Heterogeneous Neighborhood-Enhanced Graph Contrastive Learning for Recommendation

Lei Sang, Chi Zhang, Maohao Huang, Lin Mu, Yiwen Zhang, and Xindong Wu, *Fellow, IEEE*

***Abstract*—Heterogeneousself-supervised graph learning has gained considerable attention in recommender systems for its ability to capture diverse semantic and structural relationships in real-world data. Contrastive learning enhances representation learning by maximizing agreement between positive pairs while distinguishing negative ones in cross-views. However, two key challenges remain: 1) noise, such as false negatives, that degrades representation quality; and 2) lack of cross-view alignment causes biased and inconsistent representations. To address these challenges, we propose heterogeneous neighborhood-enhanced graph contrastive learning for recommendation (HNGCL). HNGCL ensures cross-view consistency through alignment and uniformity losses, encouraging embeddings that are both well-aligned and uniformly distributed across views, thereby enhancing generalization and discriminative power. To mitigate noise, HNGCL introduces a neighborhood-enhanced strategy that integrates collaborative neighbors to generate high-quality positive pairs, reducing false negatives and suppressing noise propagation. By leveraging heterogeneous graph structures and cross-view contrastive learning, HNGCL effectively captures intricate semantic and structural patterns, producing robust feature representations. Extensive experiments on real-world datasets demonstrate that HNGCL significantly outperforms state-of-the-art methods in recall and normalized discounted cumulative gain (NDCG), showcasing its effectiveness in overcoming these challenges and advancing recommendation performance. Our code for the model implementation is available at https://github.com/zhangchi107/HNGCL.**

***Index Terms*—Contrastive learning, heterogeneous graph neural networks (GNNs), recommender system.**

I. Introduction

RECOMMENDER systems play a critical role in connecting users to a vast array of information, products, and services, particularly in domains such as e-commerce and social media [1], [2]. Among the various approaches, collaborative filtering (CF) [3] has long served as a cornerstone in personalized recommendation tasks [4], [5]. However, traditional CF methods primarily rely on shallow, single-hop user–item interactions [6], which makes them susceptible to well-known issues such as the cold-start problem [7] and data sparsity [8]. To address these limitations, recent advances have introduced graph neural networks (GNNs) into recommender systems, effectively leveraging graph-based structural information to exploit higher-order connectivity between users and items [9]. While GNN-based methods have shown significant improvements in recommendation quality, they still largely operate on homogeneous user–item bipartite graphs, which restricts their ability to capture the rich, heterogeneous information inherent in real-world datasets.

To overcome the limitations of homogeneous graphs and leverage rich semantic relations in real-world data, heterogeneous information networks (HINs) have emerged as a powerful tool for recommendation tasks. HINs comprise multiple types of nodes and edges, enabling the modeling of complex relationships across users, items, and auxiliary entities such as categories or tags [10]. A distinguishing feature of HINs lies in their ability to capture high-order semantics through meta-paths, which represent meaningful composite relations between different types of nodes [11]. For example, in a movie recommendation scenario illustrated in Fig. 1, the relationship between two users can be inferred through "user–movie–user" (UMU) and "user–movie–actor–movie–user" (UMAMU), which reveal the semantic relations of: 1) sharing the same movie; and 2) being connected through actors who have performed in both movies. These paths uncover hidden connections, enriching the semantic context for recommendations. Building upon HINs, recent studies have introduced heterogeneous graph neural networks (HGNNs) to aggregate information along meta-paths, refining node embeddings to capture both structural and semantic richness [12]. Models such as HAN [13] use attention mechanisms to adaptively learn the importance of different nodes and meta-paths, demonstrating strong performance in recommendation scenarios. Additionally, incorporating social network information into meta-paths can further enhance the semantic richness captured by HINs. Social ties, such as friendships or shared interests, provide additional contextual signals that complement user–item interactions, enabling more accurate and robust recommendations [14], [15].

Building on the success of HGNNs and advances in contrastive learning from computer vision and related fields [16], recent research has explored the integration of contrastive learning into HGNNs in recommendation tasks. For example, HGCL [17] applies cross-view contrastive learning by

Received 20 January 2025; revised 16 May 2025 and 3 July 2025; accepted 7 July 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62206002 and Grant 62272001; and in part by Anhui Provincial Natural Science Foundation under Grant 2208085QF195 and Grant 2308085MF221. *(Corresponding author: Yiwen Zhang.)*

Lei Sang, Chi Zhang, Maohao Huang, Lin Mu, and Yiwen Zhang are with the School of Computer Science and Technology, Anhui University, Hefei 230601, China (e-mail: sanglei@ahu.edu.cn; zhangchi@stu.ahu.edu.cn; e23301222@stu.ahu.edu.cn; mulin@ahu.edu.cn; zhangyiwen@ahu.edu.cn).

Xindong Wu is with the Key Laboratory of Knowledge Engineering with Big Data (the Ministry of Education of China), Hefei University of Technology, Hefei 230601, China (e-mail: xwu@hfut.edu.cn).

Digital Object Identifier 10.1109/TCSS.2025.3588471

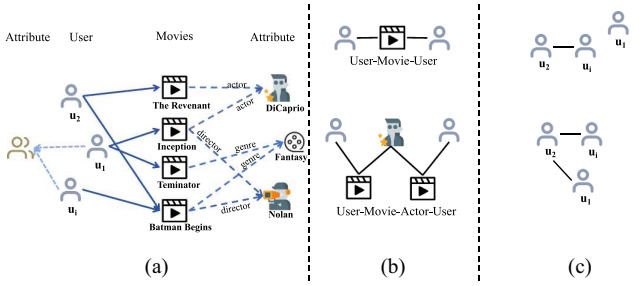

Fig. 1. (a) Our general idea of introducing HINs to model user historical information. (b) Some meta-paths of HINs. (c) Subgraph based on meta-paths.

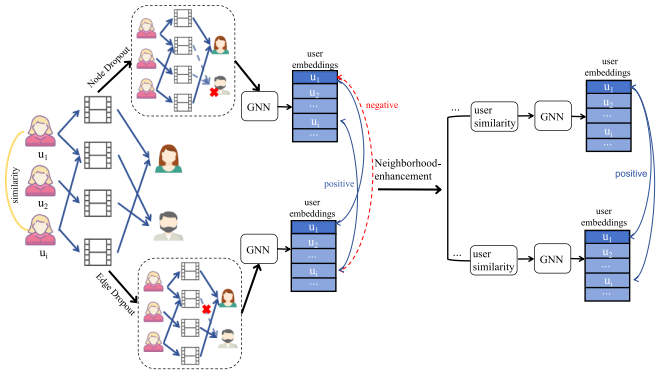

Fig. 2. Left: The original adjacency graph is augmented into two views via random perturbations, which are independently encoded by GNNs to generate node embeddings. Standard contrastive loss constructs positive pairs such as u_1-u_1 and u_i-u_i, and treats pairs like u_1-u_i as negative samples, which may mistakenly treat collaborative neighbors u_i as false negatives. Right: Neighborhood enhancement alleviates this by identifying similar neighbors as additional positive pairs.

leveraging meta-path-based semantic views to align user and item embeddings from different perspectives, thereby capturing richer semantic relationships. This approach enforces consistency across semantic spaces and leads to more robust and expressive representations. Despite these advances, conventional contrastive learning methods commonly rely on random data augmentation to generate positive and negative pairs for GNN-based embedding learning. Such augmentation strategies often produce two augmented graph views through node and edge dropout, which are then encoded by GNNs. For instance, as illustrated in Fig. 2, positive pairs (e.g., u_1-u_1 and u_i-u_i) and negative pairs (e.g., u_1-u_i) are selected accordingly. Yet, this process may separate anchor nodes (user nodes and item nodes) from their true collaborative neighbors (interacted neighbors and nearest neighbors), causing interest-aligned nodes to be mistakenly treated as false negatives and ultimately degrading representation quality. In Fig. 2, u_1 and u_i not only share common movie preferences, but also exhibit potential similarity through their affinity for the same actor. However, they may be incorrectly treated as a negative sample pair, thereby introducing noise. Although contrastive learning methods have demonstrated strong performance, they still face the following two drawbacks.

1) *CH1: How to Mitigate the Noise Generated by Contrastive Learning?* Most existing contrastive learning augmentation strategies, such as node dropout or edge perturbation, primarily focus on generating diverse views of the graph to improve generalization. However, these approaches may push anchor nodes away from collaborative neighbors, the resulting representations can be distorted by semantic noise, which negatively impacts the quality of learned embeddings and the overall recommendation performance [16], [18]. Therefore, it is crucial to develop augmentation mechanism that explicitly preserves the integrity of meaningful neighborhood structures and maintains the proximity of anchor nodes to their collaborative neighbors within the embedding space.
2) *CH2: How to Achieve Cross-View Consistency?* Existing research on cross-view contrastive learning in recommendation primarily focuses on leveraging multiview information to enhance representation quality and overall model performance. However, most works overlook two key aspects: alignment, which ensures that representations of the same entity from different views are similar, and uniformity, which ensures that representations are evenly distributed in the embedding space [12]. Alignment ensures that semantically similar representations from different views are closely clustered, while uniformity enhances the expressiveness of embeddings and preserves information by ensuring they are well-distributed over the hypersphere, thereby maximizing mutual information. The lack of proper alignment and uniformity can lead to cross-view inconsistencies and introduce representational biases, undermining the generalization and discriminative power of the learned embeddings and ultimately degrading recommendation accuracy. Therefore, achieving both alignment and uniformity within the hypersphere is crucial for mitigating contrastive noise and preserving the semantic consistency of learned representations.

In response to the two aforementioned challenges, we propose a novel model called heterogeneous neighborhood-enhanced graph contrastive learning (HNGCL). Specifically, to tackle *CH1*—the mitigation of noise introduced by contrastive learning—we design a cross-view neighborhood-enhanced mechanism based on HINs, which explicitly distinguishes between different node categories and improves recommendation accuracy [19]. This mechanism considers both interacted neighbors and nearest neighbors of an anchor node as positive samples, preventing the model from inadvertently pushing away collaborative neighbors during contrastive optimization. Furthermore, we adopt a variant of the InfoNCE loss [20] to construct more discriminative positive pairs across dual views and to alleviate the negative impact of improper negative sample selection. For *CH2*, to achieve cross-view consistency, we incorporate both alignment loss and uniformity loss to guide the optimization of the model. These two objectives, defined on the unit hypersphere, play complementary roles in enhancing the generalizability and discriminability of learned representations in contrastive learning [21], [22], [23]. Specifically, alignment encourages semantically similar instances to be mapped to proximate locations in the embedding space, while uniformity

ensures that representations are evenly distributed on the hypersphere, thereby preserving maximum information and avoiding collapse. By simultaneously optimizing these two losses, HNGCL effectively mitigates contrastive noise and preserves the semantic integrity of learned representations, leading to improved recommendation performance. We summarize the main contributions as follows.

1) We propose a novel and effective contrastive learning framework named HNGCL for recommendation tasks. By leveraging multiview information and jointly optimizing alignment and uniformity losses, HNGCL significantly improves the robustness and accuracy of learned representations.
2) To minimize the interference of noise and enhance the influence of positive sample pairs, we introduce a neighborhood-enhanced mechanism and evaluate the effectiveness of our method.
3) We conduct top-K recommendation evaluation experiments on three real-world datasets. Experimental results exhibit that HNGCL consistently outperforms state-of-the-art baselines, including both GNN-based and contrastive learning recommendation models.

II. Related Work

A. GNN-Based Recommendation

GNNs have emerged as powerful tools in recommender systems due to their ability to model complex user–item interactions via graph structures. Unlike traditional collaborative filtering and matrix factorization methods, which are often limited in capturing higher-order connectivity, GNN-based models leverage the topology of user–item graphs to learn more expressive representations [24]. Empirical evidence suggests that they deliver superior performance across various recommender systems [25], [26]. For instance, graph convolutional networks (GCNs) [27], [28] are widely adopted for their capacity to capture multihop dependencies. Building on this, NGCF [29] models high-order connectivity through layered neighborhood aggregation, while LightGCN [30] simplifies the traditional GCN by removing feature transformation and nonlinearities, focusing purely on the propagation of user–item interaction signals. This lightweight design has inspired numerous subsequent works and laid the foundation for integrating GNNs with contrastive learning in recommendation tasks [31]. Further developments, such as DGCF [32] and SVD-GCN [33], demonstrate the influence of graph topology modeling on recommendation effectiveness. These models introduce disentangled and spectral designs, respectively, to better exploit the structural semantics of user–item graphs. Collectively, these studies highlight the significance of GNN architectures in learning from interaction patterns and provide a structural underpinning for more advanced learning mechanisms such as contrastive learning.

B. Contrastive Learning for Recommendation

Contrastive learning has recently become a prominent technique for representation learning, aiming to pull semantically similar samples closer and push dissimilar ones apart in the embedding space. Originally successful in computer vision and natural language processing [20], [34], contrastive learning has been increasingly adopted in recommendation to address challenges such as data sparsity and noisy supervision. For example, SGL [35] introduces self-supervised contrastive signals by constructing multiple perturbed views of a user–item graph, substantially improving performance on sparse datasets and long-tail recommendations. NCL [31] further refines positive pair construction by incorporating latent semantic neighbors, thereby expanding the scope of meaningful contrastive signals. However, many early works assume that nodes not directly connected are negative, leading to semantic noise. To mitigate this, NESCL [16] treats collaborative neighbors as positive samples, effectively mitigating the false negative problem in user–item graphs by leveraging the inherent structure of collaborative filtering. Recently, contrastive learning has been extended to heterogeneous recommendation scenarios. For example, GCLHANRec [18] introduces hybrid noise-aware augmentation strategies to distinguish between true and false negatives, thereby improving robustness in heterogeneous graph-based recommendation. HGCL [17] integrates heterogeneous information network (HIN) semantics into the contrastive learning framework by employing meta-network structures, enabling personalized knowledge transfer and adaptive contrastive enhancement to address the problem of data sparsity. RecDCL [36] further improves contrastive learning in this context by combining batch-wise and feature-wise contrastive strategies, effectively tackling both data sparsity and representational redundancy. These advancements demonstrate the growing recognition of contrastive learning as a robust paradigm for improving recommendation performance under limited supervision and complex graph structures in recommender systems.

C. Heterogeneous GNNs

Heterogeneous information networks (HINs) capture richer semantics by modeling multiple types of nodes and edges, which better reflect real-world user–item interactions. Early works such as HERec [37] integrate HIN embeddings into matrix factorization, showing significant performance gains in sparse settings. HGAT [38] introduces attention mechanisms to weigh the importance of different node types, capturing fine-grained semantic relations. Building on these ideas, heterogeneous graph neural networks (HGNNs) [39], [40], [41], [42] have become a key focus area [43], combining meta-path semantics with deep graph models. For example, HAN [13] employs hierarchical attention to select important node-level and path-level features, while HeCo [12] leverages self-supervised contrastive signals across semantic views to enhance representation quality. Despite their promise, HGNN-based methods still face challenges in effectively aligning multitype semantics and scaling to large graphs. Moreover, existing contrastive learning models often focus on homogeneous user–item graphs, leaving the rich structure of HINs underutilized. This motivates the need for frameworks that can integrate heterogeneous semantics with robust contrastive learning mechanisms.

In addition to traditional HIN-based models, recent studies have leveraged hypergraph neural networks to model complex higher-order relationships in recommendation. Khan et al. [44] proposed a framework for session-based social recommendations using heterogeneous hypergraph structures. Similarly, Khan et al. [45] introduced a model capturing diverse relations such as item and user similarity through hypergraph motifs. By incorporating attentive aggregation mechanisms, these approaches build on the strengths of HGNNs, demonstrating their potential to handle complex interactions and improve recommendation performance.

III. Preliminaries

A. Graph Collaborative Filtering for Recommendation

Graph-based collaborative filtering models primarily rely on users' historical behaviors to uncover user preferences and capture the characteristics of items they have interacted with [29], [46]. We define $\mathcal{U}$ and $\mathcal{I}$ as the sets of users and items respectively. The observed user–item interactions are represented by a binary matrix $\mathbf{R} \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{I}|}$, where each entry $\mathbf{R}_{ui} = 1$ indicates that user $u \in \mathcal{U}$ has interacted with item $i \in \mathcal{I}$, and $\mathbf{R}_{ui} = 0$ otherwise. The adjacency matrix $\mathbf{A}$ of the bipartite user–item interaction graph is defined as

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{R} \\ \mathbf{R}^\top & \mathbf{0} \end{bmatrix} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{I}|)\times(|\mathcal{I}|+|\mathcal{U}|)}$ is the adjacency matrix used to learn low-dimensional embedding representations for users and items based on historical interactions, enabling the model to predict the likelihood of future user–item interactions.

In the graph collaborative filtering framework, multilayer graph convolutional networks (GCNs) are employed to capture high-order collaborative signals from the user–item interaction graph and encode them into low-dimensional embedding representations. Each GCN layer aggregates feature information from neighboring nodes and updates node features in a parameterized manner. The layer-wise propagation rule is defined as

$$\mathbf{E}^{(l+1)} = \sigma\left(\hat{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\hat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{E}^{(l)}\mathbf{W}^{(l)}\right) \tag{2}$$

where $\mathbf{E}^{(l)}$ denotes the node embedding matrix at the lth layer, and $\mathbf{W}^{(l)}$ is the trainable weight matrix of that layer. The adjacency matrix with added self-loops is defined as $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, where $\mathbf{A}$ is the original adjacency matrix and $\mathbf{I}$ is the identity matrix. The symmetric normalization term $\hat{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\hat{\mathbf{D}}^{-\frac{1}{2}}$ is used to prevent numerical instability and over-smoothing during message propagation. And $\sigma(\cdot)$ denotes a nonlinear activation function.

After the propagation process, we obtain the final node representations, which are then used to compute predicted preference scores between users and items [47]. Let $\mathbf{z}_u \in \mathbb{R}^d$ and $\mathbf{z}_i \in \mathbb{R}^d$ denote the final embedding vectors of user u and item i, respectively, where d is the dimensionality of the latent space. These embeddings are directly derived from the output of the feature propagation mechanism described in (2), where the multilayer GNN iteratively aggregates neighborhood information to learn meaningful representations for each node in the user–item graph. The predicted preference score of user u for item i is calculated via the inner product

$$s(u, i) = \mathbf{z}_u^\top \mathbf{z}_i \tag{3}$$

where a higher score $s(u, i)$ indicates a greater likelihood of interaction.

TABLE I
Notation and Description

Notation	Description
$\mathcal{G}$	Heterogeneous graph
$\mathcal{V}$, $\mathcal{E}$	The set of nodes and edges in the graph
$\mathcal{U}$, $\mathcal{I}$	The set of users and the set of items
$\mathcal{A}$, $\mathcal{R}$	The set of node types and edge types
$\mathbf{R} \in \mathbb{R}^{\lvert\mathcal{U}\rvert\times\lvert\mathcal{I}\rvert}$	User-item interaction matrix
Φ	Meta-path
ϕ	Node type mapping function
ψ	Edge type mapping function
$\mathcal{G}_{\mathcal{V}}^{\Phi}$	Meta-path-based subgraph over nodes of a specific type
$\mathcal{N}$	The union of users and items
$\mathcal{N}_i$	The set of nodes that have interacted with node i
$\mathcal{S}_i$	Nearest neighbors of node i in the embedding space
τ	Temperature parameter in contrastive loss
γ	Hyperparameter controlling the uniformity term
β	Hyperparameter for cross-view contrastive loss
α	Weight of the overall loss function

B. Definitions

Heterogeneous information networks (HINs) incorporate diverse semantic information and exhibit complex structural characteristics. The subsequent discussion formally defines several key concepts in heterogeneous graphs. The important notations used throughout this article are summarized in Table I.

Heterogeneous Graph (HG): A heterogeneous graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{R}, \phi, \psi)$, where $\mathcal{V}$ and $\mathcal{E}$ denote the sets of nodes and edges, respectively. The node type mapping function is $\phi : \mathcal{V} \to \mathcal{A}$, and the edge type mapping function is $\psi : \mathcal{E} \to \mathcal{R}$. Here, $\mathcal{A}$ and $\mathcal{R}$ represent the sets of node types and edge types, respectively. The condition $|\mathcal{A}| + |\mathcal{R}| > 2$ ensures the heterogeneity of the graph.

Meta-Path: Meta-paths are widely used to capture complex semantic relationships between nodes in a heterogeneous graph. Given a heterogeneous graph $\mathcal{G}$, a meta-path Φ is defined as a sequence in the form $\mathcal{A}_1 \xrightarrow{\mathcal{R}_1} \mathcal{A}_2 \xrightarrow{\mathcal{R}_2} \cdots \xrightarrow{\mathcal{R}_{i-1}} \mathcal{A}_i$, which describes a composite relation between nodes v_1 and v_i.

As illustrated in the heterogeneous graph constructed on the DoubanMovie dataset in Fig. 1, there exist various types of nodes (e.g., User, Movie, Actor, Director, and Genre) and edges (e.g., interactions between users and movies, connections between users via actors). Generally, diverse meta-paths can reflect different types or levels of semantic relevance between entities. For instance, users u_2 and u_i connected via the meta-path UMU share the same movie, while the meta-path UMAMU indicates that users have engaged with movies acted by the same actor, thereby capturing a higher-order semantic relation.

Meta-Path-Based Subgraph: A meta-path is defined as a sequence of edge types from the edge type set $\mathcal{R}$, which describes

a composite relation between different node types. Given a node type $\mathcal{A}_t \in \mathcal{A}$, let $\mathcal{V}_{\mathcal{A}_t}$ denote the set of nodes of type $\mathcal{A}_t$. For each node $v \in \mathcal{V}_{\mathcal{A}_t}$, we define $\mathcal{E}_v^{\Phi}$ as the set of edges that connect v to other nodes via the meta-path Φ. By traversing all nodes in $\mathcal{V}_{\mathcal{A}_t}$ and aggregating their meta-path-based connections, we collect the union of all such edges as $\mathcal{E} = \bigcup_{v \in \mathcal{V}_{\mathcal{A}_t}} \mathcal{E}_v^{\Phi}$. We then construct a meta-path-based subgraph $\mathcal{G}_{\mathcal{V}_{\mathcal{A}_t}}^{\Phi} = (\mathcal{V}_{\mathcal{A}_t}, \mathcal{E})$, which captures the structural semantics defined by the meta-path Φ among nodes of type $\mathcal{A}_t$. Taking Fig. 1 as an example, we specify the meta-path Φ as UMU, and the node type $\mathcal{A}_t$ as "Movie". By traversing all movie nodes and collecting their connections through the meta-path "UMU", we obtain the subgraph $\mathcal{G}_U^{UMU}$, which reflects coauthorship relationships between books.

IV. Proposed Model

In this section, we present the proposed heterogeneous neighborhood-enhanced graph contrastive learning (HNGCL) model, whose overall framework is illustrated in Fig. 3. The model consists of the following key components: 1) contrastive view construction; 2) alignment and uniformity representation optimization; and 3) a heterogeneous neighborhood-enhanced mechanism. We first obtain node embeddings from two distinct views. These embeddings are then optimized by aligning and uniformly distributing them. Specifically, alignment encourages embeddings of similar representations to be close in the latent space, thereby enhancing the model's ability to capture personalized preferences and item characteristics. In contrast, uniformity enforces a well-dispersed embedding distribution across the hypersphere, which reduces representational redundancy and improves the generalization capability of the model. To further enrich the positive sample space and mitigate the effects of data sparsity, we introduce a neighborhood-enhanced strategy that incorporates anchor nodes, interacted neighbors, and nearest neighbors as positive pairs for contrastive learning. This design ensures that both explicit interactions and implicit semantic similarities are fully leveraged. Finally, we apply a contrastive loss function to jointly optimize representation learning and the recommendation objective. This enables the model to learn meaningful user–item representations that are both semantically discriminative and structurally consistent.

A. Contrastive View Construction

We first introduce the heterogeneous cross-view contrastive learning framework. The essence of cross-view contrastive learning lies in integrating information from multiple structural or semantic sources to learn more robust and generalizable node representations. We leverage this principle to construct two complementary views: *user–item interaction view*, which captures the explicit collaborative signals between users and items, and *meta-path-based view*, which captures higher-order semantic relationships through heterogeneous structures.

User–Item Interaction View: Based on the observed interactions, we construct the matrix $\mathcal{A}$ as described in Section III-A, and define the user–item bipartite graph $\mathcal{G}_{ui} = \{(u, i) | u \in U, i \in I\}$, which captures direct interaction relationships between users and items. Specifically, we first construct a symmetric normalized adjacency matrix $\hat{\mathcal{A}}$ from $\mathcal{A}$

$$\hat{\mathcal{A}} = \mathbf{D}^{-\frac{1}{2}} \mathcal{A} \mathbf{D}^{-\frac{1}{2}} \tag{4}$$

where $\mathbf{D}$ is the diagonal degree matrix of $\mathcal{A}$.

We then perform message propagation to aggregate neighborhood information over L layers. The embeddings are iteratively updated as follows:

$$H^{(l+1)} = \hat{\mathcal{A}} H^{(l)} \tag{5}$$

where $H^{(l)}$ denotes the embeddings at the lth layer, and the initial embeddings $H^{(0)}$ are the input user and item features $H^{(0)} = \left[H_u^{(0)}; H_i^{(0)}\right]$, with $H_u^{(0)} \in \mathbb{R}^{|U| \times d}$ and $H_i^{(0)} \in \mathbb{R}^{|I| \times d}$.

After L layers of propagation, we obtain the embeddings for users and items by summing the outputs of all layers

$$H = \sum_{l=0}^{L} H^{(l)}. \tag{6}$$

Then split H into the final user and item embeddings $E_u \in \mathbb{R}^{|U| \times d}$ and $E_i \in \mathbb{R}^{|I| \times d}$, respectively. This view captures direct collaborative signals and is efficient in modeling large-scale interactions.

Data Preprocessing for Meta-path-Based View: Before constructing the meta-path-based view, we preprocess the original heterogeneous graph to filter unsuitable data and generate meaningful subgraphs. Specifically, given a set of candidate meta-paths $\{\Phi_1, \Phi_2, \ldots, \Phi_p\}$, we extract subgraphs for users and items corresponding to each meta-path. For instance, in Fig. 1 the meta-path "UMU" connects users who have interacted with the same movie, while the meta-path "UMAMU" captures higher-order semantic relationships via actor. During preprocessing, nodes without any connections (i.e., isolated nodes) are removed to ensure that the resulting subgraphs are semantically meaningful. This preprocessing step ensures that the generated views are representative and free of noise, thereby improving the robustness of the model.

Meta-Path-Based View: To model complex user preferences and item attributes, we leverage multihop connections between various node and edge types (i.e., meta-paths) to model the preferences of user u and the attributes of item i, constructing a meta-path-based view as the contrastive view. The selection of meta-paths is crucial for capturing semantic relationships in heterogeneous graphs, and we choose widely used meta-paths tailored to dataset characteristics. Additionally, our semantic-level attention mechanism dynamically adjusts the importance of each meta-path, enhancing robustness and generalizability. Given a set of candidate meta-paths $\{\Phi_1, \Phi_2, \ldots, \Phi_p\}$, we obtain the corresponding user and item subgraphs, denoted as $\mathcal{G}_u^{\Phi_k}$ and $\mathcal{G}_i^{\Phi_k}$, respectively. Let $\mathcal{N}_u^{\Phi_k}$ and $\mathcal{N}_i^{\Phi_k}$ represent the meta-path-based neighbors of user u and item i. We apply a graph convolutional operation over each subgraph to aggregate the semantic information

$$h_u^{\Phi_k} = \sum_{v \in \mathcal{N}_u^{\Phi_k}} \frac{1}{\sqrt{|\mathcal{N}_u^{\Phi_k}||\mathcal{N}_v^{\Phi_k}|}} h_v$$

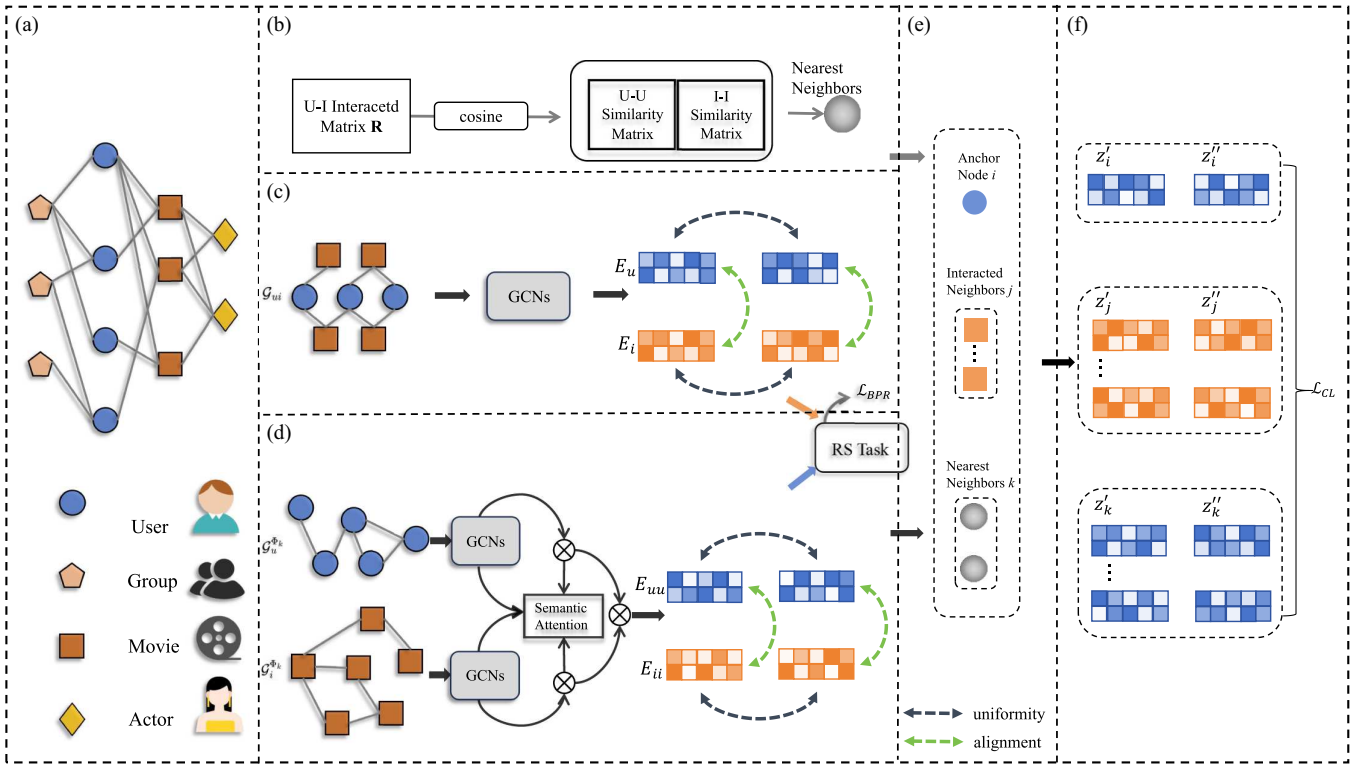

Fig. 3. Overall framework of heterogeneous neighborhood-enhanced graph contrastive learning (HNGCL). (a) Heterogeneous graph is constructed based on the DoubanMovie dataset. (b) Nearest neighbors of nodes are identified using cosine similarity computed from the user–item interaction matrix. (c) User–item interaction view is generated by extracting user–movie interaction information, resulting in the embedding matrices $E_u, E_i \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{I}|)\times D}$, which are then optimized via alignment and uniformity objectives. (d) By aggregating meta-path information to generate the meta-path-based view, it obtains the embeddings matrix $E_{uu}, E_{ii} \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{I}|)\times D}$, then aligns and uniforms obtained embeddings. (e) Collaborative neighbors are collected using both the interaction matrix and the similarity matrix from step (c) to serve as data augmentation. (f) Before performing cross-view contrastive learning, the embeddings of all users and items are indexed to construct positive sample pairs.

$$h_i^{\Phi_k} = \sum_{j \in \mathcal{N}_i^{\Phi_k}} \frac{1}{\sqrt{|\mathcal{N}_i^{\Phi_k}||\mathcal{N}_j^{\Phi_k}|}} h_j \tag{7}$$

where h_v and h_j denote the feature embeddings of neighbor nodes v and j.

After obtaining specific meta-path embeddings $\{h_u^{\Phi_1}, \ldots, h_u^{\Phi_p}\}$ and $\{h_i^{\Phi_1}, \ldots, h_i^{\Phi_p}\}$, we apply a semantic-level attention mechanism to adaptively fuse them into final representations E_{uu} and E_{ii}

$$\begin{aligned} E_{uu} &= \sum_{k=1}^{p} \beta_{\Phi_k} \cdot h_u^{\Phi_k} \\ E_{ii} &= \sum_{k=1}^{p} \beta_{\Phi_k} \cdot h_i^{\Phi_k} \end{aligned} \tag{8}$$

where β_{Φ_k} denotes the attention weight indicating the importance of meta-path Φ_k for the current node. These attention weights are computed as

$$\beta_{\Phi_k} = \frac{\exp\left(w^\top \tanh\left(W_s h^{\Phi_k}\right)\right)}{\sum_{l=1}^{p} \exp\left(w^\top \tanh\left(W_s h^{\Phi_l}\right)\right)} \tag{9}$$

where W_s and w are learnable parameters of the attention network, and h^{Φ_k} is the meta-path-level embedding for either user or item.

As illustrated in Fig. 2, semantically related nodes, such as two users u_1 and u_2, without direct interactions (i.e., they have not watch the same movies, just be connected through semantic associations via actor), can be connected via a meta-path like "UMAMU", capturing high-order semantic similarity that cannot be observed directly from user–item interactions.

Meta-paths, such as "UMU" and "UMAMU" in Fig. 2, provide a powerful means of capturing high-order semantic relationships in heterogeneous graphs. However, as the graph size and node type diversity increase, defining meaningful meta-paths becomes increasingly complex and time-intensive. Furthermore, the selection of meta-paths often relies heavily on domain-specific knowledge, which may limit the model's generalizability to other datasets or application domains.

By combining data preprocessing, meta-path selection, and the semantic-level attention mechanism, our framework effectively integrates heterogeneous semantics and collaborative signals. To address these challenges, we identify a set of meta-paths tailored to the characteristics of the datasets, which are widely used in recommender systems and effectively capture user preferences and behavioral patterns. Additionally, the use of a semantic-level attention mechanism allows our model to adaptively weigh the importance of each meta-path, reducing the impact of suboptimal meta-path

selection and enhancing the robustness of learned representations. By integrating the meta-path-based view with the user–item interaction view, our framework effectively combines heterogeneous semantics with collaborative signals, reducing dependency on domain-specific meta-path design and improving generalizability.

By combining data preprocessing, meta-path selection, and the semantic-level attention mechanism, our framework effectively integrates heterogeneous semantics and collaborative signals.

B. Optimization of Alignment and Uniformity

To effectively integrate the user–item interaction view and the meta-path-based view, we incorporate two complementary objectives: *alignment* and *uniformity*. These objectives jointly guide the model to learn user and item representations that are both semantically consistent across views and well-distributed in the embedding space. Specifically, we align the representations of the same user or item across different views, encourages the embeddings of the same user or item from different views to be close in the representation space, thereby bridging the semantic gap between structural and semantic information. Meanwhile, the uniformity objective promotes a uniform distribution of representations on the hypersphere, which helps mitigate over-clustering and enhances generalization. Alignment encourages closeness between user and item embeddings across views, while uniformity promotes a balanced distribution to avoid over-clustering. These objectives prevent the model from overfitting to specific user or item groups and contribute to fairer and more robust recommendation performance.

The alignment loss is defined as follows:

$$\mathcal{L}_{\text{align}} = \mathop{\mathbb{E}}_{(u,i)\sim p_{\text{pos}}} ||\widetilde{E}_u - \widetilde{E}_i||^2 + \mathop{\mathbb{E}}_{(u,i)\sim p_{\text{pos}}} ||\widetilde{E}_{uu} - \widetilde{E}_{ii}||^2 \tag{10}$$

where $p_{\text{pos}}(\cdot)$ denotes the distribution of positive user–item pairs, $\widetilde{E}_u$ and $\widetilde{E}_i$ are the normalized embeddings from the interaction view, and $\widetilde{E}_{uu}$ and $\widetilde{E}_{ii}$ are from the meta-path-based view.

To prevent the learned embeddings from collapsing into a narrow subspace, we employ the following uniformity loss:

$$\begin{aligned}\mathcal{L}_{\text{uniform}} = &\log \mathop{\mathbb{E}}_{u,u'\sim p_{\text{user}}} e^{-2||\widetilde{E}_u - \widetilde{E}'_u||^2}/2 \\ &+ \log \mathop{\mathbb{E}}_{i,i'\sim p_{\text{item}}} e^{-2||\widetilde{E}_i - \widetilde{E}'_i|^2|}/2 \\ &+ \log \mathop{\mathbb{E}}_{u,u'\sim p_{\text{user}}} e^{-2||\widetilde{E}_{uu} - \widetilde{E}'_{uu}||^2}/2 \\ &+ \log \mathop{\mathbb{E}}_{i,i'\sim p_{\text{item}}} e^{-2||\widetilde{E}_{ii} - \widetilde{E}'_{ii}||^2}/2\end{aligned} \tag{11}$$

where $p_{\text{user}}(\cdot)$ and $p_{\text{item}}(\cdot)$ denote the sampling distributions over users and items, respectively.

By jointly optimizing the alignment and uniformity losses, the model is encouraged to produce embeddings that are not only semantically aligned across views but also well-dispersed, thus improving both expressiveness and generalization of the user and item representations.

Unlike traditional collaborative filtering (CF) models that rely on negative sampling—where the selection and quality of negative samples significantly influence the discriminative power of the learned representations—the proposed loss functions operate solely on positive sample pairs. This design avoids the risk of mistakenly treating semantically relevant neighbors as false negatives, thereby preserving potentially valuable semantic relationships that may reflect users' latent interests.

C. Neighborhood-Enhanced Mechanism

To expand the source of positive samples, we adopt a novel strategy: neighborhood-enhanced mechanism. This strategy enriches positive samples composed of anchor nodes and their directly interacted nodes, further incorporating their top-ranked collaborative neighbors to enrich the positive sample modeling. Since the interacted neighbors N_i of the anchor node i can be easily obtain through user–item interaction graph $\mathcal{G}_{ui}$, we focus on introducing how to generate its nearest neighbors in HINs. We use cosine similarity to measure the proximity between nodes. For example, in the case of item–item similarity, the similarity between items i and j is defined as

$$\text{sim}(i,j) = \frac{|\mathcal{N}_i \cap \mathcal{N}_j|}{\sqrt{|\mathcal{N}_i| \cdot |\mathcal{N}_j|}} \tag{12}$$

where $\mathcal{N}_i$ and $\mathcal{N}_j$ represent the sets of users who have interacted with items i and j, respectively. The numerator counts the number of users who interacted with both items, and the denominator normalizes the score.

By selecting the top-K most similar nodes as collaborative neighbors for each anchor node, we construct a richer set of positive sample pairs. These pairs consist of the anchor and its collaborative neighbors sampled from $\mathcal{S}_i$. This neighborhood-enhanced mechanism effectively enriches the self-supervised signal, allowing the model to capture fine-grained semantic relationships and collaborative patterns that reflect user preferences. As a result, it improves the model's ability to learn personalized and robust representations.

Contrastive loss quantifies the disparity in similarity between positive and negative sample pairs, guiding the model to bring similar samples closer in the embedding space while pushing dissimilar ones farther apart. Therefore, the effectiveness of contrastive learning heavily depends on the proper construction of positive and negative sample pairs. We design two different views to learn richer data and robust feature representations. Through the neighborhood-enhanced mechanism, we include both anchor nodes and collaborative neighbors in the modeling of positive sample pairs, thereby obtaining richer semantic relationships and latent interaction patterns.

We use a variant of the InfoNCE loss, which is based on the principle of noise contrastive estimation, to evaluate the similarity between positive sample pairs. Our goal is to minimize the separation between anchor nodes and their collaborative neighbors. For each anchor node i, we construct two views of its embedding: $\mathbf{z}'_i$ and $\mathbf{z}''_i$, where $\mathbf{z}'_i \in \mathbb{R}^d$ and $\mathbf{z}''_i \in \mathbb{R}^d$ are obtained from two different perspectives. Similarly, for each interacted neighbor $j \in \mathcal{N}_i$ and each collaborative neighbor $k \in \mathcal{S}_i$,

we obtain their representations $\mathbf{z}'_j, \mathbf{z}''_j$ and $\mathbf{z}'_k, \mathbf{z}''_k$, respectively. Unlike traditional approaches that rely on randomly sampled negatives—which may introduce false negatives—we compute similarity scores against all nodes in the batch. This avoids the risk of treating semantically similar nodes as negatives and better preserves meaningful relationships.

The overall contrastive loss is defined as

$$\begin{aligned}\mathcal{L}_{\text{CL}} = &- \sum_{i\in\mathcal{N}} \log \frac{\exp(\mathbf{z}'_i \cdot \mathbf{z}''_i/\tau)}{\sum_{m\in\mathcal{N}} \exp(\mathbf{z}'_i \cdot \mathbf{z}''_m/\tau)} \\ &- \sum_{i\in\mathcal{N}} \log \sum_{j\in\mathcal{N}_i} \frac{\exp(\mathbf{z}'_j \cdot \mathbf{z}''_i/\tau)}{\sum_{m\in\mathcal{N}} \exp(\mathbf{z}'_j \cdot \mathbf{z}''_m/\tau)} \\ &- \sum_{i\in\mathcal{N}} \log \sum_{k\in\mathcal{S}_i} \frac{\text{sim}(i,k)\cdot\exp(\mathbf{z}'_k \cdot \mathbf{z}''_i/\tau)}{\sum_{m\in\mathcal{N}} \exp(\mathbf{z}'_k \cdot \mathbf{z}''_m/\tau)} \end{aligned} \tag{13}$$

where τ is a temperature hyperparameter that controls the sharpness of the distribution, and $\text{sim}(i,k)$ is a similarity weighting term between anchor node i and its collaborative neighbor k.

By jointly optimizing the contrastive loss over these three components, the model effectively learns more expressive and semantically aligned representations.

D. Overall Loss Functions of Our Proposed Model

In this section, we present the overall loss function of our proposed model, which integrates four components: contrastive loss, Bayesian personalized ranking (BPR) loss [48], alignment loss and uniformity loss. Each component serves a specific optimization objectives purpose, collectively driving the model to learn expressive, consistent, and personalized user–item representations.

BPR Loss: Given the fact that most data in recommendations is implicit feedback (such as user clicks, browsing, or purchasing behavior), ranking-based objectives are more suitable for capturing user preferences. We employ the BPR loss function to refine the preference ranking of users for items

$$\mathcal{L}_{\text{BPR}} = -\frac{1}{|\mathcal{N}|} \sum_{(u,i)\in\mathcal{N}} \log\left(\sigma\left(s(u,i) - s(u,i^-)\right)\right) \tag{14}$$

where $\sigma(\cdot)$ denotes the sigmoid function, $s(u,i)$ is the predicted matching score between user u and item i, and i^- is a negative item sampled from the set of items that user u has not interacted with.

Therefore, the final loss function is presented below

$$\mathcal{L}_{\text{HNGCL}} = \alpha \cdot \left(\mathcal{L}_{\text{CL}} + \beta\left(\mathcal{L}_{\text{align}} + \gamma \cdot \mathcal{L}_{\text{uniform}}\right)\right) + \mathcal{L}_{\text{BPR}} \tag{15}$$

where α is a hyperparameter that balances the contribution of the self-supervised learning component with the supervised BPR loss. The hyperparameters β and γ control the relative importance of the alignment and uniformity regularization terms within the self-supervised objective.

The setting of hyperparameters will be discussed in the experimental section.

TABLE II
DATA SPARSITY AND STATISTICAL DETAILS

Data	User	Item	Interaction	Sparsity
Yelp	19 239	14 284	198 397	99.91%
DoubanBook	13 024	22 347	792 062	99.73%
DoubanMovie	13 367	12 677	1 068 278	99.37%

V. EXPERIMENT

In this section, we conduct experiments on three real-world recommendation datasets to evaluate the performance of HNGCL and investigate the following research questions.

RQ1: How does HNGCL perform in recommendation tasks compared with various baselines methods?
RQ2: How does each key component of HNGCL contribute to recommendation performance?
RQ3: How robust is HNGCL when exposed to noisy interactions?
RQ4: How does HNGCL combat data sparsity?
RQ5: How do distinct parameter configurations impact the performance of HNGCL?

The experimental setup plays a crucial role in ensuring the replicability and reliability of the research. Following, we initiate the discussion with the datasets, baseline models, evaluation metrics, and implementation. Then, we address each of the aforementioned questions in turn.

A. Experimental Setup

1) Datasets and Metrics: Three publicly available datasets from different domains are utilized in our experiments: Yelp [49], DoubanBook [50], and DoubanMovie [51]. These datasets vary significantly in size and sparsity levels and are summarized in Table II. These datasets are derived from real-world applications, such as Yelp for restaurant reviews and DoubanBook and DoubanMovie for book and movie preferences, respectively. As widely used benchmarks in the recommendation system community, they provide a reliable proxy for simulating user behaviors and preferences in specific domains. For evaluation metrics, we employ two common metrics, Recall@K and NDCG@K, with K respectively designated as 10 and 20. These are standard metrics designed to assess the relevance and ranking quality of recommendations. A high normalized discounted cumulative gain (NDCG) score, for instance, indicates that the system effectively prioritizes relevant items, which is critical for meeting user expectations.

2) Baselines: To evaluate the performance of HNGCL, we compare it against several classical collaborative filtering models. Additionally, since our proposed loss function is built upon GNN architectures, we also include several state-of-the-art GNN-based collaborative filtering models for comparison. Descriptions of the baseline models are provided below.

HAN [13] is a GNN-based model tailored for heterogeneous graphs. It employs both node-level and semantic-level attention

mechanisms to aggregate neighborhood information and learns node embeddings using meta-path-based attention encoders.

LightGCN [30] is a GCN-based collaborative filtering model that simplifies traditional GCNs by removing feature transformation and nonlinear activation. It updates node embeddings solely through neighborhood aggregation.

DGCF [32] performs disentangled representation learning for users and items by modeling an intention-aware interaction graph, which helps capture diverse user preferences.

BPR [48] is a classical latent factor-based CF model that optimizes pairwise ranking for implicit feedback, and performs well on large-scale recommendation tasks.

HeCo [12] is a self-supervised contrastive learning framework for heterogeneous graphs. It constructs dual views based on the network schema and meta-paths to perform cross-view contrast, enabling collaborative supervision between views.

SMIN [52] is a self-supervised model for social recommendation. It leverages HGNNs to capture rich heterogeneous semantics and improves representation learning through multitask contrastive objectives.

NCL [31] is a contrastive learning framework that incorporates prototype-based supervision. It aligns users and items with their corresponding prototypes to enhance representation quality.

HGCL [17] is a heterogeneous graph contrastive learning method tailored for recommender systems. It captures semantic signals across different relations and enhances node embeddings through cross-view contrastive learning.

RecDCL [36] introduces a dual contrastive learning framework that combines batch-wise and feature-wise contrastive strategies to reduce representation redundancy and improve robustness.

NESCL [16] enhances contrastive learning in recommendation by selecting collaborative neighbors of anchor nodes as positives and designing two supervised contrastive loss functions to improve recommendation accuracy.

Parameter Settings: The experiments are implemented using PyTorch. For baseline models, we follow the parameter settings reported in the original articles and perform additional fine-tuning to achieve optimal performance. For each model, we fix the embedding dimension at 128, and set the batch size to 1024. We employ the early stopping strategy to avert overfitting, where training is terminated if the performance on the validation set (measured by Recall@10 and Recall@20) does not improve for 20 consecutive epochs. For our proposed HNGCL model, the parameter settings are as follows: the L_2 regularization coefficient is fixed at 0.0001 for all three datasets, the learning rate is tuned within the range [0.0005, 0.001], the number of GNN and GCN layers is chosen from [1], [2], [3], [4], the hyperparameter α is varied within the range [0.01, 0.2], and the hyperparameter β and γ are tuned within the range [0.0001, 1].

B. Performance Comparison (RQ1)

The comparative performance of HNGCL relative to nine benchmarks over the past 5 years on three datasets is detailed in Table III. The results prove that HNGCL outperforms all baseline methods in Recall@10, NDCG@10, Recall@20, and NDCG@20 on all datasets. Compared with suboptimal model, HNGCL achieves the most remarkable performance improvement on the Yelp dataset, with gains of 18.77%, 7.74%, 13.00%, and 11.82% in the four metrics, respectively.

The superior performance of HNGCL can be attributed to the following three key components:

1) *Contrastive Learning Based on HGNNs:* HNGCL introduces a contrastive view construction strategy that fully encodes semantic information through meta-paths. By integrating the user–item interaction view with a meta-path-based semantic view, it effectively captures the heterogeneous semantics of different node and edge types in HINs, thereby enhancing recommendation performance.
2) *Optimizing Alignment and Uniformity:* Contrastive loss enhances representation learning, however, it may inadvertently weaken the model's ability to preserve the proximity of collaborative neighbors in the embedding space. To address this, HNGCL explicitly enforces alignment and uniformity between embeddings obtained from distinct views, ensuring consistent semantic understanding across views and tasks, and promotes the discovery of latent relational structures, ultimately improving the quality of learned representations.
3) *Neighborhood-Enhanced Mechanism:* To enrich the set of positive samples, we employ the neighborhood-enhanced mechanism, which incorporates collaborative neighbors of these anchor nodes into the positive sample set. This strategy enhances semantic relevance and reduces the impact of noisy or ambiguous samples.

Notably, as shown in Table III, the experimental results exhibit that even on the Yelp dataset, which exhibits the highest sparsity level (99.91%, compared with 99.73% on DoubanBook and 99.37% on DoubanMovie), HNGCL can still outperform the current best contrastive learning models NESCL and NCL. This result highlights the superior stability and robustness of HNGCL under extremely sparse conditions. Overall, the experimental results demonstrate that HNGCL consistently outperforms state-of-the-art baselines across multiple benchmarks, especially under conditions of high data sparsity and heterogeneous relational structures. This validates the effectiveness of our proposed heterogeneous neighborhood-enhanced contrastive learning framework.

The improvements of the HNGCL model on the DoubanMovie dataset [51] are greater than those on Yelp [49] and DoubanBook [50], which can be attributed to several factors. First, the DoubanMovie dataset is less sparse, with denser user–item interactions, enabling the model to more effectively capture associations between users and items. Second, the higher frequency of interactions in DoubanMovie amplifies user similarity and item associations, helping the model better learn user preferences and item features, thereby improving recommendation accuracy. Finally, HNGCL's dual-view design—integrating the user–item interaction view and the meta-path-based semantic view—proves particularly effective for datasets such as DoubanMovie, where richer interaction

TABLE III
TOP-K RECOMMENDATION PERFORMANCE OF BASELINE MODELS ON THREE REAL-WORLD DATASETS

Dataset	Metric	HAN (2019)	LightGCN (2020)	DGCF (2020)	BPR (2021)	SMIN (2021)	NCL (2022)	HGCL (2023)	RecDCL (2024)	NESCL (2024)	HNGCL (Ours)	Improv.
Yelp	R@10	0.0339	0.0603	0.0452	0.0395	0.0571	0.0594	0.0631	0.0563	0.0618	**0.0734**	16.32%
	N@10	0.0407	0.0465	0.0329	0.0297	0.0430	0.0517	0.0507	0.0418	0.0477	**0.0557**	7.74%
	R@20	0.0511	0.1008	0.0976	0.0696	0.0868	0.0922	0.0959	0.0948	0.0976	**0.1139**	13.00%
	N@20	0.0281	0.0609	0.0588	0.0398	0.0496	0.0608	0.0602	0.0576	0.0588	**0.0681**	11.82%
DoubanBook	R@10	0.0786	0.1173	0.1140	0.1024	0.0931	0.1043	0.1030	0.1168	0.1272	**0.1392**	9.43%
	N@10	0.0967	0.1358	0.1325	0.1190	0.1130	0.1470	0.1210	0.1355	0.1566	**0.1610**	2.81%
	R@20	0.1140	0.1692	0.1628	0.1402	0.1189	0.1485	0.1373	0.1310	0.1772	**0.1913**	7.96%
	N@20	0.1022	0.1446	0.1401	0.1209	0.1017	0.1493	0.1117	0.1080	0.1617	**0.1663**	2.84%
DoubanMovie	R@10	0.1107	0.1430	0.1391	0.1171	0.1274	0.1106	0.1302	0.1159	0.1309	**0.1526**	9.23%
	N@10	0.1680	0.2037	0.1947	0.1668	0.1810	0.1980	0.1827	0.1495	0.1700	**0.2195**	6.71%
	R@20	0.1754	0.2135	0.2089	0.1849	0.1897	0.1705	0.2026	0.1898	0.1943	**0.2264**	6.04%
	N@20	0.1738	0.2085	0.2016	0.1784	0.1855	0.1969	0.1949	0.1852	0.1771	**0.2156**	3.41%

Note: The optimal outcomes in the experiments are indicated in bold, and the suboptimal results are highlighted with underlines. "Improv." shows the improvement of HNGCL over the suboptimal results. Our HNGCL outperforms the existing baselines.

information allows the complementary views to capture both structural and semantic relationships more comprehensively. This robust modeling of diverse relationships explains the superior performance on DoubanMovie compared with the other datasets.

To better understand the underlying reasons behind this performance gain, we further analyze how HNGCL addresses the limitations of baseline models, particularly issues such as overfitting and sensitivity to noise. Compared with baseline models such as LightGCN and NCL, which may suffer from overfitting in sparse and noisy environments due to limited semantic modeling or lack of regularization, HNGCL demonstrates improved generalizability. The incorporation of contrastive learning with alignment and uniformity constraints acts as a form of regularization, discouraging overfitting by encouraging consistent embeddings across heterogeneous views. Furthermore, the neighborhood-enhanced mechanism enriches the positive sample set, reducing the reliance on limited or noisy interactions and improving robustness.

C. Ablation Study (RQ2)

We conduct an ablation study to validate the effectiveness of several key components in HNGCL. Specifically, we evaluate the contribution of each part and visualize the results.

W/o CL: This variant discards the contrastive loss, thereby disabling the functionality of cross-view contrastive learning.

W/o NE: This variant excludes the neighborhood-enhanced (NE) module, performing only cross-view contrastive learning without incorporating collaborative neighbors into the positive sample set.

W/o A&U: This variant discards the optimization of alignment and uniformity (A&U), while retaining cross-view contrastive learning and neighborhood-enhanced positive sampling.

Table IV and Fig. 4 present the recommendation performance of HNGCL and its variants across three datasets, using Recall@10 and NDCG@10 as evaluation metrics. Analysis of these data reveals that HNGCL consistently outperforms its variants in all datasets, underscoring the importance of contrastive learning based on HGNNs, alignment and uniformity optimization, and the neighborhood-enhanced mechanism in recommender systems. The experimental results indicate that the proposed components play a vital role in effectively integrating heterogeneous information, enhancing representation learning through semantic-aware data augmentation, and mitigating the negative impact of noise. Collectively, they contribute to the superior performance and robustness of HNGCL.

Additionally, the performance of the W/o A&U variant exhibits the lowest on Yelp dataset, demonstrating the significant impact of alignment and uniformity in this context. Compared with the other two datasets, this performance gap may be attributed to the higher data sparsity in Yelp dataset. In highly sparse datasets, the reduced presence of noise and irrelevant

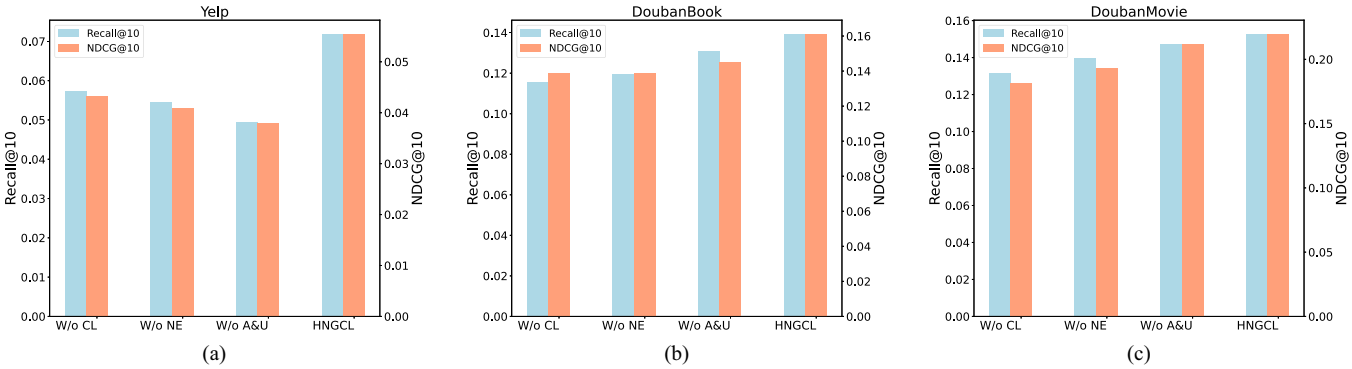

Fig. 4. Analysis of ablation experiments on three datasets for recall@10 and NDCG@10. (a) Yelp. (b) DoubanBook. (c) DoubanMovie.

TABLE IV
ABLATION EXPERIMENT IN HNGCL

Dataset	Yelp		DoubanBook		DoubanMovie	
Metric	R@10	N@10	R@10	N@10	R@10	N@10
W/o CL	0.0573	0.0433	0.1154	0.1390	0.1314	0.1815
W/o NE	0.0546	0.0410	0.1194	0.1391	0.1397	0.1935
W/o A&U	0.0494	0.0380	0.1310	0.1449	0.1473	0.2119
HNGCL	0.0734	0.0557	0.1392	0.1610	0.1526	0.2195

features allows the key features to stand out more distinctly. Consequently, feature representations are more likely to realize a uniform distribution on the hypersphere, and enforcing alignment can more effectively capture critical patterns, thereby enhancing recommendation performance.

Specifically, removing alignment and uniformity losses leads to each view learning representations independently, which can result in biased and inconsistent embeddings for the same node across different views. This misalignment hinders the model's ability to integrate complementary information from multiple views, thereby diminishing its capacity to capture the true underlying data structure and resulting in suboptimal downstream performance. By introducing alignment and uniformity losses, the embeddings of the same node from different views are explicitly encouraged to be mapped closer together in the latent space. This not only reduces representational bias, but also enhances consistency and compatibility between views, facilitating more effective information fusion and improved generalization ability. Our ablation results further confirm that models equipped with alignment and uniformity losses consistently achieve higher accuracy and robustness. These findings highlight the crucial role of cross-view alignment in generating unbiased, consistent, and informative node representations.

D. Robustness Analysis (RQ3)

To simulate real-world scenarios where data may be incomplete or missing, we randomly remove 20% of the original training data to evaluate the robustness of HNGCL. Furthermore, to account for potential data errors or inaccuracies in practical applications, we introduce random noise into the user–item interaction data at rates of 20% and 40%, respectively, to assess model performance under varying levels of data contamination. Eventually, we evaluate all models on the original, unmodified test set using Recall@10 and NDCG@10 as evaluation metrics. The detailed experimental results refer to Fig. 5, where the bar chart represents the optimal performance of each model w.r.t. Recall@10 (left y-axis), and the line chart w.r.t. NDCG@10 (right y-axis).

By comparing the performance of the original model trained on a complete, noise-free dataset with that of the robustness-tested model (i.e., trained on datasets with missing data and injected noise), we focus on the HNGCL model's resistance to disturbances, providing insights into its overall robustness.

1) Although the introduction of noise adversely affects the performance of all models, HNGCL consistently outperforms the baseline methods. This result indicates that the methods adopted by the HNGCL are effective in filtering out redundant or noisy information, thereby demonstrating excellent noise resistance and stability.
2) The performance of HNGCL declines most significantly on the Yelp dataset, indicating that the model's capacity to differentiate various categories or user behaviors is more susceptible to being affected by the introduction of noise from irrelevant features in highly sparse datasets.

In summary, HNGCL maintains superior performance and high recommendation quality in the presence of data incompleteness and noise, confirming its robustness and reliability.

E. Data Sparsity Analysis (RQ4)

Due to the fact that most users only interact with a minority of items, it is challenging to enhance the expressiveness of the recommender systems by generating high-quality representations. To evaluate the effectiveness of HNGCL in mitigating the issue of data sparsity, we conduct sparsity experiments on three datasets and compare them with commonly used baseline models, including LightGCN. NDCG@10 is adopted as the primary evaluation metric for assessing model performance. According to [29], we categorize users into four groups based on distinct levels of sparsity, determined by the count of user

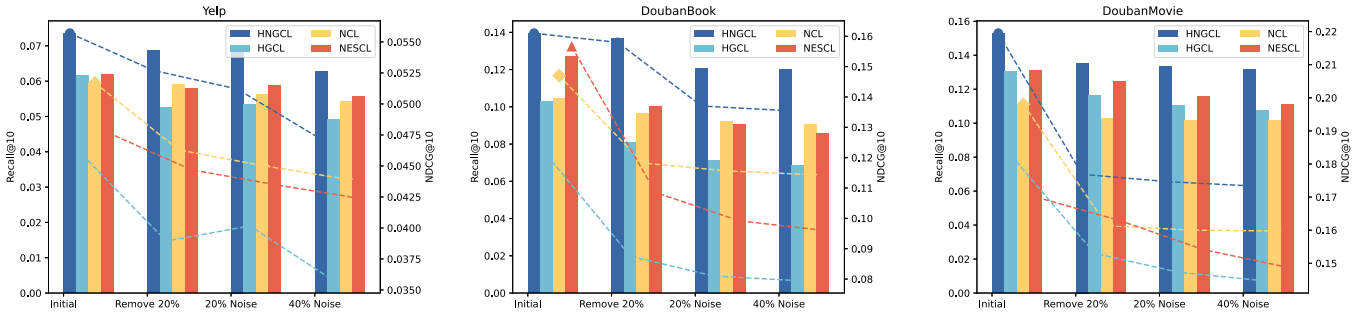

Fig. 5. Performance comparison w.r.t. the original model, data incompleteness and noise.

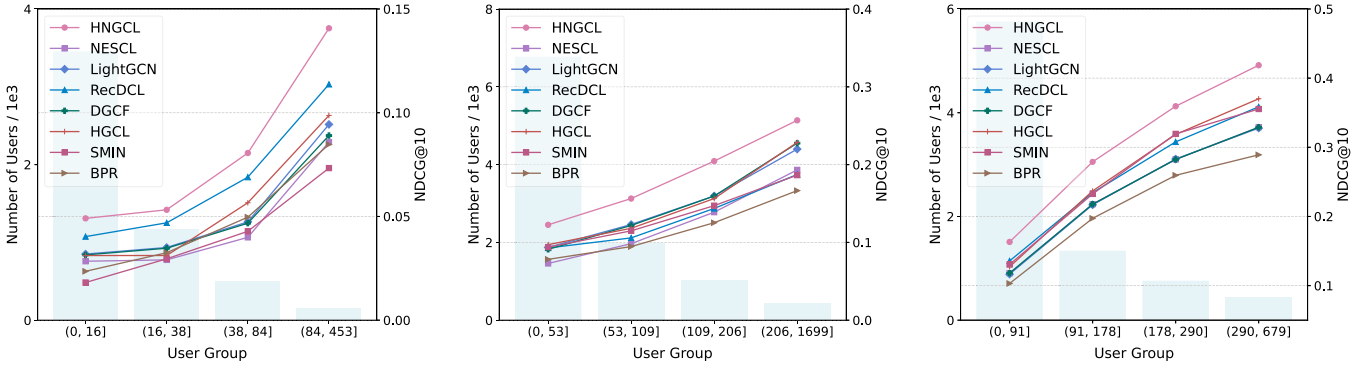

Fig. 6. Comparison of sparsity levels on different user groups. The background bar chart represents the number of users in each interval (left y-axis), and the line chart indicates the optimal performance of each model w.r.t. NDCG@10 (right y-axis). The x-axis represents the interaction intervals between each user group. (a) Yelp. (b) DoubanBook. (c) DoubanMovie.

interactions. Each group maintains a consistent count of total interactions. Taking the Yelp dataset, which has the highest data sparsity, as an example, the four intervals for user interaction intervals are (0, 16], (16, 38], (38, 84], and (84, 453]. Fig. 6 shows the NDCG@10 evaluation for divergent user groups on three datasets, and Recall@10 also shows a similar trend. It shows that the performance of all models significantly promotes when the amount of interactions increases, indicating that the scale of user–item interactions will greatly affect the quality of the learned representations. HNGCL consistently outperforms other baseline models on all datasets, validating its effectiveness in handling sparse data scenarios.

F. Parameter Analysis (RQ5)

1) Loss Function Coefficient: In this section, we investigate the critical hyperparameters α, β, and γ in our proposed loss function and conduct a systematic evaluation to examine how these hyperparameters influence the performance. The experiments are based on three real-world datasets: Yelp, DoubanBook, and DoubanMovie. We analyze the impact of each hyperparameter on model performance based on the results presented in Fig. 7.

Analysis of Hyperparameter α: Initially, we fix β and γ at their default values of 1 and select different values of α from the candidate list [0.01, 0.05, 0.1, 0.15, 0.2] to evaluate model performance. On the Yelp dataset, model performance initially decreases with increasing α, then slightly improves, but never surpasses the performance achieved at $\alpha = 0.01$. On the DoubanBook dataset, performance initially rises to a peak before declining, indicating an optimal comprehensive performance when $\alpha = 0.1$. In contrast, on the DoubanMovie dataset, performance exhibits a consistent upward trend, achieving the best result when $\alpha = 0.2$.

Analysis of Hyperparameter β: Based on the optimal value of α, we further investigate the impact of hyperparameter β by varying its value in the range [0.0001, 0.001, 0.01, 0.1, 1], while keeping γ fixed at 1. Unlike α, model performance gradually improves across all three datasets as β increases, achieving the best comprehensive performance when $\beta = 1$.

Analysis of hyperparameter γ: Lastly, with the optimal values of α and β fixed, we assess model performance by selecting different values of γ from the candidate list [0.0001, 0.001, 0.01, 0.1, 1]. On the Yelp dataset, the impact of γ mirrors that of β, with performance steadily improving and reaching its peak at $\gamma = 1$. On the DoubanBook and DoubanMovie datasets, performance initially declines and then improves, ultimately achieving the best results at $\gamma = 1$.

In summary, to achieve optimal model performance across various datasets, we have determined the optimal hyperparameter configurations for α, β, and γ as follows: for the Yelp dataset, the values are 0.01, 1, and 1, respectively; for the

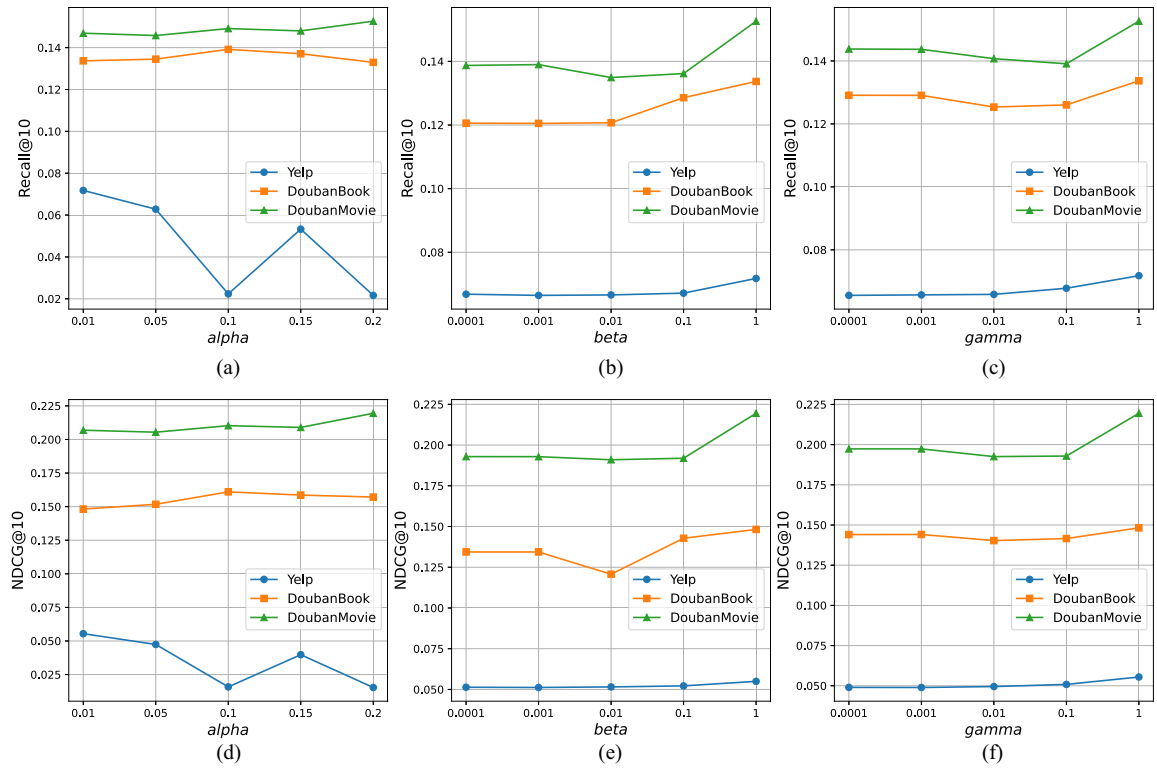

Fig. 7. Recall@10 and NDCG@10 comparison w.r.t. different hyperparameters on three real-world datasets. (a) and (d) Alpha. (b) and (e) Beta. (c) and (f) Gamma.

DoubanBook dataset, they are 0.1, 1, and 1; and for the DoubanMovie dataset, they are 0.2, 1, and 1. These results suggest that as the dataset size increases, assigning greater weight to the contrastive loss component is beneficial for enhancing overall performance.

2) Parameter Analysis: We investigate the effects of four major parameters: embedding dimension, learning rate, the number of GNN layers, and the number of GCN layers. Experiments are conducted on the Yelp, DoubanBook, and DoubanMovie datasets, with Recall@10 and NDCG@10 used as evaluation metrics.

Embedding Dimension. We set the embedding dimensions to [16, 32, 64, 128, 256]. As shown in Fig. 8(a) and (e), the metrics of Recall@10 and NDCG@10 improve significantly as the increase of embedding dimensions. This phenomenon can be attributed to higher-dimensional embeddings' ability to capture and represent complex relationships and fine-grained relationships within the data. Nonetheless, as the dimensions increase, the computational complexity of the model also rises, potentially leading to overfitting or noise interference. To strike a balance between performance and efficiency, and to maintain consistency with existing research, in both the HNGCL and each baseline model, the embedding dimension is fixed at 128, following classical baselines. This setting provides sufficient expressive power to capture the complex relationships between nodes while avoiding issues such as overfitting or excessive computational cost that can arise from higher-dimensional embeddings. It strikes a good balance between efficiency and effectiveness.

Learning Rate: The learning rate is a critical factor that significantly affects both model convergence and overall performance. The learning rate for different datasets is determined based on tuning results. We select a relatively small learning rate to ensure the stable convergence of the model while avoiding issues such as gradient explosion or oscillation. Specifically, we evaluate five distinct learning rates: [0.0005, 0.001, 0.005, 0.01, 0.05]. The results showed that HNGCL achieved a good balance between performance and convergence speed when the learning rates are set to 0.0005 for the Yelp and DoubanMovie datasets, and 0.001 for the DoubanBook dataset. Fig. 8(b) and (f) illustrate that the model achieves optimal performance with a learning rate of 0.0005 on the Yelp and DoubanMovie datasets, while on the DoubanBook dataset, the peak performance is attained with a learning rate of 0.001. An excessively high learning rate may hinder converge or lead to gradient explosion, whereas an overly low learning rate might trap the model in local minima or lead to vanishing gradients. Considering efficiency and performance, we assign the learning rate of 0.0005 for the Yelp and DoubanMovie datasets, and 0.001 for the DoubanBook dataset.

GNN Layer: The results in Fig. 8(c) and (g) demonstrate the impact of GNN layer count from 1 to 4 on model performance. On the Yelp dataset, the model achieves optimal comprehensive performance with two GNN layers, whereas for the DoubanBook and DoubanMovie datasets, the best performance is obtained with just one layer. Although increasing the number of GNN layers enables the model to aggregate information from more distant neighbors and capture more complex graph

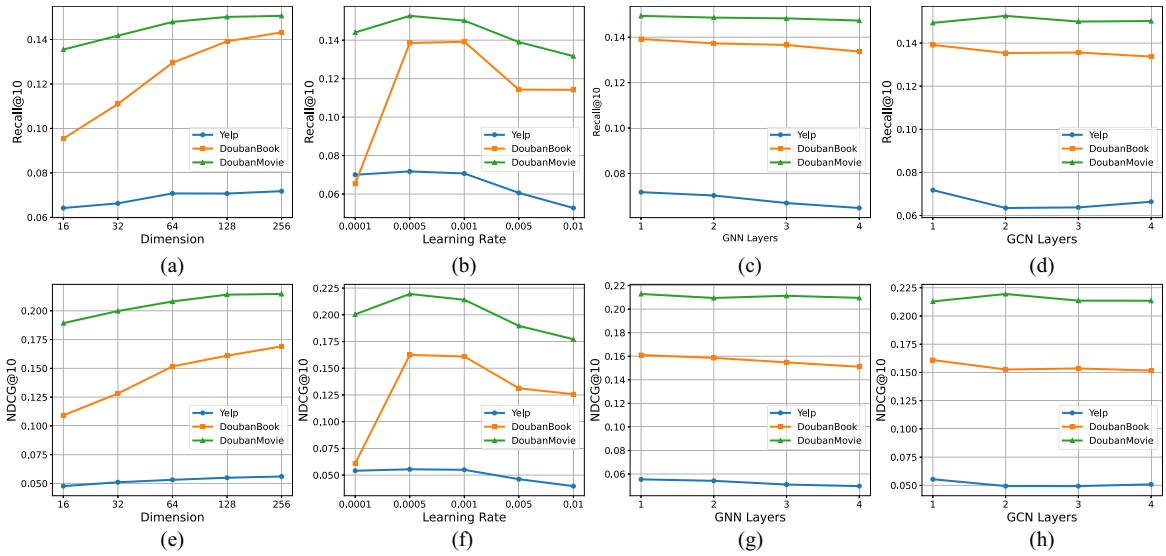

Fig. 8. Recall@10 and NDCG@10 on three real-world datasets w.r.t dimension, learning rate, GNN layers, and GCN layers. (a) and (d) Dimension. (b) and (f) Learning rate. (c) and (g) GNN layers. (d) and (h) GCN layers.

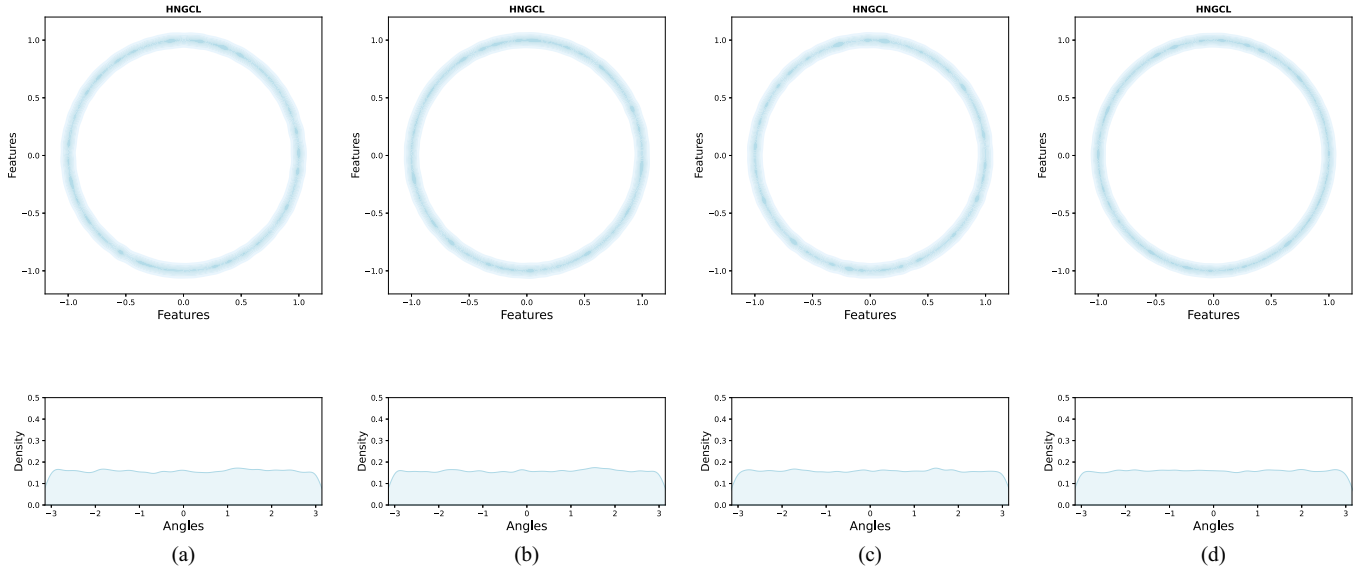

Fig. 9. Distribution of embedding representations learned from DoubanBook dataset. (a) CL only. (b) Only add NE mechanism. (c) Only add A&U. (d) HNGCL.

structures, it also introduces greater model complexity, which may degrade performance due to over-smoothing or overfitting. Therefore, we set the number of GNN layers to 1 for DoubanBook and DoubanMovie, and 2 for Yelp.

In contrast, the DoubanMovie dataset reaches peak performance with a 2-layer GCN architecture. This may be attributed to the fact that deeper attention-based networks increase model complexity, which can lead to overfitting, noise accumulation, and a diminished ability to capture essential features. Therefore, we set the number of GCN layers to 1 for Yelp and DoubanBook, and 2 for DoubanMovie.

GCN Layer: We further investigate the impact of varying the number of GCN layers on model performance. As shown in Fig. 8(d) and (h), while deeper GCN architectures can capture higher-order neighbors information and richer semantics, the best performance is achieved with a single GCN layer on the Yelp and DoubanBook datasets. In contrast, the DoubanMovie dataset reaches peak performance with a 2-layer GCN architecture. This may be attributed to the fact that deeper attention-based networks increase model complexity, which can lead to overfitting, noise accumulation, and a diminished ability to capture essential features. Therefore, we set the number of GCN layers to 1 for Yelp and DoubanBook, and 2 for DoubanMovie.

3) Embedding Visualization: We employ the nonparametric Gaussian kernel density estimation (KDE) [53] method to visualize the embeddings generated by HNGCL. Specifically,

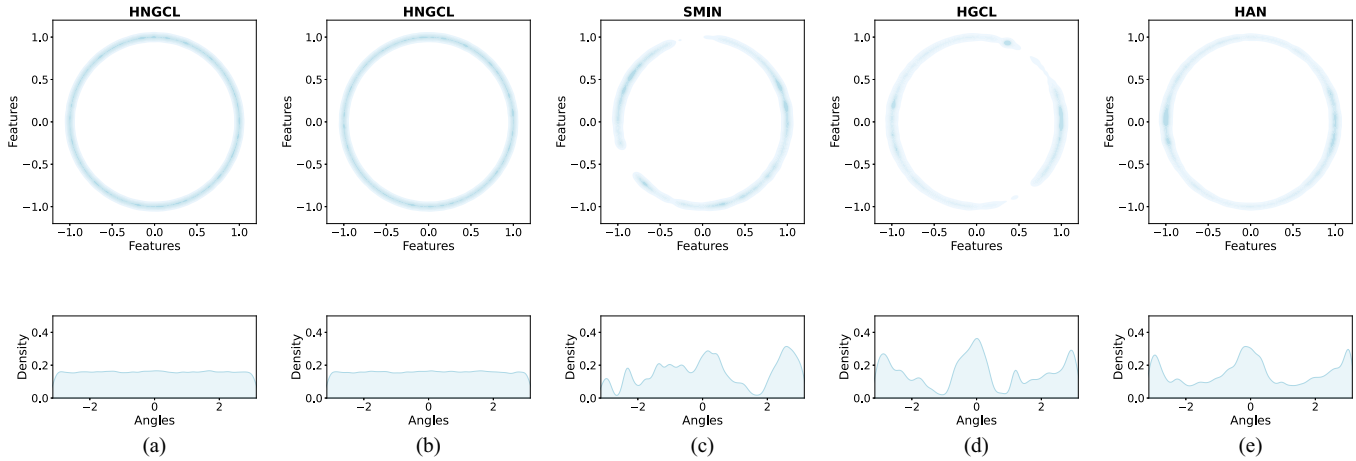

Fig. 10. Comparison of representations distribution on DoubanBook dataset. (a) HNGCL. (b) 40%noise HNGCL. (c) SMIN. (d) HGCL. (e) HAN.

we randomly select 2000 learned representations from DoubanBook dataset and project onto the unit hypersphere using t-SNE [54]. As shown in Fig. 9, we compare four configurations: 1) contrastive learning only; 2) contrastive learning with the neighborhood-enhanced mechanism; 3) contrastive learning with alignment and uniformity optimization; and 4) the full model incorporating all modules. By correlating the results from Table IV with the feature distributions depicted in Fig. 9, it is evident that incorporating the neighborhood-enhanced mechanism and optimizing for alignment and uniformity can improve the quality of the learned representations. In particular, incorporating alignment and uniformity losses brings embeddings of the same nodes from different views closer together, while also promoting a more uniform distribution on the hypersphere. These enhancements lead to more discriminative and structured embedding distributions, ultimately improving the model's generalization ability compared with using contrastive learning alone. In Fig. 9, angular density estimation further confirms that alignment and uniformity losses help produce smoother, less peaked density curves, indicating reduced over-clustering and enhanced cross-view consistency.

To further elucidate the advantages of the HNGCL model, we compare it and its noise-perturbed variant (with 40% noise) against three baseline models: HAN, HGCL, and SMIN. For each method, 4000 learned representations are extracted from the DoubanBook dataset at the point when the respective model achieves its best performance. To more clearly display the feature distribution, we also visualize the angular density estimation for each point. As shown in Fig. 10, the embeddings produced by the baseline models exhibit concentrated density along certain arcs, with sharp peaks in their corresponding angular density curves. In contrast, the feature distribution of the HNGCL model is more uniform, and its angular density estimation curves are comparatively smoother. Combining the results from Table V, even when noise is introduced, HNGCL still demonstrates a more desirable feature distribution compared with the baseline models. These phenomena may be attributed to the unique positive sample pair construction mechanism of HNGCL and the optimization objectives focused on alignment and uniformity. Together, these components promote a more uniform distribution of nodes and avoiding excessive aggregation of nodes within the feature space.

TABLE V
EFFECT OF NOISE ON HNGCL

Dataset	Yelp		DoubanBook		DoubanMovie	
Metric	R@10	N@10	R@10	N@10	R@10	N@10
Original data	0.0734	0.0557	0.1392	0.1610	0.1526	0.2195
Remove 20%data	0.0688	0.0527	0.1366	0.1580	0.1355	0.1769
Add 20%noise	0.0683	0.0512	0.1204	0.1370	0.1336	0.1747
Add 40%noise	0.0627	0.0472	0.1198	0.1355	0.1317	0.1734

In summary, the HNGCL model has demonstrated its superiority in terms of the uniformity of feature distribution and the smoothness of angular density estimation. These properties not only indicate the effectiveness of the model in capturing local features but also underscore its robustness in global feature representation. The experimental results further confirm the potential of HNGCL in handling complex graph-structured data, particularly in application scenarios that require strong generalization capability.

VI. CONCLUSION

In this article, we propose a novel model, HNGCL, designed for recommendation tasks. HNGCL leverages heterogeneous relationships to generate user–item interaction view and meta-path-based view. By optimizing alignment and uniformity across views, HNGCL learns consistent and discriminative representations. Additionally, a neighborhood-enhanced contrastive strategy is introduced to generate high-quality positive sample pairs, mitigating the noise caused by anchor nodes drifting away from collaborative neighbors.

Despite its effectiveness, HNGCL still faces limitations. Its scalability on large-scale graphs requires further improvement, and like many GNN-based models, it may underperform in cold-start scenarios. Furthermore, the reliance on meta-paths introduces challenges, such as the need for domain-specific knowledge and potential difficulty in capturing complex higher-order relationships. Hypergraph-based methods, which can naturally model higher-order interactions through hyperedges, represent a promising alternative to meta-paths. Additionally, while our study demonstrates the model's effectiveness on standardized datasets, it does not directly validate user satisfaction in real-world scenarios. In future work, we will explore integrating hypergraph-based approaches or incorporating social network information into the meta-paths to improve the model's flexibility and generalizability. We aim to further exploit the rich semantic relationships embedded in socially-aware meta-paths to enhance recommendation quality. Additionally, we plan to design experiments that adapt to evolving user preferences, investigate methods to improve scalability, and develop solutions for cold-start scenarios, while also conducting user studies to bridge the gap between offline evaluation and real-world applicability.

REFERENCES

[1] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 173–182.

[2] X. Cai, W. Guo, M. Zhao, Z. Cui, and J. Chen, "A knowledge graph-based many-objective model for explainable social recommendation," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 6, pp. 3021–3030, Dec. 2023.

[3] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 426–434.

[4] H. Zhang, F. Shen, W. Liu, X. He, H. Luan, and T.-S. Chua, "Discrete collaborative filtering," in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2016, pp. 325–334.

[5] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The Adaptive Web: methods and Strategies of Web Personalization*, Berlin, Germany: Springer, 2007, pp. 291–324.

[6] A. Khelloufi et al., "A multimodal latent-features-based service recommendation system for the social internet of things," *IEEE Trans. Computat. Social Syst.*, vol. 11, no. 4, pp. 5388–5403, Aug. 2024.

[7] Y. Bi, L. Song, M. Yao, Z. Wu, J. Wang, and J. Xiao, "A heterogeneous information network based cross domain insurance recommendation system for cold start users," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. retrieval*, 2020, pp. 2211–2220.

[8] J. Li, M. Jing, K. Lu, L. Zhu, Y. Yang, and Z. Huang, "From zero-shot learning to cold-start recommendation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 4189–4196.

[9] C. Gao, X. Wang, X. He, and Y. Li, "Graph neural networks for recommender system," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, 2022, pp. 1623–1625.

[10] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous graph transformer," in *Proc. web Conf.*, 2020, pp. 2704–2710.

[11] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Heterogeneous information networks: the past, the present, and the future," *Proc. VLDB Endowment*, vol. 15, no, p. 12. 2022.

[12] X. Wang, N. Liu, H. Han, and C. Shi, "Self-supervised heterogeneous graph neural network with co-contrastive learning," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, 2021, pp. 1726–1736.

[13] X. Wang et al., "Heterogeneous graph attention network," in *Proc. World Wide Web Conf.*, 2019, pp. 2022–2032.

[14] L. Sang, M. Xu, S. Qian, M. Martin, P. Li, and X. Wu, "Context-dependent propagating-based video recommendation in multimodal heterogeneous information networks," *IEEE Trans. Multimedia*, vol. 23, pp. 2019–2032, 2021.

[15] L. Sang, Y. Wang, Y. Zhang, Y. Zhang, and X. Wu, "Intent-guided heterogeneous graph contrastive learning for recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 4, pp. 1915–1929, Apr. 2025.

[16] P. Sun, L. Wu, K. Zhang, X. Chen, and M. Wang, "Neighborhood-enhanced supervised contrastive learning for collaborative filtering," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 5, pp. 2069–2081, May 2023.

[17] M. Chen, C. Huang, L. Xia, W. Wei, Y. Xu, and R. Luo, "Heterogeneous graph contrastive learning for recommendation," in *Proc. 16th ACM Int. Conf. Web Search Data Mining*, 2023, pp. 544–552.

[18] K. Zhu, T. Qin, X. Wang, Z. Chen, and J. Ding, "Graph contrastive learning with hybrid noise augmentation for recommendation," in *Proc. Int. Conf. Adv. Data Mining IEEE Int. Symp. Spread Spectr. Tech. Appl.*, Springer, 2023, pp. 325–339.

[19] W. Chen, Y. Zhang, H. Li, L. Sang, and Y. Zhang, "Dual-domain collaborative denoising for social recommendation," *IEEE Trans. Computat. Social Syst.*, early access, 2025.

[20] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework contrastive learning visual representations," in *Proc. Int. Conf. Mach. Learn. (PMLR)*, 2020, pp. 1597–1607.

[21] C. Wang et al., "Towards representation alignment and uniformity in collaborative filtering," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, 2022, pp. 1816–1825.

[22] L. Sang, W. Fei, Y. Zhang, Y. Huang, and Y. Zhang, "Heterogeneous adaptive preference learning for recommendation," *ACM Trans. Recomm. Syst.*, Apr. 2024.

[23] L. Sang, Y. Wang, Y. Zhang, and X. Wu, "Denoising heterogeneous graph pre-training framework for recommendation," *ACM Trans. Inf. Syst.*, Dec. 2024, doi: 10.1145/3706632.

[24] J. Hu, B. Hooi, S. Qian, Q. Fang, and C. Xu, "MGDCF: Distance learning via Markov graph diffusion for neural collaborative filtering," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 7, pp. 3281–3296, Jul. 2024.

[25] J. Yu, H. Yin, J. Li, M. Gao, Z. Huang, and L. Cui, "Enhancing social recommendation with adversarial graph convolutional networks," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3727–3739, Aug. 2020.

[26] J. Yu, H. Yin, J. Li, Q. Wang, N. Q. V. Hung, and X. Zhang, "Self-supervised multi-channel hypergraph convolutional network for social recommendation," in *Proc. web Conf.*, 2021, pp. 413–424.

[27] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2017.

[28] J. Hu, B. Hooi, B. He, and Y. Wei, "Modality-independent graph neural networks with global transformers for multimodal recommendation," 2024, *arXiv:2412.13994*.

[29] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 165–174.

[30] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and powering graph convolution network for recommendation," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2020, pp. 639–648.

[31] Z. Lin, C. Tian, Y. Hou, and W. X. Zhao, "Improving graph collaborative filtering with neighborhood-enriched contrastive learning," in *Proc. ACM Web Conf.*, 2022, pp. 2320–2329.

[32] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, and T.-S. Chua, "Disentangled graph collaborative filtering," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. retrieval*, 2020, pp. 1001–1010.

[33] S. Peng, K. Sugiyama, and T. Mine, "SVD-GCN: A simplified graph convolution paradigm for recommendation," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, 2022, pp. 1625–1634.

[34] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9729–9738.

[35] J. Wu et al., "Self-supervised graph learning for recommendation," in *Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. retrieval*, 2021, pp. 726–735.

[36] D. Zhang et al., "RECDCL: Dual contrastive learning for recommendation," in *Proc. ACM Web Conf.*, 2024, pp. 3655–3666.

[37] C. Shi, B. Hu, W. X. Zhao, and S. Y. Philip, "Heterogeneous information network embedding for recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 2, pp. 357–370, Feb. 2018.

[38] H. Linmei, T. Yang, C. Shi, H. Ji, and X. Li, "Heterogeneous graph attention networks for semi-supervised short text classification," in *Proc.*

Conf. Empirical Methods Nat. Lang. Process. 9th Int. Joint Conf. Nat. Lang. Process. (EMNLP-IJCNLP), 2019, pp. 4821–4830.

[39] Z. Wang, Q. Li, D. Yu, X. Han, X.-Z. Gao, and S. Shen, "Heterogeneous graph contrastive multi-view learning," in *Proc. SIAM Int. Conf. Data Mining (SDM)*, Philadelphia, PA, USA: SIAM, 2023, pp. 136–144.

[40] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, "Heterogeneous graph neural network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 793–803.

[41] L. Sang, M. Xu, S. Qian, and X. Wu, "Adversarial heterogeneous graph neural network for robust recommendation," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 5, pp. 2660–2671, Oct. 2023.

[42] J. Hu, B. Hooi, and B. He, "Efficient heterogeneous graph learning via random projection," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 12, pp. 8093–8107, Dec. 2024.

[43] S. Xu, et al., "Topic-aware heterogeneous graph neural network for link prediction," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, 2021, pp. 2261–2270.

[44] B. Khan, J. Wu, J. Yang, M. K. Hayat, and S. Xue, "A unified hypergraph framework for inter and intra-session dynamics in session-based social recommendations," *IEEE Trans. Big Data*, early access, 2025.

[45] B. Khan, J. Wu, J. Yang, and X. Ma, "Heterogeneous hypergraph neural network for social recommendation using attention network," *ACM Trans. Recomm. Syst.*, vol. 3, no. 3, Mar. 2025.

[46] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, "Graph convolutional neural networks for web-scale recommender systems," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery & Data Mining*, 2018, pp. 974–983.

[47] K. Mao, J. Zhu, X. Xiao, B. Lu, Z. Wang, and X. He, "UltraGCN: ultra simplification of graph convolutional networks for recommendation," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, 2021, pp. 1253–1262.

[48] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "Bpr: Bayesian personalized ranking from implicit feedback," in *Proc. 25th Conf. Uncertainty Artif. Intell. (UAI)*, Montreal, Quebec, Canada: AUAI Press, 2009, pp. 452–461.

[49] H. Wang, K. Zhou, X. Zhao, J. Wang, and J.-R. Wen, "Curriculum pre-training heterogeneous subgraph transformer for top-n recommendation," *ACM Trans. Inf. Syst.*, vol. 41, no. 1, pp. 1–28, 2023.

[50] J. Yu, H. Yin, M. Gao, X. Xia, X. Zhang, and N. Q. Viet Hung, "Socially-aware self-supervised tri-training for recommendation," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, 2021, pp. 2084–2092.

[51] W. Yu and S. Li, "Recommender systems based on multiple social networks correlation," *Future Gener. Comput. Syst.*, vol. 87, pp. 312–327, 2018.

[52] X. Long, et al., "Social recommendation with self-supervised metagraph informax network," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, 2021, pp. 1160–1169.

[53] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," 2010, *arXiv:1011.2602*.

[54] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, Nov. 2008.

Lei Sang received the Ph.D. degree in computer science from the Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia, in 2021.

Currently, he is a Lecturer with the School of Computer Science and Technology, Anhui University, Anhui, China. His research interests include natural language processing, data mining, and recommender systems.

Chi Zhang received the bachelor's degree in computer science and technology from the Southeast University Chengxian College, Nanjing, China, in 2023. She is currently working toward the master's degree in computer technology with Anhui University's School of Computer Science and Technology, Anhui, China.

Her research interests include graph neural network, recommender systems, and data mining.

Maohao Huang received the bachelor's degree in computer science and technology from Huizhou University, Huizhou, China, in 2023. He is currently working toward the master's degree in computer technology with Anhui University's School of Computer Science and Technology, Anhui, China.

His research interests include graph neural network, recommender systems, and data mining.

Lin Mu received the Ph.D. degree in computer science from the University of Science and Technology of China, Hefei, China, in 2021.

Currently, he is a Lecturer with the School of Computer Science and Technology, Anhui University, Hefei. His research interests include information extraction, natural language processing, and large language models (LLM).

Yiwen Zhang received the Ph.D. degree in management science and engineering from Hefei University of Technology, Hefei, China, in 2013.

Currently, he is a Full Professor with the School of Computer Science and Technology, Anhui University, Hefei. His research interests include service computing, recommender systems, and big data analytics.

Dr. Zhang has published more than 100 papers in highly regarded conferences and journals, including IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON SERVICES COMPUTING, *ACM Transactions on Information Systems*, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *ACM Transactions on Knowledge Discovery from Data*, *Special Interest Group on Information Retrieval*, *American Journal of Computational Linguistics*, etc.

Xindong Wu (Fellow, IEEE) received the B.S. and M.S. degrees in computer science from Hefei University of Technology, Hefei, China, in 1987, and the Ph.D. degree in artificial intelligence from the University of Edinburgh, Edinburgh, U.K., in 1993.

Currently, he is the Director and Professor with the Key Laboratory of Knowledge Engineering with Big Data (the Ministry of Education of China), Hefei University of Technology, and a Senior Research Scientist with the Research Center for Knowledge Engineering, Zhejiang Lab, Hangzhou, China. His research interests include data mining, knowledge engineering, big data analytics, and marketing intelligence.

Dr. Wu is a Foreign Member of Russian Academy of Engineering and a Fellow of AAAS.