A Social Recommendation Model Based on Cross-View Contrastive Learning and Multi-Head Attention for Multi-Rating Fusion

Rui Chen[a,*], Zhuo Dai[a], Wei Lu[b], Yanbu Guo[a], Weizhi Meng[c], Pu Li[a], Min Huang[a], Xiangjie Kong[d]

[a] *College of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450000, Henan, China*
[b] *School of Information, Xian University of Finance and Economics, Xian, 710000, Shaanxi , China*
[c] *Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, 2800, Capital Region, Denmark*
[d] *School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, 310014, Zhejiang, China*

Abstract

In recent years, social recommender systems have become a hot research field. Contrastive learning effectively enhances the expression of user representations by modeling the consistency of representations between interactive views and social views, thereby improving recommendation performance. This paper proposes a social recommendation model based on cross-view contrastive learning, which employs a multi-head attention mechanism to fuse multi-rating information (MVCLMH). It adaptively assigns weights to multiple views, making more effective use of rich social relationship and social trust information to alleviate data sparsity. Within the rating view, interaction-aware noise (with direction-preserving constraints) is introduced to enhance model robustness. We conduct comparative learning from both rating and social perspectives, and construct comparative tasks within and across rating and social contexts. Experimental results on real-world datasets demonstrate that MVCLMH improves recommendation accuracy and outperforms exist-

*Corresponding author

Email addresses: ruichen@zzuli.edu.cn (Rui Chen), luweiufe@xaufe.edu.cn (Wei Lu), weizhi.meng@ieee.org (Weizhi Meng), xjkong@iee.org (Xiangjie Kong)

ing models in terms of Recall@10, Recall@30, NDCG@10, NDCG@30, as well as scalability for large-scale applications.

Keywords: Contrastive Learning; Graph Neural Networks (GNNs); Social Recommender Systems

1. INTRODUCTION

Social recommender systems aim to leverage social information to better capture user preferences. The underlying principle is that users' online behaviors are not independent, but rather interconnected through their social ties [1, 2, 3]. Traditional recommendation approaches are primarily based on Collaborative Filtering (CF), which models user preferences by mining historical user–item interaction data [4]. However, such methods usually rely on relatively dense interaction matrices and thus struggle to capture deep interest patterns when the data is sparse. Moreover, they excessively depend on historical behaviors, failing to reflect the subtle differences and multidimensional nature of user preferences [5]. These approaches often overlook the roles of social relations, contextual signals, and multimodal information, which ultimately limits their performance in complex recommendation scenarios.

To alleviate the aforementioned issues, early studies extended matrix factorization and collaborative filtering. With the rise of social networks, Ma et al. [6] and Jamali and Esteret al. [7] introduced social regularization to improve prediction accuracy after embedding trust propagation into matrix decomposition. More recently, deep learning frameworks have drawn significant attention due to their ability to learn nonlinear representations [8]. The integration of deep learning with graph neural networks (GNNs) has further enhanced modeling capabilities. He et al. [9] proposed neural collaborative filtering, whereas Fan et al. [10] and Wang et al. [11] leveraged GNNs to capture social diffusion effects, enabling recommender systems to exploit social relations more effectively. However, social connections are not always equivalent to interest similarity. Directly incorporating social links may introduce noise and semantic conflicts, which can lead to unstable performance.

To alleviate data sparsity and noise issues, contrastive learning has recently emerged as a promising self-supervised paradigm in recommender systems[11, 12]. By aligning representations across different semantic spaces while disentangling irrelevant information, contrastive learning provides an

effective mechanism for enhancing robustness under sparse interactions[11, 12, 13]. Chen et al. [14] proposed the SimCLR framework, establishing a general paradigm for self-supervised contrastive learning. Building upon this foundation, Mu et al.[15] and Li et al.[16] explored behavior perturbation and graph-structured contrastive learning in recommendation scenarios, significantly improving representation robustness. Nevertheless, existing contrastive learning–based social recommendation methods still lack systematic solutions for deep integration of rating and social information, effective semantic alignment across heterogeneous views, and conflict resolution between social relations and behavioral preferences.

Accurate prediction of user preferences and generation of personalized recommendations remain as critical challenges in recommendation systems [17]. By synthesizing existing studies, we identify four pressing limitations that remain unresolved. (1) Most research simplifies multi-level ratings into binary signals when modeling user–item interactions. This coarse-grained approach overlooks the intensity of user preferences, resulting in recommendations that lack subtlety. (2) Although incorporating social relations helps alleviate sparsity, the significant discrepancy between real-world social networks and interest graphs often introduces noisy information when forcefully fused, and may even compromise the learning of core interaction signals [6, 9]. (3) Existing multi-view fusion methods largely rely on static weighting or simple concatenation, which limits their ability to flexibly adapt to semantic differences and potential redundancy across views [16, 18]. (4) Current contrastive learning methods mainly focus on enforcing consistency within individual views, while insufficiently addressing deep semantic alignment across different views [12, 19]. This limitation reduces their effectiveness in multi-modal and multi-relational scenarios.

To address the aforementioned challenges, this paper proposes a social recommendation model named MVCLMH (Multi-View Contrastive Learning with Multi-Head Attention for Multi-Rating Fusion). The core idea of MVCLMH is to simultaneously model the rating view and the social view, leveraging a multi-head attention mechanism to achieve dynamic fusion of multi-relational embeddings, thereby overcoming the limitations of static fusion methods [18]. For modeling user preferences, we design a gated mechanism with residual connections that allows the model to adaptively select the proportion of information from interaction and social domains, enhancing stability in scenarios with sparse social relations[9]. Meanwhile, to mitigate sparsity and long-tail distribution issues, we introduce an Interaction-

aware Direction-preserving Noise (IDN) mechanism, which dynamically adjusts noise intensity with sign-consistency constraints to improve the robustness of representations [20]. In terms of contrastive learning, we further incorporate a cross-view contrastive loss, enforcing the representations of the same node under different views to align in the embedding space, thereby effectively bridging the semantic gap between social and rating views [12, 13].

The main contributions of this work can be summarized as follows:

(1) We propose a preference modeling framework that integrates multi-head attention with a gated mechanism, addressing the limitation of existing methods that often simplify multi-rating data into binary interactions and overlook differences in preference intensity. This framework dynamically allocates weights to different relations at a fine-grained level and adaptively fuses interaction and social information, enabling more accurate modeling of users' multidimensional interests.

(2) We introduce an Interaction-aware Direction-preserving Noise (IDN) perturbation strategy to enhance robustness in scenarios with sparse social relations and long-tail distributions. By dynamically adjusting noise based on node activity and enforcing sign-consistency constraints, this strategy effectively mitigates overfitting for high-frequency interactions and information loss for low-frequency nodes, thereby improving the stability and generalization of embedding representations.

(3) We develop a cross-view contrastive learning mechanism to overcome the limitation of existing contrastive learning approaches that focus only on intra-view consistency and lack semantic alignment across views. This mechanism establishes explicit contrastive constraints between rating and social views, forcing the representations of the same node across different views to align in the embedding space. As a result, it promotes deep integration and complementarity of multi-modal information, significantly enhancing recommendation accuracy and robustness.

MVCLMH integrates methods explored and proven effective in existing research, forming a complete framework. This framework leverages the multi-rating convolution concept from the MCLA model, employing a multi-head attention mechanism to dynamically fuse multi-rating information. It utilizes a noise mechanism with orientation-preserving perturbations for cross-view contrastive learning, enabling better explicit alignment of ratings and social representations within a shared embedding space. Instead of using individual components in isolation, this architecture systematically integrates them under a common optimization objective, jointly addressing issues such as rat-

ing strength modeling, semantic conflicts in social behavior, and multi-view geometric inconsistencies.

Overall, the MVCLMH model significantly outperforms existing methods in both theory and methodology. It constructs a unified multi-perspective recommendation framework that not only mitigates the impact of data sparsity and social noise but also resolves alignment and conflict issues between different perspectives. Experimental results demonstrate that MVCLMH outperforms existing state-of-the-art baseline models on multiple real-world datasets, validating its effectiveness and practical application value.

2. Related Work

2.1. Social Recommendation Models

Social influence plays a critical role in shaping user preferences, making it an indispensable factor in modern recommender systems. Social recommendation leverages social relation graphs to enhance recommendation performance, effectively alleviating data sparsity and improving prediction accuracy [15, 16, 21]. SocialMF [7] decomposes the social trust network to represent users in two distinct low-dimensional latent feature spaces. DiffNet [22] and its advanced version DiffNet++ [11] simulate the iterative social influence process using a progressive layer-wise diffusion mechanism. Yu et al. [23] introduce a deep adversarial framework leveraging GCNs to capture the heterogeneous strengths of social relationships. He et al. [24] proposed Neural Collaborative Filtering (NCF), combining neural networks with collaborative filtering to strengthen the modeling of nonlinear interactions. Although NCF was not specifically designed for social recommendation, it laid the foundation for subsequent social recommendation methods based on graph neural networks (GNNs).

In recent years, joint modeling of implicit and explicit social relations has received considerable attention. Taheri et al. [25] proposed Hell TrustSVD, which automatically extracts social relations to enhance TrustSVD and performs well even in the absence of explicit trust labels. Wang et al. [26] introduced the SERec model, simulating users' information acquisition paths via a social exposure mechanism, outperforming the traditional "friend preference" assumption. . However, most existing social recommendation models fail to evaluate whether the social information adequately reflects the influence of user preferences [19].

2.2. Applications of Contrastive Learning in Recommender Systems

Contrastive learning provides powerful self-supervised signals for recommender systems, eliminating the need for explicit labels [27, 28]. To address sparsity and label noise issue [29, 30]. SGL [31] employs multiple strategies, such as node dropout, edge dropout and random walk, to generate various views, thereby augmenting node representations from graph structure perspectives. Wang et al. [19] The combination of dual variational inference and graph contrastive learning provides a new framework for social recommender systems. NCL [32] introduces a novel structure-contrastive objective, treating users or items alongside their structural neighbors as positive contrastive pairs. Qin et al. [18] proposes a contrastive learning framework that captures and optimizes user intent by utilizing multiple interaction subsequences, thereby improving the accuracy and personalization of the recommendation model. These studies highlight the significant potential of combining contrastive learning with graph structures and multi-view designs, yet challenges remain in effectively integrating social and rating embeddings.

2.3. Multi-View Fusion and Representation Learning

Multi-view fusion aims to establish synergistic relationships among heterogeneous information while mitigating noise. Liu et al. [33] introduced MVGCN, which combines gated mechanisms with multi-channel graph neural networks to selectively integrate information from different views. HAMNet [15] dynamically perceives association weights between users and neighboring nodes based on graph attention mechanisms, accurately capturing local interaction patterns. Zhang et al. [34] further designed a contrastive fusion framework to align multi-modal information such as visual and content features. Although existing studies have explored multi-view fusion from various perspectives, systematic methods for addressing semantic conflicts and alignment between rating and social graphs remain limited. In particular, there is no mature solution in the context of contrastive learning. Motivated by this, the present work combines multi-view fusion with contrastive learning, proposing a multi-head attention-based semantic alignment method over social and rating multi-graphs.

3. Proposed Model

From a technical perspective, MCLA employs static multi-relation decomposition and adaptive fusion. While it incorporates adaptive noise, it

lacks sign constraints and focuses solely on intra-view comparisons, resulting in issues such as neglecting rating intensity differences and insufficient cross-view alignment. MVCLMH addresses this deficiency through multi-head attention, dynamic weighting of multiple ratings, and bidirectional cross-view comparisons. SimGCL uses fixed-intensity Gaussian noise perturbation in a single view as its core, but the non-directional nature of this noise can easily lead to semantic drift. MVCLMH's IDN mechanism effectively preserves the embedded geometry through dynamic intensity adjustment and sign constraints. XSimGCL emphasizes intra-view cross-layer noise fusion and comparison, but does not involve social views or multi-view fusion. MVCLMH constructs a rating-social dual-view comparison framework, integrating dual-domain information through a gating fusion mechanism. DiffNet++ relies on social influence diffusion mechanisms but lacks dedicated noise strategies and contrastive learning. Directly injecting social information can easily introduce noise. MVCLMH, on the other hand, suppresses noise through IDN and reduces semantic conflicts through cross-view constraints, comprehensively optimizing model performance.

The overall architecture of the MVCLMH model is illustrated in Figure 1. It mainly consists of three components: (1) a multi-head attention module for fusing multi-relational information, (2) a gated mechanism module for integrating user preferences, and (3) a multi-view contrastive learning module.

The original observation matrix used to describe the user-item rating association within the interaction domain, where each entry r_{ui} denotes the true rating of user u on item i ; if there is no interaction between user u and item i , then $r_{ui} = 0$. Following the idea of partitioning the rating matrix in the MCLA [35] model, the original rating matrix can be divided into $\{A_1^r, A_2^r, A_3^r, ..., A_a^r\} \in \mathbb{R}^{m \times n}$, and similarly, the social relation matrix can be partitioned into $\{A_1^s, A_2^s, A_3^s, ..., A_b^s\} \in \mathbb{R}^{m \times m}$, where m representing the number of users and n representing the number of items, a and b present the number of relation types in the interaction domain and social domain, respectively, and their values are determined by the relation structure of the specific dataset. We denote the d dimensional user embeddings in the interaction and social domains as $p_r \in \mathbb{R}^{m \times d}$ and $p_s \in \mathbb{R}^{m \times d}$, respectively, and the d dimensional item embeddings in the interaction domain as $q_r \in \mathbb{R}^{n \times d}$. The goal of Top-N recommendation is to predict the probability $\hat{r}_{u,i}$ of interaction between user u and an item i from the candidate set of unobserved items.

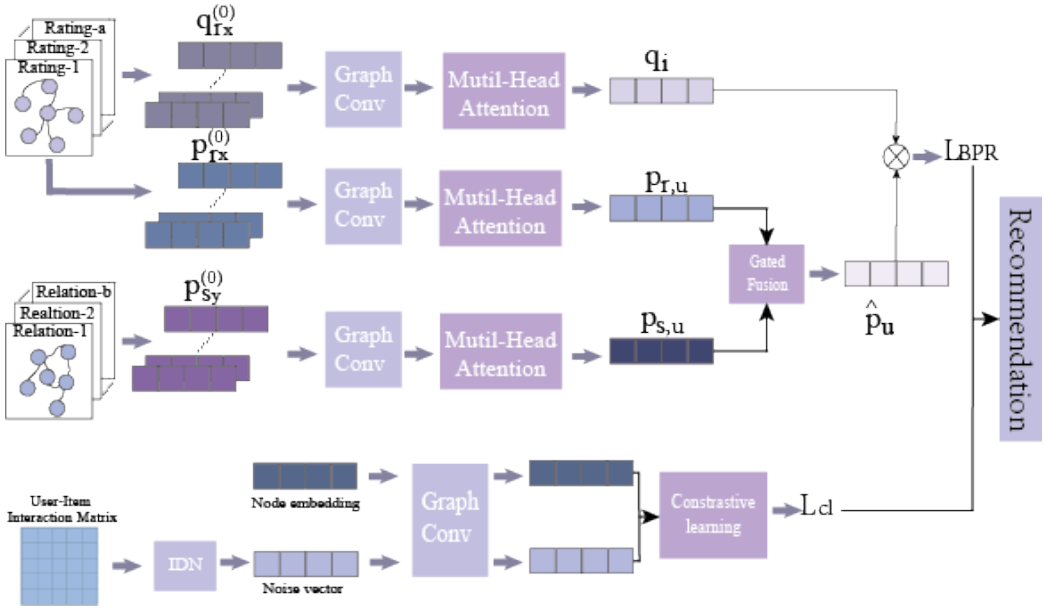

Figure 1: Architecture of the MVCLMH model

3.1. Multi-Head Attention for Multi-Level Rating Embedding Fusion

In the field of computer vision, splitting the original data into multiple dimensions is often employed to help models better capture key features [36]. This idea of decomposition has also been applied in recommender systems, enhancing the model's ability to capture user preferences and improving overall performance [37].

To finely model each user's preferences for items with different ratings and the influence of various social relations on user preferences, we first preform a structured decomposition of the input data: In the interaction domain, the original user-item rating matrix is divided according to different rating levels $\{A_1^r, A_2^r, A_3^r, ..., A_a^r\}$, in teh social domain, the original social relationship matrix is split according to relationship type $\{A_1^s, A_2^s, A_3^s, ..., A_b^s\}$.

Each decomposed view is treated as an independent view. We employ a simple yet effective Graph Neural Network (GNN), LightGCN [38], to learn embeddings for different relations. Specifically, for user u and item i under different relations in the two domains, the graph convolution operation is defined as follows:

$$p_{r_x,u}^{(k+1)} = \sum_{i \in N_u^r x} \frac{1}{\sqrt{|N_u^{r_x}|}\sqrt{|N_i^{r_x}|}} q_{r_x,i}^{(k)} \tag{1}$$

$$p_{s_y,u}^{(k+1)} = \sum_{j \in N_u^{s_y}} \frac{1}{|N_u^{s_y}|} p_{s_y,j}^{(k)} \tag{2}$$

$$q_{r_x,i}^{(k+1)} = \sum_{u \in N_i^{r_x}} \frac{1}{\sqrt{|N_i^{r_x}|}\sqrt{|N_u^{r_x}|}} p_{r_x,u}^{(k)} \tag{3}$$

where $p_{r_x,u}^{(k)}$, $q_{r_x,i}^{(k)}$ and $p_{s_y,u}^{(k)}$ denote the embeddings of user u and item i at the k -th convolution layer under relation x in the interaction domain, and the embedding of user u at the k-th convolution layer under relation y in the social domain, respectively. $N_u^{r_x}$ and $N_i^{r_x}$ represent the neighborhoods of user u and item i under relation x in the interaction domain, while $N_u^{s_y}$ represents the neighborhood of user u under relation y in the social domain.

$$E_{r_x}^{final} = \frac{1}{K+1} \sum_{k=0}^{K} E_{r_x}^{k} \tag{4}$$

$$p_{r_x,u} = split\left(E_{r_x}^{final}[1, 2, 3, ..., m]\right) \tag{5}$$

$$q_{r_x,i} = split\left(E_{r_x}^{final}[m+1, m+2, m+3, ..., m+n]\right) \tag{6}$$

where $E_{r_x}^{k}$ denotes the embedding matrix at the k -th layer for relation x in the interaction domain, with $E_{r_x}^{k} \in \mathbb{R}^{(m+n)\times d}$. $E_{r_x}^{final}$ represents the embedding matrix for relation x in the interaction domain after the layer-wise averaging operation. The operator $split\,(*)$ denotes the decomposition of embeddings into user and item parts, where $p_{r_x,u}$ and $q_{r_x,i}$ represent the embeddings of user u and item i under relation x in the interaction domain, respectively. Similarly, the embedding of user u in the social domain under relation y can be obtained as $p_{s_y,u}$.

Thus, we can obtain the multi-relational embeddings of a user across the two domains: $\{p_{r_1,u}, p_{r_2,u}, ..., p_{r_a,u}, p_{s_1,u}, p_{s_2,u}, ..., p_{s_b,u}\}$, and similarly, the multi-relational embeddings of an item in the interaction domain: $\{q_{r_1,i}, q_{r_2,i}, ..., q_{r_a,i}\}$, These embeddings represent the user's (or item's) preferences across different rating levels and social relations.For a given user, the rating values of items influence their overall preference for the items. To better capture the user's comprehensive preference in the interaction domain, we employ a multi-head

attention mechanism to aggregate the multi-relational embeddings. The aggregation is formally defined as follows:

$$X = stack\left(p_{r_1,u}, p_{r_2,u}, ..., p_{r_a,u}\right) \tag{7}$$

$$Q = reshape(X \cdot W^Q), K = reshape(X \cdot W^K), V = reshape(X \cdot W^V) \tag{8}$$

$$A^h_{r_x,u} = softmax\left(\frac{Q^h_{r_x,u} \cdot \left(K^h_{r_x,u}\right)^T}{\sqrt{d_k}}\right) \tag{9}$$

$$\mathcal{O}_{r_x,u} = Concat\left(A^1_{r_x,u}V, A^2_{r_x,u}V, ..., A^h_{r_x,u}V\right) W^O \tag{10}$$

$$\alpha_{r_x,u} = \frac{1}{H}\sum_{i=1}^{H} mean - row\left(A^i_{r_x,u}\right) \tag{11}$$

$$p_{r,u} = \sum_{r_x}^{\{p_{r_1,u}, p_{r_2,u}, ..., p_{r_a,u}\}} \left(\alpha_{r_x,u} p_{r_x,u} + \mathcal{O}_{r_x,u}\right) \tag{12}$$

where $stack\,(*)$ denotes the stacking operation, which concatenates the original multi-view embeddings into a three-dimensional tensor $X \in \mathbb{R}^{m \times a \times d}$. The operator $reshape\,(*)$ represents the linear projection and head-splitting operation, mapping each view's embeddings to Q, K, V matrices for multi-head attention. $W^Q, W^K, W^V, W^O \in \mathbb{R}^{d \times d}$ are learnable parameters. $A^h_{r_x,u}$ denotes the attention scores of user u under relation x in the interaction domain for the h-th head. $Concat\,(*)$ represents the multi-head concatenation operation, and $\mathcal{O}_{r_x,u}$ is the aggregated feature vector obtained after merging all heads. $mean - row\left(A^i_{r_x,u}\right)$ computes the row-wise average of the attention matrix $A^i_{r_x,u}$, resulting in an -dimensional vector representing the global attention distribution of the m -th head. $\alpha_{r_x,u}$ denotes the weight of the view, obtained by summing and averaging the multi-head attention scores. $p_{r,u}$ represents the final user embedding in the interaction domain after aggregating multi-level ratings. Similarly, we can obtain the user embedding in the social domain $p_{s,u}$ and the item embedding in the interaction domain $q_{i,r}$. The process of integrating multi-level rating embeddings using the multi-head attention mechanism is illustrated in Figure 2.

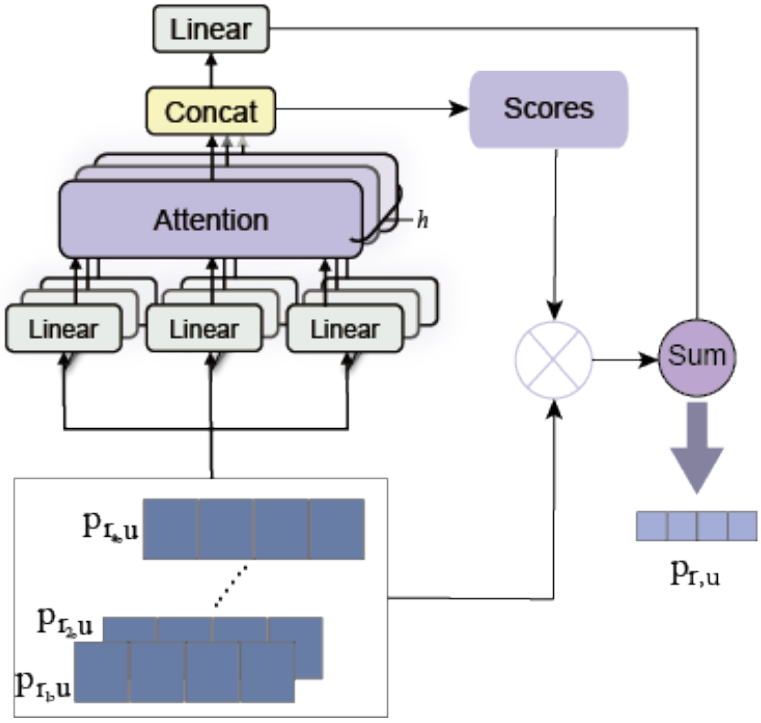

Figure 2: The process of integrating multi-level rating embeddings using the multi-head attention mechanism

3.2. Gated Mechanism for Preference Fusion

The preference feature vectors of user u in the interaction and social domains, $p_{r,u}$ and $p_{s,u}$, capture user preferences from different perspectives. Existing preference fusion methods, such as simple concatenation, mean aggregation, or static weighting, fail to adequately handle the sparsity of social relations. As a result, these approaches cannot effectively integrate social and interaction-based preferences, leading to suboptimal recommendation performance. To better enhance the aggregation of user information across both interaction and social domains, we design and employ a gated mechanism for user preference fusion. The formulation is as follows:

$$\hat{p}_u = g \cdot p_{r,u} + (1 - g) \cdot p_{s,u} + \lambda_u p_{r,u} \tag{13}$$

where $\hat{p}_u$ denotes the final user feature vector obtained after fusing the interaction and social domain embeddings. $\lambda_u p_{r,u}$ represents a residual term, where λ_u is a parameter used to control the residual weights. g is the gate

vector that controls the contribution from each domain. The computation of the gate vector is defined as follows:

$$g = \sigma(W(p_{r,u} \oplus p_{s,u}) + \gamma) \tag{14}$$

where σ denotes the Sigmoid activation function, W represents the learnable weight parameters, $\oplus$ denotes the concatenation of two vectors, and γ is the bias term.

The gate vector g is a vector rather than a scalar, allowing the model to dynamically determine the contribution from different sources for each embedding dimension. To further enhance the stability of the fused user preference representation, we add a residual term at the end of the fusion process and manually adjust its weight. This design significantly improves the model's robustness and fault tolerance during preference aggregation. The process of Gating mechanism integration is illustrated in Figure 3.

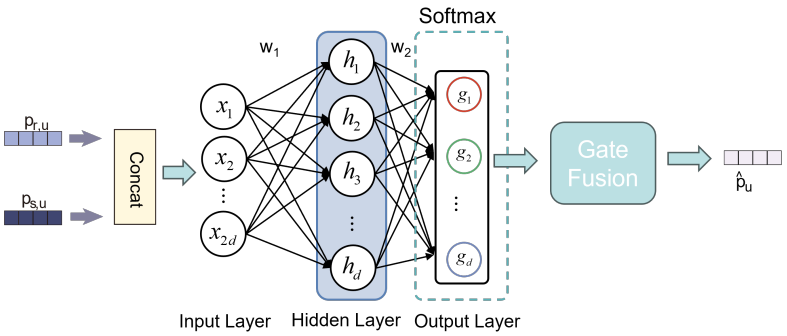

Figure 3: The process of Gating mechanism integration

3.3. Cross-View Contrastive Learning

Since the rating information in the interaction domain is decomposed, while this helps the model better capture user preferences, it also further increases data sparsity. To mitigate the resulting decrease in model robustness, we adopt a strategy similar to SimGCL [27], introducing random noise perturbations at the node level.

To address the three core challenges commonly observed in recommendation datasets—long-tail distribution imbalance, multi-view heterogeneity, and the social–behavior semantic gap—we propose an Interaction-aware Direction-preserving Noise (IDN) mechanism. This mechanism innovatively integrates

three designs: dynamic noise intensity adjustment, view-specific noise configuration, and sign consistency constraints. Specifically, An adaptive nonlinear mapping based on node interaction frequency applies strong perturbations to highly active nodes to suppress overfitting, and weak perturbations to sparse nodes to preserve semantic integrity. Independent noise magnitude ranges are configured for different rating views to accommodate the intrinsic characteristics of each view. A sign function ensures that the noise vectors remain geometrically aligned with the original embedding space, preserving the topological structure of the representations. The generation of noise perturbations is formally defined as follows:

$$\eta_* = \epsilon_{down} + \frac{(c_* - c_{min})}{(c_{max} - c_{min})} (\epsilon_{up} - \epsilon_{down}) \tag{15}$$

where ϵ_{down} and ϵ_{up} denote the lower and upper bounds of the random noise generation interval, c_{min} and c_{max} represent the minimum and maximum values of user–item interactions, and η_* Indicates the strength of the generated user or item noise vector. During the subsequent contrastive learning process, the generated noise vector is added to the original node embeddings to perform data augmentation. The operation is formally defined as follows:

$$p_{r,u}^{(1,k)} = p_{r,u} + sign(p_{r,u}) \cdot n' \cdot \eta_u \tag{16}$$

$$p_{r,u}^{(2,k)} = p_{r,u} + sign(p_{r,u}) \cdot n'' \cdot \eta_u \tag{17}$$

where $p_{r,u}$ denotes the original embedding vector of user u. The operator $sign\,(p_{r,u})$ ensures that the noise preserves the directional sign of the original embedding. $n^{'}$ and $n^{''} \sim \mathcal{N}\,(0,1)$ are Gaussian random noise vectors used for generating the perturbation. The specific process is shown in Figure 4.

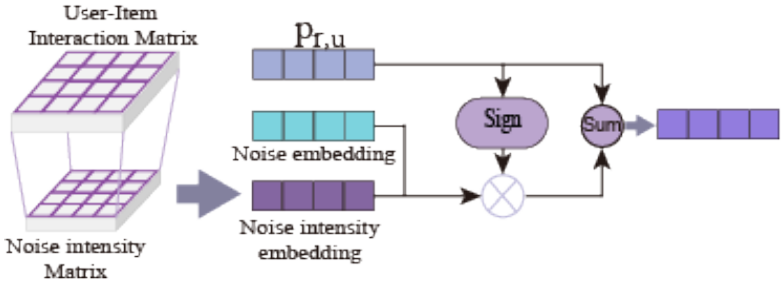

Figure 4: The process of noise generation and injection

The proposed IDN can be regarded as a principled consistency regularizer for heterogeneous semantic spaces. By combining intra-view contrastive learning with cross-view constraints, it mitigates gradient conflicts between scoring and social objectives and promotes stable convergence. Specifically, IDN preserves embedding orientation to confine perturbations within local neighborhoods, preventing semantic drift and abrupt topological changes, while adaptively scaling noise intensity according to interaction frequency to accommodate long-tailed data distributions. Together with explicit noise boundary constraints, this controlled perturbation strategy balances robustness, semantic consistency, and optimization stability, enabling MVCLMH to achieve reliable representation learning beyond heuristic noise injection or empirical parameter tuning.

Compared with traditional multi-view contrastive learning, which primarily focuses on intra-view representation consistency, cross-view contrastive learning introduces an explicit cross-view contrastive loss to enforce that the embeddings of the same node across different views are pulled closer in the embedding space. This mechanism enables deep semantic alignment across heterogeneous views, effectively addressing three major limitations of conventional methods: view-specific representation isolation, gradient attenuation in sparse views, and modality-level semantic gaps. It also demonstrates significant advantages in exploiting complementary information and handling long-tail data.

Therefore, in our model, we extend traditional multi-view contrastive learning by incorporating a cross-view contrastive learning mechanism, which is formally defined as follows:

$$\mathcal{L}_B(X, Y; w) = \frac{1}{B} \sum_{f=1}^{B} w_f^{(*)} \left[-s(x_f, y_f) + \log \left(\sum_{j=1}^{B} e^{s(x_f, y_j)} - e^{s(x_f, y_f)} + \varepsilon \right) \right] \tag{18}$$

$$\mathcal{L}_{cl} = \sum_{k\in\{r_1,\ldots,r_a\}} \left[\mathcal{L}_{Nu}\left(p_{r,u}^{(1,k)}, p_{r,u}^{(2,k)}; w^{(u)}\right) + \mathcal{L}_{Ni}\left(q_{r,i}^{(1,k)}, q_{r,i}^{(2,k)}; w^{(i)}\right)\right] + \lambda_1 \sum_{x,y\in\{r_1,\ldots,r_a\},x\neq y} \left[\mathcal{L}_{Nu}\left(p_{r,u}^{(1,x)}, p_{r,u}^{(2,y)}; w^{(u)}\right)\right] + \lambda_2 \sum_{x,y\in\{r_1,\ldots,r_a\},x\neq y} \left[\mathcal{L}_{Ni}\left(q_{r,i}^{(1,x)}, q_{r,i}^{(2,y)}; w^{(i)}\right)\right] + \lambda_3 \sum_{k\in\{r_1,\ldots,r_a\}} \left[\mathcal{L}_{Nu}\left(p_{r,u}^{(1,k)}, p_{s,u}^{(k)}; w^{(u)}\right)\right] \quad (19)$$

$$s(\delta,\theta) = \frac{\left(\frac{\delta}{\|\delta\|}\right)^T \left(\frac{\theta}{\|\theta\|}\right)}{\tau} \quad (20)$$

$$w_f^{(*)} = \frac{1}{\log\left(c_f^{(*)} + 1\right) + \varepsilon} \quad (21)$$

where $s(\delta,\theta)$ denotes the normalized similarity coefficient between δ and θ. $w_f^{(*)}$ represent the weighting information of the f-th user or item samples, respectively, while $c_f^{(*)}$ denote the interaction counts of the f-th user or item samples. $\tau > 0$ is the temperature parameter, $\varepsilon > 0$ is a numerical stability term, and $\lambda_1, \lambda_2, \lambda_3$ are the weight parameters in the cross-view contrastive learning objective. The specific process of cross-view contrastive learning is illustrated in Figure 3.

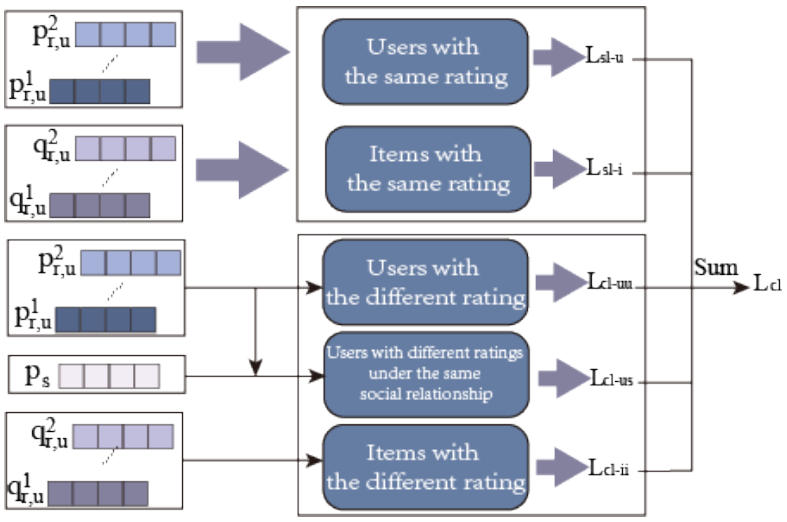

Figure 5: The specific process of cross-view contrastive learning

3.4. Prediction and Model Optimization

For a user u and an item i the interaction probability is given by the inner product $\hat{r}_{u,i} = \hat{p}_u^T q_i$ of their embedding vectors, In optimizing the main recommendation task, we adopt Bayesian Personalized Ranking (BPR) as the core objective function. Finally, the MVCLMH model is optimized using a joint training strategy:

$$\mathcal{L} = \mathcal{L}_{BPR} + \mu_1 \mathcal{L}_{cl} + \mu_2 \|\Theta\|_2 \tag{22}$$

where Θ denotes the regularization loss term of the model, used to constrain model parameters and alleviate overfitting, while μ_1 and μ_2 are hyperparameters controlling the contributions of the contrastive learning loss and the regularization loss, respectively.

4. Experiments

4.1. Datasets

To evaluate the performance of our model, we conducted experiments on three real-world datasets: Douban-Music, Epinions, and Yelp. In all three datasets, the ratings in the interaction domain range from 1 to 5, so we divided the rating data into five distinct rating views. Each user in these

datasets has between one and two social connections. The dataset statistic are show Table 1.

Table 1: dataset statistic

Dataset	Douban-Music	Epinions	Yelp
User	1112	119365	15887
Item	10542	43055	10256
Ratings	70029	198126	744923
Density	0.597%	0.004%	0.457%
Relations	7	6	7

4.2. Baselines

We selected several representative baseline models for performance comparison, including MF and NeuMF based on matrix factorization, DiffNet and LightGCN based on graph neural networks, and MHCN, SEPT, SimGCL, and XSimGCL based on self-supervised learning. A brief description of each model is as follows:

MF [39]: By decomposing the high-dimensional sparse user-item interaction matrix into a low-dimensional latent vector matrix of users and items, and predicting user preferences for items using the vector inner product, this method efficiently solves the data sparsity problem and captures potential associations.

NeuMF [24]:This model uses neural networks as its core, mapping users and items to dense vectors through an embedding layer. It leverages three instantiated models—GMF, MLP, and NeuMF (a fusion of the two)—to flexibly capture the linear and nonlinear interactions of latent features. Log loss is used to adapt to implicit feedback scenarios.

DiffNet [22]: By simulating the recursive and dynamic social influence diffusion process in social networks through a hierarchical influence propagation structure, and combining user (item) free embedding with feature vectors to construct a hybrid embedding, it takes into account both time and storage efficiency and is suitable for scenarios where user/item attributes or social networks are missing.

LightGCN [38]: By removing feature transformations and nonlinear activation operations that are useless for collaborative filtering in GCN, only the core component of neighborhood aggregation is retained, and layer combination is achieved by combining the weighted summation of each layer's embedding.

MHCN [40]: By capturing various high-order user relationships through multi-channel hypergraph modeling and leveraging self-supervised learning to maximize hierarchical mutual information to compensate for the loss of aggregated information, this method demonstrates superior performance compared to existing methods in both general recommendation and cold-start recommendation tasks.

SEPT [41]: By combining social relationships to construct a three-view encoder and adopting a self-supervised triple training mechanism, user representations are continuously reinforced from pseudo-labels generated by other users to improve recommendation performance.

SimGCL [27]: Instead of graph structure enhancement, we construct a contrast view by adding random uniform noise to the embedding space, which adjusts the uniformity of representation and improves the recommendation effect in a simpler and more effective way.

XSimGCL [42]: By replacing cumbersome graph augmentation with simple noise perturbations and completing recommendation and contrastive learning in a single forward pass through cross-layer contrast, a more uniform representation of users and items can be learned efficiently.

MCLA[35]: A social recommendation model based on multi-relational graph contrastive learning and adaptive strategies. It decomposes rating data into multi-relational graphs, designs adaptive fusion strategies, and employs adaptive noise contrastive learning to effectively leverage rating diversity and mitigate data sparsity.

4.3. Overall Performance

To evaluate the performance of all models in Top-N recommendation, we employed two widely used metrics: NDCG@N and Recall@N, with N=10,30. All reported results are the averages over six repeated experiments.

In the experiments, all models were optimized using the Adam optimizer. The parameter settings are as follows: embedding dimensions for all datasets were set to 128, the batch size was 2048, and the number of convolutional layers in both the interaction and social domains was 3. Table 2 presents the results validating the effectiveness of the models, with the best-performing methods highlighted in bold.

Table 2: Overall Performance Comparison

Datasets	Baseline	MF		GNN		SSL					ours	Improv
		MF	NeuMF	DiffNet	LightGCN	MHCN	SEPT	SimGCL	XSimGCL	MCLA	MVCLMH	
Douban-Music	Recall@10	0.02817	0.02758	0.02324	0.03153	0.02934	0.03162	0.03440	0.03459	**0.03803**	0.04112	8.13%
	NDCG@10	0.04250	0.03678	0.03426	0.05145	0.04562	0.04798	0.05502	0.05634	**0.05721**	0.05971	4.37%
	Recall@30	0.05953	0.05604	0.05342	0.07505	0.06200	0.06892	0.07662	0.07571	**0.07952**	0.08174	2.79%
	NDCG@30	0.04780	0.04226	0.04026	0.05963	0.05022	0.05296	0.06216	0.06267	**0.06436**	0.06436	2.10%
Epinions	Recall@10	0.00526	0.01108	0.01062	0.01435	0.01752	0.01467	0.01748	0.01496	**0.02024**	0.02368	16.70%
	NDCG@10	0.00287	0.00604	0.00726	0.00854	0.01020	0.00899	0.01052	0.00892	**0.01172**	0.01348	15.02%
	Recall@30	0.01141	0.02732	0.01834	0.02625	0.03460	0.02695	0.03255	0.02826	**0.03821**	0.04268	11.70%
	NDCG@30	0.00439	0.00914	0.00916	0.01166	0.01450	0.01208	0.01433	0.01231	**0.01625**	0.01834	12.86%
Yelp	Recall@10	0.06525	0.06444	0.07670	0.07482	0.07911	0.08639	0.08718	0.08774	**0.09323**	0.10012	7.39%
	NDCG@10	0.07087	0.06858	0.08838	0.08347	0.08889	0.09353	0.09744	0.09814	**0.10417**	0.11173	7.26%
	Recall@30	0.14324	0.15232	0.16738	0.16039	0.16757	0.17627	0.18094	0.18018	**0.18686**	0.19437	4.02%
	NDCG@30	0.09656	0.09838	0.11655	0.11033	0.11634	0.12331	0.12661	0.12722	**0.13372**	0.14081	5.30%

From the experimental results, the following conclusions can be drawn:

(1) Self-supervised learning (SSL) models (e.g., SimGCL, XSimGCL, MCLA) consistently outperform traditional matrix factorization (MF) and graph neural network (GNN) models across all three datasets in most cases. MF-based models (such as MF and NeuMF) rank lowest across all metrics, as they rely solely on the basic interaction matrix and cannot capture complex relationships. GNN-based models (such as LightGCN and DiffNet) can leverage graph structures to enhance representation power, but their performance is unstable when facing data sparsity and noise. In contrast, SSL models enhance data robustness through contrastive learning, with their core advantage lying in generating high-quality negative samples or aligning views, effectively uncovering implicit relationships. This improvement is especially pronounced in sparse scenarios (e.g., Epinions), confirming that self-supervised mechanisms are a key pathway for improving recommendation performance.

(2) Models explicitly incorporating social relations (e.g., DiffNet, MHCN) fail to consistently improve performance and are notably weaker than behavior-driven SSL models. For instance, DiffNet achieves a lower Recall@10 than the baseline GNN model LightGCN on Douban-Music, indicating that directly integrating social edges may introduce noise. MHCN slightly outperforms some baselines on Recall@10 in Epinions but still lags behind SSL models on the same dataset and exhibits a larger gap on Yelp. The fundamental reason is the low alignment between real social networks and actual user interests; forcibly integrating social information may disrupt the learning of core interaction signals. In contrast, SSL models relying solely on user–item interactions enhance generalization via self-supervision and avoid the unreliability of social data, establishing a more stable advantage.

Our model achieves the best performance across all datasets and metrics,

attributable to the multi-view contrastive learning framework and sparsity-specific designs. On one hand, by aligning multi-view representations of user–item interactions and social context simultaneously and applying contrastive loss to enforce consistency across views, the model significantly enhances noise resistance. On the other hand, integrating high-order graph structure propagation with self-supervised enhancement effectively mitigates overfitting caused by sparse data. The core innovation lies in coupling structural modeling with multi-view contrastive learning, forming a unified paradigm that combines robustness with expressive power.

To test the significance of the model, we performed hypothesis tests on three datasets using RecaLL@10 and NDCG@10 as evaluation criteria. The specific results are shown in Table 3.

Table 3: Hypothesis test results on three datasets

Dataset	Recall@10	p-value	NDCG@10	p-value
Douban-Music	0.04112±9.32E-4	4E-4	0.05971±1.4E-3	7.5E-3
Epinions	0.02368±1.6E-3	7.3E-4	0.01348±9.9E-4	7.6E-3
Yelp	0.10012±3.54E-3	4.9E-3	0.11173±2.08E-3	2.8E-4

The results in Table 3 show that for both Recall@10 and NDCG@10, all p-values are significantly lower than 0.01, indicating that the observed performance improvement is statistically significant and not due to random noise. Furthermore, the relatively small standard deviation demonstrates the stability and robustness of the proposed model across multiple runs. The consistently significant improvements on the Douban Music, Epinions, and Yelp datasets further validate the strong generalization ability of our method under different data sparsity levels.

4.4. Ablation Study

We conducted a further analysis of the different components of the proposed MVCLMH model, which mainly consists of three parts: multi-head attention for multi-rating fusion, gated fusion with residual connections for user preference aggregation, and cross-view contrastive learning. Accordingly, we created three variants of the model. (1) Replacing the multi-head attention fusion in MVCLMH with a simple weighted fusion mechanism, named MVCLMH-MH. (2) Removing the residual term in the gated fusion component of MVCLMH, named MVCLMH-C. (3) Removing the cross-view contrastive learning component from MVCLMH, named MCLMH.

The specific results of these ablation experiments are presented in Table 4.

Table 4: Results of the Ablation Study

Method		MVCLMH-MH	MVCLMH-C	MVCLMH-V	MVCLMH
Douban-music	Recall@10	0.03563	0.04007	0.03847	**0.04112**
	NDCG@10	0.05704	0.05841	0.05794	**0.05971**
Epinions	Recall@10	0.01942	0.02173	0.02057	**0.02259**
	NDCG@10	0.01187	0.01285	0.01164	**0.01136**
Yelp	Recall@10	0.08849	0.09974	0.09561	**0.10012**
	NDCG@10	0.09973	0.10853	0.10374	**0.11173**

Analysis of the ablation study results indicates that the three core components of the proposed MVCLMH model—the multi-head attention for rating fusion, the gated preference fusion module with residual connections, and the cross-view contrastive learning mechanism—each contribute indispensably to performance improvement. Removing the multi-head attention mechanism (MVCLMH-MH) results in the most pronounced performance degradation, confirming its superiority in dynamically modeling complex dependencies among ratings. The exclusion of cross-view contrastive learning (MVCLMH-V) leads to the next most significant yet still noticeable performance loss, highlighting its critical role in enhancing representation robustness for recommendation accuracy. Although removing the residual connection in the gated fusion module (MVCLMH-C) has a relatively smaller but consistent effect, it validates the effectiveness of the residual structure in stabilizing information flow and mitigating gradient issues. The complete MVCLMH model achieves the best performance across all datasets and evaluation metrics, fully demonstrating the synergistic benefits and technical necessity of its overall architectural design.

To demonstrate the necessity of our proposed multi-head attention mechanism and residual-gated fusion module for user and social information fusion, we designed the following components for ablation experiments: (1) fusion using mean averaging; (2) fusion using Concat+MLP; (3) fusion using single-head attention mechanism; and (4) residual removal using a gating mechanism. Recall@10 and NDCG@10 were used as evaluation metrics. The specific experimental results are shown in Table 5.

Table 5: Experimental results of different fusion strategies

Fusion Strategy	Douban-Music Recall@10	Douban-Music NDCG@10	Epinions Recall@10	Epinions NDCG@10	Yelp Recall@10	YelpNDCG@10
Mean Fusion	0.03879	0.05702	0.02003	0.01074	0.09274	0.09749
Concat + MLP	0.04035	0.05837	0.02095	0.01157	0.09453	0.10357
Single-Head Attention	0.03928	0.05755	0.02105	0.01198	0.09752	0.10493
Gate (w/o Residual)	0.04075	0.05862	0.02214	0.01277	0.09885	0.10974
MVCLMH	**0.04112**	**0.05971**	**0.02368**	**0.01348**	**0.10012**	**0.11173**

The results in Table 5 show that the progressive enhancement fusion strategy can continuously improve the performance of all datasets. Simple static fusion methods offer limited benefits, while attention-based and gating mechanisms significantly improve recommendation performance. Notably, MVCLMH achieves the best performance on all three datasets, indicating that the combination of multi-head attention and gating fusion can effectively capture heterogeneous user preferences and suppress noisy information.

To investigate the impact of the number of attention heads in the multi-head fusion module, we varied the number of attention heads to $\{1, 2, 4, 8\}$, while keeping other hyperparameters constant.As shown in Table 4, when the number of heads in the multi-head attention fusion mechanism increases from 1 to 4, the values of the two evaluation metrics, Recall@10 and NDCG@10, increase on the three datasets mentioned above. This indicates that the recommendation performance of the model is also continuously improving, proving that multi-head attention effectively enhances representation diversity and alleviates the information entanglement problem in the multi-view fusion process.

Table 6: Impact of Attention Head Number on Recommendation Performance

Head Num	Douban-Music Recall@10	Douban-Music NDCG@10	Epinions Recall@10	Epinions NDCG@10	Yelp Recall@10	Yelp NDCG@10
1	0.03885	0.05403	0.02142	0.01087	0.09518	0.10594
2	0.04002	0.05569	0.02296	0.01143	0.09806	0.10887
4	**0.04112**	**0.05697**	**0.02368**	**0.01172**	**0.10012**	**0.11173**
8	0.04063	0.05621	0.02321	0.01136	0.09924	0.11041

On the Epinions dataset, the model performance improvement is most significant, with Recall@10 increasing by 10.55% compared to the single-head attention mechanism. This demonstrates that fine-grained subspace modeling yields greater benefits in sparse and socially driven scenarios. In contrast, the improvements on the Douban-Music and Yelp datasets are relatively gradual, indicating that marginal benefits decrease with increasing data density.

When the number of heads in the multi-head attention fusion mechanism further increases to 8, the evaluation metrics on all three datasets decrease. We analyze that this may be due to the reduced expressive power of each

attention head and redundancy between attention patterns. Overall, the results show that a consistent number of attention heads achieves the best balance between information fusion and model stability.

To further investigate the impact of the interactively perceived orientation-preserving noise (IDN) mechanism on recommendation performance, we designed a special ablation experiment to explore the contribution of this mechanism to the overall recommendation system. The experiment mainly included the following four parts: (1) using standard Gaussian noise for contrastive learning (Base-Noise); (2) removing the orientation-preserving mechanism (w/o Sign); (3) fixing the noise intensity (w/o Freq); and (4) not setting upper and lower bounds for noise generation (w/o Bound). We used Recall@10 and NDCG@10 as evaluation metrics. The specific experimental results are shown in Table 5.

Table 7: IDN important

Method	Douban Recall@10	Douban NDCG@10	Epinions Recall@10	Epinions NDCG@10	Yelp Recall@10	Yelp NDCG@10
Base-Noise	0.03479	0.05583	0.01767	0.01055	0.08796	0.09811
w/o Sign	0.03795	0.05788	0.02104	0.01199	0.09301	0.10472
w/o Freq	0.03491	0.05617	0.01841	0.01083	0.08815	0.09874
w/o Bound	0.03628	0.05701	0.01839	0.01126	0.09015	0.10244
IDN-Full	**0.04112**	**0.05971**	**0.02368**	**0.01348**	**0.10012**	**0.11173**

The data in the table demonstrates that this mechanism significantly improves the performance of the recommendation model. Compared to standard Gaussian perturbation (baseline noise), the proposed IDN exhibits superior performance across all datasets and metrics. The relative improvement is particularly significant on the sparse Epinions dataset, with improvements of 34.0% and 27.8% in Recall@10 and NDCG@10, respectively, indicating that structured perturbations provide a more effective contrast signal than pure random noise.

Removing the orientation preservation constraint leads to a significant performance degradation, suggesting that unconstrained perturbations may distort the embedding geometry and introduce semantic drift. Fixing the noise amplitude eliminates adaptive regularization between nodes with different interaction frequencies, resulting in decreased generalization performance, especially under long-tailed distributions. Furthermore, removing the noise boundary constraint introduces unstable perturbation scales, negatively impacting representation stability.

Overall, each component of the IDN contributes independently and collaboratively to the final performance, validating the necessity and effectiveness of the proposed design.

4.5. Impact of the Number of Convolutional Layers

In this model, we conducted experiments on different views in both the interaction and social domains using four different numbers of convolutional layers: $\{1, 2, 3, 4\}$. We report model performance on three datasets using Recall@10 and Recall@30 as the evaluation criteria, with the corresponding results presented in Figure 4.

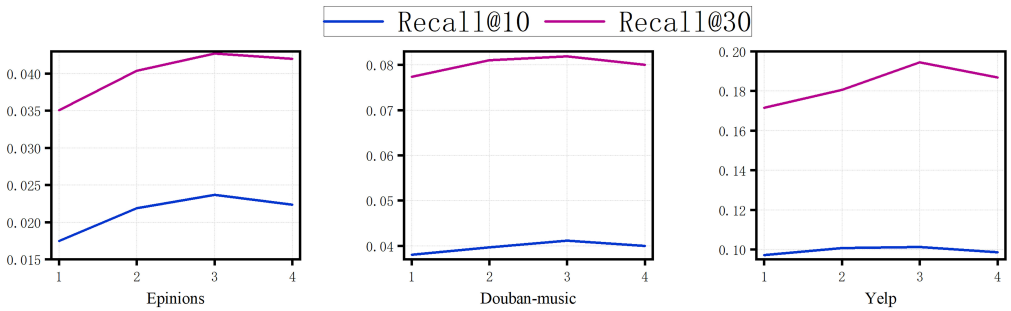

Figure 6: Impact of the Number of Convolutional Layers on Evaluation Metrics

From the experimental results, we can observe that the number of convolutional layers slightly affects the model's performance, with the effect being most pronounced on the Epinions dataset. We found that the optimal number of convolutional layers for both the interaction and social domains is three. Too few layers result in insufficient information aggregation, whereas too many layers can lead to overfitting, thereby degrading the overall model performance.

4.6. Impact of the Temperature Parameter τ

The temperature parameter $\{0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4\}$ in the contrastive learning formula (Eq20) was varied within the range. We use Recall@10 And Recall@30 The performance of the model is evaluated, and the corresponding results are shown in Figure 5.

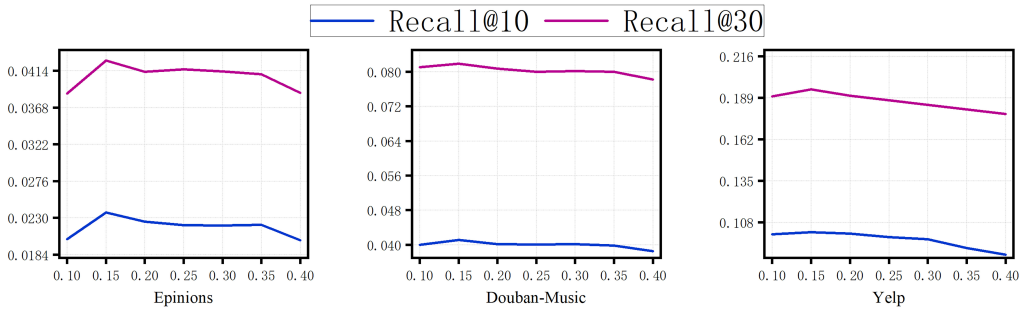

Figure 7: Impact of the temperature parameter τ on evaluation metrics in contrastive learning

From the experimental results, the temperature parameter τ in contrastive learning is identified as 0.15, yielding the best overall performance across all three datasets. When $\tau < 0.15$ the similarity distribution becomes overly sharp, excessively amplifying the separation between positive and negative samples and increasing optimization sensitivity. In contrast, when $\tau > 0.15$, the similarity distribution is over-smoothed, reducing inter-sample discrimination and weakening the effectiveness of the contrastive objective.

4.7. Impact of the Contrastive Learning Loss Weight μ_1

The parameter μ_1 in Eq(22) is used to control the weight of the contrastive learning component in the overall loss. We vary μ_1 within the set $\{0.0001, 0.0005, 0.001, 0.002, 0.003, 0.004\}$and use Recall@10 And Recall@30 As the main evaluation index, the corresponding results are shown in Figure 6.

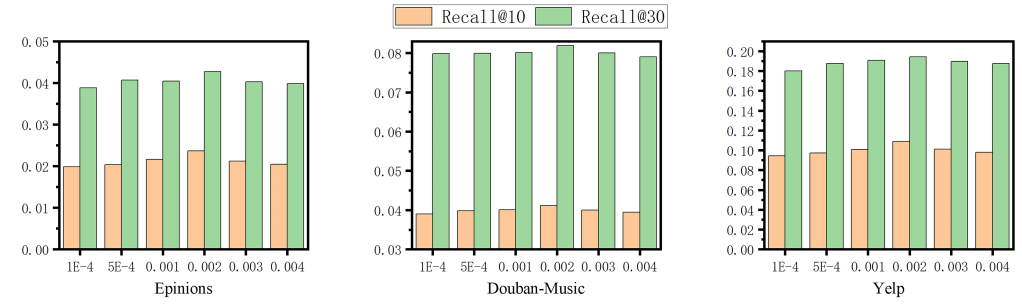

Figure 8: Impact effect of the contrastive learning loss weight μ_1

The experimental results show that the optimal contrastive loss weight μ_1 is 0.002, at which the model achieves the highest overall performance. When $\mu_1 > 0.002$, excessive regularization enforces overly uniform embeddings and degrades discriminability; when $\mu_1 < 0.002$, the self-supervised signal becomes insufficient to provide effective representation enhancement. Both deviations lead to a measurable performance decline.

4.8. Impact of learning rate on model performance

The learning rate determines the step size for each iteration. We vary the learning rate within the set $\{0.0001, 0.001, 0.002, 0.01, 0.1\}$ and use Recall@10 And Recall@30 As the main evaluation index, the corresponding results are shown in Figure 7.

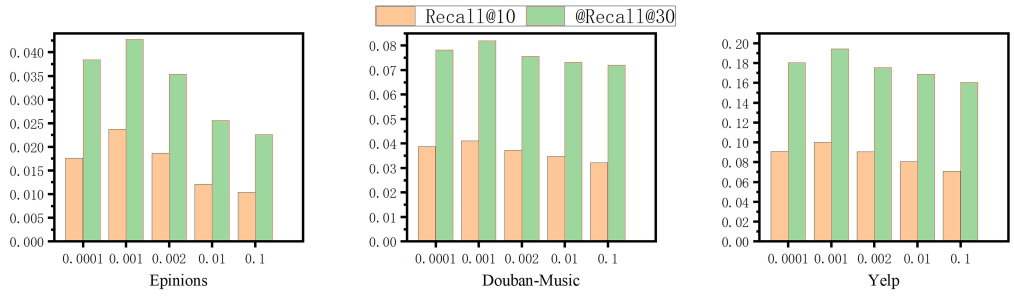

Figure 9: Impact effect of learning rate on model performance

The model achieves optimal performance on all three datasets at a learning rate of 0.001. Under a fixed number of training epochs, learning rates below this value slow convergence and prevent full optimization, whereas higher learning rates induce unstable updates and hinder convergence. Both deviations result in a consistent decline in evaluation metrics.

4.9. Noise Boundary Range Sensitivity Analysis

To investigate the impact of noise boundary range sensitivity, we conducted multiple experiments on three datasets with different noise boundaries, using Recall@10 and Recall@30 as evaluation metrics. The specific experimental results are shown in Figure 10.

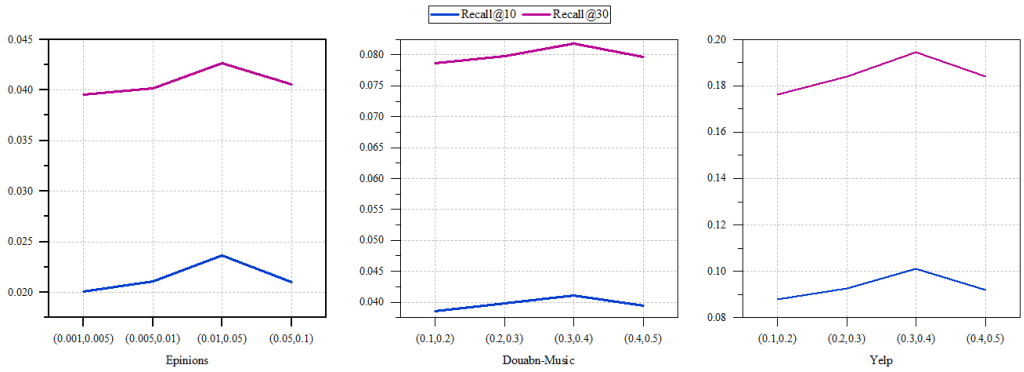

Figure 10: Impact of noise boundary

The experimental results in the figure above show that the optimal performance range varies across the three different datasets, but overall, the results exhibit a similar trend. A moderate perturbation range yields the best results; too little noise fails to provide sufficient regularization, while excessive noise compromises semantic consistency. This indicates that IDN is not particularly sensitive to precise boundary selection and can maintain stable performance within a reasonable range.

4.10. Impact of cross-view comparative learning of component weights

To investigate the impact of the three different weights($\lambda_1, \lambda_2, \lambda_3$) of the cross-view contrastive learning loss in Equation 19 on the overall performance of the model, we took different values for three parameters on three datasets to further explore the contribution of each part to the whole. We used Recall@10 and Recall@30 as evaluation metrics. The specific experimental results are shown in Figures 11.

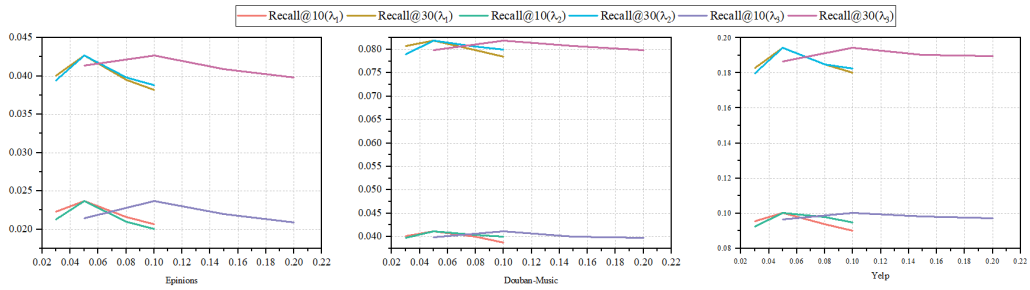

Figure 11: Impact of λ_1, λ_2, λ_3

Figures 11 illustrate the impact of the three weighting parameters λ_1, λ_2 and λ_3 in the cross-view contrastive learning loss (Eq. 19) on Recall@10 and Recall@30 across three datasets. It can be observed that all three parameters exhibit a consistent unimodal trend, where the performance first increases and then decreases as the weight grows. Specifically, the optimal performance is achieved when $\lambda_1 = 0.05$, $\lambda_2 = 0.05$ and $\lambda_3 = 0.1$. This phenomenon indicates that moderate cross-view consistency regularization effectively enhances representation robustness and reduces noise interference, while overly large weights may over-constrain the embedding space and weaken view diversity, leading to performance degradation.

Furthermore, λ_3 plays a crucial role in transferring structural knowledge from the social graph to the sparse rating space, which is essential for mitigating data sparsity and improving generalization. If the weight is too small, the cross-graph supervision signal may be masked by the main recommendation loss, resulting in limited performance improvement. Therefore, assigning a larger optimal value to λ_3 helps to strengthen the propagation of cross-structural information and stabilize the representation learning process.

5. Conclusion and Future Work

5.1. Conclusion

This paper presents the MVCLMH model, a cross-view contrastive learning framework based on multi-head attention and gated fusion mechanisms. The model is designed to deeply capture user preference information and enhance the effectiveness of information integration. By integrating cross-view contrastive learning, it captures comprehensive user preference signals while mitigating the challenges posed by data sparsity.

5.2. Future Work

Most contemporary recommendation models, including mvclmh, use self supervised learning to introduce auxiliary supervision signals to alleviate data sparsity. However, such strategies usually have limited adaptability and rely heavily on manually designed self-monitoring targets, which may lead to poor performance under highly sparse conditions. In future work, we plan to further explore and develop methods that more effectively mitigate the challenges of data sparsity, thereby enhancing the generalization ability of social recommendation models across diverse domains.

6. Acknowledge

This work was supported by the Major Science and Technology Programs in Henan Province[No.241100210100], by the National Natural Science Foundation of China [No. 62072416], by the Key Research and Development Special Project of Henan Province[No. 252102211070, No.232102211051, No. 242102210107 , No. 252102210127], by the Key Research and Development Program of Shaanxi Province [No. 2024GX-YBXM-545], and by the Key·Scientific Research Projects·of HigherEducation·Institutions in·Henan Province.(24B520038).

References

[1] Qing Meng, Bo Liu, Hengyuan Zhang, Xuheng Sun, Jiuxin Cao, and Roy Ka-Wei Lee. Temporal-aware and multifaceted social contexts modeling for social recommendation. *Knowledge-Based Systems*, 248:108923, 2022.

[2] Ljubisa Bojic. Ai alignment: Assessing the global impact of recommender systems. *Futures*, 160:103383, 2024.

[3] Xiwei Wang, Siguleng Wuji, Mali Li, Yutong Liu, and Ran Luo. Social impact of recommendation algorithm in crisis: Forming algorithmic experience through group information interaction and algorithm task fit. *Information Processing Management*, 63(1):104323, 2026.

[4] Azadeh Faroughi, Parham Moradi, and Mahdi Jalili. Enhancing recommender systems through imputation and social-aware graph convolutional neural network. *Neural Networks*, 184:107071, 2025.

[5] Saman Forouzandeh, Pavel N. Krivitsky, and Rohitash Chandra. Multiview graph dual-attention deep learning and contrastive learning for multi-criteria recommender systems. *Expert Systems with Applications*, 291:128378, 2025.

[6] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. Recommender systems with social regularization. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, page 287–296, New York, NY, USA, 2011. Association for Computing Machinery.

[7] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, page 135–142, New York, NY, USA, 2010. Association for Computing Machinery.

[8] Duantengchuan Li, Jiayao Lu, Zhihao Wang, Jingxiong Wang, Xiaoguang Wang, Fobo Shi, and Yu Liu. Recommender system based on noise enhancement and multi-view graph contrastive learning. *Applied Soft Computing*, 177:113220, 2025.

[9] Syed Tauhid Ullah Shah, Fazlullah Khan, Shirin Yamani, Ryan Alturki, Foziah Gazzawe, and Muhammad Imran Razzak. Dsrs: Delight sequential recommender system. *Engineering Applications of Artificial Intelligence*, 142:109936, 2025.

[10] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. Graph neural networks for social recommendation. In *The World Wide Web Conference*, WWW '19, page 417–426, New York, NY, USA, 2019. Association for Computing Machinery.

[11] Le Wu, Junwei Li, Peijie Sun, Richang Hong, Yong Ge, and Meng Wang. Diffnet++: A neural influence and interest diffusion network for social recommendation, 2021.

[12] Yi Yang, Shaopeng Guan, and Xiaoyang Wen. Enhancing robustness in implicit feedback recommender systems with subgraph contrastive learning. *Information Processing Management*, 62(1):103962, 2025.

[13] Jie Wang, Ziang Niu, Zheng Wang, Liaoyuan Tang, Bo Yan, Rong Wang, and Feiping Nie. Selective-relaxed contrastive learning for hyperspectral image classification with noisy labels. *Pattern Recognition*, 171:112298, 2026.

[14] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020.

[15] Wang Guang and Liu Jihui. Adaptive memory-augmented graph attention network for social recommendation. In *2025 IEEE 5th International*

Conference on Electronic Technology, Communication and Information (ICETCI), pages 133–138. IEEE, 2025.

[16] Haiying Li, Huihui Wang, Shunmei Meng, and Xingguo Chen. Graph contrastive learning fornbsp;multi-behavior recommendation. In *Advanced Data Mining and Applications: 20th International Conference, ADMA 2024, Sydney, NSW, Australia, December 3–5, 2024, Proceedings, Part VI*, page 34–48, Berlin, Heidelberg, 2024. Springer-Verlag.

[17] Yanjun Xu, Chunqi Tian, Wei Wang, and Lizhi Bai. Semantic analysis-based recommender system using sequential clustering and convolutional neural network. *Engineering Applications of Artificial Intelligence*, 161:112196, 2025.

[18] Xiuyuan Qin, Huanhuan Yuan, Pengpeng Zhao, Guanfeng Liu, Fuzhen Zhuang, and Victor S Sheng. Intent contrastive learning with cross subsequences for sequential recommendation. In *Proceedings of the 17th ACM international conference on web search and data mining*, pages 548–556, 2024.

[19] Yifan Wang, Fei Xiong, Zhiyuan Zhang, Shirui Pan, Liang Wang, and Hongshu Chen. Dual variational graph contrastive learning for social recommendation. *Knowledge-Based Systems*, 327:114132, 2025.

[20] Paulo Roberto de Souza and Frederico Araújo Durão. Exploiting social capital for improving personalized recommendations in online social networks. *Expert Systems with Applications*, 246:123098, 2024.

[21] Ataus Samad and Vandana Bhatia. Fedgclrec: Federated graph contrastive learning framework for social influence recommendations. *Expert Systems with Applications*, 297:129294, 2026.

[22] Le Wu, Peijie Sun, Yanjie Fu, Richang Hong, Xiting Wang, and Meng Wang. A neural influence diffusion model for social recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, page 235–244, New York, NY, USA, 2019. Association for Computing Machinery.

[23] Junliang Yu, Hongzhi Yin, Jundong Li, Min Gao, Zi Huang, and Lizhen Cui. Enhancing social recommendation with adversarial graph convolutional networks. *IEEE Transactions on knowledge and data engineering*, 34(8):3727–3739, 2020.

[24] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, page 173–182, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee.

[25] Qian Li, Xiangmeng Wang, Zhichao Wang, and Guandong Xu. Be causal: De-biasing social network confounding in recommendation. *ACM Transactions on Knowledge Discovery from Data*, 17(1):1–23, 2023.

[26] Yichen Liu, Qianqian Ren, Shengxi Fu, and Yong Liu. Kan-infused social recommendation: A contrastive graph learning approach with bidirectional feature fusion. *Information Fusion*, 125:103448, 2026.

[27] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pages 1294–1303, 2022.

[28] Menghan Wang, Xiaolin Zheng, Yang Yang, and Kun Zhang. Collaborative filtering with social exposure: A modular approach to social recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[29] Wenjie Wang, Fuli Feng, Xiangnan He, Liqiang Nie, and Tat-Seng Chua. Denoising implicit feedback for recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, WSDM '21, page 373–381, New York, NY, USA, 2021. Association for Computing Machinery.

[30] Wenjie Wang, Fuli Feng, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*,

SIGIR '21, page 1288–1297, New York, NY, USA, 2021. Association for Computing Machinery.

[31] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. Self-supervised graph learning for recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 726–735, New York, NY, USA, 2021. Association for Computing Machinery.

[32] Zihan Lin, Changxin Tian, Yupeng Hou, and Wayne Xin Zhao. Improving graph collaborative filtering with neighborhood-enriched contrastive learning. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 2320–2329, New York, NY, USA, 2022. Association for Computing Machinery.

[33] Jiarun Fu, Rong Gao, Yonghong Yu, Jia Wu, Jing Li, Donghua Liu, and Zhiwei Ye. Contrastive graph learning long and short-term interests for poi recommendation. *Expert Systems with Applications*, 238:121931, 2024.

[34] Wenhan Li, Zhenzhong Chen, Jiangong Wang, Taijun Liu, Hua Chen, and Gaoming Xu. A cross-architecture masked contrastive learning framework for few-shot underwater acoustic target classification. *Knowledge-Based Systems*, 328:114252, 2025.

[35] Yuhan Xia, Yufei Tang, Bohang Yang, Chenghao Liu, and Qian Tao. Multi-relation graph contrastive learning with adaptive strategy for social recommendation. *Neurocomputing*, 624:129448, 2025.

[36] Fangjing Li, Zhihai Wang, Diping Wang, Haiyang Liu, and Xinxin Ding. Dpcac: Effective multi-scale graph contrastive learning via dual-perspective clustering and adaptive contrast. *Knowledge-Based Systems*, 328:114247, 2025.

[37] Xiang Li, Changsheng Shui, Zhongying Zhao, Junyu Dong, and Yanwei Yu. Multi-channel hypergraph contrastive learning for matrix completion. *ACM Trans. Inf. Syst.*, September 2025. Just Accepted.

[38] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, YongDong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution

network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, page 639–648, New York, NY, USA, 2020. Association for Computing Machinery.

[39] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[40] Junliang Yu, Hongzhi Yin, Jundong Li, Qinyong Wang, Nguyen Quoc Viet Hung, and Xiangliang Zhang. Self-supervised multi-channel hypergraph convolutional network for social recommendation. In *Proceedings of the Web Conference 2021*, WWW '21, page 413–424, New York, NY, USA, 2021. Association for Computing Machinery.

[41] Junliang Yu, Hongzhi Yin, Min Gao, Xin Xia, Xiangliang Zhang, and Nguyen Quoc Viet Hung. Socially-aware self-supervised tri-training for recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 2084–2092, New York, NY, USA, 2021. Association for Computing Machinery.

[42] Junliang Yu, Xin Xia, Tong Chen, Lizhen Cui, Nguyen Quoc Viet Hung, and Hongzhi Yin. Xsimgcl: Towards extremely simple graph contrastive learning for recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 36(2):913–926, 2024.