Dual-View Disentangled Multi-Intent Learning for Enhanced Collaborative Filtering

SHANFAN ZHANG, Xi'an Key Laboratory of Social Intelligence and Complexity Data Processing, School of Software, Xi'an Jiaotong University, China

YONGYI LIN, School of Mathematics and Statistics, Xi'an Jiaotong University, China

YUAN RAO*, Xi'an Key Laboratory of Social Intelligence and Complexity Data Processing, School of Software, Xi'an Jiaotong University, China

CHENLONG ZHANG, Xi'an Key Laboratory of Social Intelligence and Complexity Data Processing, School of Software, Xi'an Jiaotong University, China

Disentangling user intentions from implicit feedback has become a promising strategy to enhance recommendation accuracy and interpretability. While prior methods decompose user and item representations into multiple latent subspaces, they often model intentions independently and lack explicit supervision, thus failing to capture the joint semantics that drive user–item interactions. To address these limitations, we propose *DMICF*, a unified framework that explicitly models interaction-level intent alignment while leveraging structural signals from both user and item perspectives. *DMICF* adopts a dual-view architecture that jointly encodes user–item interaction graphs from both sides, enabling bidirectional information fusion. This design enhances robustness under data sparsity by allowing the structural redundancy of one view to compensate for the limitations of the other. To model fine-grained user–item compatibility, *DMICF* introduces an intent interaction encoder that performs sub-intent alignment within each view, uncovering shared semantic structures that underlie user decisions. This localized alignment enables adaptive refinement of intent embeddings based on interaction context, thus improving the model's generalization and expressiveness, particularly in long-tail or cold-start scenarios. Furthermore, *DMICF* integrates an intent-aware scoring mechanism that aggregates compatibility signals from matched intent pairs across user and item subspaces, enabling personalized prediction grounded in semantic congruence rather than entangled representations. To facilitate semantic disentanglement, we design a discriminative training signal via multi-negative sampling and softmax normalization, which pulls together semantically aligned intent pairs while pushing apart irrelevant or noisy ones. Extensive experiments demonstrate that *DMICF* consistently delivers robust performance across datasets with diverse interaction distributions. Qualitative and quantitative results confirm that *DMICF* effectively disentangles interaction intents and automatically distinguishes between collective behavioral patterns and individual-specific preferences.

CCS Concepts: • **Do Not Use This Code** → **Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

---

*Corresponding author.

Authors' Contact Information: Shanfan Zhang, zhangsfxajtu@gmail.com, Xi'an Key Laboratory of Social Intelligence and Complexity Data Processing, School of Software, Xi'an Jiaotong University, Xi'an, Shaanxi, China; Yongyi Lin, linyongyi@stu.xjtu.edu.cn, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi, China; Yuan Rao, raoyuan@mail.xjtu.edu.cn, Xi'an Key Laboratory of Social Intelligence and Complexity Data Processing, School of Software, Xi'an Jiaotong University, Xi'an, Shaanxi, China; Chenlong Zhang, zcl2023@stu.xjtu.edu.cn, Xi'an Key Laboratory of Social Intelligence and Complexity Data Processing, School of Software, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

---

1 INTRODUCTION

Recommendation systems leverage users' implicit feedback to infer personalized preferences and retrieve content aligned with individual interests. By mitigating information overload, they have become integral to real-world applications such as short video platforms, news delivery, and e-commerce services [33]. Among various paradigms, Collaborative Filtering (CF) methods have shown strong performance by modeling high-order user–item connectivity via message passing over bipartite graphs (e.g., *LightGCN* [6], *ClusterGCF* [13], and *LightCCF* [35]). However, most existing approaches overlook the fact that user–item interactions are often driven by multiple latent, interdependent intent factors. In real-world scenarios, user behaviors are shaped by diverse, context-specific motivations—e.g., purchasing swimsuits and sunscreen may imply plans for a seaside vacation. Without properly disentangling such latent intents, user embeddings may conflate heterogeneous behavioral signals, degrading both recommendation accuracy and interpretability.

Capturing users' latent intentions facilitates semantically richer recommendations—such as suggesting swimming goggles tailored to specific needs. Nonetheless, these intentions are typically implicit, sparsely manifested in observed behaviors, and confounded by noisy interactions, posing significant challenges for intent-aware modeling. Recent efforts (e.g., *DCCF*[19], *BIGCF*[32], and *IPCCF*[10]) have proposed disentanglement-based frameworks that decompose user and item representations into multiple intent-specific subspaces. Despite these advances, two major challenges persist, limiting the effectiveness and interpretability of current intent-disentangled models:

Problem 1: Recent intent-aware recommendation models have made progress by decomposing user and item preferences into multiple latent intents. However, they frequently model user and item intents independently, neglecting the explicit modeling of semantic alignment that fundamentally governs user–item interactions. In practice, item intents correspond to latent motivations that align with specific user interests [15, 19, 32], reflecting a joint semantic structure influencing interaction likelihood. For example, a user with a strong preference for horror films is more likely to engage with titles such as *The Shining*, *Silent Hill*, or *Halloween*, while showing limited interest in emotionally warm films like *Flipped*. This illustrates the necessity of modeling intent at the granularity of user–item pairs, where alignment between user and item latent intents yields a more faithful and interpretable explanation of behavioral patterns. Explicitly modeling this collaborative intent structure enables the capture of complex cross-entity semantic dependencies often missed by isolated intent factorization. Moreover, aligning user and item intents at the interaction level promotes effective information sharing under sparse data conditions, thereby alleviating sparsity and improving generalization to unseen user–item pairs. Consequently, recommendation systems that incorporate such collaborative intent modeling are better positioned to deliver more accurate predictions and provide interpretable insights into user behavior.

Problem 2: A fundamental limitation of existing disentanglement-based recommendation models is the lack of explicit and semantically grounded supervision during the intent separation process. Typically, such models construct latent intent representations by aggregating intent signals from neighboring nodes through iterative message passing on user–item interaction graphs. Although this approach captures multi-hop relational dependencies, it often entangles heterogeneous semantics by conflating structural proximity with diverse intent patterns, thereby obscuring intent-specific representations critical for accurate recommendation. The absence of targeted supervision significantly hinders

the model's ability to disentangle and distinguish fine-grained user preferences and item characteristics, especially under conditions of data sparsity and complex behavioral patterns. Addressing this challenge requires integrating explicit supervisory signals or inductive biases that guide the learning of disentangled representations in a semantically meaningful way. Potential solutions include designing loss functions that promote orthogonality or mutual exclusivity among latent intents, or implementing behaviorally grounded alignment mechanisms to enforce semantic consistency across intent subspaces. While recent work, *IPCCF*, has attempted alignment between node embeddings derived from structural and intent-aware propagation, the lack of grounding in external semantic labels weakens these alignment objectives, limiting their effectiveness in defining clear disentanglement boundaries. As a result, learned intent representations often exhibit restricted discriminability and interpretability, undermining both recommendation accuracy and model transparency.

To address the limitations inherent in existing disentangled recommendation models, this work proposes the **D**ual-View Disentangled **M**ulti-**I**ntent Learning for Enhanced **C**ollaborative **F**iltering (*DMICF*) framework. *DMICF* adopts a dual-view modeling strategy that extracts and integrates local structural information from both user and item perspectives. Information exchange between these two views is facilitated through a dual-view fusion mechanism, enabling the model to utilize the structural context of one side to support inference when the data from the other side is limited. This design significantly enhances the model's robustness and generalization capability, particularly in sparse scenarios. Recognizing that user decisions often arise from a complex interplay of latent intentions, *DMICF* introduces an intention interaction encoder that explicitly aligns multi-granularity intent vectors of users and items within each structural view. This fine-grained alignment mechanism captures the semantic dependencies and joint distributions underlying user–item interactions, thereby facilitating a more accurate and interpretable understanding of behavioral patterns. By conducting alignment within each view—rather than merely aggregating global representations—*DMICF* enables the model to disentangle sub-intent-level semantics and to adaptively refine intent embeddings in a context-aware manner. Such a design not only expands the expressiveness of the latent intention space but also enhances model generalization, particularly under challenging conditions such as cold-start users or long-tail items. Furthermore, the localized alignment process supports dynamic modulation of intent contributions based on interaction context, allowing the model to flexibly emphasize relevant intent dimensions while suppressing irrelevant or noisy factors. These capabilities collectively strengthen both the robustness and interpretability of the model, offering a principled approach to semantic disentanglement in collaborative filtering scenarios.

Furthermore, *DMICF* introduces a novel scoring mechanism grounded in intent-level interaction modeling. In this paradigm, recommendation scores are computed by aggregating the compatibility signals from matched user–item intent pairs, with each pair aligned along a distinct sub-preference dimension. This fine-grained alignment ensures that predictions are governed by the semantic congruence between user and item intents, rather than by their holistic or entangled embeddings. By modeling intent interactions at the subspace level, the framework captures subtle variations in user interests and item attributes that are often overlooked by conventional coarse-grained embedding methods. Building upon this intent-aware foundation, *DMICF* incorporates a multi-negative sampling strategy to inject contrastive supervision into the learning process. For each observed user–item interaction, a set of negative candidates is uniformly sampled from the global item pool, regardless of the user's interaction history. The predicted scores for both positive and negative pairs are then normalized using a softmax-based transformation, which amplifies the relative significance of the positive instance while attenuating the influence of semantically irrelevant or noisy negatives. This scoring formulation induces a dual supervisory signal: positive interactions exert attractive gradients in the intent embedding space, encouraging alignment along meaningful semantic dimensions, while negative samples introduce repulsive

forces that push apart incompatible or spurious components. Through this bidirectional optimization, *DMICF* facilitates the disentanglement of user and item intents across fine-grained subspaces. Consequently, the model not only learns more interpretable and structured intent representations, but also enhances recommendation accuracy by suppressing redundant or entangled latent signals that typically degrade performance in multi-intent collaborative filtering scenarios.

The main contributions of this work are as follows:

- We introduce a novel intention alignment framework that explicitly models and aligns multi-granularity latent intentions between users and items. This design captures fine-grained interaction semantics and facilitates interpretable preference matching.
- We develop *DMICF*, a dual-view disentangled multi-intent recommendation model that integrates structural context from both user and item perspectives. The model enhances fine-grained intent alignment through a dual-view fusion design and an intent-aware scoring mechanism, reinforced by a multi-negative sampling strategy that promotes semantic contrast and discriminative learning.
- Comprehensive experiments across three real-world datasets validate the effectiveness of *DMICF*, showing competitive or superior performance compared to ten state-of-the-art methods. Further qualitative analysis, including visualizations of the learned intention representations, demonstrates that *DMICF* can effectively disentangles latent intentions and autonomously differentiates between collective and individual intent semantics.

The remainder of this paper is structured as follows. Section 2 reviews the related literature. Section 3 introduces the necessary preliminaries and provides a detailed description of the proposed model. Section 4 presents the experimental setup and evaluates the model through comprehensive numerical simulations. Finally, Section 5 summarizes the main findings and concludes the paper.

2 RELATED WORK

2.1 GNN-Based Recommendation Models

In recent years, Graph Convolutional Networks (GCNs) have attracted considerable attention in recommendation systems owing to their capability to learn expressive node representations from non-Euclidean user–item interaction graphs. For example, *NGCF* [23] employs standard GCN architectures to capture multi-hop connectivity patterns between users and items. Nevertheless, subsequent research has identified that nonlinear activation functions and feature transformation layers may adversely affect performance in ID-based collaborative filtering scenarios. To address these limitations, *GCCF* [3] introduces a linear propagation scheme that excludes nonlinearities, while *LightGCN* [6] further simplifies the architecture by removing self-connections, nonlinear activation functions, and feature transformation matrices, thereby achieving a more efficient yet effective model design. Despite these methodological advancements, GCN-based recommendation models continue to face the significant challenge that the pervasive sparsity of user–item interaction data limits the richness and generalizability of learned representations.

To address the challenges of data sparsity and to improve model robustness, recent recommendation studies have increasingly adopted self-supervised learning techniques. These methods typically aim to enhance representation quality by maximizing mutual information across multiple views or subgraphs. For instance, *SimGCL*[30] eliminates graph augmentations by injecting uniform noise into the embedding space, producing contrastive views that enhance representation uniformity and discriminability. *RGCL*[21] and *LGMRec*[5] enrich contrastive signals by generating multi-level views that incorporate both local and global features. *SimRec*[27] combines knowledge distillation and contrastive learning for adaptive structural transfer from teacher to student models. *STERLING*[9] maximizes mutual information

across co-clusters to capture intra- and inter-type semantics without relying on negative sampling. AU^+ [17] mitigates overfitting to sampling variance by injecting self-supervised contrastive signals through a minimal perturbation mechanism, enhancing generalization in alignment and uniformity objectives. *PAAC* [1] mitigates popularity bias via supervised alignment of unpopular item embeddings and popularity-aware re-weighted contrastive learning. Despite these advances, existing contrastive learning approaches often overlook the complex and multifaceted nature of user intent embedded in interaction behaviors. Most methods generate contrastive views from structural or topological variations but fail to explicitly disentangle distinct user intentions or interests. This entanglement leads to oversimplified user representations, making it difficult to capture nuanced preferences and undermining the potential benefits of contrastive objectives in recommendation.

2.2 Recommendation with Disentangled Representations

Learning disentangled representations of latent user intents from implicit feedback has emerged as a pivotal research direction in recommender systems [8]. Numerous methods have been developed to capture high-level user intents and enhance recommendation effectiveness. For example, *MacridVAE* [15] utilizes variational autoencoders to infer latent intent distributions, while *DGCF* [25] extends this approach by incorporating intent disentanglement within a graph neural network framework via embedding splitting. To exploit auxiliary information, *DisenHAN* [26] employs heterogeneous graph attention networks for disentangling user and item representations. *MIDGN*[36] and *LGD-GCN*[4] disentangle user interaction intents by jointly modeling global collaborative signals and local neighborhood structures. Knowledge-guided approaches such as *KGIN* [24], *KDR*[16] and *DIKGNN*[22] incorporate semantic signals from knowledge graphs to supervise intent disentanglement, offering better interpretability. *GDCF*[31] introduces geometric intent modeling by linking user intents with diverse conceptual geometries. *GDCCDR* [14] designs a parameter-free adaptive filter to distill informative interactions, enhancing the granularity of disentangled user intents. *UDITSR* [33] leverages user queries to supervise inherent and demand intent modeling via dual-intent translation in a unified semantic space. Recently, contrastive learning techniques have been integrated into intent modeling to further enhance representation robustness. *DCCF* [19], for instance, strengthens self-supervised signals by learning disentangled representations enriched with global contextual information. *BIGCF* [32] addresses both the individuality and collectivity of latent intents underlying user-item interactions to improve collaborative filtering performance. *IHGCL* [20] incorporates contrastive learning on heterogeneous graphs by leveraging meta-paths to capture latent user−item intents, employing a dual contrastive objective and a bottleneck autoencoder to reduce noise. *IPCCF* [10] applies contrastive learning to align structure-based and intent-aware node representations, thereby enhancing model robustness and mitigating bias. Despite these advances, existing disentangled recommendation models commonly exhibit two critical limitations. First, they lack a comprehensive understanding of the collaborative relationship between user and item intents, which often results in fragmented intent representations that fail to faithfully capture the intrinsic nature of user−item interactions. This fragmentation impairs the model's ability to accurately reflect the joint semantics driving recommendation outcomes. Second, most methods do not incorporate explicit supervision signals grounded in authentic semantic labels during the intent disentanglement process. The absence of such explicit guidance hampers the formation of clear and well-separated intent boundaries, diminishing the model's discriminative power among diverse latent intent factors and consequently limiting the granularity and precision of recommendation results.
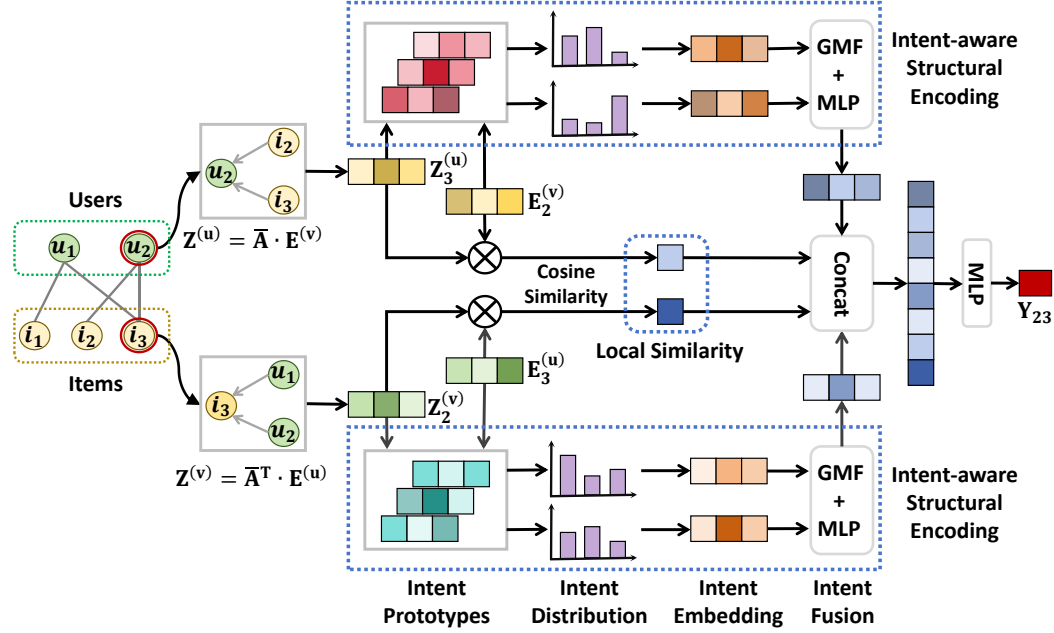
Fig. 1. Overview of the *DMICF* framework, featuring dual-view structural modeling, intent-aware encoding, and fusion modules to enable multi-granularity intent modeling and end-to-end semantic interaction prediction.

3 METHODOLOGY

In a typical recommendation setting, the problem domain consists of a set of users $\mathcal{U} = \{u_1, u_2, \ldots, u_M\}$ and a set of items $\mathcal{I} = \{i_1, i_2, \ldots, i_N\}$. User interactions with items are commonly represented by a binary interaction matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, where each entry $\mathbf{A}_{ij} \in \{0, 1\}$ indicates whether user u_i has interacted with item v_j. These interactions can also be modeled as edges $\mathcal{E}$ in a bipartite graph $\mathcal{G} = (\mathcal{U} \cup \mathcal{I}, \mathcal{E})$. The recommendation task is to predict the likelihood of interactions based on a training subset of observed edges $\mathcal{E}_{train} \subseteq \mathcal{E}$. Formally, this corresponds to estimating the conditional probability $\hat{\mathbf{A}}_{ij} = \mathbb{P}\left(\langle u_i, v_j \rangle \in \mathcal{E} \mid \mathcal{E}_{train}\right)$, where $\hat{\mathbf{A}}_{ij}$ denotes the predicted preference score reflecting the likelihood that u_i will interact with v_j.

The overall architecture of *DMICF* is illustrated in Fig. 1. Designed to address challenges such as semantic sparsity and intent entanglement, *DMICF* integrates three mutually reinforcing components that together provide a comprehensive solution for enhanced collaborative filtering:

- **Dual-view structural exploration.** To capture localized semantics from both sides of user–item interactions, *DMICF* constructs personalized subgraphs centered on users and items from the global interaction graph. This dual-view design allows the model to extract complementary structural contexts, uncovering view-specific semantic patterns and relational signals.

- **Intent Interaction Encoding.** Operating on the dual-view subgraphs, *DMICF* infers latent intent distributions for users and items to model their multifaceted characteristics. A dedicated intent interaction encoder is introduced to explicitly capture the alignment and interdependence between these distributions. This mechanism facilitates

the modeling of high-order collaborative signals and uncovers fine-grained semantic compatibilities that go beyond surface-level associations.

- **Cross-View Fusion.** To unify the interaction semantics captured from both perspectives, *DMICF* aggregates structure-aware and intent-aware representations into a joint embedding space. This unified embedding is then passed to a prediction layer for rating estimation, achieving an end-to-end alignment between structural encoding and semantic-level preference matching.

3.1 Dual-View Structural Awareness and Intent Interaction Encoding

A foundational inductive bias in collaborative filtering is the structural consistency assumption: users with analogous interaction histories tend to occupy similar neighborhoods in the user–item bipartite graph and should, accordingly, be represented by comparable embeddings. This assumption extends symmetrically to items that co-occur in overlapping user groups. To effectively leverage this structural prior, *DMICF* introduces a Dual-View Structural Modeling module, which utilizes a lightweight, one-hop propagation scheme inspired by LightGCN. This design captures localized relational semantics while avoiding the oversmoothing and computational inefficiencies commonly associated with deep message passing.

Formally, given initial user and item embeddings $\mathbf{E}^u \in \mathbb{R}^{M \times d}$ and $\mathbf{E}^v \in \mathbb{R}^{N \times d}$, the model propagates relational information via a normalized adjacency matrix $\bar{\mathbf{A}} = \mathbf{D}_u^{-1/2} \cdot \mathbf{A} \cdot \mathbf{D}_v^{-1/2} \in \mathbb{R}^{M \times N}$, where $\mathbf{D}_u \in \mathbb{R}^{M \times M}$ and $\mathbf{D}_v \in \mathbb{R}^{N \times N}$ denote the diagonal degree matrices of users and items, respectively. Structural representations are computed as:

$$\mathbf{Z}^u = \bar{\mathbf{A}} \cdot \mathbf{E}^v, \quad \mathbf{Z}^v = \bar{\mathbf{A}}^T \cdot \mathbf{E}^u \tag{1}$$

These structure-aware embeddings $\mathbf{Z}^u$ and $\mathbf{Z}^v$ serve as semantic priors for modeling latent intent distributions. Intuitively, users embedded in similar topological contexts are more likely to exhibit comparable latent intents, and the same applies to items. This induces a form of implicit semantic regularization, where structural coherence encourages the disentanglement of intent representations even in the absence of explicit labels. Consequently, the model learns interpretable and semantically meaningful latent factors, crucial for robust intent-aware recommendation.

Moreover, the dual-view architecture enhances the system's adaptability under sparse or cold-start conditions. When one modality (e.g., new users) suffers from limited interaction data, the structural signal from the other modality (e.g., frequently interacted items) can still contribute valuable information. This cross-perspective compensation mechanism promotes generalization and stability, particularly in real-world scenarios characterized by long-tail distributions and data sparsity.

To capture the inherently multi-intent nature of user behavior, *DMICF* explicitly constructs latent intent distributions from both user and item perspectives. This design is grounded in the observation that user–item interactions are typically driven by heterogeneous and entangled factors—such as the diversity of user intentions and the contextual interpretations of items. To model this complexity, *DMICF* introduces two learnable sets of latent intent prototypes, denoted as $\mathbf{C}^u = \left[\mathbf{c}_1^u, \mathbf{c}_2^u, \ldots, \mathbf{c}_K^u \right] \in \mathbb{R}^{K \times d}$ and $\mathbf{C}^v = \left[\mathbf{c}_1^v, \mathbf{c}_2^v, \ldots, \mathbf{c}_K^v \right] \in \mathbb{R}^{K \times d}$, serving as semantic anchors in the latent space (*DCCF*, *BIGCF*, *IPCCF*). From the user perspective, both users and items are softly assigned to the same set of user-intent prototypes $\mathbf{C}^u$, facilitating the modeling of how different users interpret item semantics. Specifically:

- The structure-aware embedding $\mathbf{z}_i^u$ of user i is used to compute the user's intent distribution as:

$$\mathbf{p}_i^u = \left[cos\left(\mathbf{z}_i^u, \mathbf{c}_1^u\right), cos\left(\mathbf{z}_i^u, \mathbf{c}_2^u\right), \ldots, cos\left(\mathbf{z}_i^u, \mathbf{c}_K^u\right) \right] \in \mathbb{R}^K \tag{2}$$

$$where, cos\left(\mathbf{z}_i^u, \mathbf{c}_k^u\right) = \frac{\mathbf{z}_i^u \cdot \mathbf{c}_k^u}{\left\|\mathbf{z}_i^u\right\| \cdot \left\|\mathbf{c}_k^u\right\|} \tag{3}$$

- Simultaneously, the initial embedding $\mathbf{e}_j^v$ of item j is compared with the same prototype set $\mathbf{C}^u$ to derive its user-view intent distribution:

$$\tilde{\mathbf{p}}_j^u = \left[cos\left(\mathbf{e}_j^v, \mathbf{c}_1^u\right), cos\left(\mathbf{e}_j^v, \mathbf{c}_2^u\right), \ldots, cos\left(\mathbf{e}_j^v, \mathbf{c}_K^u\right) \right] \in \mathbb{R}^K \tag{4}$$

In parallel, from the item perspective, the model symmetrically captures how items are perceived by diverse user groups via the prototype set $\mathbf{C}^v$:

- The structure-aware item embedding $\mathbf{z}_j^v$ is aligned with item-side prototypes:

$$\mathbf{q}_j^v = \left[cos\left(\mathbf{z}_j^v, \mathbf{c}_1^v\right), cos\left(\mathbf{z}_j^v, \mathbf{c}_2^v\right), \ldots, cos\left(\mathbf{z}_j^v, \mathbf{c}_K^v\right) \right] \in \mathbb{R}^K \tag{5}$$

- Correspondingly, the initial user embedding $\mathbf{e}_i^u$ is used to compute the item-view user intent distribution:

$$\tilde{\mathbf{p}}_i^v = \left[cos\left(\mathbf{e}_i^u, \mathbf{c}_1^v\right), cos\left(\mathbf{e}_i^u, \mathbf{c}_2^v\right), \ldots, cos\left(\mathbf{e}_i^u, \mathbf{c}_K^v\right) \right] \in \mathbb{R}^K \tag{6}$$

The shared-prototype strategy within each perspective enhances the model by significantly reducing the number of learnable parameters, thus improving training efficiency and mitigating redundancy. Simultaneously, it enforces a unified latent intent space that promotes stronger semantic alignment between user and item representations, enabling more coherent and interpretable disentanglement of multi-intent factors.

To bridge the semantic gap between discrete latent intent distributions and the continuous representations required for intent-aware matching, *DMICF* employs four dedicated multi-layer perceptrons (MLPs) to project user- and item-specific intent distributions into semantically expressive embedding vectors. Formally, the model computes:

$$\mathbf{h}_i^u = MLP_{user-user}\left(\mathbf{p}_i^u\right) \in \mathbb{R}^{d'}, \quad \tilde{\mathbf{h}}_j^u = MLP_{user-item}\left(\tilde{\mathbf{p}}_j^u\right) \in \mathbb{R}^{d'}$$

$$\mathbf{r}_j^v = MLP_{item-item}\left(\mathbf{q}_j^v\right) \in \mathbb{R}^{d'}, \quad \tilde{\mathbf{r}}_i^v = MLP_{item-user}\left(\tilde{\mathbf{p}}_i^v\right) \in \mathbb{R}^{d'} \tag{7}$$

Each MLP is independently parameterized and concludes with a sigmoid activation function, constraining each output dimension to the interval $[0, 1]$. This enables the model to softly activate multiple latent intents per user or item, thereby capturing the inherent multi-intent nature of user preferences and item characteristics. Such flexibility is critical for modeling fine-grained behaviors and high-variance semantic signals.

To further enhance representational interpretability and robustness, an ℓ_1-normalization is applied to user-side embeddings $\mathbf{h}_i^u$ and $\tilde{\mathbf{r}}_i^v$, projecting them onto a probability simplex:

$$\mathbf{h}_{il}^u = \frac{\mathbf{h}_{il}^u}{\sum_{t=1}^{d'} \mathbf{h}_{it}^u}, \quad \tilde{\mathbf{r}}_{il}^v = \frac{\tilde{\mathbf{r}}_{il}^v}{\sum_{t=1}^{d'} \tilde{\mathbf{r}}_{it}^v}, \quad \forall l \in \left[1, 2, \ldots, d'\right] \tag{8}$$

This normalization preserves the relative strength of each activated dimension, including low-activation intents that might otherwise be suppressed. By preventing dominant signals from overshadowing subtle preferences—especially under sparse interaction conditions—the model maintains sensitivity to nuanced intent overlaps and user-specific

variations. Consequently, this mechanism enhances personalization and diversity, while effectively alleviating popularity bias by preserving the representational integrity of less popular or weakly expressed interests.

After deriving intent-aware embeddings for both users and items under each semantic perspective, *DMICF* introduces a dual-perspective intent interaction encoder to model the fine-grained synergy between user preferences and item characteristics within the intent space. This component plays a critical role in transcending isolated representation learning, enabling the modeling of co-dependent semantic alignments between user and item intents. Specifically, for a given user i and item j, two symmetrical interaction encoders are instantiated to model their mutual intent interactions from dual perspectives. From the user perspective, the encoder fuses the user's intent embedding $\mathbf{h}_i^u$ with the item's user-view intent embedding $\tilde{\mathbf{h}}_j^u$, producing a joint representation $\mathbf{z}_{ij}^u \in \mathbb{R}^{d^*}$ that encapsulates the fused semantics:

$$\mathbf{z}_{ij}^u = InteractionEncoder\left(\mathbf{h}_i^u, \tilde{\mathbf{h}}_j^u\right) \tag{9}$$

Symmetrically, from the item's perspective, the encoder integrates the item's intent embedding $\mathbf{r}_j^v$ with the user's item-view intent embedding $\tilde{\mathbf{r}}_i^v$, yielding a counterpart representation $\mathbf{z}_{ij}^v \in \mathbb{R}^{d^*}$:

$$\mathbf{z}_{ij}^v = InteractionEncoder\left(\tilde{\mathbf{r}}_i^v, \mathbf{r}_j^v\right) \tag{10}$$

These two representations, $\mathbf{z}_{ij}^u$ and $\mathbf{z}_{ij}^v$, jointly capture the bidirectional alignment of user–item intent semantics. Their integration provides a holistic interaction signal that serves as input to the final prediction layer, enabling precise modeling of user preferences conditioned on both perspectives. To instantiate the interaction encoder, multiple fusion strategies can be employed:

- **Concatenation + MLP:** This straightforward yet expressive strategy concatenates user and item intent embeddings, followed by a multi-layer perceptron (MLP) to model their joint interaction:

$$\mathbf{z}_{ij}^u = MLP_{user-intent}\left(\left[\mathbf{h}_i^u; \tilde{\mathbf{h}}_j^u\right]\right), \quad \mathbf{z}_{ij}^v = MLP_{item-intent}\left(\left[\tilde{\mathbf{r}}_i^v; \mathbf{r}_j^v\right]\right) \tag{11}$$

- **Generalized Matrix Factorization (GMF [7]) + MLP:** To capture element-wise semantic interactions, GMF performs Hadamard product between paired embeddings, followed by a non-linear transformation:

$$\mathbf{z}_{ij}^u = MLP_{user-intent}\left(\mathbf{h}_i^u \odot \tilde{\mathbf{h}}_j^u\right), \quad \mathbf{z}_{ij}^v = MLP_{item-intent}\left(\tilde{\mathbf{r}}_i^v \odot \mathbf{r}_j^v\right) \tag{12}$$

  This design is more parameter-efficient than concatenation and better captures localized feature interactions while preserving model flexibility.

- **Cross-Attention Fusion [11]:** To model asymmetric and adaptive interactions, one viable strategy is to employ a cross-attention mechanism that dynamically computes interaction weights:

$$\mathbf{z}_{ij}^u = \text{Attn}\left(\mathbf{h}_i^u \mathbf{W}_Q^u, \tilde{\mathbf{h}}_j^u \mathbf{W}_K^u, \tilde{\mathbf{h}}_j^u \mathbf{W}_V^u\right), \quad \mathbf{z}_{ij}^v = \text{Attn}\left(\tilde{\mathbf{r}}_i^v \mathbf{W}_Q^v, \mathbf{r}_j^v \mathbf{W}_K^v, \mathbf{r}_j^v \mathbf{W}_V^v\right) \tag{13}$$

  where $\mathbf{W}_Q^u, \mathbf{W}_K^u, \mathbf{W}_V^u, \mathbf{W}_Q^v, \mathbf{W}_K^v, \mathbf{W}_V^v \in \mathbb{R}^{d' \times d^*}$ are learnable projection matrices, and the attention function is

$$\text{Attn}\left(\mathbf{Q}, \mathbf{K}, \mathbf{V}\right) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d^*}}\right)\mathbf{V} \tag{14}$$

  This formulation enhances the model's ability to capture intent-level dependencies through soft alignment. However, the additional parameterization introduces increased computational overhead and potential overfitting risks, particularly under sparse interaction scenarios.

3.2 Dual-Perspective Structural Fusion and Multi-Negative Optimization

To fully capture the dual-perspective semantics of user–item interactions, *DMICF* integrates structural information from both the user and item perspectives. Beyond generating disentangled intent-aware representations under each view—denoted as $\mathbf{z}_{ij}^u$ and $\mathbf{z}_{ij}^v$, respectively—the model further leverages local structural similarity to enhance fine-grained user–item matching. Concretely, *DMICF* computes the cosine similarity between each structure-aware embedding and its corresponding initial embedding on the same side, resulting in two scalar indicators:

$$\tilde{\mathbf{z}}_{ij}^u = cos\left(\mathbf{z}_i^u, \mathbf{e}_j^v\right), \quad \tilde{\mathbf{z}}_{ij}^v = cos\left(\mathbf{z}_j^v, \mathbf{e}_i^u\right) \tag{15}$$

which quantify the degree of local alignment between the neighborhood-enriched representations and the original semantic features. These similarity scores reflect micro-level structural coherence, enabling the model to account for personalized interpretation of interaction patterns.

To integrate these complementary signals, *DMICF* explores two structure fusion strategies:

- **Two-stage fusion:** The model first fuses the dual-perspective intent-aware embeddings $\mathbf{z}_{ij}^u$ and $\mathbf{z}_{ij}^v$ via an aggregation function-such as fusion strategies used by interaction encoder-to obtain a unified intent-level structural representation. This fused vector is then concatenated with the similarity scores $\tilde{\mathbf{z}}_{ij}^u$ and $\tilde{\mathbf{z}}_{ij}^v$ to yield the final structure-informed interaction encoding:

$$\mathbf{z}_{ij}^{struct} = \left[Fuse\left(\mathbf{z}_{ij}^u, \mathbf{z}_{ij}^v\right) \parallel \tilde{\mathbf{z}}_{ij}^u \parallel \tilde{\mathbf{z}}_{ij}^v\right] \tag{16}$$

- **Unified fusion:** Alternatively, the model directly concatenates all four signals—$\mathbf{z}_{ij}^u$, $\mathbf{z}_{ij}^v$, $\tilde{\mathbf{z}}_{ij}^u$, and $\tilde{\mathbf{z}}_{ij}^v$—into a unified structure representation:

$$\mathbf{z}_{ij}^{struct} = \left[\mathbf{z}_{ij}^u \parallel \mathbf{z}_{ij}^v \parallel \tilde{\mathbf{z}}_{ij}^u \parallel \tilde{\mathbf{z}}_{ij}^v\right] \tag{17}$$

This design preserves the raw information from both global and local views without early compression, potentially enhancing the model's flexibility in capturing nuanced user–item relationships.

Each fusion scheme offers distinct advantages. The late-fusion strategy emphasizes compactness and semantic abstraction, while early fusion retains a richer set of discriminative features at the cost of higher dimensionality. In practice, *DMICF* supports both configurations and selects the optimal strategy based on validation performance or downstream recommendation objectives.

Once the structure-informed interaction encoding $\mathbf{z}_{ij}^{struct}$ is obtained, *DMICF* proceeds to predict the user–item interaction score $\hat{\mathbf{A}}_{ij}$. The prediction is computed via a multi-layer perceptron: $\hat{\mathbf{A}}_{ij} = MLP_{predict}(\mathbf{z}_{ij}^{struct})$, which transforms the learned high-order interaction representation into a scalar relevance score.

DMICF adopts a multi-negative sampling strategy to enhance learning signals and strengthen embedding discrimination. This design choice is motivated by empirical findings from NeuMF [7], which indicate that training with only a single negative instance per positive interaction may not yield optimal performance. To address this limitation, *DMICF* forgoes the traditional pairwise Bayesian Personalized Ranking (BPR) loss [17, 18] in favor of a more direct and scalable optimization objective that can more effectively exploit multiple negative samples. Specifically, for each observed user–item interaction $\langle i, j \rangle \in \mathcal{E}_{train}$, a negative candidate set $\mathcal{N}_i$ is constructed by uniformly sampling S items from the entire item set, regardless of their historical interaction status with user i. This naïve sampling scheme is simple yet effective, and allows for flexible batch-wise implementation. To ensure that the predicted interaction score for a true link is significantly higher than that for negative samples, we compute the softmax-normalized probability of a true

interaction as:

$$\hat{\mathbf{A}}_{ij} = \frac{exp\left(\hat{\mathbf{A}}_{ij}/\tau\right)}{exp\left(\hat{\mathbf{A}}_{ij}/\tau\right) + \sum_{l \in \mathcal{N}_i} exp\left(\hat{\mathbf{A}}_{il}/\tau\right)} \tag{18}$$

where τ is a temperature hyperparameter that controls the sharpness of the probability distribution. The final training objective minimizes the squared error between the predicted scores and binary targets:

$$\mathcal{L}_{loss} = \sum_{\langle i,j \rangle \in \mathcal{E}_{train}} \left[\left(1 - \hat{\mathbf{A}}_{ij}\right)^2 + \sum_{l \in \mathcal{N}_i} \left(0 - \hat{\mathbf{A}}_{il}\right)^2\right] \tag{19}$$

Importantly, the use of zero as the ground truth label for negative samples reflects their randomly sampled nature: since $\mathcal{N}_i$ is uniformly drawn from the entire item set, it may contain false negatives (i.e., items the user actually interacted with), but the model is encouraged to learn general discrimination without overfitting to sampling noise.

By incorporating multiple negative samples per interaction, *DMICF* introduces a contrastive supervision mechanism that sharpens the discrimination between relevant and irrelevant intent signals. For each positive pair, the prediction $\hat{\mathbf{A}}_{ij}$ is optimized to approximate 1, while predictions for negatives $\hat{\mathbf{A}}_{ij}$ are simultaneously suppressed toward 0. This dual supervision enforces directional gradients on the intent subspaces: positive interactions attract intent embeddings toward the semantic core of user preference, while negative ones repel incompatible or noisy components. Through this process, *DMICF* is implicitly guided to perform intent disentanglement across both the user and item domains. The model learns to emphasize intent dimensions that are consistently activated by genuine preferences, and down-weight those that are frequently associated with negative or spurious signals. This not only enhances the interpretability of learned embeddings, but also improves recommendation performance by filtering out entangled, irrelevant, or conflicting intents that commonly plague multi-intent systems.

3.3 Time Complexity Analysis

We instantiate *DMICF* with a **GMF + MLP** design for the intent interaction encoder and a **Unified fusion** scheme for dual-view aggregation. Each training instance (including S negative samples) is processed through three sequential stages—intent prototype matching, intent embedding encoding, and intent interaction encoding—followed by dual-view fusion. The intent prototype matching phase incurs a cost of $O(4Kd)$, where cosine similarities are computed between users/items and K prototypes of dimension d in two semantic views. The resulting intent distributions are then passed through four MLPs ($MLP_{user-user}$, $MLP_{user-item}$, $MLP_{item-item}$ and $MLP_{item-user}$) of structure $[K, h, d']$, yielding an encoding cost of $O(4Khd')$. These embeddings are further processed by two deep MLPs ($MLP_{user-intent}$ and $MLP_{item-intent}$) of structure $[d', h_1, h_2, d^*]$ to model user-intent and item-intent interactions, resulting in a complexity of $O\left(2 \times (d'h_1 + h_1h_2 + h_2d^*)\right)$, which is dominated by $O\left(h_1h_2\right)$. Finally, the outputs from dual-view encoding streams are aggregated via a unified fusion network ($MLP_{predict}$) with structure $[d^*, h_3, 1]$, yielding a prediction cost of $O\left(d^*h_3\right)$. Assuming h_1h_2 is the dominant term, the total complexity per instance is $O\left((S+1)h_1h_2\right)$. Thus, for the entire training set with $|\mathcal{E}_{train}|$ interactions, the overall complexity is $O\left(|\mathcal{E}_{train}|(S+1)h_1h_2\right)$. This linear complexity ensures that *DMICF* is computationally scalable to large-scale recommendation tasks when appropriately parallelized.

3.4 Comparative Analysis on Intent Disentanglement

One of the key innovations of *DMICF* lies in its interaction-grounded, dual-view disentanglement mechanism, which enables a more semantically aligned and context-aware modeling of latent user–item intents. Unlike existing approaches that primarily focus on enhancing overall embedding consistency, *DMICF* explicitly constructs multi-granular intent

representations from both the user and item perspectives, and aligns them within a unified latent space informed by localized structural semantics. By introducing shared prototype spaces and dual-view intent encoders, the model captures bidirectional preference semantics at a fine-grained level, thereby uncovering nuanced intent patterns that are closely tied to the observed interaction behaviors.

In contrast, although *IPCCF* [10] incorporates contrastive regularization across multiple propagation paths, its intent modeling remains implicit and indirect. The model promotes consistency between representations propagated via different paths but does not offer an explicit mechanism for isolating or aligning intent-level semantics between users and items. As a result, it may fail to capture the subtle and heterogeneous motivations that underlie real-world user–item interactions.

BIGCF [32], while effective in leveraging high-order structural signals through graph-level contrastive learning, primarily focuses on distributional modeling via Gaussian embeddings. Its design aggregates both individual and collective preferences into a shared representation, but without explicit intent supervision, the resulting embeddings tend to entangle diverse semantics, thereby limiting interpretability and intent separability, especially in complex interaction graphs with long-tail behaviors.

DCCF [19] improves intent disentanglement by incorporating subspace-level contrastive constraints and learnable semantic masks, encouraging latent factors to specialize. However, its disentanglement relies on heuristic or prototype-level subspace supervision, without fully exploiting the mutual reinforcement of dual structural contexts. Moreover, the lack of direct alignment between the intent distributions of users and items constrains its ability to precisely capture bidirectional semantic dependencies inherent in collaborative filtering.

By comparison, *DMICF* achieves a more expressive and interpretable disentangled representation space through its integrated design: (i) prototype-based soft alignment under dual structural views, (ii) dedicated MLP transformations into semantic intent vectors, and (iii) interaction-level encoders that explicitly model cross-view intent alignments. This framework not only improves personalization and recommendation accuracy but also offers robust intent-level interpretability, especially under sparse or cold-start conditions.

3.5   Unified Design Benefits and Strategic Superiority of *DMICF*

Building upon the above comparative insights, the superior performance of *DMICF* can be attributed to the synergistic integration of its core design principles. *DMICF* is founded on the integration of three mutually reinforcing principles: dual-view structural modeling, interaction-grounded intent disentanglement, and discriminative representation learning through contrastive supervision. These components are orchestrated within a unified framework specifically designed to address the challenges of fine-grained, multi-intent recommendation. By capturing complementary relational patterns from both user and item perspectives, anchoring intent semantics in observed interaction behaviors, and enforcing representational separation through multi-negative signals, *DMICF* achieves superior robustness, interpretability, and accuracy. The key innovations of this design include:

- **Dual-view relational modeling**, which separately encodes user- and item-perspective interactions to capture complementary semantic structures. By constructing and fusing personalized subgraphs from both ends of the interaction graph, *DMICF* enriches structural representations and enables view-specific intent extraction.
- **Interaction-grounded, multi-granularity intent alignment**, which explicitly anchors intent disentanglement in the semantic dependencies of real user–item interactions. This ensures that the learned intent representations

Table 1. Statistics of the Experimental Datasets.

Dataset	#user	#item	#interactions	sparsity
Amazon	78,578	77,801	2,240,156	99.96%
Tmall	47,939	41,390	2,357,450	99.88%
ML-10M	69878	10195	6,999,171	99.02%

are not merely constrained by graph topology or heuristic factorization, but directly reflect meaningful behavioral semantics.

- **Multi-negative contrastive supervision**, which enhances representational discrimination by separating incompatible intent components. Through global item-level sampling, this mechanism introduces fine-grained, directionally informative signals that go beyond conventional positive-only or graph-level contrastive schemes.

Together, these components form a coherent and interpretable modeling pipeline that achieves robust performance, strong generalization, and semantic transparency. Empirical results demonstrate that *DMICF* consistently outperforms recent state-of-the-art models such as *IPCCF*, *BIGCF*, and *DCCF*, which are constrained respectively by global embedding fusion, coarse-grained graph-level modeling, or subspace heuristic supervision. In contrast, *DMICF*'s fine-grained, interaction-aware design enables it to effectively disentangle and match user–item intents across diverse structural and semantic scenarios.

4 EXPERIMENTS

To evaluate the effectiveness of *DMICF*, comprehensive experiments were conducted on three widely-used datasets to address the following research questions:

- **RQ1:** How does *DMICF* perform relative to state-of-the-art recommendation algorithms across different datasets?
- **RQ2:** How does *DMICF* perform across user groups with different interaction frequencies, particularly under data sparsity conditions?
- **RQ3:** To what extent do the key components of *DMICF* contribute to its overall performance?
- **RQ4:** Is *DMICF* capable of effectively disentangling user and item intents?
- **RQ5:** How do various hyperparameters impact the performance of *DMICF*?

4.1 Experimental Setting

4.1.1 Datasets. To validate the effectiveness and generalizability of *DMICF*, three widely recognized large-scale real-world datasets—Amazon-Book, Tmall and ML-10M—were selected, all of which have been extensively employed in prior studies. The detailed statistics of these datasets are summarized in Table 1. For fairness and consistency, the data preprocessing strictly follows the protocol of LightGCL, with datasets directly adopted from its publicly available source code.

4.1.2 Baselines. To ensure a comprehensive comparison, we evaluate representative baseline methods from three categories, detailed as follows.

(1) GNN-based recommendation methods

- *IMP-GCN* [12]. This method enhances GCN-based recommendation by partitioning users into distinct subgraphs according to their interests, enabling more targeted and effective representation learning.

- *LightGCN* [6]. This method streamlines the message passing mechanism of GCN by linearly propagating user and item embeddings over the interaction graph for collaborative filtering.

(2) Self-supervised learning approaches for recommendation

- *HCCF* [28]. This method simultaneously captures local and global collaborative interactions through a hypergraph neural network, and introduces cross-view contrastive learning to enhance representation robustness via data augmentation.
- *SHT* [29]. This method combines hypergraph learning and topology-aware Transformers under a self-supervised framework to robustly learn user/item embeddings from noisy and sparse interaction graphs.
- *LightGCL* [2]. This lightweight graph contrastive learning framework utilizes singular value decomposition to create augmented views, facilitating effective embedding contrast.
- *EGCF* [34]. This method eliminates explicit embeddings and data augmentation, enhancing user-item similarity via iterative item-based propagation and augmentation-free contrastive learning.

(3) Disentangled multi-intent recommendation systems

- *DGCF* [25]. This method constructs intent-aware graphs by modeling the distribution of intents for each interaction, enabling the learning of disentangled representations.
- *DCCF* [19]. This method leverages intent modeling and contrastive learning to identify intent prototypes for users and items, constructing multiple views of the interaction graph through learnable graph augmentations for enhanced representation learning.
- *BIGCF* [32]. This method adopts a causal perspective on user–item interactions by modeling individual intents to capture personal preferences and collective intents to reflect overall behavioral patterns.
- *IPCCF* [10]. This method enhances semantic understanding of user–item interactions by integrating direct and high-order signals through a double-helix propagation framework, and facilitates intent disentanglement via intent-aware message passing and contrastive learning.

4.1.3 Hyperparameters. The proposed *DMICF* model was implemented using PyTorch and trained with the Adam optimizer at a learning rate of 3×10^{-5}. Across all three benchmark datasets, a consistent model architecture and hyperparameter setting were adopted, with the number of latent intents fixed at $K = 32$. Four parallel multi-layer perceptrons $MLP_{user-user}$, $MLP_{user-item}$, $MLP_{item-item}$ and $MLP_{item-user}$ were used to intent distributions into expressive vector representations. Each MLP is independently parameterized with a two-layer structure of dimensions [32, 48, 80], followed by a sigmoid activation. The intent interaction encoder follows a GMF+MLP design, where $MLP_{user-intent}$ and $MLP_{item-intent}$ share a three-layer architecture with dimensions [80, 128, 64, 16]. A unified fusion mechanism combined information from both user and item perspectives. The final prediction module, $MLP_{predict}$, is a two-layer perceptron with dimensions [34, 32, 1] that outputs the predicted interaction probability. The temperature parameter τ was set to 0.2, and the number of negative samples per positive instance was set to $S = 50$ to support multi-negative training. To improve training stability and avoid overfitting, an early stopping strategy was applied based on *Recall*@40 on the validation set: training is terminated if performance does not improve for three consecutive evaluation rounds. The code used for the experiments has been publicly released and is available at https://github.com/ZINUX1998/DMICF , ensuring reproducibility and transparency of our results.

To ensure a fair comparison, all baseline models were configured with a latent embedding dimension of $d = 32$ and a batch size of 4096. For all baseline models, we adopted the officially released source code. If a dataset was originally

used in the baseline's corresponding paper, we directly followed the provided configurations and reported the results based on the official implementation. For datasets not used in the original papers, we followed the hyperparameter tuning principles described therein and conducted grid search to identify the best-performing configurations under our experimental setup. All experiments were conducted on a single NVIDIA RTX A5000 GPU with 24GB memory, ensuring consistent computational conditions across all model evaluations.

4.2 Experimental Results and Analysis

4.2.1 Performance Comparison (RQ1): We employed the widely adopted metrics *Recall@N* and *NDCG@N* (with N = 20 and 40) to evaluate the Top-N recommendation performance of *DMICF* and all baseline methods. The results are reported in Table 2, where the best performance is highlighted in blue and the second-best is underlined for clarity. From these results, we observe the following:

(1) Among the three categories of models, decoupled recommendation methods consistently achieve the best performance, followed by SSL-based approaches, while GNN-based models exhibit the weakest predictive capability. A closer analysis of these results reveals several insights: GNN-based models primarily emphasize local structural features derived from direct user–item interactions, yet they struggle in sparse data settings common in real-world recommender systems. Furthermore, they typically overlook the fact that the underlying factors driving user–item interactions are often highly entangled due to diverse user preferences, which undermines the quality of user representations learned through message propagation. In contrast, SSL-based models can partially mitigate the data sparsity issue by constructing multi-perspective interaction graphs. For instance, *LightGCL* enhances message passing with global collaborative signals, providing additional supervisory cues. However, these methods are still sensitive to noisy interactions and often fail to produce robust self-supervised signals, thereby limiting the effectiveness of contrastive learning when applied to heterogeneous or sparse interaction graphs. Decoupled recommendation methods, on the other hand, go beyond structural encoding by explicitly modeling the latent intents behind user–item interactions, enabling them to better capture nuanced user preferences. This paradigm improves the modeling of long-tail items through more informative and intent-aware self-supervision, and enhances model robustness under sparse interaction conditions.

(2) *DMICF* demonstrates consistently strong performance across all datasets and evaluation metrics. In particular, it significantly outperforms the strongest baselines, including *BIGCF* and *IPCCF*, on the Tmall and ML-10M datasets. Such improvements reflect a substantial enhancement in recommendation effectiveness, especially considering that IPCCF has already achieved substantial improvements over previous baselines by employing contrastive learning to align node representations derived from structure-based and intent-based message propagation, thereby providing direct supervision for the intent disentanglement process and making it a highly competitive benchmark.

A deeper examination of metric-level performance reveals a critical and informative pattern: *DMICF* exhibits a more substantial improvement on Recall@N than on NDCG@N. Specifically, the average performance gains on Recall@N reach 8.15% and 6.31% on Tmall and ML-10M, respectively, while the corresponding improvements in NDCG@N are only 5.30% and 4.45%. This discrepancy highlights an important characteristic of *DMICF*: it is particularly adept at identifying and retrieving relevant items, even if they are not ranked at the very top of the recommendation list. Since Recall@N treats all relevant items within the top-N equally, regardless of their position, this suggests that *DMICF* provides broader relevant-item coverage, effectively enhancing recommendation recallability. Conversely, the more moderate improvements in NDCG@N, a metric that discounts lower-ranked relevant items and emphasizes top-ranked precision, imply that the model's enhancement in fine-grained ranking accuracy, though present, is less substantial. This observation suggests that *DMICF* does not exclusively prioritize the elevation of a few highly relevant items to the

Table 2. Overall performance comparisons on Amazon-Book, Tmall and ML-10M datasets w.r.t. Recall@N (abbreviated as R@N) and NDCG@N (abbreviated as N@N).

Dataset	Metric	DGCF	IMP-GCN	LightGCN	HCCF	SHT	LightGCL	DCCF	EGCF	BIGCF	IPCCF	DMICF	impr.%	p-val
Amazon	R@20	0.0722	0.0787	0.0832	0.0813	0.0802	0.0975	0.0863	0.0911	0.0960	0.1083	0.1044	-3.74	$3.3e^{-5}$
	N@20	0.0543	0.0617	0.0640	0.0636	0.0628	0.0751	0.0661	0.0697	0.0743	0.0875	0.0797	-9.79	$7.2e^{-6}$
	R@40	0.1104	0.1172	0.1253	0.1209	0.1186	0.1468	0.1296	0.1390	0.1420	0.1530	0.1558	1.83	$5.4e^{-6}$
	N@40	0.0669	0.0744	0.0780	0.0766	0.0751	0.0913	0.0804	0.0854	0.0893	0.1020	0.0966	-5.59	$1.2e^{-5}$
Tmall	R@20	0.0605	0.0648	0.0631	0.0713	0.0694	0.0728	0.0702	0.0736	0.0753	0.0786	0.0858	9.16	$4.8e^{-6}$
	N@20	0.0414	0.0441	0.0434	0.0493	0.0478	0.0517	0.0486	0.0509	0.0536	0.0568	0.0602	5.99	$9.1e^{-6}$
	R@40	0.0963	0.1026	0.1013	0.1128	0.1102	0.1130	0.1098	0.1158	0.1181	0.1188	0.1323	11.36	$4.7e^{-5}$
	N@40	0.0539	0.0575	0.0566	0.0636	0.0621	0.0657	0.0642	0.0655	0.0690	0.0707	0.0764	8.06	$3.6e^{-6}$
ML-10M	R@20	0.1983	0.2144	0.2164	0.2427	0.2390	0.2382	OOM	0.2692	0.2762	0.2826	0.2996	6.02	$6.3e^{-5}$
	N@20	0.2318	0.2487	0.2539	0.2860	0.2778	0.2837	OOM	0.3084	0.3194	0.3382	0.3523	4.17	$1.4e^{-7}$
	R@40	0.3052	0.3280	0.3319	0.3549	0.3465	0.3264	OOM	0.3891	0.3928	0.3986	0.4249	6.60	$7.9e^{-6}$
	N@40	0.2601	0.2793	0.2827	0.3124	0.3081	0.3016	OOM	0.3383	0.3462	0.3640	0.3812	4.73	$5.2e^{-6}$

Table 3. Performance Comparison Between DMICF and IPCCF with Increasing Top-N Cutoff on the Amazon Dataset.

Model	Recall						NDCG					
	R@40	R@50	R@60	R@70	R@80	R@90	N@40	N@50	N@60	N@70	N@80	N@90
IPCCF	0.1530	0.1695	0.1845	0.1977	0.2094	0.2202	0.1020	0.1072	0.1118	0.1157	0.1192	0.1222
DMICF	0.1558	0.1758	0.1937	0.2095	0.2245	0.2379	0.0966	0.1028	0.1083	0.1129	0.1172	0.1210
impr.%	1.83	3.72	4.99	5.97	7.21	8.04	-5.59	-4.10	-3.23	-2.42	-1.71	-0.99

very top positions, but instead promotes a more balanced and inclusive retrieval strategy that distributes relevant items more evenly throughout the ranked list.

Further insights emerge from the results on the Amazon dataset. While *DMICF* does not consistently surpass IPCCF in all metrics, two observations are particularly noteworthy: (1) On the Recall@40 metric, *DMICF* outperforms *IPCCF*, achieving the best known performance on this dataset; (2) The performance gap between *DMICF* and *IPCCF* narrows as N increases, especially in the Recall@N metric. Specifically, Recall@N improves from -3.64% at $N = 20$ to +1.83% at $N = 40$, and NDCG@N improves from -9.66% to -5.59%. This trend is further confirmed in Table 3, where increasing N from 40 to 90 results in a monotonic increase in Recall@N advantage (from 1.83% to 8.04%) and a steady reduction in NDCG@N disadvantage (from -5.59% to -0.99%). These findings reflect a scaling robustness property of *DMICF*. That is, *DMICF* demonstrates increasing relative advantage as the recommendation list lengthens. This suggests that *DMICF* is especially effective at modeling long-tail user-item interactions, enabling it to retrieve more relevant items in the middle and tail regions of the ranked list. From a structural perspective, this strength can be attributed to the dual-perspective fusion and disentangled intent modeling in *DMICF*, which captures diverse and fine-grained user interests that are not limited to the most obvious top-ranked items.
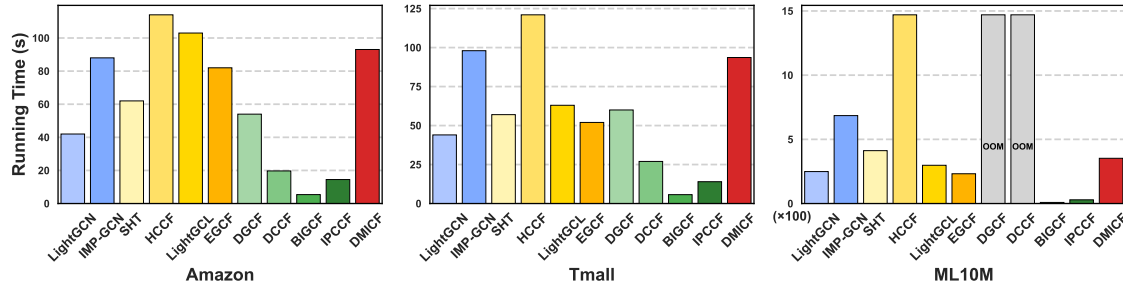
Fig. 2. Comparison of training time per epoch (in seconds) for all evaluated methods across three datasets.

Finally, to assess the computational efficiency of the proposed *DMICF* model, we conduct a detailed measurement of its training cost on all datasets. Specifically, we evaluate the time consumption per training epoch, as shown in Figure 1. The results indicate that *DMICF*'s per-epoch cost scales proportionally with the number of edges in each dataset. For instance, the training time on Amazon is close to that on Tmall, while ML-10M takes approximately three times longer due to its much denser interaction graph. When comparing per-epoch training time, *DMICF* is on par with *LightGCL*, yet significantly slower than existing disentangled models such as *DGCF*, *DCCF*, *BIGCF*, and *IPCCF*. The main reason is that these methods typically adopt customized sampling strategies that allow them to train on only a small subset of user-item interactions per epoch, thereby substantially reducing computational load and making them less sensitive to data scale. However, *DMICF* demonstrates a notable advantage in convergence speed, requiring substantially fewer epochs to reach its peak performance. Specifically, *DMICF* converges within 27, 28, and 18 epochs on Amazon, Tmall, and ML-10M, respectively. By contrast, baseline models often require nearly 200 epochs to achieve their best results. Remarkably, on ML-10M, *DMICF* reaches near-peak or even superior performance to *IPCCF* within just 9 epochs (Recall@20=0.2843, NDCG@20=0.3370, Recall@40=0.4085, NDCG@40=0.3660), whereas *IPCCF* requires 164 epochs, alongside a costly preprocessing phase to extract high-order relations. This fast convergence is primarily attributed to two key design choices: (1) The dual-perspective modeling framework, which effectively expands the receptive field without relying on deep message passing; (2) The intent-aligned interaction design, which embeds intent disentanglement directly into user–item pair modeling, providing strong supervision signals that facilitate rapid and effective disentanglement. In summary, although *DMICF* may incur a higher cost per epoch compared to some lightweight disentangled models, its rapid convergence and superior accuracy enable it to strike a favorable balance between computational cost and predictive performance, underscoring its practical value for large-scale recommendation applications.

4.2.2 Effectiveness of DMICF *across User Groups with Varying Interaction Densities (RQ2):* To evaluate the robustness of *DMICF* under varying levels of data sparsity, we stratify users into ten cohorts based on their interaction frequencies and report Recall@40 for each group. Fig. 3 presents both the statistical distribution and the corresponding performance on Amazon, Tmall, and ML-10M datasets. These datasets exhibit distinct interaction profiles: Amazon is dominated by sparse users in the $[10, 20)$ range, Tmall shows a denser distribution centered around $[40, 50)$, while ML-10M contains a long tail with most users in $[90, +)$. Across all datasets, *DMICF* achieves peak performance within the $[20, 30)$ interval, underscoring its effectiveness in mitigating data sparsity and leveraging limited feedback signals. Performance degrades gradually beyond this optimal range, which we attribute to two factors. First, the scaling of the Recall denominator with increasing interactions may suppress apparent gains, even when the number of relevant items retrieved remains constant.
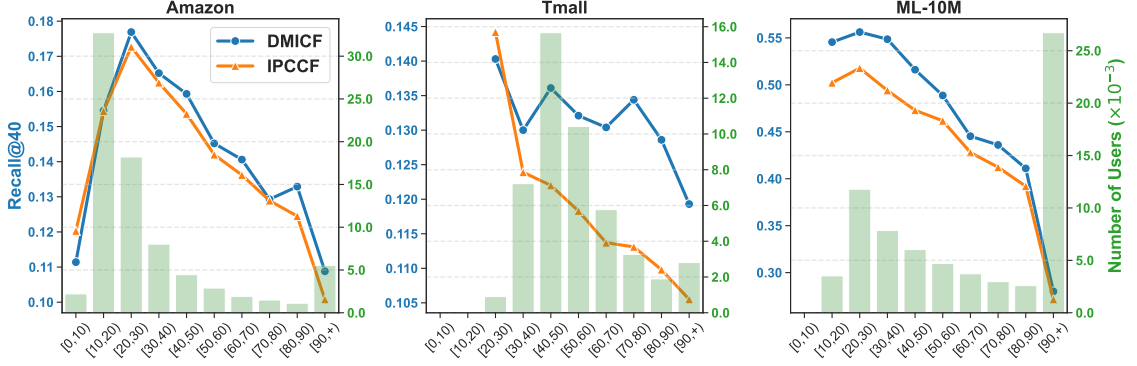
Fig. 3. Performance of *DMICF* and *IPCCF* Across User Groups with Varying Interaction Frequencies on Amazon, Tmall, and ML-10M. Bar plots (right y-axis) show the number of users per group, and line plots (left y-axis) indicate the Recall@40 performance of *DMICF* and *IPCCF* within each group. **Note:** Tmall contains no users in the $[0, 10)$ and and $[10, 20)$ groups, and ML-10M contains no users in the $[0, 10)$ group.

Second, users with extensive histories tend to exhibit more diverse and multi-intent preferences. This dispersion dilutes the semantic strength of individual intent vectors, making it more difficult for the model to capture dominant user preferences and maintain high matching accuracy.

To further contextualize these findings, we compare *DMICF* against *IPCCF* across all cohorts. Two key insights emerge: (1) Both models follow a similar trend, with performance peaking in the $[20, 30)$ range and declining thereafter, suggesting a shared sensitivity to the trade-off between signal sufficiency and intent representation dilution. (2) *DMICF* performs slightly below *IPCCF* in the sparsest cohort, likely due to *IPCCF*'s contrastive alignment between structural and intent-aware representations. However, *DMICF* consistently outperforms *IPCCF* in denser cohorts, demonstrating superior capability in modeling complex and multi-granular user intents when sufficient feedback is available. This advantage is attributed to its dual-view design and fine-grained intent disentanglement, which jointly enhance its expressiveness and generalization across varying user behaviors.

4.2.3 Ablation Study (RQ3): To evaluate the contribution of each key component in the *DMICF* framework, we conducted a comprehensive ablation study by designing and evaluating seven model variants. These variants were constructed to examine the effects of different intent interaction encoder designs and dual-view information fusion strategies, as well as the impact of removing the dual-view modeling architecture altogether.

Specifically, *DMICF* explores three alternative instantiations for the intent interaction encoder—Concatenation+MLP, GMF+MLP, and Cross Attention Fusion—and two distinct dual-view information fusion strategies: two-stage fusion and unified fusion. This gives rise to five feasible architectural variants beyond the main *DMICF* model used in our experiments, denoted as follows:

- ***Cat_two:*** Uses Concatenation+MLP for the intent interaction encoder and two-stage fusion for dual-view integration.
- ***Cat_uni:*** Uses Concatenation+MLP with unified fusion.
- ***GMF_two:*** Uses GMF+MLP with two-stage fusion.
- ***Cross_two:*** Uses Cross Attention Fusion with two-stage fusion.
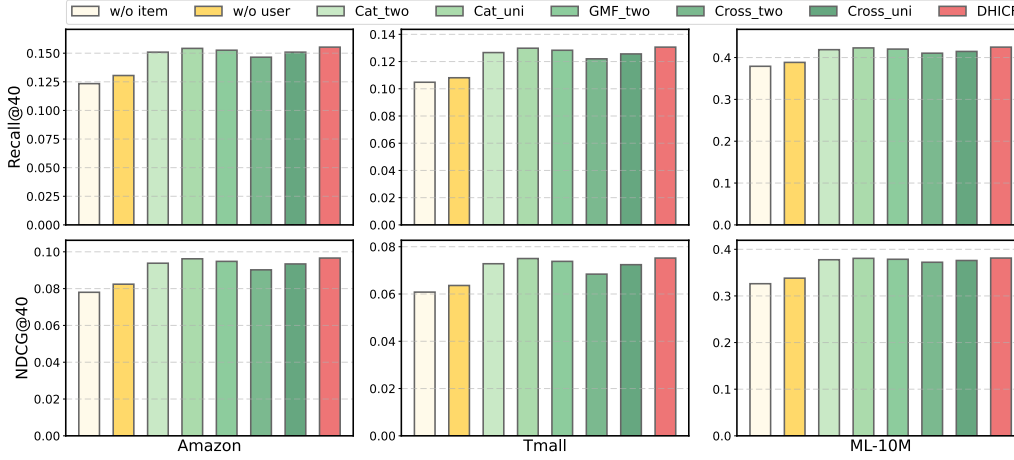- ***Cross_uni:*** Uses Cross Attention Fusion with unified fusion.

Fig. 4. Ablation Study on Key Components of *DMICF* Evaluated by Recall@40 and NDCG@40 Across Different Datasets.

To further validate the effectiveness of the dual-perspective architecture, we also implemented two single-view baselines:

- ***w/o user:*** Utilizes only the item-perspective information to predict user–item interaction probabilities.
- ***w/o item:*** Utilizes only the user-perspective information for interaction prediction.

Each variant selectively modifies or removes specific modules or mechanisms, enabling us to isolate and quantify their respective contributions to the overall recommendation performance.

The experimental results are summarized in Fig.4, focusing on two representative metrics: Recall@40 and NDCG@40. While similar trends were also observed on other metrics, we report these two for clarity and conciseness. We draw the following observations:

(1) Impact of Intent Interaction Encoder Design on Model Performance. Among the evaluated encoder designs, the GMF+MLP configuration consistently delivers the best performance across all datasets. This highlights the effectiveness of element-wise multiplicative modeling of user-item latent factors, which allows the model to better capture fine-grained intent signals. Meanwhile, Concatenation+MLP also performs competitively and introduces only a marginal performance drop compared to GMF+MLP. Notably, the variants ***Cat_two*** and ***GMF_two*** yield nearly identical results, as do ***Cat_uni*** and the original *DMICF*. However, the Cross Attention Fusion variant exhibits clear performance degradation. Specifically, compared to ***GMF_two***, ***Cross_two*** results in an average decrease in Recall@40 of 4.16%, 5.16%, and 2.39% on the Tmall, ML-10M, and Amazon datasets, respectively. Similarly, ***Cross_uni*** underperforms *DMICF* by 2.92%, 3.98%, and 2.48%, indicating that while attention-based modeling may offer expressive power, it may introduce optimization instability or overfitting when applied to sparse interaction data.

(2) Unified Fusion Demonstrates Superior Effectiveness over Two-Stage Fusion for Dual-View Integration. Across all variants, models that employ unified fusion to integrate the user and item perspectives consistently outperform their two-stage fusion counterparts. For instance, *DMICF* outperforms ***GMF_two***, ***Cross_uni*** outperforms ***Cross_two***, and ***Cat_uni*** outperforms ***Cat_two*** across all datasets. This result indicates that joint modeling of dual-view representations in a unified architecture facilitates better interaction between complementary information sources, leading to more robust user-item matching under sparse conditions. Furthermore, both GMF+MLP and the unified fusion strategy are

relatively lightweight compared to their alternatives, yet together they yield the best overall performance—highlighting a favorable balance between model complexity and recommendation effectiveness.

(3) The ablation results of the *w/o user* and *w/o item* variants provide compelling evidence that the dual-view structure in *DMICF* plays a critical role in capturing rich semantic representations of nodes within the interaction graph. By jointly modeling user- and item-centered perspectives, *DMICF* is able to better preserve structural and intent-aware signals across heterogeneous neighborhoods. This dual-view modeling substantially alleviates the impact of data sparsity, thereby leading to significant improvements in recommendation accuracy. These findings underscore the effectiveness of incorporating complementary views in enhancing the expressiveness and robustness of representation learning for recommender systems.

4.2.4 Qualitative and Quantitative Evaluation of Intent Disentanglement (RQ4): Our study integrates case-level inspection, population-level statistical characterization, and latent structure visualization on the Amazon-Book dataset, thus offering a holistic understanding of how intent semantics evolve and specialize throughout training.

To initiate our investigation from an individual-level perspective, we randomly select a target user u_4 and visualize the relevance scores associated with her learned user-perspective and item-perspective intent representations-denoted as $\mathbf{h}_4^u$ and $\tilde{\mathbf{r}}_4^v$)-at three critical training stages: initialization (epoch 0), early learning (epoch 5), and convergence (epoch 27). As shown in Fig. 5, the initial distribution of relevance scores appears flat and uninformative—an expected consequence of Xavier initialization applied to both the embedding matrices ($\mathbf{E}^u$ and $\mathbf{E}^v$) and the intent prototype matrices ($\mathbf{C}^u$ and $\mathbf{C}^v$). This baseline confirms the absence of semantic discrimination at the outset. As training progresses, a clear semantic alignment begins to emerge: certain intent dimensions gain increasingly higher relevance scores, while others are progressively suppressed. By epoch 27, the distribution becomes markedly peaked, indicating the model's ability to filter out irrelevant latent intents and concentrate representational capacity on those that are behaviorally meaningful for the user. Notably, the item-perspective representation ($\tilde{\mathbf{r}}_4^v$) evolves in parallel, reflecting an interaction-guided cross-perspective alignment process. This trajectory exemplifies the capacity of *DMICF* to capture personalized intent structures by leveraging disentangled representation learning in a dynamically adaptive manner.

To generalize our observations beyond individual users, we extend our analysis to a population-level perspective by aggregating user-specific relevance vectors into two matrices: $\mathbf{h}^u \in \mathbb{R}^{N \times d'}$ and $\tilde{\mathbf{r}}^v \in \mathbb{R}^{N \times d'}$, where each column represents the activation values of a particular intent dimension across the entire user base. This representation enables us to assess how the model distributes representational capacity across different intent factors at a global scale. Figure 6 visualizes the temporal evolution of the mean and variance of each intent dimension across training epochs. Three key phenomena emerge. First, the variance across most intent dimensions increases substantially as training progresses, suggesting that user behaviors are increasingly differentiated along distinct semantic axes. This confirms that the model effectively leverages its disentangled structure to associate different users with different subsets of intent factors. Second, the mean relevance scores begin to diverge across intent dimensions: some dimensions consistently attain higher mean values, indicating they capture globally shared preferences prevalent across the user population, while others maintain low means, reflecting their role in modeling fine-grained and personalized behaviors. This bifurcation between collective and individual semantics reflects the model's ability to represent both macro-level trends and micro-level diversity. Third, and most critically, we observe a positive correlation between the mean and variance of intent activations. Intent dimensions with higher average relevance also exhibit greater intra-intent variability, implying that even globally salient semantic factors maintain sufficient flexibility to represent diverse behavioral nuances. This characteristic is particularly important for avoiding mode collapse—a common challenge in disentangled representation learning where
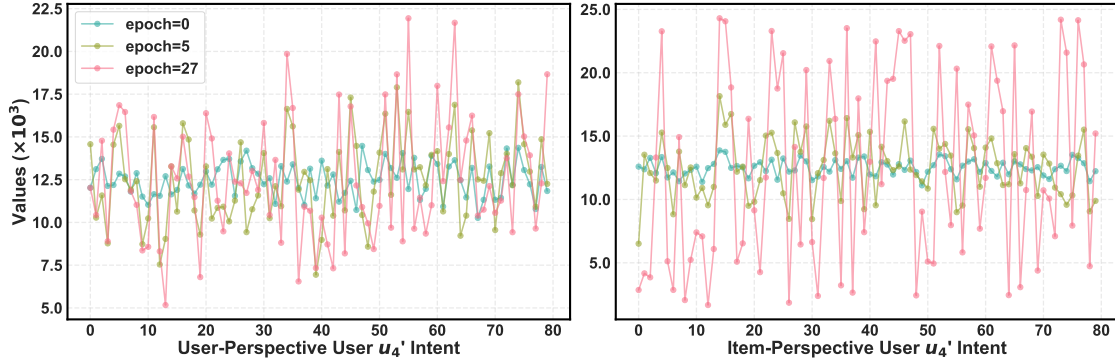
Fig. 5. Evolution of Intent Relevance Scores for a Target User u_4 ($\mathbf{h}_4^u$ and $\tilde{\mathbf{r}}_4^v$) across Training Stages.
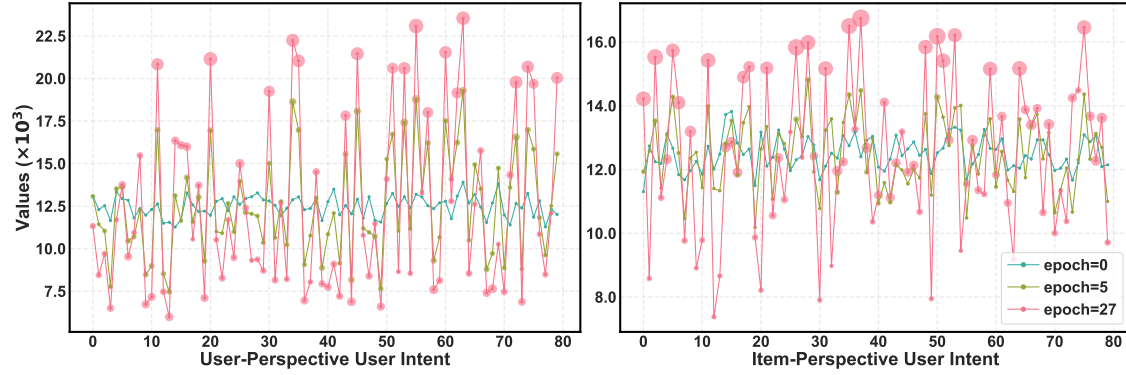

Fig. 6. Evolution of Mean and Variance for Each Intent Dimension of User Intent Relevance Scores. (Dot Size Scaled by Variance)

over-regularized factors fail to capture diversity. Instead, our model preserves intra-intent expressiveness, enabling context-sensitive personalization within even the most widely-used intent dimensions. This capability is essential for scalability in real-world recommender systems, where capturing both commonalities and idiosyncrasies is vital for delivering nuanced and user-tailored recommendations.

To further validate the generality of disentangled intent modeling, we conduct a quantitative assessment on the item side by randomly sampling 5,000 items and analyzing the statistical evolution of their intent relevance scores. Specifically, for each item, we compute the mean and variance of relevance values across the d' disentangled dimensions (i.e., $\tilde{\mathbf{h}}_j^u$ and $\mathbf{r}_j^v$). As depicted in Fig.7 and Fig.8, the initial score distributions are narrowly concentrated, reflecting a lack of semantic differentiation. However, as training progresses, both the mean and variance distributions expand considerably, indicating stronger selectivity and specialization. Interestingly, the joint distribution forms a characteristic "V"-shaped or checkmark pattern, where items with extreme average relevance scores also exhibit higher variances. This structure suggests that intent representations are not uniformly allocated; instead, *DMICF* learns to activate specific intents only when semantically aligned with item characteristics. Such behavior underscores the model's capability to disentangle item semantics in a fine-grained and behavior-driven manner—an essential property in sparse recommendation scenarios where subtle item-level distinctions are critical to accurate matching.
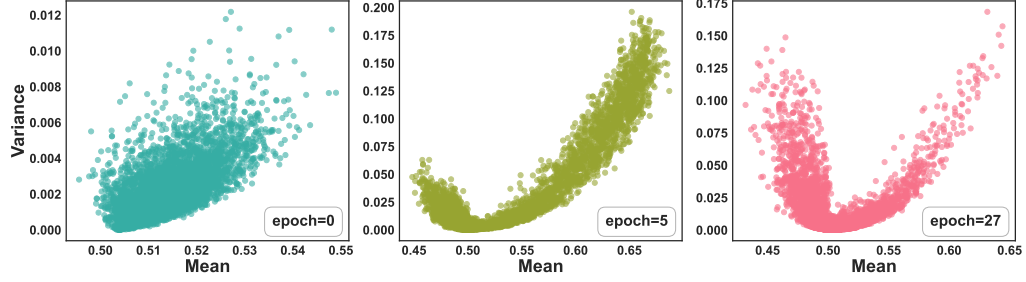
Fig. 7. User-Perspective Evolution of Item Intent Scores. The figure illustrates the training evolution of mean and variance in user-perspective item intent scores ($\tilde{\mathbf{h}}_j^u$) for 5,000 items.
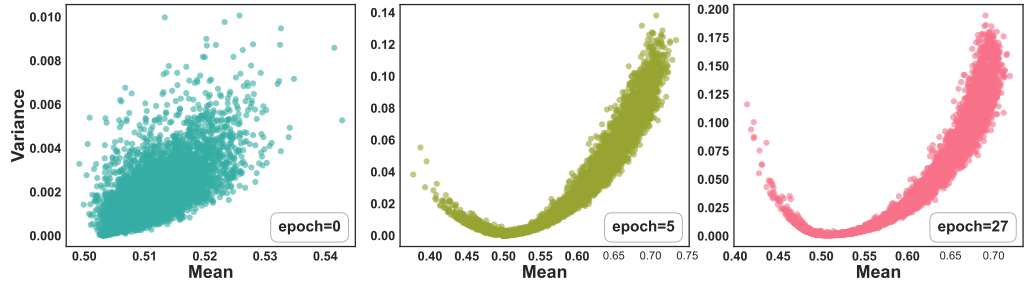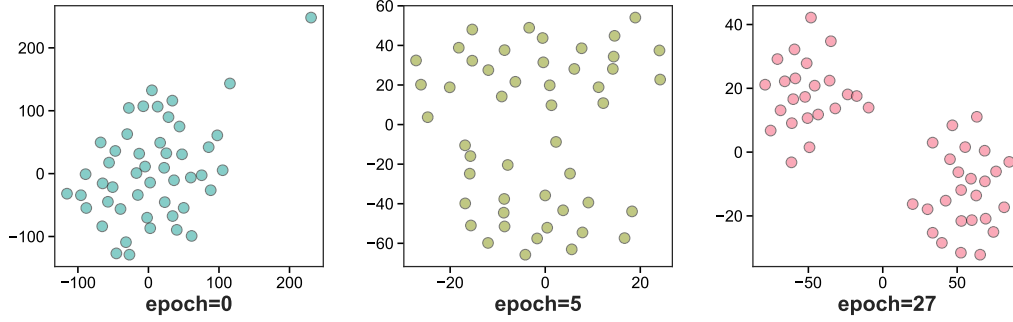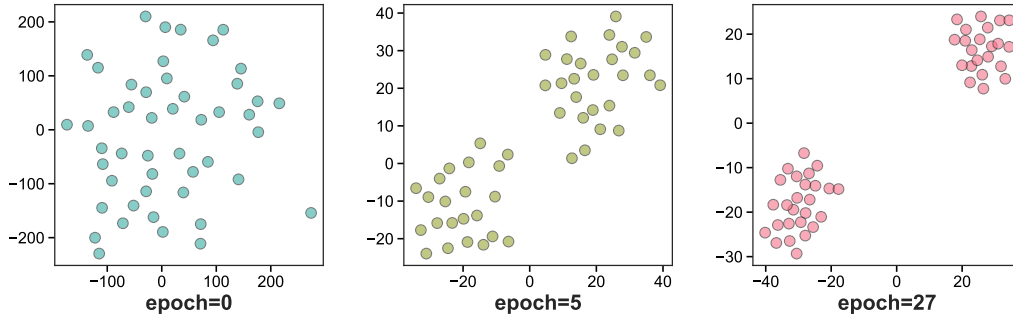

Fig. 8. Item-Perspective Evolution of Item Intent Scores. The figure illustrates the training evolution of mean and variance in item-perspective item intent scores ($\mathbf{r}_j^v$) for 5,000 items.

To elucidate the internal organization of learned intent spaces, we visualize the learned intent prototype vectors $\mathbf{C}^u$ and $\mathbf{C}^v$ via t-SNE projection in Fig.9 and Fig.10. At epoch 0, the prototypes are uniformly dispersed—consistent with random initialization—offering no meaningful geometric pattern. However, by epoch 27, two clearly separable clusters emerge in both user and item prototype spaces. This self-organized clustering reflects semantic polarization: one cluster comprises prototypes with high average relevance (cf. Fig. 6), indicative of globally shared behavioral patterns, while the other includes sparsely activated prototypes associated with fine-grained, user/item-specific signals. This emergent geometry provides compelling evidence that *DMICF* does not merely perform intent allocation heuristically, but instead learns to structurally organize the latent space into semantically coherent subregions. This organization enhances both representational capacity and interpretability, aligning with the model's core objective: to disentangle multi-faceted user–item interactions into functionally and semantically distinct intent subspaces.

In summary, our statistical and visualization analyses collectively demonstrate that *DMICF* is capable of effectively disentangling user and item intents from both the user and item perspectives. The learned intent space evolves from an initially uniform distribution to a semantically structured configuration, exhibiting clear differentiation between globally aligned and locally personalized intent dimensions. This progressive specialization confirms the model's ability to construct an intent-aware representation space that simultaneously supports global coherence and individual-level differentiation—a critical property for effective disentangled representation learning in recommender systems.

Fig. 9. Evolution of User-Perspective Intent Prototypes $\mathbf{C}^u$ via t-SNE at Different Training Stages.

Fig. 10. Evolution of Item-Perspective Intent Prototypes $\mathbf{C}^v$ via t-SNE at Different Training Stages.

4.2.5 Hyperparameter Sensitivities (RQ5): We systematically evaluates the influence of four pivotal hyperparameters on the performance of the proposed *DMICF* model: (1) the number of intent prototypes K, (2) the dimensionality of intent embeddings d', (3) the number of negative samples S, and (4) the temperature coefficient τ. Performance is primarily measured using Recall@40 across three benchmark datasets—Amazon, Tmall, and ML-10M—with similar trends observed on other metrics and thus omitted for brevity. The detailed performance variations with respect to these hyperparameters are visualized in Fig. 11.

Sensitivity Analysis on Latent Intent Granularity. The number of latent intent prototypes K fundamentally controls the granularity of user and item preference disentanglement. Extensive ablation studies varying $K \in \{16, 24, 32, 40, 48, 56, 64\}$ reveal a distinct non-monotonic performance pattern. Initially, increasing K improves model accuracy by enabling finer-grained semantic distinctions within the intent space. However, beyond a certain threshold, further increments degrade performance due to semantic redundancy and fragmentation. Insufficient prototypes limit the expressive power, causing semantically distinct user behaviors to collapse into overly coarse clusters, thereby hindering precise intent modeling. Conversely, excessive prototypes lead to overlapping or noisy latent factors, which impair the clarity of intent alignments and increase overfitting risks, especially in sparse data regimes. Parallel exploration of the intent embedding dimensionality d' indicates an optimal value around 80. Larger dimensions yield marginal or no performance gains, potentially diluting semantic specificity and introducing noise.
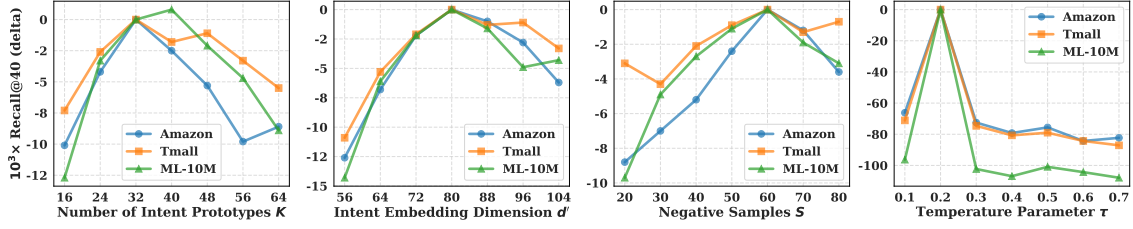
Fig. 11. parameter analysis.

A salient strength of *DMICF* lies in its end-to-end supervision mechanism that directly optimizes intent prototypes via interaction-aligned user–item pairs. This facilitates dynamic adaptation of the latent space with respect to K and d', thereby mitigating representational collapse and enhancing robustness. Empirically, the model exhibits stable performance with moderate hyperparameter variations, maintaining fluctuations within a narrow band (less than 5×10^{-3}) when $K \in [24, 48]$ and $d' \in [72, 96]$, evidencing the framework's resilience and flexibility.

Impact of Contrastive Negative Sampling. The number of negative samples S and the temperature coefficient τ critically influence the contrastive learning efficacy embedded in *DMICF*. Increasing S initially enriches the contrastive signal by exposing the model to diverse negative examples, thus enhancing discriminative power. Nevertheless, the benefits plateau and reverse beyond $S = 60$, where performance slightly deteriorates at $S = 70$ and $S = 80$. This phenomenon can be attributed to two intertwined challenges: (1) the denominator of the softmax loss (Eq. 19) grows substantially with large S, diluting the positive sample's gradient signal and thereby weakening the model's capacity to emphasize true user–item interactions; (2) uniform negative sampling increases the likelihood of incorporating false negatives—items that are relevant but treated as negatives—introducing label noise that confuses the learning process and undermines generalization.

Regarding the temperature parameter τ, our experiments identify $\tau = 0.2$ as optimal. Deviations in either direction cause notable performance drops. A lower τ (< 0.2) sharpens the probability distribution excessively, concentrating probability mass on a few samples and leading to gradient sparsity and unstable training prone to overfitting spurious correlations. Conversely, a higher τ (> 0.2) flattens the distribution, weakening the contrast between positive and negative pairs and impairing the learning of discriminative embeddings. This degeneration results in reduced supervisory effectiveness and compromises the contrastive objective's role in robust representation learning.

5 CONCLUSION

This paper presents *DMICF*, a novel dual-view disentangled multi-intent learning framework that addresses two critical limitations in existing recommendation models: the absence of collaborative intent alignment between users and items, and the lack of semantically grounded supervision during intent disentanglement. Unlike existing models that focus on embedding consistency or coarse-grained structural signals, *DMICF* leverages dual-view relational modeling to extract complementary semantic patterns from personalized user- and item-centric subgraphs. Its interaction-grounded intent interaction encoder anchors intent disentanglement directly in observed user–item behaviors, enabling fine-grained semantic alignment that reflects true decision-making processes. Additionally, *DMICF* employs multi-negative contrastive supervision, which enforces discriminative intent representations by separating incompatible semantic components, thereby improving interpretability and robustness. Extensive experiments on three representative datasets with varying interaction densities demonstrate the superior performance of *DMICF* over ten state-of-the-art baselines.

Beyond quantitative results, qualitative and quantitative analyses reveal that *DMICF* can effectively disentangle latent intent dimensions and automatically distinguish between collective and individual-specific semantics. These findings validate the model's capacity to enhance both recommendation accuracy and behavioral transparency. Future research could explore extending *DMICF* to incorporate temporal dynamics, cross-domain behavior, or multimodal feedback, further broadening its applicability and interpretability in real-world recommendation scenarios.

References

[1] Miaomiao Cai, Lei Chen, Yifan Wang, Haoyue Bai, Peijie Sun, Le Wu, Min Zhang, and Meng Wang. 2024. Popularity-Aware Alignment and Contrast for Mitigating Popularity Bias. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*. Association for Computing Machinery, New York, NY, USA, 187–198. doi:10.1145/3637528.3671824

[2] Xuheng Cai, Chao Huang, Lianghao Xia, and Xubin Ren. 2023. LightGCL: Simple Yet Effective Graph Contrastive Learning for Recommendation. In *The Eleventh International Conference on Learning Representations*.

[3] Lei Chen, Le Wu, Richang Hong, Kun Zhang, and Meng Wang. 2020. Revisiting Graph Based Collaborative Filtering: A Linear Residual Graph Convolutional Network Approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 27–34. doi:10.1609/aaai.v34i01.5330

[4] Jingwei Guo, Kaizhu Huang, Xinping Yi, and Rui Zhang. 2024. Learning Disentangled Graph Convolutional Networks Locally and Globally. *IEEE Transactions on Neural Networks and Learning Systems* 35, 3 (2024), 3640–3651. doi:10.1109/TNNLS.2022.3195336

[5] Zhiqiang Guo, Jianjun Li, Guohui Li, Chaoyang Wang, Si Shi, and Bin Ruan. 2024. LGMRec: local and global graph learning for multimodal recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 8454–8462. doi:10.1609/aaai.v38i8.28688

[6] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, YongDong Zhang, and Meng Wang. 2020. 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)*. Association for Computing Machinery, Virtual Event, China, 639–648. doi:10.1145/3397271.3401063

[7] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 173–182. doi:10.1145/3442381.3449986

[8] Dietmar Jannach and Markus Zanker. 2024. A Survey on Intent-aware Recommender Systems. *ACM Trans. Recomm. Syst.* 3, 2 (2024), 1–32. doi:10.1145/3700890

[9] Baoyu Jing, Yuchen Yan, Kaize Ding, Chanyoung Park, Yada Zhu, Huan Liu, and Hanghang Tong. 2024. STERLING: synergistic representation learning on bipartite graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 12976–12984. doi:10.1609/aaai.v38i12.29195

[10] Haojie Li, Junwei Du, Guanfeng Liu, Feng Jiang, Yan Wang, and Xiaofang Zhou. 2025. Intent Propagation Contrastive Collaborative Filtering. *IEEE Transactions on Knowledge and Data Engineering* 37, 5 (2025), 2665–2679. doi:10.1109/TKDE.2025.3543241

[11] Guanyu Lin, Chen Gao, Yu Zheng, Jianxin Chang, Yanan Niu, Yang Song, Kun Gai, Zhiheng Li, Depeng Jin, Yong Li, and Meng Wang. 2024. Mixed Attention Network for Cross-domain Sequential Recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM '24)*. Association for Computing Machinery, New York, NY, USA, 405–413. doi:10.1145/3616855.3635801

[12] Fan Liu, Zhiyong Cheng, Lei Zhu, Zan Gao, and Liqiang Nie. 2021. Interest-aware Message-Passing GCN for Recommendation. In *Proceedings of the Web Conference 2021 (WWW '21)*. Association for Computing Machinery, New York, NY, USA, 1296–1305. doi:10.1145/3442381.3449986

[13] Fan Liu, Shuai Zhao, Zhiyong Cheng, Liqiang Nie, and Mohan Kankanhalli. 2024. Cluster-Based Graph Collaborative Filtering. *ACM Trans. Inf. Syst.* 42, 6 (2024), 1–24. doi:10.1145/3687481

[14] Jing Liu, Lele Sun, Weizhi Nie, Peiguang Jing, and Yuting Su. 2024. Graph disentangled contrastive learning with personalized transfer for cross-domain recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 8769–8777. doi:10.1609/aaai.v38i8.28723

[15] Jianxin Ma, Chang Zhou, Peng Cui, Hongxia Yang, and Wenwu Zhu. 2019. Learning Disentangled Representations for Recommendation. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NIPS '19)*. 5712–5723.

[16] Shanlei Mu, Yaliang Li, Wayne Xin Zhao, Siqing Li, and Ji-Rong Wen. 2021. Knowledge-Guided Disentangled Representation Learning for Recommender Systems. *ACM Trans. Inf. Syst.* 40, 1 (2021), 1–26. doi:10.1145/3464304

[17] Zhongyu Ouyang, Chunhui Zhang, Shifu Hou, Chuxu Zhang, and Yanfang Ye. 2024. How to Improve Representation Alignment and Uniformity in Graph-Based Collaborative Filtering?. In *Proceedings of the International AAAI Conference on Web and Social Media*. 1148–1159. doi:10.1609/icwsm.v18i1.31379

[18] Yifang Qin, Wei Ju, Yiyang Gu, Ziyue Qiao, Zhiping Xiao, and Ming Zhang. 2025. PolyCF: Towards Optimal Spectral Graph Filters for Collaborative Filtering. *ACM Trans. Inf. Syst.* (2025). doi:10.1145/3728464

[19] Xubin Ren, Lianghao Xia, Jiashu Zhao, Dawei Yin, and Chao Huang. 2023. Disentangled Contrastive Collaborative Filtering. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 1137–1146. doi:10.1145/3539618.3591665

[20] Lei Sang, Yu Wang, Yi Zhang, Yiwen Zhang, and Xindong Wu. 2025. Intent-Guided Heterogeneous Graph Contrastive Learning for Recommendation. *IEEE Transactions on Knowledge and Data Engineering* 37, 4 (2025), 1915–1929. doi:10.1109/TKDE.2025.3536096

[21] Jie Shuai, Kun Zhang, Le Wu, Peijie Sun, Richang Hong, Meng Wang, and Yong Li. 2022. A Review-aware Graph Contrastive Learning Framework for Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*. 1283–1293. doi:10.1145/3477495.3531927

[22] Ke Tu, Wei Qu, Zhengwei Wu, Zhiqiang Zhang, Zhongyi Liu, Yiming Zhao, Le Wu, Jun Zhou, and Guannan Zhang. 2023. Disentangled Interest importance aware Knowledge Graph Neural Network for Fund Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 2482–2491. doi:10.1145/3583780.3614846

[23] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*. Association for Computing Machinery, New York, NY, USA, 165–174. doi:10.1145/3331184.3331267

[24] Xiang Wang, Tinglin Huang, Dingxian Wang, Yancheng Yuan, Zhenguang Liu, Xiangnan He, and Tat-Seng Chua. 2021. Learning Intents behind Interactions with Knowledge Graph for Recommendation. In *Proceedings of the Web Conference 2021 (WWW '21)*. Association for Computing Machinery, New York, NY, USA, 878–887. doi:10.1145/3442381.3450133

[25] Xiang Wang, Hongye Jin, An Zhang, Xiangnan He, Tong Xu, and Tat-Seng Chua. 2020. Disentangled Graph Collaborative Filtering. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*. Association for Computing Machinery, New York, NY, USA, 1001–1010. doi:10.1145/3397271.3401137

[26] Yifan Wang, Suyao Tang, Yuntong Lei, Weiping Song, Sheng Wang, and Ming Zhang. 2020. DisenHAN: Disentangled Heterogeneous Graph Attention Network for Recommendation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 1605–1614. doi:10.1145/3340531.3411996

[27] Lianghao Xia, Chao Huang, Jiao Shi, and Yong Xu. 2023. Graph-less Collaborative Filtering. In *Proceedings of the ACM Web Conference 2023 (WWW '23)*. 17–27. doi:10.1145/3543507.3583196

[28] Lianghao Xia, Chao Huang, Yong Xu, Jiashu Zhao, Dawei Yin, and Jimmy Huang. 2022. Hypergraph Contrastive Collaborative Filtering. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 70–79. doi:10.1145/3477495.3532058

[29] Lianghao Xia, Chao Huang, and Chuxu Zhang. 2022. Self-Supervised Hypergraph Transformer for Recommender Systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*. Association for Computing Machinery, New York, NY, USA, 2100–2109. doi:10.1145/3534678.3539473

[30] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2022. 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 (SIGIR '22)*. 1294–1303. doi:10.1145/3477495.3531937

[31] Yiding Zhang, Chaozhuo Li, Xing Xie, Xiao Wang, Chuan Shi, Yuming Liu, Hao Sun, Liangjie Zhang, Weiwei Deng, and Qi Zhang. 2022. Geometric Disentangled Collaborative Filtering. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 80–90. doi:10.1145/3477495.3531982

[32] Yi Zhang, Lei Sang, and Yiwen Zhang. 2024. Exploring the Individuality and Collectivity of Intents behind Interactions for Graph Collaborative Filtering. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*. Association for Computing Machinery, New York, NY, USA, 1253–1262. doi:10.1145/3626772.3657738

[33] Yuting Zhang, Yiqing Wu, Ruidong Han, Ying Sun, Yongchun Zhu, Xiang Li, Wei Lin, Fuzhen Zhuang, Zhulin An, and Yongjun Xu. 2024. Unified Dual-Intent Translation for Joint Modeling of Search and Recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*. Association for Computing Machinery, New York, NY, USA, 6291–6300. doi:10.1145/3637528.3671519

[34] Yi Zhang, Yiwen Zhang, Lei Sang, and Victor S. Sheng. 2024. Simplify to the Limit! Embedding-Less Graph Collaborative Filtering for Recommender Systems. *ACM Trans. Inf. Syst.* 43, 1 (2024), 1–30. doi:10.1145/3701230

[35] Yu Zhang, Yiwen Zhang, Yi Zhang, Lei Sang, and Yun Yang. 2025. Unveiling Contrastive Learning' Capability of Neighborhood Aggregation for Collaborative Filtering. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25)*.

[36] Sen Zhao, Wei Wei, Ding Zou, and Xianling Mao. 2022. Multi-view intent disentangle graph networks for bundle recommendation. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*. 4379–4387. doi:10.1609/aaai.v36i4.20359