Bridging Short Videos and Streamers with Multi-Graph Contrastive Learning for Live Streaming Recommendation

Changle Qu
Gaoling School of Artificial Intelligence
Renmin University of China
Beijing, China
changlequ@ruc.edu.cn

Liqin Zhao
Yanan Niu
Kuaishou Technology
Beijing, China
{zhaoliqin,niuyanan}@kuaishou.com

Xiao Zhang*
Jun Xu
Gaoling School of Artificial Intelligence
Renmin University of China
Beijing, China
{zhangx89,junxu}@ruc.edu.cn

Abstract

Recently, live streaming services have seen a surge in popularity, prompting many platforms to offer both short video and live streaming services to meet the diverse needs of users and streamers. This has resulted in a close connection between short videos and live streaming within these platforms. Incorporating short video data into live streaming recommendation through cross-domain approaches can effectively mitigate the sparsity of live streaming gifting data. However, existing cross-domain recommendation methods primarily focus on transferring information across domains through overlapping users or items, while overlooking the strong connection between non-overlapping short videos and streamers. In this paper, we propose **MGCCDR**, a **M**ulti-**G**raph **C**ontrastive learning framework for **C**ross-**D**omain **R**ecommendation, which leverages both overlapping users and non-overlapping items to enhance information transfer. Specifically, we first learn global representations from a global graph to establish connections between streamers and short videos. Subsequently, we construct three bipartite graphs among users, authors, and videos and introduce multi-graph learning to capture preferences within the target domain view, the source domain view, and the cross-domain view. Additionally, to address the varying contributions of each graph to the final recommendation task, we design an attention-based method to effectively integrate these representations, facilitating the information aggregation across domains. Extensive experiments on both commercial and public datasets demonstrate that our MGCCDR significantly outperforms the state-of-the-art methods.

CCS Concepts

• **Information systems → Recommender systems**.

Keywords

Live Streaming Recommendation, Cross-Domain, GNN

*Corresponding author: Xiao Zhang. Work partially done at Engineering Research Center of Next-Generation Intelligent Search and Recommendation, Ministry of Education.

SIGIR '25, Padua, Italy

ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3729914

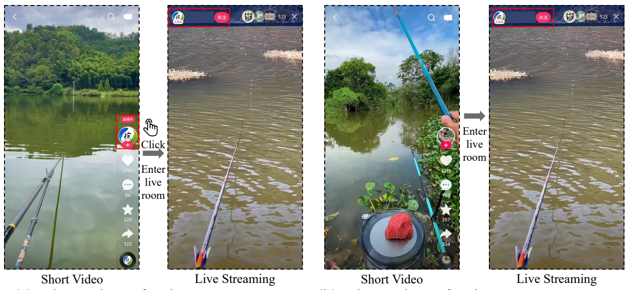
ACM Reference Format:
Changle Qu, Liqin Zhao, Yanan Niu, Xiao Zhang, and Jun Xu. 2025. Bridging Short Videos and Streamers with Multi-Graph Contrastive Learning for Live Streaming Recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3726302.3729914

1 Introduction

Nowadays, with the rapid advancement of mobile internet technology, streaming recommender systems have seamlessly integrated into people's daily lives [6–9, 21, 34, 45, 47, 50, 51]. Through live streaming platforms, streamers can share their experiences in real time, while users can engage through real-time interactions like sending virtual gifts. Accurately modeling user preferences and recommending suitable streamers can improve the experience for both users and streamers, which is crucial for the growth of live streaming platforms. However, due to the high cost of virtual gifting, gifting behavior data is extremely sparse, making it challenging to model user preferences based solely on live streaming data [5]. Given that many live streaming platforms, such as YouTube, TikTok, and Kwai, also offer short video services, an intuitive idea is to leverage cross-domain recommendation methods to capture user preferences from the abundant short video data to assist in live streaming recommendation, which has been shown to effectively mitigate the issue of data sparsity in recommender system [20, 22, 24, 25, 46].

In recent years, significant efforts have been dedicated to enhancing the performance of cross-domain recommendation [30, 43, 54]. Several studies first obtain user and item representations from each domain separately and then design sophisticated information transfer modules to integrate these representations into a final output. For example, CoNet [18] employs a cross-connection network, while BiTGCF [23] uses a feature fusion module to achieve knowledge transfer. Since user interests in the source and target domains are not entirely aligned, simplistic feature fusion may lead to negative transfer [1, 22, 29, 49], where irrelevant or conflicting information from the source domain can degrade the model's performance in the target domain. DisenCDR [4] addresses this issue by separating each user's shared and domain-specific representations and leveraging Kullback-Leibler (KL) divergence to enhance the distinction between user representations within each domain.

Despite their effectiveness, most previous works rely solely on overlapping users or items between the source and target domains for information transfer, **overlooking the strong connections that may exist between non-overlapping items**, which are



(a) The author of video is a streamer. (b) The author of video is not a streamer.

Figure 1: Illustration of user behaviors on the platform. (a) When users view a short video published by a streamer, they may click on the streamer's profile picture to enter the live streaming room. (b) When the author of the short video is not a streamer, users may explore related topics from the video and find relevant streamers to watch their live streams.

particularly significant between live streaming and short videos: (1) Streamers can either publish short videos or host live streams, and according to statistics from a popular live streaming platform, **70% of streamers share short videos**, which are often closely related to their live streams or personal content. (2) Furthermore, although certain authors of short videos are not streamers, they often **share clips that are directly derived from the live sessions of specific streamers**. This overlap suggests that user behaviors in live streaming may be influenced by their habits related to short videos. For example, Figure 1 illustrates some user behaviors on the platform. When the author of the short video that captures the user's interest is also a streamer, the user may click on the streamer's profile picture to enter the live streaming room and interact with the streamer. Even if the author is not a streamer, users may still seek out streamers whose content is related to the short videos they have enjoyed. Consequently, a comprehensive exploration of the relationships between short videos and streamers can not only improve the accuracy of user interest modeling, but also enhance the representation of streamers, thereby facilitating effective knowledge transfer across different domains.

In this paper, we propose **MGCCDR**, a **M**ulti-**G**raph **C**ontrastive learning framework for **C**ross-**D**omain **R**ecommendation aimed at enhancing information transfer and aggregation across domains by leveraging non-overlapping short videos and streamers. Specifically, to investigate the relationships between streamers and short videos, we first learn global representations of both entities from a global graph constructed using a combination of live streaming and short video data. Utilizing these representations, we identify significant connections by matching streamers and short videos from dual perspectives. Building upon this foundation, we further construct three bipartite graphs among users, streamers, and short videos. To fully leverage the information from each graph, we perform message propagation across all three, thereby generating user and streamer representations from various perspectives. This approach enables us to capture preferences within the target domain view, the source domain view, and the cross-domain view, collectively facilitating effective information transfer. Recognizing that each graph encapsulates distinct information and contributes differently to the overall recommendation task, we propose an attention-guided multi-graph fusion mechanism to integrate these representations effectively. In particular, we acknowledge that user interests may differ across domains, and the user-streamer graph provides a more relevant and less noisy view for the target domain. Consequently, we utilize the representations derived from this graph as the key to calculate attention probabilities, which serve as fusion weights to seamlessly integrate the multi-graph representations. In this way, we can extract valuable insights from diverse perspectives, thereby promoting cross-domain information integration and enhancing the representations of both users and streamers.

In the training process, we jointly optimize the entire framework using self-supervised contrastive learning loss and BPR loss. In summary, our major contributions are as follows:

- We highlight that there are significant connections between non-overlapping items in the cross-domain scenario of live streaming and short videos, which is a prevalent scenario in real-world recommendations but has yet to be thoroughly investigated in the existing literature.
- We propose MGCCDR, a novel cross-domain recommendation method that leverages multi-graph learning and an attention-guided multi-graph fusion mechanism to achieve more effective cross-domain information transfer and integration.
- Extensive experiments on both commercial and public datasets demonstrate the superior performance of MGCCDR. Notably, MGCCDR is not limited to short video and live streaming scenarios but is also applicable to general cross-domain recommendation tasks.

2 Related Work

2.1 Live Streaming Recommendation

Recently, live streaming, as an emerging form of social media, has drawn increasing attention from researchers. Unlike traditional recommendation scenarios that focus on suggesting items to users, live streaming recommendation aims to connect users with streamers they may be interested in. LiveRec [32] models the repeated consumption relationship between users and streamers, taking temporal dynamics into account during recommendation. eLiveRec [53] employs a disentangled encoder module to learn both cross-domain shared intents and domain-specific intents, enhancing e-commerce live streaming recommendations. Given the multi-modal nature of live streaming, which includes text, images, and audio, methods such as MTA [40] and ContentCTR [10] explore how to effectively fuse multi-modal information. Beyond multi-modal fusion, MMBee [12] introduces a graph representation learning approach based on meta-paths to enrich the representations of both users and streamers. Furthermore, some recent studies focus on capturing the dynamic changes within live streams. For instance, Sliver [21] modifies the data stream format of live streams, while KuaiHL [11] proposes a method for predicting highlight moments in live streaming sessions. Building on these works, Moment&Cross [5] enhances live streaming recommendations by incorporating short video data to assist in capturing user preferences more comprehensively. However, Moment&Cross adopts a user-centered approach, overlooking the critical role of streamers in cross-domain scenarios. Therefore, there remains significant room for improvement in leveraging abundant short video data to alleviate the data sparsity issue in live streaming recommendation.

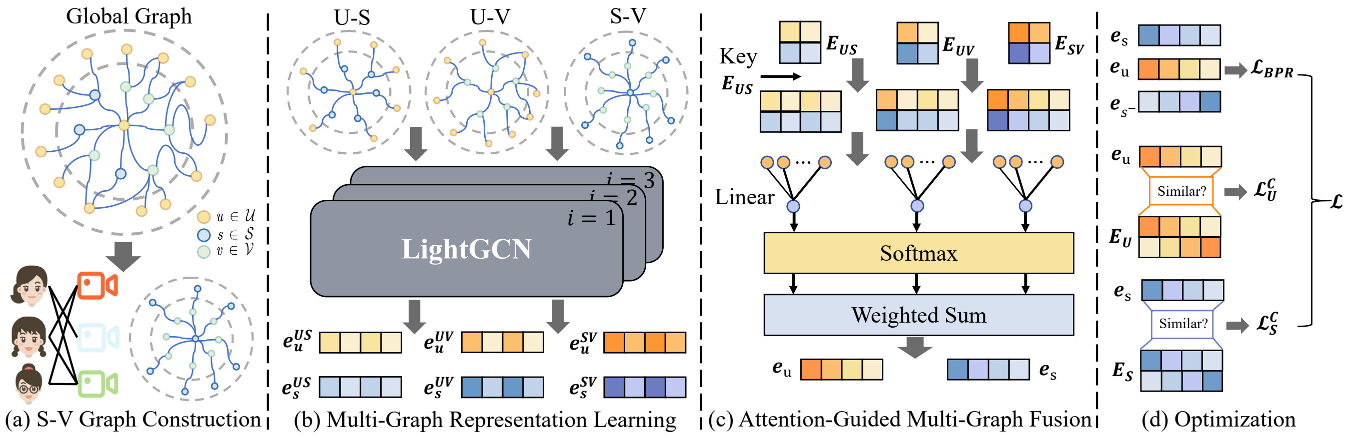


Figure 2: Overall framework of MGCCDR: a multi-graph contrastive learning framework for cross-domain recommendation.

2.2 Cross-Domain Recommendation

Cross-domain recommendation aims to enhance recommendation performance in the target domain by leveraging rich information from the source domain, effectively addressing the data sparsity problem in recommender systems [2, 14, 42, 57, 59]. Early cross-domain recommendation methods assume that auxiliary user-item interactions could assist in modeling users in the target domain, extending traditional single-domain recommendation models to multi-domain scenarios [35]. More recently, some approaches focus on transferring user/item representations from the source domain to improve recommendations in the target domain by designing various transfer modules to integrate the representations learned within their respective domains [18, 23, 28]. Meanwhile, recognizing that user interests across domains are not always identical, several studies adopt a disentangled representation learning paradigm to capture both domain-invariant and domain-specific features [4, 13, 22, 48, 52], effectively mitigating the negative transfer problem. Despite these advancements, existing methods primarily focus on information transfer between overlapping users or items, while overlooking the potential knowledge transfer between non-overlapping items. Our work aims to address these limitations by constructing an extra bipartite graph to connect items from the source and target domains and employing an attention-based multi-view fusion mechanism to enhance cross-domain recommendation.

3 Problem Formulation

Formally, given a set of users denoted as $\mathcal{U} = \{u_1, u_2, \dots, u_M\}$, a set of streamers represented by $\mathcal{S} = \{s_1, s_2, \dots, s_N\}$, and a set of short videos indicated by $\mathcal{V} = \{v_1, v_2, \dots, v_O\}$, where M, N, and O correspond to the number of users, streamers, and short videos, respectively. Let $\mathrm{X}_{M\times N} = \{x_{us} \mid u \in \mathcal{U}, s \in \mathcal{S}\}$ and $\mathrm{Y}_{M\times O} = \{y_{uv} \mid u \in \mathcal{U}, v \in \mathcal{V}\}$ represent the user-streamer interactions in live streaming domain and user-video interactions in short video domain, respectively. $x_{us}, y_{uv} \in \{0, 1\}$, where 1 denotes an interaction between user-streamer or user-video pair. Additionally, to explore the relationship between streamers and short videos, we also construct a binary matrix $\mathrm{Z}_{N\times O} = \{z_{sv} \mid s \in \mathcal{S}, v \in \mathcal{V}\}$, where $z_{sv} \in \{0, 1\}$, with $z_{sv} = 1$ indicating that the short video v is published by the streamer s. However, since many short video authors are not necessarily streamers, the matrix $\mathrm{Z}_{N\times O}$ is highly sparse. We will introduce how to uncover the underlying relationships between streamers and short videos to enrich $\mathrm{Z}_{N\times O}$ in Section 4.2. The goal of this paper is to explore how to leverage rich information from the short video domain to enhance live streaming recommendation, specifically by learning from $\mathrm{X}_{M\times N}$, $\mathrm{Y}_{M\times O}$, and $\mathrm{Z}_{N\times O}$ to predict unseen user-streamer interactions.

4 Our Approach: MGCCDR

4.1 Overview of MGCCDR

The framework of MGCCDR is illustrated in Figure 2. First, to enable cross-domain information transfer through non-overlapping items, we explore the relationships between streamers and short videos using a global graph that encompasses users, streamers, and short videos. By identifying closely related streamer-video pairs, we construct the Streamer-Video (S-V) graph. Simultaneously, we leverage user-streamer interactions to build the User-Streamer (U-S) graph and user-video interactions to construct the User-Video (U-V) graph. Then we apply multi-graph representation learning on the three graphs to capture user and streamer representations from different perspectives. Finally, we propose an attention-guided multi-graph fusion method that calculates attention probabilities between the representations obtained from each graph and the target graph representations. These attention probabilities serve as fusion weights, enabling the effective fusion of multi-graph representations. The entire model is optimized in a joint manner through a combination of BPR loss and contrastive learning loss.

4.2 Streamer-Video Graph Construction

As discussed in Section 1, some short videos are not created by streamers. However, these videos may consist of clips from a particular streamer's live content or be closely related to a specific streamer. To explore the strong connections between these non-overlapping items and enrich the sparse matrix $\mathrm{Z}_{N\times O}$, we integrate all user-streamer and user-video interactions to construct a global graph, where both streamers and videos are regarded as the same type of items. Then, we apply LightGCN [16] to learn the global representations of users, streamers, and videos. The similarity distribution between streamers and their respective published short

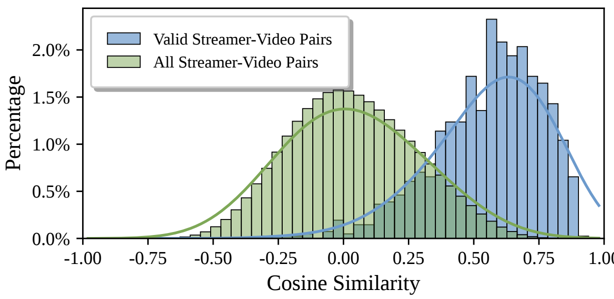


Figure 3: Cosine similarity distribution of valid streamer-video pairs and all streamer-video pairs. Valid streamer-video pairs refer to pairs where the streamer is the author of the short video, while all streamer-video pairs include all possible pairings between streamers and short videos.

videos, as well as the overall similarity distribution between all streamers and all short videos, is illustrated in Figure 3. It can be observed that the peak of the similarity distribution between streamers and their published short videos is significantly higher than that of the global similarity distribution. This validates the strong connection between streamers and their published short videos and highlights the rationale for exploring streamer-video relationships within the global graph. We enrich the matrix $\mathbf{Z}_{N\times O}$ between short videos and streamers from dual perspectives:

Short Video Perspective: We first enrich $\mathbf{Z}_{N\times O}$ from the short video perspective by identifying the most similar streamer $s \in \mathcal{S}$ for each short video $v \in \mathcal{V}$ whose publisher is not a streamer and setting z_{sv} to 1 accordingly, which can be formally defined as:

$$z_{sv} = \begin{cases} 1, & \text{if } s = \arg\max_{s'\in\mathcal{S}} \mathrm{sim}(\mathbf{e}_{s'}^G, \mathbf{e}_v^G), \\ z_{sv}, & \text{otherwise,} \end{cases} \tag{1}$$

where $\mathbf{e}_{s'}^G$ and $\mathbf{e}_v^G$ are the learned global representations of streamer s' and video v, and $\mathrm{sim}(\cdot,\cdot)$ represents the cosine similarity function.

Streamer Perspective: To further enrich $\mathbf{Z}_{N\times O}$, we take the streamer perspective by identifying short videos with high similarity to each streamer. Specifically, for each streamer $s \in \mathcal{S}$ with fewer than K associated short videos, we rank and identify short videos based on their similarity to the streamer. A short video v is selected if its similarity with the streamer satisfies $\mathrm{sim}(\mathbf{e}_s^G, \mathbf{e}_v^G) \geq \phi$, where ϕ is a predefined threshold set as the average similarity between each streamer and their published short videos in the initial matrix $\mathbf{Z}_{N\times O}$. We iteratively add the top-ranked short video to the matrix by setting $z_{sv} = 1$ for the corresponding pairs until the number of associated short videos reaches K. After the above process, if a streamer still has no associated short videos, we add the short video with the highest similarity to the matrix.

Through this dual-perspective relationship mining, we construct a bipartite graph between streamers and short videos, laying a solid foundation for the subsequent multi-graph representation learning and cross-domain information transfer.

4.3 Multi-Graph Representation Learning

By constructing bipartite graphs from interactions in both the live streaming domain and the short video domain, along with the previously built streamer-video bipartite graph, we obtain three bipartite graphs in total. By formulating these three graphs, we can capture target domain information, source domain information, and potential cross-domain fusion information, respectively. Through these distinct views, we learn user and streamer representations from different perspectives, thereby facilitating information transfer between the source and target domains. The representation learning process for the three graphs is detailed as follows.

4.3.1 User-Streamer Graph. The first graph we construct is the U-S graph, which aims to directly capture user and streamer representations within the live streaming domain. Following previous works [26, 27, 31, 37], we also employ the simple yet effective GNN-based collaborative filtering model, LightGCN, to capture the complex relationships between users and streamers by iteratively aggregating neighboring information over I layers. In the i-th layer, the aggregation proceeding is depicted as:

$$\begin{aligned} \mathbf{e}_u^{US(i)} &= \sum_{s\in\mathcal{N}_u^{US}} \frac{1}{\sqrt{|\mathcal{N}_u^{US}||\mathcal{N}_s^{US}|}} \mathbf{e}_i^{US(i-1)}, \\ \mathbf{e}_s^{US(i)} &= \sum_{u\in\mathcal{N}_s^{US}} \frac{1}{\sqrt{|\mathcal{N}_u^{US}||\mathcal{N}_s^{US}|}} \mathbf{e}_u^{US(i-1)}, \end{aligned} \tag{2}$$

where $\mathbf{e}_u^{US(i)}$ and $\mathbf{e}_s^{US(i)}$ represent embeddings of user u and streamer s at the i-th layer, $\mathcal{N}_u^{US}$ and $\mathcal{N}_s^{US}$ are the neighbors of user u and streamer s in the U-S graph. Then, we perform mean pooling over the representations from the 0-th layer to the I-th layer to obtain the final user and streamer representations, denoted as $\mathbf{e}_u^{US}$ and $\mathbf{e}_s^{US}$, respectively, in the U-S graph:

$$\mathbf{e}_u^{US} = \frac{1}{I}\sum_{i=0}^{I} \mathbf{e}_u^{US(i)}, \quad \mathbf{e}_s^{US} = \frac{1}{I}\sum_{i=0}^{I} \mathbf{e}_s^{US(i)}. \tag{3}$$

4.3.2 User-Video Graph. The second graph we construct is the U-V graph, which aims to leverage user-video interactions in the short video domain to learn the representations of users and short videos. Subsequently, we indirectly obtain the representation of streamers through the established S-V graph. Specifically, we first perform I iterations of information propagation on the U-V graph using LightGCN to obtain the representations of users and short videos. For each layer i, the information propagation is denoted as:

$$\begin{aligned} \mathbf{e}_u^{UV(i)} &= \sum_{v\in\mathcal{N}_u^{UV}} \frac{1}{\sqrt{|\mathcal{N}_u^{UV}||\mathcal{N}_v^{UV}|}} \mathbf{e}_v^{UV(i-1)}, \\ \mathbf{e}_v^{UV(i)} &= \sum_{u\in\mathcal{N}_v^{UV}} \frac{1}{\sqrt{|\mathcal{N}_u^{UV}||\mathcal{N}_v^{UV}|}} \mathbf{e}_u^{UV(i-1)}, \end{aligned} \tag{4}$$

where $\mathbf{e}_u^{UV(i)}$ and $\mathbf{e}_v^{UV(i)}$ represent embeddings of user u and short video v at the i-th layer, $\mathcal{N}_u^{UV}$ and $\mathcal{N}_v^{UV}$ are the neighbors of user u and short video v in the U-V graph. Similar to the U-S graph learning, we also perform mean pooling over the representations of each layer to obtain the final user and short video representations, denoted as $\mathbf{e}_u^{UV}$ and $\mathbf{e}_v^{UV}$, respectively, in the U-V graph:

$$\mathbf{e}_u^{UV} = \frac{1}{I}\sum_{i=0}^{I} \mathbf{e}_u^{UV(i)}, \quad \mathbf{e}_v^{UV} = \frac{1}{I}\sum_{i=0}^{I} \mathbf{e}_v^{UV(i)}. \tag{5}$$

Since our goal is to enhance live streaming recommendation by leveraging rich data from the short video domain, specifically to obtain better representations of users and streamers, we aggregate the representations of short videos associated with each streamer through the previously constructed S-V graph to obtain the streamer representation $\mathbf{e}_s^{UV}$ in U-V graph:

$$\mathbf{e}_s^{UV} = \frac{1}{\left|\mathcal{N}_s^{SV}\right|} \sum_{v \in \mathcal{N}_s^{SV}} \mathbf{e}_v^{UV}, \tag{6}$$

where $\mathcal{N}_s^{SV}$ represents the set of short videos associated with streamer s in S-V graph.

By learning on the U-V graph, we explore the relationships between users and short videos, obtaining user preference representations in the short video domain. Additionally, we derive fine-grained streamer representations by aggregating the associated short video embeddings, providing complementary representation information from a perspective different from that of the U-S graph.

4.3.3 Streamer-Video Graph. The third graph we construct is the S-V graph, which aims to explore the relationships between streamers and short videos, thereby enabling cross-domain information transfer across non-overlapping items. To further leverage the S-V graph for information transfer from the short video domain to the live streaming domain, in addition to aggregating short video representations into streamer representations in the U-V graph, we also apply LightGCN on the S-V graph to perform iterative aggregation of neighboring information across I layers. The graph propagation in the i-th layer is depicted as:

$$\begin{aligned} \mathbf{e}_s^{SV(i)} &= \sum_{v \in \mathcal{N}_s^{SV}} \frac{1}{\sqrt{|\mathcal{N}_s^{SV}||\mathcal{N}_v^{SV}|}} \mathbf{e}_v^{SV(i-1)}, \\ \mathbf{e}_v^{SV(i)} &= \sum_{s \in \mathcal{N}_v^{SV}} \frac{1}{\sqrt{|\mathcal{N}_s^{SV}||\mathcal{N}_v^{SV}|}} \mathbf{e}_s^{SV(i-1)}, \end{aligned} \tag{7}$$

where $\mathbf{e}_s^{SV(i)}$ and $\mathbf{e}_v^{SV(i)}$ represent embeddings of streamer s and short video v at the i-th layer, $\mathcal{N}_s^{SV}$ and $\mathcal{N}_v^{SV}$ are the neighbors of streamer s and short video v in the S-V graph. Similar to Equation 3 and Equation 5, we also perform mean pooling to obtain the representations $\mathbf{e}_s^{SV}$ and $\mathbf{e}_v^{SV}$ in the S-V graph:

$$\mathbf{e}_s^{SV} = \frac{1}{I} \sum_{i=0}^{I} \mathbf{e}_s^{SV(i)}, \quad \mathbf{e}_v^{SV} = \frac{1}{I} \sum_{i=0}^{I} \mathbf{e}_v^{SV(i)}. \tag{8}$$

Then we aggregate the representations of short videos that each user interacts with through the U-V graph to obtain the user representation $\mathbf{e}_u^{SV}$ in the S-V graph:

$$\mathbf{e}_u^{SV} = \frac{1}{\left|\mathcal{N}_u^{UV}\right|} \sum_{v \in \mathcal{N}_u^{UV}} \mathbf{e}_v^{SV}, \tag{9}$$

where $\mathcal{N}_u^{UV}$ represents the set of short videos interacted by user u in U-V graph.

In summary, our multi-graph learning framework obtains three sets of representations from different perspectives: $\mathbf{e}_u^{US}$ and $\mathbf{e}_s^{US}$ from the user-streamer graph, $\mathbf{e}_u^{UV}$ and $\mathbf{e}_s^{UV}$ from the user-video graph, and $\mathbf{e}_u^{SV}$ and $\mathbf{e}_s^{SV}$ from the streamer-video graph.

4.4 Attention-Guided Multi-Graph Fusion

After obtaining diverse representations through multi-graph learning, our objective is to effectively fuse these representations into the final user representation $\mathbf{e}_u$ and streamer representation $\mathbf{e}_s$.

Considering that only the U-S graph represents real user interactions in the live streaming domain, while the U-V and S-V graphs serve as auxiliary views, a straightforward fusion with uniform weights may result in negative transfer. This is attributable to the potential divergence in user interests between the live streaming and short video domains. Some existing methods attempt to mitigate this issue by assigning distinct fixed weights to each graph during fusion [26]. However, this approach imposes a significant optimization burden due to the extensive number of hyper-parameters involved. Furthermore, considering the considerable variability in user behaviors, it is imperative that the contribution of each graph to the final representation is tailored to the individual characteristics of each user and streamer. Consequently, the fusion coefficients should be personalized for each user and streamer.

To address these challenges, we propose Attention-Guided Multi-Graph Fusion (AMF), which effectively integrates representations from multi-graph based on an attention mechanism. We use the representations obtained from the U-S graph as the key to compute the attention probabilities with respect to the three sets of representations, and these probabilities are employed as fusion weights Specifically, for each graph $g \in \{US, UV, SV\}$, we first calculate the attention scores as follows:

$$r_u^g = \mathbf{Linear}(\mathbf{e}_u^g \| \mathbf{e}_u^{US}), \quad r_s^g = \mathbf{Linear}(\mathbf{e}_s^g \| \mathbf{e}_s^{US}), \tag{10}$$

where r_u^g and r_s^g represent the attention scores for user u and streamer s in graph g, respectively. The operator '$\|$' denotes the concatenation of two vectors.

Then we obtain attention probabilities using softmax function:

$$\lambda_u^g = \frac{\exp\left(r_u^g\right)}{\sum_{g \in \{US,UV,SV\}} \exp\left(r_u^g\right)}, \quad \lambda_s^g = \frac{\exp\left(r_s^g\right)}{\sum_{g \in \{US,UV,SV\}} \exp\left(r_s^g\right)}, \tag{11}$$

where λ_u^g and λ_s^g represent the attention probabilities for user u and streamer s in graph g, respectively.

By integrating the three sets of representations along with their corresponding attention weights, we obtain the final user and streamer representations $\mathbf{e}_u$ and $\mathbf{e}_s$:

$$\mathbf{e}_u = \sum\nolimits_{g \in \{US,UV,SV\}} \lambda_u^g \mathbf{e}_u^g, \quad \mathbf{e}_s = \sum\nolimits_{g \in \{US,UV,SV\}} \lambda_s^g \mathbf{e}_s^g. \tag{12}$$

AMF leverages an attention mechanism that utilizes the representations of the target domain as the key to calculate fusion weights, which not only reduces the number of hyper-parameters but also assigns personalized weights for each user and streamer, thereby facilitating the effective integration of multi-graph representations.

4.5 Prediction and Optimization.

4.5.1 Prediction. After obtaining the final user and streamer representations, we utilize the inner-product to calculate the preference score $\hat{y}_{u,s}$ for user u on streamer s:

$$\hat{y}_{u,s} = \mathbf{e}_u \cdot \mathbf{e}_s. \tag{13}$$

4.5.2 Optimization. For model optimization, we apply the standard BPR loss [33] as the main loss, which is widely used in the recommender system. The objective function is formally depicted as:

$$\mathcal{L}_{\text{BPR}} = \sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{N}_u^{US}} -\log \sigma\left(\hat{y}_{u,s} - \hat{y}_{u,s^-}\right), \tag{14}$$

where s^- represents the negative streamer that is randomly sampled from streamers that users have not interacted.

Algorithm 1 The Learning Algorithm of MGCCDR

Input: User-Streamer interactions matrix $\mathbf{X}_{M\times N}$, User-Video interactions matrix $\mathbf{Y}_{M\times O}$, Streamer-Video matrix $\mathbf{Z}_{N\times O}$, predefined threshold ϕ, layer number I, contrastive loss coefficient λ, temperature τ.
Output: MGCCDR Model with learnable parameters θ.

// Streamer-Video Graph Construction:
while LightGCN not Convergence **do**
  for $i = 1$ **to** I **do**
    Conduct message propagation in global graph
  end for
  Calculate BPR loss using Eq. (14)
  Update model parameter using Adam
end while
Get global representations: e_s^G and e_v^G and construct streamer-video graph
// Multi-Graph Representation Learning:
while MGCCDR not Convergence **do**
  for $i = 1$ **to** I **do**
    Conduct message propagation using Eq. (2), Eq. (4), and Eq. (7)
  end for
  Calculate $\mathbf{e}_u^{US}$, $\mathbf{e}_s^{US}$, $\mathbf{e}_u^{UV}$, $\mathbf{e}_s^{UV}$, $\mathbf{e}_u^{SV}$, and $\mathbf{e}_s^{SV}$ using Eq. (3), Eq. (5), Eq. (6), Eq. (8), and Eq. (9)
  // Attention-Guided Multi-Graph Fusion:
  Calculate $\mathbf{e}_u$ and $\mathbf{e}_s$ using Eq. (10), Eq. (11), and Eq. (12)
  // Optimization:
  Calculate BPR loss $\mathcal{L}_{\mathrm{BPR}}$ using Eq. (14)
  Calculate contrastive loss $\mathcal{L}_{\mathcal{U}}^{C}$ and $\mathcal{L}_{\mathcal{S}}^{C}$ using Eq. (15) and Eq. (16)
  Calculate total loss $\mathcal{L}$ using Eq. (17)
  Update model parameter using Adam
end while
return θ

Given the remarkable success of contrastive learning in the recommender system [39, 41, 44], we also incorporate a contrastive learning loss in addition to the BPR loss. Specifically, inspired by [44], we generate different representations for the same user and streamer by adding a noise vector to their original embeddings. We then optimize the model using the InfoNCE [15] loss:

$$\mathcal{L}_{\mathcal{U}}^{C} = -\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} \log \frac{\exp\left(\mathrm{sim}(\mathbf{e}_u, \mathbf{e}_u^{'})/\tau\right)}{\sum_{u^-\in\mathcal{U}} \exp\left(\mathrm{sim}(\mathbf{e}_u, \mathbf{e}_{u^-})/\tau\right)}, \tag{15}$$

$$\mathcal{L}_{\mathcal{S}}^{C} = -\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} \log \frac{\exp\left(\mathrm{sim}(\mathbf{e}_s, \mathbf{e}_s^{'})/\tau\right)}{\sum_{s^-\in\mathcal{S}} \exp\left(\mathrm{sim}(\mathbf{e}_s, \mathbf{e}_{s^-})/\tau\right)}, \tag{16}$$

where $(\mathbf{e}_u, \mathbf{e}_u^{'})$ and $(\mathbf{e}_s, \mathbf{e}_s^{'})$ represent the original and augmented representations of the same user u and streamer s, respectively, and τ denotes the temperature parameter.

Based on the BPR loss and the contrastive loss, we optimize our proposed MGCCDR by the joint loss $\mathcal{L}$:

$$\mathcal{L} = \mathcal{L}_{\mathrm{BPR}} + \lambda\left(\mathcal{L}_{\mathcal{U}}^{C} + \mathcal{L}_{\mathcal{S}}^{C}\right), \tag{17}$$

where λ is the hyper-parameter to balance the two loss functions.

The learning algorithm is presented in Algorithm 1.

5 Experiments

In this section, we first describe the experimental setups and then conduct an extensive evaluation and analysis of the proposed MGCCDR.

Table 1: Statistics of the experimental datasets.

Dataset	Domain	#Users	#Items	#Interactions	Sparsity
Comerical	Video	16,625	114,305	2,738,031	99.85%
	Live		10,980	180,445	99.90%
Douban	Movie	12,905	18,416	1,427,309	99.40%
	Book		24,646	552,794	99.82%
	Movie	12,425	18,102	1,341,810	99.40%
	Music		32,472	857,689	99.79%

5.1 Experimental Setups

5.1.1 Datasets. We conduct extensive experiments on two datasets: one collected from a popular live streaming platform and the other a well-known publicly available Douban dataset. The statistics of these datasets are shown in Table 1.

Commerical dataset. This dataset is constructed based on behavior logs of $16,625$ users who engaged with both live streaming and short video services on a popular live streaming platform over a one-week period in 2024. It includes each user's historical gifting behaviors in live streaming services and their click interactions with short video services. For data pre-processing, we group interaction records by user and apply filtering criteria to retain users with at least 5 interactions in the live streaming domain and at least 20 interactions in the short video domain. Additionally, we keep items with at least 5 interactions in the live streaming domain and at least 10 interactions in the short video domain.

Douban dataset[1]**.** To demonstrate that our methodology extends beyond live streaming and short video services and can be applied to various cross-domain scenarios, we also conduct experiments on the Douban dataset. Specifically, we select the three largest domains: movie, book, and music, to build cross-domain recommendation scenarios. We use the movie domain, which has more interaction records, as the source domain, and the book and music domains, which have fewer interactions, as the target domains. Accordingly, we construct two cross-domain recommendation tasks: *movie* → *book* and *movie* → *music.*

5.1.2 Evaluation Protocol. Following previous works [19, 36], we apply the leave-one-out splitting strategy for evaluation. To better validate our approach, the test set consists of the last interacted item and all uninteracted items for each user, which provides a more comprehensive evaluation compared to randomly sampling a subset of uninteracted items. The results of our MGCCDR and baselines are evaluated by HR@$\{10, 20, 40\}$ and NDCG@$\{10, 20, 40\}$.

5.1.3 Baselines. To demonstrate the effectiveness of our proposed method, we compare MGCCDR with with two categories of competitive baselines: single-domain methods and cross-domain methods.

Single-domain methods. We first compare MGCCDR with the following single-domain methods: (1) **BPRMF** [33] is a famous method based on matrix factorization and optimized with the BPR loss. (2) **NeuMF** [17] is a representative neural CF model that utilizes GMF and MLP simultaneously to learn representations. (3) **NGCF** [38] is a neural collaborative filtering method that leverages GNNs to capture high-order connections and collaborative signals. (4) **LightGCN** [16] removes the feature transformation and

[1] https://recbole.s3-accelerate.amazonaws.com/CrossDomain/Douban.zip

Table 2: Performance comparison between MGCCDR and the baselines on three cross-domain recommendation scenarios. The best and second-best performance methods are highlighted in bold and underlined fonts, respectively. * represents improvements over the second-best methods are significant (t-test with p-value < 0.01).

Datasets	Domain	Metrics	Single-domain Methods				Cross-domain Methods							Ours	%Improv
			BPRMF	NeuMF	NGCF	LightGCN	CMF	EMCDR	CoNet	DCDCSR	DTCDR	BiTGCF	DisenCDR	MGCCDR	
Comerical	Video → Live	HR@10	0.0471	0.0409	0.0513	0.0529	0.0829	0.0726	0.0373	0.0801	0.0524	0.0882	0.0583	**0.0983***	+11.45%
		NDCG@10	0.0245	0.0217	0.0267	0.0271	0.0430	0.0385	0.0184	0.0418	0.0263	0.0459	0.0291	**0.0513***	+8.38%
		HR@20	0.0729	0.0646	0.0792	0.0782	0.1264	0.1100	0.0608	0.1220	0.0852	0.1371	0.0907	**0.1486***	+6.16%
		NDCG@20	0.0310	0.0276	0.0337	0.0334	0.0540	0.0480	0.0243	0.0523	0.0345	0.0582	0.0372	**0.0640***	+11.76%
		HR@40	0.1083	0.0962	0.1178	0.1140	0.1844	0.1664	0.1003	0.1805	0.1316	0.2044	0.1422	**0.2170***	+9.96%
		NDCG@40	0.0382	0.0341	0.0415	0.0407	0.0658	0.0594	0.0323	0.0642	0.0439	0.0719	0.0477	**0.0780***	+8.48%
Douban	Movie → Book	HR@10	0.0555	0.0590	0.0580	0.0636	0.0655	0.0566	0.0471	0.0521	0.0465	0.0667	0.0335	**0.0863***	+29.38%
		NDCG@10	0.0293	0.0296	0.0311	0.0345	0.0341	0.0290	0.0241	0.0273	0.0239	0.0360	0.0160	**0.0467***	+29.72%
		HR@20	0.0855	0.0902	0.0896	0.0979	0.0966	0.0882	0.0721	0.0827	0.0769	0.1014	0.0570	**0.1246***	+22.87%
		NDCG@20	0.0369	0.0374	0.0391	0.0431	0.0419	0.0369	0.0304	0.0350	0.0316	0.0447	0.0219	**0.0563***	+25.95%
		HR@40	0.1317	0.1338	0.1338	0.1466	0.1408	0.1306	0.1077	0.1234	0.1166	0.1499	0.0890	**0.1712***	+14.20%
		NDCG@40	0.0463	0.0463	0.0481	0.0530	0.0509	0.0456	0.0376	0.0433	0.0396	0.0546	0.0285	**0.0658***	+20.51%
	Movie → Music	HR@10	0.0453	0.0488	0.0492	0.0460	0.0516	0.0456	0.0393	0.0473	0.0402	0.0617	0.0223	**0.0767***	+24.31%
		NDCG@10	0.0236	0.0259	0.0258	0.0233	0.0274	0.0229	0.0206	0.0258	0.0200	0.0334	0.0109	**0.0424***	+26.94%
		HR@20	0.0719	0.0744	0.0773	0.0707	0.0809	0.0724	0.0616	0.0751	0.0666	0.0929	0.0402	**0.1138***	+22.49%
		NDCG@20	0.0302	0.0323	0.0329	0.0295	0.0348	0.0296	0.0262	0.0328	0.0266	0.0412	0.0155	**0.0517***	+25.48%
		HR@40	0.1082	0.1136	0.1161	0.1090	0.1235	0.1115	0.0998	0.1128	0.1055	0.1364	0.0631	**0.1556***	+14.07%
		NDCG@40	0.0377	0.0403	0.0408	0.0372	0.0434	0.0376	0.0340	0.0405	0.0345	0.0501	0.0201	**0.0603***	+20.35%

non-linear activation components in vanilla GCN to simplify the model structure. Following previous works [2, 3, 57], we merge all interactions of both domains as one domain to train them.

Cross-domain methods. We also compare MGCCDR with the following cross-domain methods: (1) **CMF** [35] enhances traditional matrix factorization by jointly factorizing interaction matrices from multiple domains. (2) **CoNet** [18] facilitates cross-domain knowledge transfer by introducing cross-connection units within MLPs. (3) **EMCDR** [28] addresses the cross-domain cold-start problem by learning a mapping function to align user embeddings across domains. (4) **DCDCSR** [59] combines MF and DNN to map user and item latent factors across domains for cross-system recommendations. (5) **DTCDR** [58] leverages multi-task learning with an adaptable embedding-sharing strategy to improve cross-domain user and item representations. (6) **BiTGCF** [23] integrates graph collaborative filtering with bidirectional knowledge transfer to enhance recommendations across both domains. (7) **DisenCDR** [4] disentangles user representations into domain-shared and domain-specific components, transferring only the shared features.

5.1.4 Implementation Details. We implement all baseline methods using the RecBole [55, 56] library, except for DisenCDR, which is directly implemented using the released codes by the authors. To ensure a fair comparison, we set the embedding size of all methods to 64 and use a batch size of 2048. We optimize all these methods with Adam optimizer and carefully tune the learning rate among $\{1e{-}3, 5e{-}3, 1e{-}4, 5e{-}4, 1e{-}5\}$, the weight decay among $\{1e{-}5, 1e{-}6, 1e{-}7\}$, as well as the layer number I among $\{1, 2, 3\}$. We set the similarity threshold ϕ to 0.6, 0.5, and 0.4 for the three cross-domain recommendation scenarios, and the value of K to 25. All experiments are conducted on NVIDIA Tesla V100 32G GPUs. Our code is available at https://github.com/quchangle1/MGCCDR.

5.2 Experimental Results

We compare our proposed MGCCDR with the aforementioned baselines on Commercial and Douban datasets, as shown in Table 2. From the experimental results, we have the following observations:

• Firstly, compared to BPRMF and NeuMF, both NGCF and LightGCN achieve better results. This demonstrates that GNN-based methods, by incorporating higher-order neighborhood information, are more effective at capturing collaborative filtering signals, thereby learning improved user and streamer representations.

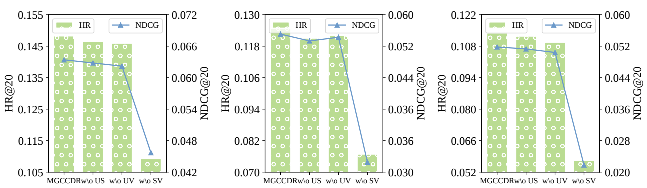
• Secondly, compared to single-domain recommendation methods, cross-domain methods achieve better results. This indicates the effectiveness of information transfer across domains, as data from the source domain can also reflect user preferences, helping to more accurately capture their preferences in the target domain.

• Nevertheless, it is noteworthy that not all cross-domain methods demonstrate superior performance compared to single-domain methods. This suggests that the mere amalgamation of user preferences from both the source and target domains may not only be ineffective in improving performance but may also lead to negative transfer. Therefore, designing effective strategies for information transfer and fusion is crucial for cross-domain recommendation.

• Finally, compared to all baseline methods, MGCCDR achieves the best results across three cross-domain scenarios, significantly surpassing SOTA methods. This demonstrates the effectiveness of MGCCDR, which can be attributed to its capacity for multi-graph learning by uncovering relationships between non-overlapping items and the implementation of an effective fusion approach. In this way, MGCCDR effectively captures valuable information from source domain, target domain, and cross-domain interactions, thereby facilitating information transfer and fusion, which ultimately enhances the representations of both users and streamers.

Table 3: Ablation study of the proposed MGCCDR on three cross-domain recommendation scenarios. "*w/o*" represents "without". AMF is short for Attention-Guided Multi-Graph Fusion. CL is short for Contrastive Loss.

Methods	Commerical		Douban			
	Video → Live		Movie → Book		Movie → Music	
	H@20	N@20	H@20	N@20	H@20	N@20
MGCCDR	**0.1486**	**0.0640**	**0.1246**	**0.0563**	**0.1138**	**0.0517**
w/o AMF	0.1159	0.0489	0.0966	0.0417	0.0733	0.0301
w/o Noise	0.1269	0.0537	0.1217	0.0553	0.1113	0.0515
w/o CL	0.0782	0.0322	0.0390	0.0168	0.0377	0.0153



(a) Video → Live. (b) Movie → Book. (c) Movie → Music.

Figure 4: Ablation study on the impact of each graph in MGCCDR framework on three cross-domain recommendation scenarios. SV, UV, and US represent the cross-domain graph, source domain graph, and target domain graph, respectively.

5.3 Ablation Study.

We conduct ablation studies by systematically removing each component at a time from MGCCDR. The results presented in Table 3 highlight the significance of each element:

***w/o* AMF** refers to a variant that performs multi-graph representation fusion using a straightforward, uniform-weighted strategy. This substitution leads to a notable decline in performance, primarily due to the misalignment of user interests across domains, as well as the differential contributions of each graph to the overall recommendation task. It can be intuitively understood that the graph constructed from user interactions within the target domain is characterized by reduced noise and assumes a more pivotal role in the final recommendation process. This observation validates the effectiveness of our proposed AMF, which leverages the target domain graph as a key to calculate attention probabilities, thereby employing them as weights for more effective fusion.

***w/o* Noise** is a variant that removes the noise-based data augmentation component. This omission also results in a noticeable performance drop, indicating that the incorporation of noise for data augmentation effectively mitigates the issue of representation degradation, particularly for low-degree nodes, thereby enabling the model to learn more robust and informative representations. Additionally, it is noted that the extent of performance decline varies across different datasets. In the Douban dataset, the performance drop is relatively minor, whereas the decline is much more pronounced in the commercial dataset. This suggests that commercial datasets are more sensitive to noise-based augmentation and benefit more from it, likely due to their diverse user behavior and sparse interaction patterns. Thus, a higher ratio of noise injection might be required to enhance representation learning in such scenarios.

***w/o* CL** represents the variant where the model is optimized solely with BPR loss, without the joint optimization of contrastive loss and BPR loss. The removal of contrastive loss significantly diminishes performance, demonstrating the necessity of incorporating contrastive learning to enhance representation learning. Moreover, our analysis reveals that the performance drop is most pronounced in the two cross-domain scenarios within the Douban dataset, indicating that the importance of contrastive learning varies across different datasets. This observation implies that contrastive learning is particularly vital in datasets characterized by more extensive user-item interactions in the target domain.

Impact of Each Graph in MGCCDR. In addition to the above ablation studies, to investigate the impact of each graph within our multi-graph representation learning framework, we compare MGCCDR with its variants where each graph was individually removed: *w/o* SV, *w/o* UV, and *w/o* US. As shown in Figure 4, the performance drops in all cases, indicating the essential role of each graph in facilitating effective cross-domain information transfer. Notably, the removal of the U-V and S-V graphs results in minor performance degradation, whereas the exclusion of the U-S graph leads to a substantial decline in performance. This observation emphasizes the significance of the U-S graph and further validates our decision to utilize the representations obtained from the US graph as the key in the AMF mechanism.

5.4 Further Analysis

In this section, we conduct more detailed experiments to investigate the effectiveness of MGCCDR, providing an in-depth analysis of how and why MGCCDR achieves state-of-the-art performance.

5.4.1 S-V Graph and Cross-Domain Interest Transfer Analysis. The proposed model, MGCCDR, distinguishes itself from other cross-domain methodologies by establishing connections between non-overlapping items across the source and target domains, thereby constructing corresponding graphs to facilitate information transfer. To investigate the underlying factors contributing to the superior performance of MGCCDR, we calculate the proportion of instances in which the S-V graph aids in capturing user cross-domain interest transfer. Specifically, we compute the proportion of users within the test set who have viewed short videos related to a particular streamer prior to interacting with that streamer. The proportions for the three cross-domain scenarios are 15.2%, 32.32%, and 35%, respectively. These results suggest that the S-V graph can, to some extent, assist in identifying user cross-domain interest transfer. Furthermore, we note that the proportion in the Douban dataset exceeds that of the commercial dataset, which explains why MGCCDR achieves a greater performance improvement on the Douban dataset compared to the commercial dataset.

5.4.2 Performance Comparison w.r.t Different Data Sparsity. To assess the efficacy of our model in addressing the issue of data sparsity within the target domain, we categorize the test users into four groups based on sparsity degrees, *i.e.,* the number of interactions in the target domain, and then analyze the performance of the model across these groups. The experimental results are presented in Figure 5, where it is evident that MGCCDR consistently surpasses the most robust baseline, BiTGCF, across all groups. This finding

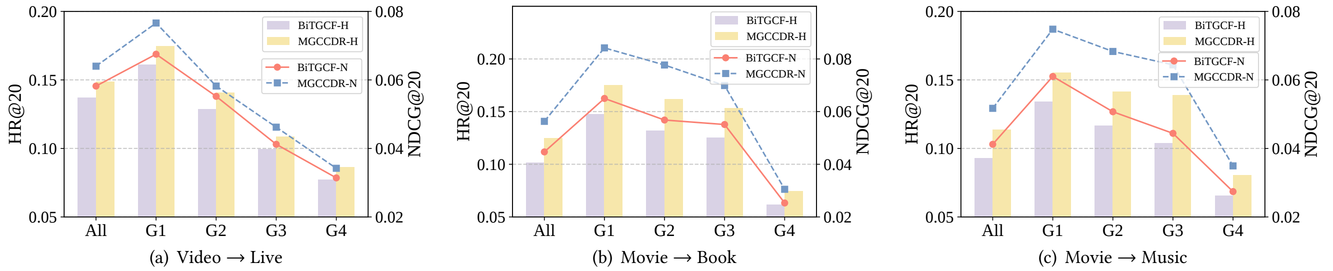


Figure 5: Performance comparison regarding different level of data sparsity on three cross-domain recommendation scenarios. The X-axis represents grouping by user interactions from low to high, which are G1, G2, G3, and G4, respectively.

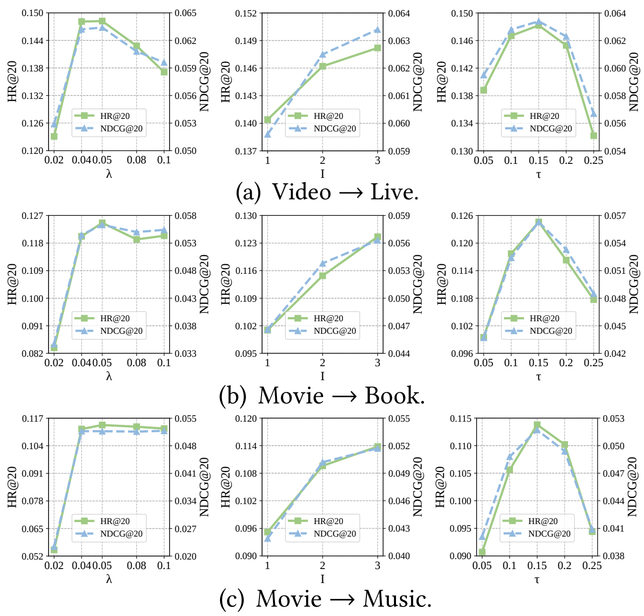


Figure 6: Hyper-parameter analysis w.r.t. λ, I, τ on three cross-domain recommendation scenarios.

underscores the robustness of MGCCDR and its ability to adapt effectively to a range of scenarios. Furthermore, we observe that both BiTGCF and MGCCDR perform even better in groups with sparse interactions compared to those with abundant interactions. This phenomenon may be attributed to differences in the amount of source-domain data or the challenge of capturing users' recent interests when their preferences are highly diverse. In such instances, GNN-based methods that do not account for interaction sequences tend to encounter difficulties.

5.4.3 Hyper-parameter Analysis. We further conduct experiment to investigate several key hyper-parameters introduced in MGCCDR.

Impact of the coefficient λ. We change the coefficient of contrastive loss λ from 0.02 to 0.1 and present the results in Figure 6. The results indicate a significant improvement in performance as the parameter λ is increased from 0.02 to 0.05, indicating that contrastive learning helps to learn better representations. However, further increasing λ from 0.05 to 0.1 yields varying trends across datasets—performance decreases, first drops and then rises, or remains stable. This is likely because an extremely large coefficient may disrupt the BPR loss, leading to performance degradation.

Impact of the layer number I. We evaluate MGCCDR with different layer numbers $I = 1, 2$, and 3. The results presented in Figure 6 show that the performance of MGCCDR improves progressively with an increase in I, which suggests that MGCCDR effectively mitigates the over-smoothing problem commonly associated with excessive layers in graph neural networks by introducing noise for data augmentation. Consequently, MGCCDR is capable of capturing deeper neighbor information and ultimately achieving superior representations of users and streamers.

Impact of the temperature τ. We vary the temperature τ from 0.05 to 0.25 and present the results in Figure 6. The results indicate that MGCCDR is highly sensitive to the τ, with notable fluctuations in performance corresponding to different values of τ. Specifically, as τ is increased from 0.05 to 0.15, performance improves because a very small τ makes contrastive learning overly focused on optimizing hard negatives. Conversely, further increasing τ from 0.15 to 0.25 leads to a performance drop, as an excessively large τ may fail to distinguish hard negatives from other negative samples. Thus, selecting an appropriate τ is crucial for contrastive learning.

6 Conclusion

We propose MGCCDR, a novel cross-domain live streaming recommendation method that bridges short videos and streamers through multi-graph contrastive learning. Unlike previous works that transfer cross-domain information through overlapping users or items, we emphasize the strong ties between non-overlapping short videos and streamers for cross-domain knowledge transfer. Specifically, MGCCDR first leverages multi-graph learning to capture representations from the source domain, target domain, and cross-domain separately. Subsequently, we apply attention-based aggregation to extract useful information from different graphs, facilitating the transfer and integration of information across domains. Extensive experimental results demonstrate the effectiveness of MGCCDR.

Acknowledgments

This work was partially supported by the National Science and Technology Major Project (2022ZD0114802), the National Natural Science Foundation of China (No. 62376275,62472426), Intelligent Social Governance Interdisciplinary Platform, Major Innovation & Planning Interdisciplinary Platform for the "Double-First Class" Initiative, Renmin University of China. Supported by PCC@RUC, fund for building world-class universities (disciplines) of Renmin University of China.

References

[1] Zhicheng An, Zhexu Gu, Li Yu, Ke Tu, Zhengwei Wu, Binbin Hu, Zhiqiang Zhang, Lihong Gu, and Jinjie Gu. 2024. DDCDR: A Disentangle-based Distillation Framework for Cross-Domain Recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4764–4773.

[2] Jiangxia Cao, Xin Cong, Jiawei Sheng, Tingwen Liu, and Bin Wang. 2022. Contrastive cross-domain sequential recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 138–147.

[3] Jiangxia Cao, Shaoshuai Li, Bowen Yu, Xiaobo Guo, Tingwen Liu, and Bin Wang. 2023. Towards universal cross-domain recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 78–86.

[4] Jiangxia Cao, Xixun Lin, Xin Cong, Jing Ya, Tingwen Liu, and Bin Wang. 2022. Disencdr: Learning disentangled representations for cross-domain recommendation. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval*. 267–277.

[5] Jiangxia Cao, Shen Wang, Yue Li, Shenghui Wang, Jian Tang, Shiyao Wang, Shuang Yang, Zhaojie Liu, and Guorui Zhou. 2024. Moment&Cross: Next-Generation Real-Time Cross-Domain CTR Prediction for Live-Streaming Recommendation at Kuaishou. *arXiv preprint arXiv:2408.05709* (2024).

[6] Xinyue Chen, Si Chen, Xu Wang, and Yun Huang. 2021. " I was afraid, but now I enjoy being a streamer!" Understanding the Challenges and Prospects of Using Live Streaming for Online Education. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–32.

[7] Sunhao Dai, Changle Qu, Sirui Chen, Xiao Zhang, and Jun Xu. 2024. Recode: Modeling repeat consumption with neural ode. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2599–2603.

[8] Sunhao Dai, Ninglu Shao, Jieming Zhu, Xiao Zhang, Zhenhua Dong, Jun Xu, Quanyu Dai, and Ji-Rong Wen. 2024. Modeling user attention in music recommendation. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*. IEEE, 761–774.

[9] Sunhao Dai, Yuqi Zhou, Jun Xu, and Ji-Rong Wen. 2023. Dually Enhanced Delayed Feedback Modeling for Streaming Conversion Rate Prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 390–399.

[10] Jiaxin Deng, Dong Shen, Shiyao Wang, Xiangyu Wu, Fan Yang, Guorui Zhou, and Gaofeng Meng. 2023. ContentCTR: Frame-level Live Streaming Click-Through Rate Prediction with Multimodal Transformer. *arXiv preprint arXiv:2306.14392* (2023).

[11] Jiaxin Deng, Shiyao Wang, Dong Shen, Liqin Zhao, Fan Yang, Guorui Zhou, and Gaofeng Meng. 2024. A Multimodal Transformer for Live Streaming Highlight Prediction. In *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.

[12] Jiaxin Deng, Shiyao Wang, Yuchen Wang, Jiansong Qi, Liqin Zhao, Guorui Zhou, and Gaofeng Meng. 2024. MMBee: Live Streaming Gift-Sending Recommendations via Multi-Modal Fusion and Behaviour Expansion. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4896–4905.

[13] Jing Du, Zesheng Ye, Bin Guo, Zhiwen Yu, and Lina Yao. 2024. Identifiability of cross-domain recommendation via causal subspace disentanglement. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval*. 2091–2101.

[14] Sheng Gao, Hao Luo, Da Chen, Shantao Li, Patrick Gallinari, and Jun Guo. 2013. Cross-domain recommendation via cluster-level latent factor model. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part II 13*. Springer, 161–176.

[15] Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 297–304.

[16] 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*. 639–648.

[17] 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*. 173–182.

[18] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. Conet: Collaborative cross networks for cross-domain recommendation. In *Proceedings of the 27th ACM international conference on information and knowledge management*. 667–676.

[19] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*. IEEE, 197–206.

[20] Hanyu Li, Weizhi Ma, Peijie Sun, Jiayu Li, Cunxiang Yin, Yancheng He, Guoqiang Xu, Min Zhang, and Shaoping Ma. 2024. Aiming at the target: Filter collaborative information for cross-domain recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2081–2090.

[21] Fengqi Liang, Baigong Zheng, Liqin Zhao, Guorui Zhou, Qian Wang, and Yanan Niu. 2024. Ensure Timeliness and Accuracy: A Novel Sliding Window Data Stream Paradigm for Live Streaming Recommendation. *arXiv preprint arXiv:2402.14399* (2024).

[22] 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*, Vol. 38. 8769–8777.

[23] Meng Liu, Jianjun Li, Guohui Li, and Peng Pan. 2020. Cross domain recommendation via bi-directional transfer graph collaborative filtering networks. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 885–894.

[24] Haokai Ma, Ruobing Xie, Lei Meng, Xin Chen, Xu Zhang, Leyu Lin, and Jie Zhou. 2023. Exploring False Hard Negative Sample in Cross-Domain Recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems* (Singapore, Singapore) *(RecSys '23)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3604915.3608791

[25] Haokai Ma, Ruobing Xie, Lei Meng, Xin Chen, Xu Zhang, Leyu Lin, and Jie Zhou. 2023. Triple Sequence Learning for Cross-domain Recommendation. *ACM Trans. Inf. Syst. (TOIS)* (2023).

[26] Yunshan Ma, Yingzhi He, Xiang Wang, Yinwei Wei, Xiaoyu Du, Yuyangzi Fu, and Tat-Seng Chua. 2024. MultiCBR: Multi-view Contrastive Learning for Bundle Recommendation. *ACM Transactions on Information Systems* 42, 4 (2024), 1–23.

[27] Yunshan Ma, Yingzhi He, An Zhang, Xiang Wang, and Tat-Seng Chua. 2022. Crosscbr: Cross-view contrastive learning for bundle recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1233–1241.

[28] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-domain recommendation: An embedding and mapping approach.. In *IJCAI*, Vol. 17. 2464–2470.

[29] Cuong Nguyen, Tal Hassner, Matthias Seeger, and Cedric Archambeau. 2020. Leep: A new measure to evaluate transferability of learned representations. In *International Conference on Machine Learning*. PMLR, 7294–7305.

[30] Chung Park, Taesan Kim, Hyungjun Yoon, Junui Hong, Yelim Yu, Mincheol Cho, Minsung Choi, and Jaegul Choo. 2024. Pacer and Runner: Cooperative Learning Framework between Single-and Cross-Domain Sequential Recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2071–2080.

[31] Changle Qu, Sunhao Dai, Xiaochi Wei, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, Jun Xu, and Ji-Rong Wen. 2024. Towards completeness-oriented tool retrieval for large language models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 1930–1940.

[32] Jérémie Rappaz, Julian McAuley, and Karl Aberer. 2021. Recommendation on live-streaming platforms: Dynamic availability and repeat consumption. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 390–399.

[33] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. 452–461.

[34] Chenglei Shen, Xiao Zhang, Wei Wei, and Jun Xu. 2023. HyperBandit: Contextual bandit with hypernetwork for time-varying user preferences in streaming recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM 2023)*. 2239–2248.

[35] Ajit P Singh and Geoffrey J Gordon. 2008. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 650–658.

[36] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1441–1450.

[37] Jiakai Tang, Sunhao Dai, Zexu Sun, Xu Chen, Jun Xu, Wenhui Yu, Lantao Hu, Peng Jiang, and Han Li. 2024. Towards Robust Recommendation via Decision Boundary-aware Graph Contrastive Learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2854–2865.

[38] 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*. 165–174.

[39] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. 2021. Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 726–735.

[40] Dinghao Xi, Liumin Tang, Runyu Chen, and Wei Xu. 2023. A multimodal time-series method for gifting prediction in live streaming platforms. *Information Processing & Management* 60, 3 (2023), 103254.

[41] Ruobing Xie, Qi Liu, Liangdong Wang, Shukai Liu, Bo Zhang, and Leyu Lin. 2022. Contrastive cross-domain recommendation in matching. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 4226–4236.

[42] Wenhao Yang, Yingchun Jian, Yibo Wang, Shiyin Lu, Lei Shen, Bing Wang, Haihong Tang, and Lijun Zhang. 2024. Not All Embeddings are Created Equal: Towards Robust Cross-domain Recommendation via Contrastive Learning. In *Proceedings of the ACM on Web Conference 2024*. 3195–3206.

[43] Xin Yang, Heng Chang, Zhijian Lai, Jinze Yang, Xingrun Li, Yu Lu, Shuaiqiang Wang, Dawei Yin, and Erxue Min. 2024. Hyperbolic Contrastive Learning for Cross-Domain Recommendation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 2920–2929.

[44] 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*. 1294–1303.

[45] Sanshi Yu, Zhuoxuan Jiang, Dong-Dong Chen, Shanshan Feng, Dongsheng Li, Qi Liu, and Jinfeng Yi. 2021. Leveraging tripartite interaction information from live stream e-commerce for improving product recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 3886–3894.

[46] Changshuo Zhang, Teng Shi, Xiao Zhang, Qi Liu, Ruobing Xie, Jun Xu, and Ji-Rong Wen. 2024. Modeling Domain and Feedback Transitions for Cross-Domain Sequential Recommendation. *arXiv preprint arXiv:2408.08209* (2024).

[47] Kepu Zhang, Teng Shi, Sunhao Dai, Xiao Zhang, Yinfeng Li, Jing Lu, Xiaoxue Zang, Yang Song, and Jun Xu. 2024. SAQRec: Aligning Recommender Systems to User Satisfaction via Questionnaire Feedback. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 3165–3175.

[48] Ruohan Zhang, Tianzi Zang, Yanmin Zhu, Chunyang Wang, Ke Wang, and Jiadi Yu. 2023. Disentangled contrastive learning for cross-domain recommendation. In *International Conference on Database Systems for Advanced Applications*. Springer, 163–178.

[49] Wen Zhang, Lingfei Deng, Lei Zhang, and Dongrui Wu. 2022. A survey on negative transfer. *IEEE/CAA Journal of Automatica Sinica* 10, 2 (2022), 305–329.

[50] Xiao Zhang, Sunhao Dai, Jun Xu, Zhenhua Dong, Quanyu Dai, and Ji-Rong Wen. 2022. Counteracting user attention bias in music streaming recommendation via reward modification. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2022)*. 2504–2514.

[51] Xiao Zhang, Haonan Jia, Hanjing Su, Wenhan Wang, Jun Xu, and Ji-Rong Wen. 2021. Counterfactual reward modification for streaming recommendation with delayed feedback. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2021)*. 41–50.

[52] Xinyue Zhang, Jingjing Li, Hongzu Su, Lei Zhu, and Heng Tao Shen. 2023. Multi-level attention-based domain disentanglement for BCDR. *ACM Transactions on Information Systems* 41, 4 (2023), 1–24.

[53] Yixin Zhang, Yong Liu, Hao Xiong, Yi Liu, Fuqiang Yu, Wei He, Yonghui Xu, Lizhen Cui, and Chunyan Miao. 2023. Cross-domain disentangled learning for e-commerce live streaming recommendation. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. IEEE, 2955–2968.

[54] Yuxi Zhang, Ji Zhang, Feiyang Xu, Lvying Chen, Bohan Li, Lei Guo, and Hongzhi Yin. 2024. Preference Prototype-Aware Learning for Universal Cross-Domain Recommendation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 3290–3299.

[55] Wayne Xin Zhao, Yupeng Hou, Xingyu Pan, Chen Yang, Zeyu Zhang, Zihan Lin, Jingsen Zhang, Shuqing Bian, Jiakai Tang, Wenqi Sun, et al. 2022. RecBole 2.0: towards a more up-to-date recommendation library. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4722–4726.

[56] Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, et al. 2021. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms. In *proceedings of the 30th acm international conference on information & knowledge management*. 4653–4664.

[57] Jiawei Zheng, Hao Gu, Chonggang Song, Dandan Lin, Lingling Yi, and Chuan Chen. 2023. Dual Interests-Aligned Graph Auto-Encoders for Cross-domain Recommendation in WeChat. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 4988–4994.

[58] Feng Zhu, Chaochao Chen, Yan Wang, Guanfeng Liu, and Xiaolin Zheng. 2019. Dtcdr: A framework for dual-target cross-domain recommendation. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1533–1542.

[59] Feng Zhu, Yan Wang, Chaochao Chen, Guanfeng Liu, Mehmet Orgun, and Jia Wu. 2020. A deep framework for cross-domain and cross-system recommendations. In *IJCAI*.