2023)	0.5063	0.6582	0.3924	0.4304	0.3138	0.4105	0.2539	0.2855	0.3435	0.4078	0.2672	0.2871
DisenPOI (Qin et al., 2023)	0.4209	0.4988	0.3402	0.3652	0.2808	0.3420	0.2307	0.2506	0.2940	0.3634	0.2401	0.2594
GETNext (Yang et al., 2022)	0.5130	0.6147	0.3730	0.4019	0.3517	0.4294	0.2889	0.3163	0.2916	0.3690	0.2119	0.2342
CLSPRec (Duan et al., 2023)	0.5368	0.6368	0.3811	0.4175	0.3545	0.4093	0.2794	0.2942	0.3545	0.4352	0.2653	0.2870
DCLS (Liu et al., 2024)	0.6842	0.7017	0.5255	0.5472	0.4382	0.4894	0.3592	0.3747	0.4357	0.4980	0.3385	0.3599
COSTA (Lei et al., 2025)	0.6000	0.6947	0.4866	0.5088	0.3615	0.4177	0.2837	0.2995	0.3539	0.4290	0.2672	0.2871
FHCRec (Chen et al., 2025)	0.6000	0.6842	0.4257	0.4481	0.4219	0.4873	0.3369	0.3562	0.3917	0.4658	0.2877	0.3099
DST-GN (ours)	**0.7895**	**0.8421**	**0.6152**	**0.6246**	**0.5506**	**0.6108**	**0.4260**	**0.4442**	**0.5626**	**0.6160**	**0.4267**	**0.4425**
improve ↑	15.39%	20.01%	17.07%	14.14%	25.65%	24.81%	18.60%	18.55%	29.13%	23.69%	26.06%	22.95%

Experiment

Datasets and Evaluation Metrics

To evaluate DST-GN, we utilized three longitudinal Foursquare benchmarks—Phoenix (PHO), Singapore (SIN), and New York City (NYC)—covering records from April 2012 to September 2013. Following standard protocols to ensure data quality, we removed POIs with fewer than 10 interactions and trajectories with fewer than 3 check-ins, while excluding inactive users with fewer than 5 trajectories. The refined data was partitioned using a chronological 8:1:1 split to preserve the temporal integrity of human mobility patterns.

Performance is quantified using Hit Rate (HR@K) and Normalized Discounted Cumulative Gain (NDCG@K) for $K \in \{5, 10\}$. While HR@K measures the frequency with which the ground-truth POI appears in the top-K list, NDCG@K provides a position-sensitive evaluation, rewarding the model for ranking the correct destination higher to reflect more efficient decision-making.

Baselines

We compare DST-GN with a wide range of state-of-the-art methods for next POI recommendation, including sequential, graph-based, and contrastive learning models.

- **CTLE** (Lin et al., 2021) is a Transformer-based sequential model that captures long-range temporal dependencies through pre-training.
- **STAR-HiT** (Xie & Chen, 2023) models hierarchical user preferences by jointly learning short-term and long-term sequential patterns with attention mechanisms.
- **GETNext** (Yang et al., 2022) constructs a global transition graph to model collective mobility patterns and enhance next-location prediction.
- **DisenPOI** (Qin et al., 2023) disentangles geographical and semantic factors within graph representations to better capture location transition behaviors.
- **CLSPRec** (Duan et al., 2023) introduces contrastive learning with trajectory-level augmentations to improve representation learning under sparse data.
- **DCLS** (Liu et al., 2024) employs dual contrastive objectives to jointly model user–POI interactions and sequential dynamics.
- **COSTA** (Lei et al., 2025) leverages cross-view contrastive learning to align spatial–temporal representations and mitigate data sparsity.
- **FHCRec** (Chen et al., 2025) enhances recommendation performance by integrating feature-level and hierarchical contrastive learning signals.

Overall Performance

Table 1 presents a comprehensive performance comparison between DST-GN and several state-of-the-art baselines across three real-world datasets. Our proposed model consistently outperforms all competing methods across every evaluation metric, achieving significant improvements in both predictive accuracy (HR) and ranking quality (NDCG). Specifically, DST-GN achieves an average improvement of over 20% across all metrics compared to the best-performing baselines, establishing strong new performance benchmarks.

The empirical results reveal clear patterns in the evolution of mobility modeling. Sequential models such as CTLE and STAR-HiT show the weaker performance, suggesting that modeling mobility as a linear sequence fails to capture the inherently non-linear and context-dependent nature of human movement. Graph-based approaches, including GETNext and DisenPOI, achieve better results by explicitly modeling spatial correlations and transition structures. However, their reliance on a single unified graph representation often entangles

heterogeneous factors, limiting their ability to distinguish between different mobility cues. In contrast, DST-GN adopts a disentangled multi-graph architecture that separately models spatial, temporal and category, enabling more fine-grained and robust representations. Although contrastive learning-based models such as CLSPRec, DCLS, COSTA, and FHCRec improve representation quality through data augmentation and alignment objectives, they lack explicit structural separation among different contextual signals. The consistently superior performance of DST-GN demonstrates that jointly leveraging independent graph views and cross-view contrastive alignment is crucial for constructing a coherent and effective mobility representation.

Ablation Experiments

To assess the contribution of each key component in DST-GN, we conduct ablation studies on the NYC dataset by comparing the full model with three variants:

- **w/o CL**: removes the cross-view contrastive learning module;
- **w/o Trans**: removes the bias-aware Spatio-Temporal Transformer;
- **w/o Global**: removes the Collective Flow Enhancer.

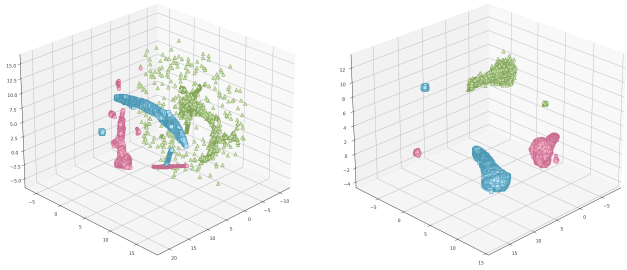
The results in Table 2 indicate that all components contribute substantially to the overall performance. Removing the contrastive learning module (**w/o CL**) results in a notable performance drop, particularly in terms of NDCG. This suggests that cross-view contrastive learning is crucial for aligning the disentangled graph representations; without it, spatial and category signals become weakly coupled, leading to less coherent mobility representations. The **w/o Trans** variant also exhibits consistent degradation, demonstrating the importance of modeling fine-grained sequential dependencies. Although graph structures capture global transition patterns, the Spatio-Temporal Transformer enables dynamic contextual weighting of historical behaviors, which is essential for integrating static POI identities with temporal and spatial constraints. The largest performance degradation is observed in the **w/o Global** variant, highlighting the critical role of collective mobility information. This finding confirms that individual interaction history alone is often insufficient under sparse settings. By incorporating population-level transition regularities, the global component provides a form of social guidance that effectively compensates for ambiguous or limited personal trajectories.

Table 2: Ablation study results on NYC.

Model Variant	HR@10	NDCG@10
w/o CL	0.6010	0.3904
w/o Trans	0.6093	0.4175
w/o Global	0.5609	0.3886
DST-GN	**0.6160**	**0.4425**

Effectiveness of Disentanglement

To empirically evaluate the impact of the cross-view contrastive learning module, we visualized the latent distributions of the spatial, temporal and category embeddings using UMAP. As illustrated in Figure. 3, the integration of the contrastive objective fundamentally reshapes the model's representation space. In the absence of contrastive learning (Figure. 3a), the embeddings from the three encoders exhibit a highly entangled and fragmented distribution, suggesting a "cognitive misalignment" where heterogeneous signals fail to converge into a unified representation. Conversely, with the introduction of the contrastive module (Figure. 3b), the three views form compact, well-defined clusters with clear inter-class separation. This transformation demonstrates that our module acts as a critical alignment mechanism, maximizing mutual information across views to extract invariant behavioral signatures while successfully disentangling context-specific noise. Consequently, the resulting representations are more discriminative and robust, providing a superior computational basis for predicting complex human mobility patterns.



(a) Without contrastive learning (b) With contrastive learning

Figure 3: UMAP visualization of spatial, temporal, and category embeddings. (a) Without contrastive learning, the embeddings exhibit a fragmented, overlapping distribution. (b) With contrastive learning, the embeddings converge into compact, well-separated clusters.

Conclusion

In this paper, we propose DST-GN to address signal entanglement and data sparsity in next-POI recommendation. By disentangling spatial, temporal, and category influences through multi-view graph learning and cross-view contrastive alignment, DST-GN enables more fine-grained modeling of human mobility. Extensive experiments on three real-world datasets demonstrate its consistent superiority over state-of-the-art baselines. Beyond empirical performance, DST-GN provides an interpretable modeling perspective in which heterogeneous contextual cues are separated and then integrated to support effective mobility prediction. Future work will explore finer-grained temporal dynamics and multimodal contextual information to further enhance the expressiveness and applicability of the proposed framework.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62272189, 62372206, 62402193), the Hubei Provincial Natural Science Foundation of China (No. 2024AFB421) and the Fundamental Research Funds for the Central Universities, Central China Normal University (No. XJ2026001501)

References

Acharya, M., & Mohbey, K. K. (2025). Exploring the evolution, progress, and future of point-of-interest recommendation over location-based social network: A comprehensive review. *GeoInformatica*, *29*(3), 305–350.

Cai, D., Qian, S., Fang, Q., Hu, J., & Xu, C. (2023). User cold-start recommendation via inductive heterogeneous graph neural network. *ACM Transactions on Information Systems*, *41*(3), 1–27.

Chen, J., Sang, Y., Zhang, P.-F., Wang, J., Qu, J., & Li, Z. (2025). Enhancing long-and short-term representations for next poi recommendations via frequency and hierarchical contrastive learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *39*, 11472–11480.

Duan, C., Fan, W., Zhou, W., Liu, H., & Wen, J. (2023). Clsprec: Contrastive learning of long and short-term preferences for next poi recommendation. *Proceedings of the 32nd acm international conference on information and knowledge management*, 473–482.

Feng, S., Meng, F., Chen, L., Shang, S., & Ong, Y. S. (2024). Rotan: A rotation-based temporal attention network for time-specific next poi recommendation. *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, 759–770.

He, X., He, W., Liu, Y., Lu, X., Xiao, Y., & Liu, Y. (2024). Imnext: Irregular interval attention and multi-task learning for next poi recommendation. *Knowledge-Based Systems*, *293*, 111674.

Huang, T., Pan, X., Cai, X., Zhang, Y., & Yuan, X. (2024). Learning time slot preferences via mobility tree for next poi recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*, 8535–8543.

Huang, Z., Xu, S., Wang, M., Wu, H., Xu, Y., & Jin, Y. (2024). Human mobility prediction with causal and spatial-constrained multi-task network. *EPJ Data Science*, *13*(1), 22.

Lei, Y., Shen, L., Sun, Z., He, T., Feng, S., & Liu, G. (2025). Costa: Contrastive spatial and temporal debiasing framework for next poi recommendation. *Neural Networks*, 107212.

Li, Y., Shan, Y., Liu, Y., Wang, H., Wang, W., Wang, Y., & Li, R. (2025). Personalized federated recommendation for cold-start users via adaptive knowledge fusion. *Proceedings of the ACM on Web Conference 2025*, 2700–2709.

Lin, Y., Wan, H., Guo, S., & Lin, Y. (2021). Pre-training context and time aware location embeddings from spatial-temporal trajectories for user next location prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*, 4241–4248.

Liu, Z., Deng, J., Zhang, D., Shen, G., Kong, X., et al. (2024). Dual contrastive learning for next poi recommendation with long and short-term trajectory modeling. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*.

Maas, A. L., Hannun, A. Y., Ng, A. Y., et al. (2013). Rectifier nonlinearities improve neural network acoustic models. *Proc. icml*, *30*, 3.

Qin, Y., Wang, Y., Sun, F., Ju, W., Hou, X., Wang, Z., Cheng, J., Lei, J., & Zhang, M. (2023). Disenpoi: Disentangling sequential and geographical influence for point-of-interest recommendation. *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 508–516.

Sanner, S., Balog, K., Radlinski, F., Wedin, B., & Dixon, L. (2023). Large language models are competitive near cold-start recommenders for language-and item-based preferences. *Proceedings of the 17th ACM conference on recommender systems*, 890–896.

Song, C., Ren, Z., & Lu, L. (2025). Integrating personalized spatio-temporal clustering for next poi recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *39*, 12550–12558.

Sun, T., Fu, K., Huang, W., Zhao, K., Gong, Y., & Chen, M. (2024). Going where, by whom, and at what time: Next location prediction considering user preference and temporal regularity. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2784–2793.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al. (2017). Graph attention networks. *stat*, *1050*(20), 10–48550.

Xie, J., & Chen, Z. (2023). Hierarchical transformer with spatio-temporal context aggregation for next point-of-interest recommendation. *ACM Transactions on Information Systems*, *42*(2), 1–30.

Yang, S., Liu, J., & Zhao, K. (2022). Getnext: Trajectory flow map enhanced transformer for next poi recommendation. *Proceedings of the 45th International ACM SIGIR Conference on research and development in information retrieval*, 1144–1153.

Yuan, H., & Hernandez, A. A. (2023). User cold start problem in recommendation systems: A systematic review. *IEEE access*, *11*, 136958–136977.

Zhang, Q., Yang, P., Yu, J., Wang, H., He, X., Yiu, S.-M., & Yin, H. (2025). A survey on point-of-interest recommendation: Models, architectures, and security. *IEEE Transactions on Knowledge and Data Engineering*.