PRELIMINARY PREPRINT VERSION – DO NOT CITE
This is not the version of record. The final version will appear in the Proceedings of IJCAI-ECAI 2026.

Disentangled Hypergraph Network with Implicit Structure Learning for Mobility Social Relationship Inference

Jingjing Zhu[2] , **Xiang Li**[2] , **Dongliang Chen**[2] , **Haobing Liu**[2] , **Yuan Cao**[2] , **Yanwei Yu**[1,2*]
[1]State Key Laboratory of Physical Oceanography, Ocean University of China
[2]Faculty of Information Science and Engineering, Ocean University of China
{zhujingjing, lixiang1202}@stu.ouc.edu.cn, {chendongliang, haobingliu, cy8661, yuyanwei}@ouc.edu.cn

Abstract

Inferring social relationships from users' mobile data holds significant value for personalized recommendations. Most methods model user interactions based on co-occurrence records, achieving impressive success in capturing social signals. However, despite these advancements, current techniques still face three primary challenges. First, compressing higher-order co-occurrences into a single representation space easily leads to semantic entanglement. Second, relying on deterministic topology overlooks unobserved homophily. Third, the intrinsic characteristics of the users themselves are frequently ignored. To address these challenges, we propose a **D**isentangled **H**ypergraph Network with **I**mplicit **S**tructure **L**earning for mobility social relationship inference, named **DHISL**. Specifically, it first captures individual spatiotemporal characteristics through feature-guided representation initialization, and then adopts a dual-branch framework: the explicit branch models social and spatiotemporal preference signals, while the implicit branch adaptively learns a sparse channel-wise relational structure via differentiable structure learning. Furthermore, we introduce dual constraints in a multi-channel disentangled space to achieve intra-channel consistency and inter-channel exclusivity. Experiments on four real-world datasets demonstrate that DHISL outperforms state-of-the-art models, achieving average improvements of 6.34% in ROCAUC and 6.39% in PRAUC. The source code is available at https://github.com/Tilamisu-zz/DHISL.

1 Introduction

With the widespread application of intelligent location-based social networks (LBSNs, *e.g.*, Foursquare), a large amount of user mobility check-in data has been generated. Mining social relationships among users from these data is of great value for a wide range of spatiotemporal applications [Dai *et al.*, 2022; Li *et al.*, 2024; Baik and Yoo, 2025]. It not only helps reveal the coupling mechanisms between social ties and human mobility, but also directly supports key scenarios such as friend recommendation [Liu *et al.*, 2022; Li *et al.*, 2023a; Wang *et al.*, 2025], and location recommendations [Yang *et al.*, 2022a; Wu *et al.*, 2023; Darraz *et al.*, 2025], which increasingly rely on advanced graph representation learning to capture high-order user interactions and mitigate data sparsity [Li *et al.*, 2025a; Li *et al.*, 2025b].

Intuitively, there exists a strong correlation between human mobility patterns and social bonds: individuals who frequently visit the same places are more likely to be socially connected [Cho *et al.*, 2011; Yang *et al.*, 2019]. Motivated by this intuition, researchers have evolved from early frequency-based co-occurrence statistical methods [Pham *et al.*, 2016] to today's sophisticated graph neural network approaches. A large number of studies transform mobility check-in data into graph structures for representation learning, ranging from co-occurrence/encounter-based user graphs and user–location bipartite graphs to heterogeneous graphs/heterogeneous hypergraphs [Wang *et al.*, 2019; Fu *et al.*, 2024; Yan *et al.*, 2025] that incorporate location semantics and temporal factors. Some recent frontier studies further advance this area by optimizing the graph structure. Despite substantial progress, identifying reliable social connections from sparse and noisy mobility data remains challenging:

First, existing methods compress captured friendship information or high-order co-occurrences directly into a single representation space [Yang et al.*, 2020; Xia* et al.*, 2022], ignoring the diverse latent intentions underlying the same behavior.* Specifically, the social relationships in reality are driven by various distinct underlying factors such as social circles or living habits. Consequently, a single representation struggles to distinguish these heterogeneous social ties effectively (strong ties based on deep interests and weak ties caused by superficial overlap in life trajectories). Moreover, there are significant differences in the intentions of interaction behaviors at the same location (active socializing based on common interests and passive encounters due to geographical convenience), and a single-channel modeling approach cannot effectively decouple these underlying factors.

Second, most methods overly rely on deterministic topological structures constructed from explicit interactions, overlooking unobserved homophily that extends beyond physical co-occurrence interactions. To be specific, the check-in

*Corresponding author: Yanwei Yu.

data is partially observable, and many real-world offline social interactions are not recorded [Qin *et al.*, 2023]. At this point, the rigid and deterministic topology causes the model to wrongly classify true friends who lack direct spatiotemporal co-presence evidence as weak ties or even unrelated. Although some studies attempt to alleviate interaction sparsity through predefined heuristic rules [Qin *et al.*, 2024], this static graph-construction paradigm still struggles to dynamically uncover deep and nonlinear latent social connections.

Third, existing research primarily constructs models from the perspective of interaction relationships, underestimating the significant impact of users' own spatio-temporal behavioral preferences on inferring social relationships [Desai et al.*, 2017].* Specifically, the maintenance of social relationships is reflected not only in explicit co-occurrences but also in the stable similarities of user behavioral patterns [Wang *et al.*, 2011; Fan *et al.*, 2017], such as daily activity radius, temporal routines, and location sequence transition patterns. Nevertheless, the existing methods often focus on complex network interaction learning while neglecting to explicitly utilize the inherent spatiotemporal attributes of individual users as an independent and powerful feature.

To address these challenges, we propose a Disentangled Hypergraph Network with Implicit Structure Learning (DHISL). *First*, we construct an initial representation comprising three dimensions. By integrating sequential mobility embeddings, global spatial statistics, and user identity embeddings, we reinforce the representation of users' intrinsic spatio-temporal preferences. *Second*, to overcome the limitations brought about by relying solely on visible physical interactions, we design a parallel dual-branch architecture. The explicit branch constructs a hypergraph structure, utilizing hyperedges to capture observed partial friendships and high-order co-occurrence relations. Conversely, the implicit branch uses a differentiable graph structure learner to dynamically infer and recover latent connections that should exist but remain unobserved. *Third*, to address the semantic confusion arising from compressing complex relationships into a single channel, we model users within a multi-channel disentangled space and propose a dual-constraint strategy. On the one hand, we employ an independence constraint to enforce orthogonality across different channels. On the other hand, we utilize a channel-aware contrastive learning mechanism to force the precise alignment of explicit high-order relationships and implicit latent structures within specific semantic channels. This design of "intra-channel consistency and inter-channel exclusivity" enables the model to automatically decompose entangled social signals. Experimental results show that our model outperforms state-of-the-art (SOTA) methods, achieving up to an 8.60% improvement in PRAUC across the four datasets.

In summary, our contributions are as follows:

- We propose a dual-branch explicit-implicit framework. This framework first explicitly enhances user spatio-temporal preference representations through a feature-guided initialization strategy. It then synergizes explicit hypergraphs with an implicit graph structure learner to dynamically complete social relationships.
- We design a multi-channel disentanglement mechanism with dual constraints to resolve the semantic confusion inherent in complex social relationships. By introducing independence constraints and channel-aware contrastive learning, we decompose entangled social signals into independent driving factors.
- We conduct extensive experiments on real-world mobility datasets. The experimental results verify that DHISL outperforms existing SOTA baselines on the social relationship inference task.

2 Related Work

Social relationship inference from user mobility data aims to mine latent social ties from users' spatiotemporal check-in records. With the advancement of representation learning techniques, research in this area has evolved from traditional feature engineering to sophisticated graph neural networks.

Graph embedding-based methods. Early methods for mobility-based social relationship inference mainly treated the problem as a classification task based on handcrafted features. For example, the EBM [Pham *et al.*, 2013] uses entropy measurement to quantify the diversity of co-occurrences, thereby inferring the probability of friendship. The PGT framework [Wang *et al.*, 2014] further integrates individual, global, and temporal factors to describe the similarity of user mobility patterns. With the advent of network embedding, researchers began to learn low-dimensional user representations from mobility traces. Walk2friends [Backes *et al.*, 2017] constructs a user-location bipartite graph and utilizes random walks to capture co-occurrence relationships. Similarly, Emb-cat [Yu *et al.*, 2018] builds a User Meeting Graph and employs hierarchical sampling to model direct interactions. While these methods pioneered the use of mobility data for social relationship inference, they largely rely on shallow structural features or linear transitions, limiting their ability to capture the complex, non-linear spatiotemporal semantics inherent in human social behaviors.

GNN-based methods. To better capture the semantic richness of mobility data, GNNs have become the mainstream solution for social relationship inference. These methods typically construct heterogeneous graphs connecting users, locations, and check-ins. For example, Heter-GCN [Wu *et al.*, 2019] models multiple types of relations and applies GCNs for semi-supervised link prediction. MVMN [Zhang *et al.*, 2020] introduces a multi-view matching network that fuses spatial, temporal, and social views to learn robust user representations. Addressing the noise prevalent in check-ins, SRINet [Qin *et al.*, 2023] proposes a graph structure learning framework to prune irrelevant edges in the user meeting graph. More recently, MRGAN [Qin *et al.*, 2024] advances this line of research by constructing a multi-relational graph that explicitly models both direct encounters and indirect preference similarities. However, a fundamental limitation of these approaches is their reliance on *pairwise* edges.

Hypergraph-based methods. To address the limitations of pairwise modeling, researchers began to use hypergraph neural networks [Gao *et al.*, 2020; Yang *et al.*, 2022b].

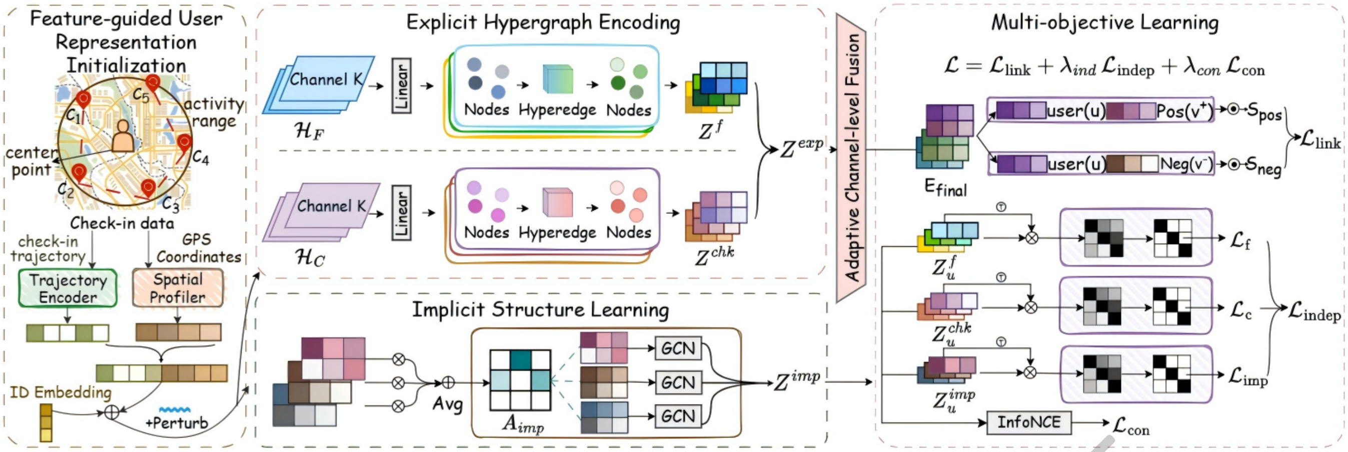

Figure 1: Overall framework of the proposed model.

LBSN2Vec++ [Yang *et al.*, 2020] models check-ins as hyperedges connecting users, POIs, and time slots, learning embeddings via hypergraph random walks. H^3GNN [Li *et al.*, 2025c] further advances this by employing trajectory hyperedges and hyperbolic embedding to accommodate the scale-free nature of social networks. However, most existing hypergraph methods rely on static construction rules, ignoring the implicit relations captured by recent works like MR-GAN [Qin *et al.*, 2024]. More importantly, these methods usually encode diverse user behaviors into a holistic embedding, resulting in entangled representations.

Different from prior studies, our DHISL combines implicit structure learning and disentanglement modeling based on the hypergraph, and is capable of effectively capturing high-order social and interest-driven intentions.

3 Problem Definition

Definition 1 (Check-in Record)**.** *Let $\mathcal{U} = \{u_1, u_2, ..., u_N\}$ denote the set of N users, and $\mathcal{P} = \{p_1, p_2, ..., p_M\}$ denote the set of M points of interest. A check-in record is denoted as $c = \langle u, t, p \rangle$, where $u \in \mathcal{U}$ represents the user, $p \in \mathcal{P}$ represents the visited location, and t is the timestamp. This constitutes the minimal atomic unit of mobility data.*

Definition 2 (Mobility Trajectory)**.** *Based on check-in records, the mobility trajectory of user u is defined as a time-ordered sequence $T_u = \langle c_1, c_2, ..., c_{|T_u|} \rangle$.*

Problem (Mobility Social Relationship Inference)**.** *Given a set of users $\mathcal{U}$, the goal is to learn an inference function $\mathcal{F}(u_i, u_j) \to \hat{y}_{ij} \in \{0, 1\}$ based on their mobility trajectories T_u and partially observed friendships. Here, $\hat{y}_{ij} = 1$ indicates that u_i and u_j are predicted to be friends, and $\hat{y}_{ij} = 0$ indicates they are non-friends.*

4 Methodology

DHISL uses feature-guided initialization, synergizing explicit hypergraphs and implicit structure learning to capture observed co-occurrences and latent homophily. Adaptive fusion and dual constraints then disentangle signals for accurate inference.

4.1 Feature-guided User Representation Initialization

For each user node, we aim to construct an informative initial representation before graph message passing, so as to capture users' mobility dynamics, spatial distribution, and individual-specific signals that are difficult to observe explicitly.

Mobility Sequence Embedding Learning. To encode the temporal dynamics of user mobility, we adopt a self-supervised sequential pretraining strategy. Following RNN-based next-location prediction models [Sun *et al.*, 2019], we encode each user's trajectory with a GRU and pretrain it with a next-location prediction objective. By optimizing the next-location prediction objective, we extract the final hidden state of the GRU for each user, compressing sequential dependencies and transition patterns into a fixed-size latent vector:

$$\mathbf{e}_u^{\text{seq}} = \text{GRU}\Big(\{\mathbf{E}_{\text{loc}}\,[p_{u,t}]\}_{t=1}^{|T_u|}\Big)_{\text{final}}, \tag{1}$$

where $\mathbf{E}_{\text{loc}} \in \mathbb{R}^{|M| \times d}$ is a learnable location embedding matrix that maps the discrete location token $p_{u,t}$ of user u at step t into a d-dimensional dense vector, M is the total number of locations.

Global Spatial Information Statistics. We calculate the statistical features $x_u^{\text{stat}} \in \mathbb{R}^4$ based on the GPS coordinates of the users' check-in locations, thereby capturing the central nature and spatial distribution of the users' activities from a global statistical perspective:

$$\mathbf{x}_u^{\text{stat}} = \big[\mu_{\text{lat}}(u),\ \mu_{\text{lon}}(u),\ \sigma_{\text{lat}}(u),\ \sigma_{\text{lon}}(u)\big]. \tag{2}$$

Then, these statistical data are normalized to obtain $\hat{x}_u^{\text{stat}} \in \mathbb{R}^4$. By concatenating them with the sequential mobility embedding, we derive the raw content feature vector of user u:

$$f_u = \big[\mathbf{e}_u^{\text{seq}} \parallel \hat{x}_u^{\text{stat}}\big]. \tag{3}$$

Feature–identity Hybrid Multi-channel Initialization. To construct semantically discriminative and disentangled multi-channel representations, we first map the raw feature vector f_u into K channel-specific subspaces (each with dimensionality D). Concretely, we employ a learnable linear projection matrix $\mathbf{W}_1$ to expand the feature space, and reshape the output into a multi-channel feature embedding, thereby explicitly isolating different latent factors of user behavior:

$$\mathbf{E}_u = \mathrm{Reshape}\big(\mathbf{W}_1 f_u + b_1\big), \tag{4}$$

where $\mathbf{E}_u \in \mathbb{R}^{K\times D}$, and b_1 denotes the learnable bias vector within the linear projection layer.

In addition, we introduce a learnable identity embedding for each user, $\mathbf{E}_{id} \in \mathbb{R}^{K\times D}$ (initialized with Xavier initialization), and fuse it with the feature-projected embedding to obtain the initial node representation:

$$\mathbf{E} = \mathrm{Mean}\left(\mathbf{E}_u, \mathbf{E}_{id}\right), \tag{5}$$

stacking the initial representations of all users yields the node feature tensor $\mathbf{E}\in \mathbb{R}^{N\times K\times D}$. Motivated by [Xu and Chen, 2025; Cao *et al.*, 2025], we inject uniform noise $\Delta(u) \sim \mathcal{U}(-\epsilon, \epsilon)$ into the initial embedding during training to enhance the robustness against noise. Then we use the perturbed representation $\mathbf{X} = \mathbf{E} + \Delta$ as the input to subsequent message passing. This random perturbation serves as a regularizer, preventing the model from overfitting to spurious feature correlations.

4.2 Explicit Hypergraph Encoding

Explicit hypergraph encoding aims to explicitly inject *directly observable* high-order interaction structures (friendship relations and spatiotemporal check-in groups) in mobility check-in scenarios into the model in a structured manner. Specifically, we construct a *friendship-relation view* and a *spatiotemporal check-in view*. Under the disentangled-channel framework, we perform hypergraph convolutional encoding for each view separately to obtain view-specific independent representations. We then adopt semantic attention to adaptively fuse different explicit views, thereby obtaining the explicit representation.

Multi-view Explicit Hypergraph Construction. Given that real-world social relations are typically sparse and locally clustered (*i.e.*, a user's friends are not necessarily directly connected to each other), we employ ego-hyperedges anchored at the central user to better match the social structure in reality. In particular, we construct the *friendship ego-hypergraph* $\mathcal{H}_F$. This hypergraph does not assume that any two users within the same hyperedge are directly connected by friendship. It treats a user's ego-network as a unified semantic set. Let $\mathcal{E}_f \subseteq \mathcal{U} \times \mathcal{U}$ denote the observed friendship edge set, and let $\mathcal{N}_u^F$ be the first-order friend set of user u:

$$\mathcal{N}_u^F = \{i \mid (u, i) \in \mathcal{E}_f\}. \tag{6}$$

Based on this, we construct a friendship hyperedge e_u^F centered at user u as follows:

$$e_u^F = \{u\} \cup \mathcal{N}_F(u). \tag{7}$$

This anchor-centered hypergraph introduces a structural inductive bias via "node→hyperedge→node" propagation. By aggregating neighbor features into a latent hyperedge representation, it encodes the local friendship context. This context is then propagated back to refine node embeddings, enabling the model to capture the shared influence of a friendship circle without requiring dense interconnections among friends.

In addition, to capture higher-order co-occurrence relations of users under the same time and to complement the social graph from a spatiotemporal behavioral perspective, we construct a spatiotemporal check-in hypergraph $\mathcal{H}_C$. Specifically, we treat each *time-slice-location* context as a hyperedge and connect all users who check in under the same context to this hyperedge, thereby explicitly modeling users' spatiotemporal routine patterns. Given a check-in record (u, p, t), we first discretize the timestamp t into a time bucket $\tau(t)$ (partitioned into four time periods), and define the corresponding spatiotemporal context key as:

$$\mathrm{key}(u, p, t) = p \parallel \tau(t). \tag{8}$$

Then we aggregate all check-in records according to $\mathrm{key}(\cdot)$, where each distinct $p \parallel \tau(t)$ corresponds to a hyperedge $e \in \mathcal{E}_{\mathrm{chk}}$, which contains all users who checked in within that *time-slice-location* context. Accordingly, for the check-in view, we construct a binary incidence matrix $\mathbf{H}^{chk} \in \mathbb{R}^{N\times|\mathcal{E}_{\mathrm{chk}}|}$:

$$\mathbf{H}^{chk}_{u,e} = 1 \Leftrightarrow u \in e. \tag{9}$$

This construction offers a clear advantage in characterizing *collective synchrony*: conventional graph modeling typically decomposes a group co-occurrence of size k into $\binom{k}{2}$ pairwise connections, which inevitably loses the holistic semantics of "a group appearing together under the same spatiotemporal context". In contrast, the hypergraph $\mathcal{H}_C$ models such group co-occurrence directly as a single hyperedge and preserves it as an indivisible higher-order unit, thereby more faithfully representing event-level group interaction structures.

Disentangled Hypergraph Representation Learning. Inspired by [Ma *et al.*, 2019; Lai *et al.*, 2024], we propose a multi-view disentanglement framework to address the feature entanglement issue caused by single-vector embeddings in social relationship inference. This framework reconstructs user representations from a holistic space into K mutually independent semantic channels, each characterizing behavioral factors driven by different latent social interests. Specifically, for any explicit view v, we adopt the classical two-step hypergraph propagation paradigm to perform message passing independently within each channel, obtaining the view-specific output.

We first aggregate channel-wise node representations to the hyperedge side, and then propagate them back to the node side, with normalization by hyperedge degrees and node degrees. For each channel $k \in \{1, \ldots, K\}$, let $\mathbf{X}_k = \mathbf{X}_{:,k,:} \in \mathbb{R}^{N\times D}$ denote the slice of the input tensor on the k-th channel. The propagation is formulated as:

$$\mathbf{X}'_k = (\mathbf{D}_v)^{-1}\mathbf{H}(\mathbf{D}_e)^{-1}(\mathbf{H})^\top \mathbf{X}_k, \tag{10}$$

where $\mathbf{X}'_k$ is the updated representation propagated back to the node side, $\mathbf{D}_v$ and $\mathbf{D}_e$ are the diagonal degree matrices of nodes and hyperedges.

By stacking all $\mathbf{X}'_k$ along the channel dimension, we obtain the output of this view $\mathbf{Z} \in \mathbb{R}^{N\times K\times D}$, which is then stacked along the view dimension as $\mathbf{Z}_{exp} = \mathrm{stack}\big(\mathbf{Z}^f, \mathbf{Z}^{chk}\big) \in \mathbb{R}^{N\times V\times K\times D}$. During the propagation process, we first perform independent intra-view propagation and then carry out inter-view fusion, so that different explicit relations do not interfere with each other.

4.3 Implicit Structure Learning

Given that users may be social friends yet have no observable physical interactions/co-occurrence traces, we introduce implicit graph learning to dynamically recover potential links in the representation space.

Sparse Graph Learner. We first apply a shared linear mapping $\phi(\cdot)$ to the features of each channel, and perform intra-channel ℓ_2 normalization to obtain unit vectors:

$$\hat{\mathbf{X}} = \mathrm{Norm}\big(\phi(\mathbf{X})\big), \tag{11}$$

where $\hat{\mathbf{X}} \in \mathbb{R}^{N\times K\times d'}$, d' is the dimension of the projected hidden space, and $\mathrm{Norm}(\cdot)$ denotes ℓ_2 normalization along the last dimension.

Then, we measure the geometric proximity between users via the channel-averaged cosine similarity as follows:

$$s_{ij} = \frac{1}{K}\sum_{k=1}^{K} \langle \hat{x}_{i,k}, \hat{x}_{j,k} \rangle, \tag{12}$$

where $\hat{x}_{i,k} \in \mathbb{R}^{d'}$ denotes the normalized feature vector of user i in the k-th channel, and s_{ij} represents the aggregate similarity score between user i and user j.

To select the most meaningful connections from complex relations, we adopt a sparsification strategy that combines Top-κ neighbor retrieval with γ-threshold filtering:

$$\tilde{a}_{ij} = \begin{cases} s_{ij}, & j \in \mathcal{N}_\kappa(i) \ \wedge \ s_{ij} > \epsilon, \\ 0, & \text{otherwise}, \end{cases} \tag{13}$$

where $\tilde{a}_{ij}$ represents the sparse edge weight in the learned implicit graph, $\mathcal{N}_\kappa(i)$ denotes the candidate neighbor set consisting of the top-κ most similar users to user i, and ϵ serves as a threshold to prune weak connections.

Finally, we normalize the rows to generate the implicit adjacency matrix $\mathbf{A}_{\text{imp}}$:

$$a_{ij} = \frac{\tilde{a}_{ij}}{\sum_{j'} \tilde{a}_{ij'} + \delta}, \tag{14}$$

where a_{ij} represents the element in the i-th row and j-th column of the matrix, denoting the normalized edge weight that indicates the implicit influence of user j on user i, and δ is a constant.

Lightweight Disentangled Graph Convolution. After obtaining $\mathbf{A}_{\text{imp}}$, we perform implicit collaborative signal propagation. Since the channels are disentangled, we apply the graph convolution independently for each channel k with a shared weight matrix. The holistic implicit representation is obtained by stacking the convolution outputs:

$$\mathbf{Z}_{\text{imp}} = \text{Stack}\left(\{\sigma\left(\mathbf{A}_{\text{imp}}\mathbf{X}_k\mathbf{W}_2\right)\}_{k=1}^{K}\right), \tag{15}$$

where $\mathbf{Z}_{\text{imp}} \in \mathbb{R}^{N\times K\times D}$ is the implicit representation tensor, $\mathbf{W}_2 \in \mathbb{R}^{D\times D}$ is the shared learnable weight matrix, and $\sigma(\cdot)$ denotes the LeakyReLU activation function.

It is worth emphasizing that the "disentanglement" of our model does not rely on complicated routing. Instead, we explicitly constrain inter-channel correlations via an independence loss, encouraging different channels to capture semantically complementary and weakly correlated interest factors, achieving stable representation disentanglement while maintaining computational efficiency.

4.4 Adaptive Channel-level Fusion

The explicit hypergraph view aggregates *observable structured interactions* based on established friendship relations and physical facts, whereas the implicit graph view is designed to mine *latent global affinities* in the absence of explicit physical interactions. To dynamically fuse these two complementary sources of information, we further introduce an adaptive gating mechanism.

Different from the standard practice of learning a single weight per user, we implement a fine-grained channel-level gating mechanism. For each user u and channel k, the model adaptively learns a fusion coefficient to balance the explicit and implicit views:

$$\begin{aligned} \boldsymbol{\alpha} &= \sigma\Big(\big[\mathbf{Z}_{\text{exp}} \,\|\, \mathbf{Z}_{\text{imp}}\big]\mathbf{W}_g + b_g\Big), \\ \mathbf{Z}_{\text{final}} &= \boldsymbol{\alpha} \odot \mathbf{Z}_{\text{exp}} + (\mathbf{1} - \boldsymbol{\alpha}) \odot \mathbf{Z}_{\text{imp}}, \end{aligned} \tag{16}$$

where $\mathbf{W}_g \in \mathbb{R}^{2D\times 1}$ is the learnable weight matrix, and $b_g \in \mathbb{R}^1$ is a bias vector.

4.5 Multi-objective Optimization

To effectively capture user preferences while ensuring high-quality disentangled representations, we employ a joint optimization strategy.

Social Link Optimization. To accurately infer potential social relationships, we employ the Bayesian Personalized Ranking (BPR) loss to optimize the structural ranking of user pairs. This objective maximizes the margin between the valid social connections and the non-existing ones:

$$\mathcal{L}_{\text{link}} = -\frac{1}{|\mathcal{E}|}\sum_{(u,v^+)\in\mathcal{E}} \ln\sigma\big(s(u,v^+) - s(u,v^-)\big), \tag{17}$$

where $\mathcal{E}$ denotes the set of ground-truth friendship edges, and (u, v^+) represents an observed social tie. For each positive pair, we sample a negative user v^- who has no interaction with u. $s(u, v)$ denotes the predicted affinity score based on the final fused embedding, and $\sigma(\cdot)$ is the sigmoid function.

Independence Constraint. Merely partitioning representations into K channels does not guarantee semantic independence. To prevent feature redundancy, we impose a Gram-matrix-based orthogonality constraint on the channel representations, averaged over the friendship, check-in, and implicit views:

$$\mathcal{L}_{\text{indep}} = \mathrm{Mean}_{V\in\{\text{f},\text{chk},\text{imp}\}}\left(\mathbb{E}_u\left\|\mathbf{Z}_u^V(\mathbf{Z}_u^V)^\top - \mathbf{I}_K\right\|_F\right), \tag{18}$$

where $\mathbf{Z}_u^V \in \mathbb{R}^{K\times D}$ denotes the user representation from a specific view, and $\mathbf{I}_K$ is the $K \times K$ identity matrix.

Minimizing the Frobenius norm $\|\cdot\|_F$ forces the off-diagonal elements of the correlation matrix to zero, thereby encouraging K latent channels to be pairwise independent.

Multi-view Contrastive Alignment. To bridge the semantic gap between explicit and implicit signals, we introduce a channel-aware contrastive loss. This loss aligns the k-th

Dataset	City		Country	
	KL	SP	CL	JP
#users	7,339	5,725	5,822	8,716
#check-ins	794,472	423,288	1,074,512	2,655,899
#POIs	84,553	48,851	107,123	343,195
#friends	33,172	18,411	97,305	67,672

Table 1: Statistics of datasets

channel of explicit view with the corresponding k-th channel of implicit view while pushing away mismatched channels:

$$\mathcal{L}_{\text{con}} = \mathbb{E}_{(a,b)}\mathbb{E}_u \ \text{CE}\left(\frac{\tilde{\mathbf{V}}_u^{(a)}(\tilde{\mathbf{V}}_u^{(b)})^\top}{\tau},\ \mathbf{y}_{\text{diag}}\right), \tag{19}$$

where $\tilde{\mathbf{V}}$ denotes the ℓ_2-normalized channel embeddings, and τ is the temperature parameter controlling the sharpness of the distribution. $\text{CE}(\cdot)$ represents the Cross-Entropy loss.

The target $\mathbf{y}_{\text{diag}}$ enforces a same-index alignment (i.e., the diagonal entries of the similarity matrix are maximized), ensuring semantic consistency across views.

Joint Objective. The final objective function is a weighted combination of the aforementioned losses:

$$\mathcal{L} = \mathcal{L}_{\text{link}} + \lambda_{ind}\mathcal{L}_{\text{indep}} + \lambda_{con}\mathcal{L}_{\text{con}}, \tag{20}$$

where λ_{ind} and λ_{con} are hyperparameters that control the strength of the independence constraint and the contrastive alignment, respectively.

5 Experiments

In this section, we perform model evaluation to investigate the effectiveness of our DHISL and baseline methods on four real-world datasets. Our experiments aim to answer the research questions as follows:

- **RQ1:** What is the performance of our proposed DHISL as compared to various SOTA social relationship inference methods?
- **RQ2:** How do the key components contribute to the performance?
- **RQ3:** How do the key hyperparameters influence the performance of the proposed DHISL?

5.1 Datasets

We use four real datasets from Foursquare [Yang *et al.*, 2019], namely Kuala Lumpur (KL), São Paulo (SP), Chile (CL), and Japan (JP), to evaluate the performance of methods. Note that we do not discard inactive users with few check-ins. Table 1 shows our data statistics.

5.2 Baselines

We compare our DHISL with two categories of baselines: (1) Pairwise Graph Learning Methods: MVMN [Zhang *et al.*, 2020]; MSC-LBSN [Trung *et al.*, 2022]; HG-MAE [Tian *et al.*, 2023]; SRINet [Qin *et al.*, 2023]; MR-GAN [Qin *et al.*, 2024]. (2) Hypergraph Neural Networks: LBSN2Vec++ [Yang *et al.*, 2020]; HHGNN [Li *et al.*, 2022]; HMGCL [Li *et al.*, 2023b]; H^{3}GNN[Li *et al.*, 2025c].

5.3 Experiment Settings and Evaluation Metrics

In our experiments, for each dataset, we sample 25% of the friendships in the social network as the training set, which is used to construct social relations in the friendship hypergraph. We then randomly sample an additional 5% of the friendships as the validation set, and use the remaining 70% as the test set. Derived from [Yu *et al.*, 2018; Qin *et al.*, 2024], in our evaluation, we employ the widely used Area Under the ROC Curve (ROCAUC) and the Area Under the Precision-Recall Curve (PRAUC) to evaluate the performance of various methods.

5.4 Experiment Results

Performance Comparison (RQ1)

The comparative results of social relationship inference on four real-world datasets are presented in Table 2. As can be observed, DHISL achieves optimal performance across all evaluation metrics. Compared with the SOTA baselines, DHISL achieves maximum improvements of 7.54% in ROCAUC and 8.60% in PRAUC.

Across all datasets, DHISL achieves an average improvement of 6.34% in ROCAUC and an average gain of 6.39% in PRAUC, mainly because the multi-channel disentanglement mechanism separates intrinsic social signals from complex friendship relationships and temporal preferences. Due to the inherent sparsity and partial observability of mobility data, all models face the problem of missing latent social connections. However, DHISL consistently outperforms SOTA methods, even on the SP dataset with sparse interactions, it achieves a 5.75% improvement in ROCAUC. This demonstrates that the proposed implicit structure learning module effectively recovers unobserved relationships.

Although H^{3}GNN and DHISL both use hypergraph modeling to capture high-order correlations, DHISL averages ROCAUC and PRAUC higher by 9.61% and 9.37%, respectively. This is because while traditional hypergraph methods have the advantage of modeling group interactions, they cannot avoid the semantic entanglement of diverse user behaviors, indicating that explicit disentanglement with independence constraints can significantly improve performance, which further confirms the necessity of separating independent latent factors for accurate social relationship inference.

Ablation Study (RQ2)

To analyze the effectiveness of the key components in our model, we conduct the following ablation studies. We design three variants to verify the contribution of different modules to model performance, consisting of:

- ***w/o Disen*** removes the independence and consistency constraints, and sets the number of channels to 1.
- ***w/o Implicit*** removes the implicit structure learning branch.
- ***w/o Feat*** removes the feature-guided user representation initialization strategy. It uses random initialization.

The results of ablation studies on SP and CL datasets are summarized in Figure 2. From the results, three conclusions are made as follows:

Method	KL		SP		CL		JP	
	ROCAUC	PRAUC	ROCAUC	PRAUC	ROCAUC	PRAUC	ROCAUC	PRAUC
MVMN	0.7468	0.7786	0.7308	0.7615	0.7926	0.8060	0.7382	0.7483
LBSN2Vec++	0.7128	0.7562	0.7258	0.7395	0.7756	0.7884	0.7454	0.7528
MSC-LBSN	0.7493	0.7896	0.7482	0.7861	0.7935	0.7986	0.7698	0.7723
HHGNN	0.7585	0.7943	0.7525	0.7910	0.8056	0.8037	0.7952	0.8063
HGMAE	0.7856	0.7550	0.7709	0.7858	0.7974	0.7935	0.8042	0.8158
HMGCL	0.7754	0.8063	0.7732	0.8085	0.8037	0.8080	0.8029	0.8176
SRINet	0.7782	0.7994	0.7805	0.8093	0.8152	0.8113	0.8016	0.8224
MRGAN	<u>0.8137</u>	<u>0.8371</u>	0.7836	0.8189	<u>0.8563</u>	<u>0.8533</u>	<u>0.8537</u>	<u>0.8604</u>
H^3GNN	0.7847	0.8051	<u>0.7844</u>	<u>0.8218</u>	0.8151	0.8136	0.8253	0.8411
Ours	**0.8694**	**0.8813**	**0.8295**	**0.8632**	**0.9215**	**0.9267**	**0.8982**	**0.9174**
Improvement	6.85%	5.28%	5.75%	5.04%	7.54%	8.60%	5.21%	6.62%

Table 2: Performance comparison of all models on four real-world datasets. Improvement denotes the improvement of the best results compared with the second-best results. The best results are shown in **bold** and the best among baselines is <u>underlined</u>.

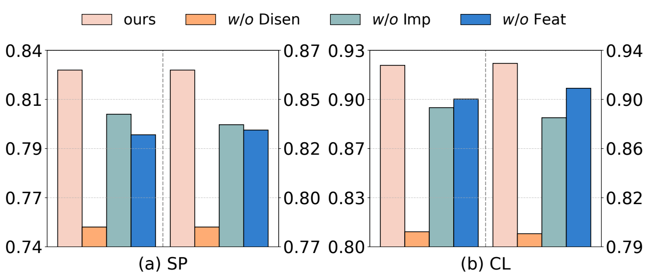

Figure 2: Ablation study under different datasets (the dashed line separates the ROCAUC on the left from the PRAUC on the right).

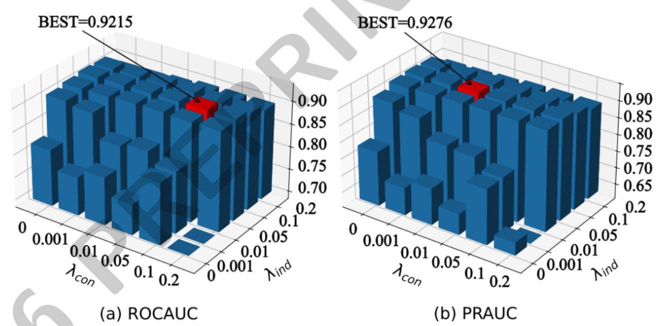

Figure 3: Parameter sensitivity analysis of λ_{ind} and λ_{con}, taking the CL dataset as an example.

(1) Three variants are consistently worse than DHISL in terms of ROCAUC and PRAUC, indicating the effectiveness and necessity of the key components. (2) With regard to ***w/o Disen***, the performance experiences a significant decline compared to that of DHISL (the ROCAUC drops from 0.82 to 0.75 on the SP dataset). This phenomenon underscores the pivotal role of the multi-channel disentanglement mechanism in social relationship inference. (3) ***w/o Feat*** performs worse than DHISL, specifically showing a distinct gap in ROCAUC. This suggests that the employment of feature-guided initialization effectively captures the intrinsic spatiotemporal characteristics and mobility dynamics of users, providing a more informative starting point than random initialization. (4) The gap between DHISL and ***w/o Implicit*** shows that implicit structure learning is critical. By dynamically recovering latent connections, DHISL can better complement sparse explicit interactions, further improving model inference accuracy.

Parameter Sensitivity Analysis (RQ3)

To answer **RQ3**, we investigate the sensitivity of λ_{ind} and λ_{con} by performing a grid search with values in $\{0, 0.001, 0.01, 0.05, 0.1, 0.2\}$. As visualized in Figure 3, the performance graph demonstrates that λ_{ind} is the primary driver of representation quality. A sharp phase transition occurs when λ_{ind} increases from 0 to 0.01, where ROCAUC surges from approximately 0.824 to over 0.914. These results support our hypothesis that without explicit orthogonality constraints, latent channels tend to collapse into redundant features, severely limiting expressiveness. Regarding λ_{con}, its efficacy is strictly conditional on the degree of disentanglement. In the entangled regime ($\lambda_{ind} < 0.01$), strongly enforcing consistency proves detrimental, as it likely propagates noise and leads to over-smoothing of indistinct embeddings. However, in the disentangled regime ($\lambda_{ind} \geq 0.01$), λ_{con} serves as a crucial regularizer. By ensuring the consistency of the diverse semantics captured from different perspectives, the robustness of the model is enhanced. Ultimately, the model achieves peak performance (ROCAUC 0.9215) within the "optimal region" where $\lambda_{ind} = 0.05$ and $\lambda_{con} = 0.1$, confirming that DHISL effectively leverages the synergy between disentangled feature learning and cross-view consistency.

6 Conclusion

We propose DHISL for mobility social relation inference. It employs a dual-branch architecture, synergizing explicit hypergraphs with implicit structure learning to capture latent homophily, while using multi-channel disentanglement to resolve semantic entanglement. Experiments on four datasets show DHISL outperforms SOTA methods, improving ROCAUC by 6.34% and PRAUC by 6.39%.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China under Grant Nos. 62176243, and 62302120, the Natural Science Foundation of Shandong Province under Grant No. ZR2024MF128, China Postdoctoral Science Foundation under Grant No. 2025M773472, and the Postdoctoral Fellowship Program of CPSF under Grant No. GZC20252288.

References

[Backes *et al.*, 2017] Michael Backes, Mathias Humbert, Jun Pang, and Yang Zhang. walk2friends: Inferring social links from mobility profiles. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1943–1957, 2017.

[Baik and Yoo, 2025] Dayoung Baik and Jaejun Yoo. Dynamic-aware spatio-temporal representation learning for dynamic mri reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 174–184. Springer, 2025.

[Cao *et al.*, 2025] Chang Cao, Han Li, Yulong Wang, Rui Wu, and Hong Chen. Adversarial training for graph convolutional networks: stability and generalization analysis. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 4797–4805, 2025.

[Cho *et al.*, 2011] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090, 2011.

[Dai *et al.*, 2022] Shaojie Dai, Yanwei Yu, Hao Fan, and Junyu Dong. Spatio-temporal representation learning with social tie for personalized poi recommendation. *Data Science and Engineering*, 7(1):44–56, 2022.

[Darraz *et al.*, 2025] Nossayba Darraz, Ikram Karabila, Anas El-Ansari, Nabil Alami, and Mostafa El Mallahi. Integrated sentiment analysis with bert for enhanced hybrid recommendation systems. *Expert Systems with Applications*, 261:125533, 2025.

[Desai *et al.*, 2017] Deshana Desai, Harsh Nisar, and Rishabh Bhardawaj. Role of temporal diversity in inferring social ties based on spatio-temporal data. In *Proceedings of the 4th ACM IKDD Conferences on Data Sciences*, pages 1–8, 2017.

[Fan *et al.*, 2017] Chao Fan, Yiding Liu, Junming Huang, Zhihai Rong, and Tao Zhou. Correlation between social proximity and mobility similarity. *Scientific reports*, 7(1):11975, 2017.

[Fu *et al.*, 2024] Chaofan Fu, Pengyang Yu, Yanwei Yu, Chao Huang, Zhongying Zhao, and Junyu Dong. Mhgcn+: multiplex heterogeneous graph convolutional network. *ACM Transactions on Intelligent Systems and Technology*, 15(3):1–25, 2024.

[Gao *et al.*, 2020] Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph learning: Methods and practices. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2548–2566, 2020.

[Lai *et al.*, 2024] Yantong Lai, Yijun Su, Lingwei Wei, Tianqi He, Haitao Wang, Gaode Chen, Daren Zha, Qiang Liu, and Xingxing Wang. Disentangled contrastive hypergraph learning for next poi recommendation. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*, pages 1452–1462, 2024.

[Li *et al.*, 2022] Yongkang Li, Zipei Fan, Jixiao Zhang, Dengheng Shi, Tianqi Xu, Du Yin, Jinliang Deng, and Xuan Song. Heterogeneous hypergraph neural network for friend recommendation with human mobility. In *Proceedings of the 31st ACM international conference on information & knowledge management*, pages 4209–4213, 2022.

[Li *et al.*, 2023a] Xiao Li, Li Sun, Mengjie Ling, and Yan Peng. A survey of graph neural network based recommendation in social networks. *Neurocomputing*, 549:126441, 2023.

[Li *et al.*, 2023b] Yongkang Li, Zipei Fan, Du Yin, Renhe Jiang, Jinliang Deng, and Xuan Song. Hmgcl: Heterogeneous multigraph contrastive learning for lbsn friend recommendation. *World Wide Web*, 26(4):1625–1648, 2023.

[Li *et al.*, 2024] Shuzhe Li, Wei Chen, Bin Wang, Chao Huang, Yanwei Yu, and Junyu Dong. Mcn4rec: Multi-level collaborative neural network for next location recommendation. *ACM Transactions on Information Systems*, 42(4):1–26, 2024.

[Li *et al.*, 2025a] Xiang Li, Chaofan Fu, Zhongying Zhao, Guangjie Zheng, Chao Huang, Yanwei Yu, and Junyu Dong. Dual-channel multiplex graph neural networks for recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 37(6):3327–3341, 2025.

[Li *et al.*, 2025b] Xiang Li, Changsheng Shui, Zhongying Zhao, Junyu Dong, and Yanwei Yu. Multi-channel hypergraph contrastive learning for matrix completion. *ACM Transactions on Information Systems*, 44(1):1–27, 2025.

[Li *et al.*, 2025c] Yongkang Li, Zipei Fan, and Xuan Song. Heterogeneous hyperbolic hypergraph neural network for friend recommendation in location-based social networks. *ACM Transactions on Knowledge Discovery from Data*, 19(3):1–29, 2025.

[Liu *et al.*, 2022] Hai Liu, Chao Zheng, Duantengchuan Li, Zhaoli Zhang, Ke Lin, Xiaoxuan Shen, Neal N Xiong, and Jiazhang Wang. Multi-perspective social recommendation method with graph representation learning. *Neurocomputing*, 468:469–481, 2022.

[Ma *et al.*, 2019] Jianxin Ma, Peng Cui, Kun Kuang, Xin Wang, and Wenwu Zhu. Disentangled graph convolutional networks. In *International conference on machine learning*, pages 4212–4221. PMLR, 2019.

[Pham *et al.*, 2013] Huy Pham, Cyrus Shahabi, and Yan Liu. Ebm: an entropy-based model to infer social strength from spatiotemporal data. In *Proceedings of the 2013 ACM SIGMOD international conference on management of data*, pages 265–276, 2013.

[Pham *et al.*, 2016] Huy Pham, Cyrus Shahabi, and Yan Liu. Inferring social strength from spatiotemporal data. *ACM Transactions on Database Systems (TODS)*, 41(1):1–47, 2016.

[Qin *et al.*, 2023] Guangming Qin, Lexue Song, Yanwei Yu, Chao Huang, Wenzhe Jia, Yuan Cao, and Junyu Dong. Graph structure learning on user mobility data for social relationship inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4578–4586, 2023.

[Qin *et al.*, 2024] Guangming Qin, Jianpeng Qi, Bin Wang, Guiyuan Jiang, Yanwei Yu, and Junyu Dong. Multi-relational graph attention network for social relationship inference from human mobility data. In *33rd International Joint Conference on Artificial Intelligence (IJCAI'24)*, pages 2315–2323. International Joint Conferences on Artificial Intelligence Organization, 2024.

[Sun *et al.*, 2019] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pages 1441–1450, 2019.

[Tian *et al.*, 2023] Yijun Tian, Kaiwen Dong, Chunhui Zhang, Chuxu Zhang, and Nitesh V Chawla. Heterogeneous graph masked autoencoders. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 9997–10005, 2023.

[Trung *et al.*, 2022] Huynh Thanh Trung, Tong Van Vinh, Nguyen Thanh Tam, Jun Jo, Hongzhi Yin, and Nguyen Quoc Viet Hung. Learning holistic interactions in lbsns with high-order, dynamic, and multi-role contexts. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):5002–5016, 2022.

[Wang *et al.*, 2011] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1100–1108, 2011.

[Wang *et al.*, 2014] Hongjian Wang, Zhenhui Li, and Wang-Chien Lee. Pgt: Measuring mobility relationship using personal, global and temporal factors. In *2014 IEEE International Conference on Data Mining*, pages 570–579. IEEE, 2014.

[Wang *et al.*, 2019] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *The world wide web conference*, pages 2022–2032, 2019.

[Wang *et al.*, 2025] Qiwei Wang, Dandan Lin, Wenqing Lin, and Ziming Wu. Frog: Effective friend recommendation in online games via modality-aware user preferences. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2822–2826, 2025.

[Wu *et al.*, 2019] Yongji Wu, Defu Lian, Shuowei Jin, and Enhong Chen. Graph convolutional networks on user mobility heterogeneous graphs for social relationship inference. In *IJCAI*, pages 3898–3904, 2019.

[Wu *et al.*, 2023] Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. Personalized news recommendation: Methods and challenges. *ACM Transactions on Information Systems*, 41(1):1–50, 2023.

[Xia *et al.*, 2022] Lianghao Xia, Chao Huang, Yong Xu, Jiashu Zhao, Dawei Yin, and Jimmy Huang. Hypergraph contrastive collaborative filtering. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval*, pages 70–79, 2022.

[Xu and Chen, 2025] Long Xu and Honghui Chen. Fa-gcl: Feature-augmented graph contrastive learning method. *Neural Networks*, page 108045, 2025.

[Yan *et al.*, 2025] Ziang Yan, Xingyu Zhao, Hanqing Ma, Wei Chen, Jianpeng Qi, Yanwei Yu, and Junyu Dong. Correlation-attention masked temporal transformer for user identity linkage using heterogeneous mobility data. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pages 12999–13007, 2025.

[Yang *et al.*, 2019] Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudre-Mauroux. Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In *The world wide web conference*, pages 2147–2157, 2019.

[Yang *et al.*, 2020] Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudré-Mauroux. Lbsn2vec++: Heterogeneous hypergraph embedding for location-based social networks. *IEEE Transactions on Knowledge and Data Engineering*, 34(4):1843–1855, 2020.

[Yang *et al.*, 2022a] Song Yang, Jiamou Liu, and Kaiqi Zhao. Getnext: Trajectory flow map enhanced transformer for next poi recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on research and development in information retrieval*, pages 1144–1153, 2022.

[Yang *et al.*, 2022b] Yuhao Yang, Chao Huang, Lianghao Xia, Yuxuan Liang, Yanwei Yu, and Chenliang Li. Multi-behavior hypergraph-enhanced transformer for sequential recommendation. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 2263–2274, 2022.

[Yu *et al.*, 2018] Yanwei Yu, Hongjian Wang, and Zhenhui Li. Inferring mobility relationship via graph embedding. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–21, 2018.

[Zhang *et al.*, 2020] Wei Zhang, Xin Lai, and Jianyong Wang. Social link inference via multiview matching network from spatiotemporal trajectories. *IEEE transactions on neural networks and learning systems*, 34(4):1720–1731, 2020.