Causal Disentanglement-Enhanced Diffusion Denoising for Social Recommendation

SHIXIAO YANG, School of Computer Science and Technology, Beijing Institute of Technology, China
ZHIDA QIN, School of Computer Science and Technology, Beijing Institute of Technology, China
ENJUN DU, School of Cyberspace Science and Technology, Beijing Institute of Technology, China
HAOYAN FU, School of Computer Science and Technology, Beijing Institute of Technology, China
HAOYAO ZHANG, School of Computer Science and Technology, Beijing Institute of Technology, China
PENGZHAN ZHOU, College of Computer Science, Chongqing University, China
TIANYU HUANG, School of Computer Science and Technology, Beijing Institute of Technology, China
GANGYI DING, School of Computer Science and Technology, Beijing Institute of Technology, China

In recent years, social recommendation systems have emerged as a pivotal technology for enhancing recommendation accuracy by leveraging user social homophily and influence. Despite that lots of works have been devoted to this area, existing works still struggle to extract the beneficial structural information from social relationships that is beneficial for recommendations and neglect the inherent popularity bias in the social networks, which leads to suboptimal recommendation performances. To address these challenges, we propose a novel framework termed Causal Disentanglement-Enhanced Diffusion Denoising for Social Recommendation (CaDDiSR). This framework first employs causal graphs to disentangle the complexities of social relationships, generating user representations with high-order structures, which are subsequently used as inputs to a diffusion process to effectively denoise social networks and retain social signals beneficial for recommendation tasks. Furthermore, the framework integrates a bidirectional knowledge distillation mechanism, which balances user representations between social and recommendation contexts, thereby facilitating the effective fusion of their respective advantages while simultaneously mitigating noise interference and enhancing overall system performance. Finally, cross-domain contrastive learning is utilized to optimize user and item representations, ensuring consistency in recommendation performance across diverse scenarios. Experimental results on multiple real-world datasets demonstrate that CaDDiSR significantly outperforms existing baseline models, substantiating its superior performance.

CCS Concepts: • **Information systems → Recommender systems**.

Additional Key Words and Phrases: Graph Neural Network, Contrastive Learning, Social Networks, Recommendation System

---

Authors' addresses: Shixiao Yang, ysx144_51@bit.edu.cn, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China; Zhida Qin, zanderqin@bit.edu.cn, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China; Enjun Du, enjundu.cs@gmail.com, School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing, China; Haoyan Fu, haoyan-fu@bit.edu.cn, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China; Haoyao Zhang, zhanghaoyao@bit.edu.cn, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China; Pengzhan Zhou, pzzhou@cqu.edu.cn, College of Computer Science, Chongqing University, Chongqing, China; Tianyu Huang, huangtianyu@bit.edu.cn, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China; Gangyi Ding, dgy@bit.edu.cn, School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China.

Manuscript submitted to ACM

ACM Reference Format:
Shixiao Yang, Zhida Qin, Enjun Du, Haoyan Fu, Haoyao Zhang, Pengzhan Zhou, Tianyu Huang, and Gangyi Ding. 2018. Causal Disentanglement-Enhanced Diffusion Denoising for Social Recommendation. 1, 1 (January 2018), 22 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In recent years, social platforms [15] have experienced widespread adoption and proliferation on the Internet. Correspondingly, the Social Recommender System (SocialRS) has emerged as a novel recommendation technology that leverages the social homophily [25] and social influence [24] between users. SocialRS aims to fully harness such social information to establish more accurate user profiles and interest preference models.

The forefront of research on SocialSR lies in the GNN (Graph Neural Networks)-based recommendation systems. Specifically, a series of works focus on distill the social neighbor's interests to achieve better user embedding and preference modeling. The DiffNet [38] and DiffNet++ [37] methods inject the user-latent interests and higher-order influences into the embedding learning by efficiently modeling the interest and modeling diffusion process. The DcRec [35] model utilize the contrastive mechanism between the social domain and interaction domain to transfer social knowledge and enhance user representations. Different with them, the works of GDMSR [27] and DSL [31] turn to calculate the informativeness of each social relations and denoise the irrelevant user connections. Another series of works pay attention to sparsity issue in social networks. The work SEPT [46] apply the self-supervised tri-training framework to encode the augmented social graphs and user-item interaction graph, which interactively improves the user representation learning process. The MHCN [48] model use the local user social structure and interaction graph to construct the refined hypergraph and obtain more comprehensive user representations. In summary, the key of the GNN-based SocialRS is to incorporate the beneficial social information into the user-item interaction graphs and alleviate the severe sparsity phenomenon in interaction and social domains.

Table 1. Intersection relationship between users/items with hot/cold start in three different datasets.

Datasets	Douban-book				Epinions				Yelp2018			
Segment	IP-User	IC-User	P-Item	N-Item	IP-User	IC-User	P-Item	N-Item	IP-User	IC-User	P-Item	N-Item
SP-User	**0.2338**	0.0645	0.6637	0.0551	**0.3940**	0.0163	0.6625	0.0350	**0.3369**	0.0476	0.6685	0.0512
SC-User	0.0431	0.2184	0.7627	0.0391	0.0177	0.2758	0.7228	0.0299	0.0372	0.2342	0.6463	0.0592
IP-User	-	-	0.6768	0.0533	-	-	0.6808	0.0325	-	-	0.6543	0.0535
IC-User	-	-	0.7714	0.0403	-	-	0.6906	0.0353	-	-	0.6616	0.0580

Despite significant advancements in previous studies, two critical issues in SocialRS remain unresolved. First, existing social relationship modeling approaches often fail to effectively filter out irrelevant or noisy social connections. In general, Social connections between users are driven by a variety of factors, many of which are unrelated to shared item preferences. To illustrate this issue, we analyze three datasets: Douban-book, Epinions, and Yelp2018. We categorize users into four types based on their interaction and social degrees within the top 20%: Social Popular Users (SP-User), Social Cold Users (SC-User), Interaction Popular Users (IP-User), and Interaction Cold Users (IC-User). Similarly, items are categorized as either Popular Items (P-Item) or Niche Items (N-Item). As indicated in Table 1, more than 60% of IP Users and SP Users do not overlap across the three datasets. In other words, most of the social connections are likely to be irrelevant with the common item interests, which highlighting the complexity and diversity of social connections. For example, in the Douban-book dataset, users may establish social connections based on shared interests in a particular book, movie, or music. However, such connections may not extend to other domains, such as shopping habits or movie preferences, where substantial differences may exist. Although some denoise works have made some efforts, they simply use the similarity measure to characterize the latent interests of users. As we have explained, when similarity measures are employed to group users based on these social connections, the resulting "similarity" may not be derived from common preferences relevant to recommendation, thus offering limited value to the recommendation system. As a result, the aforementioned methods are inadequate in processing complex social networks and miss valuable signals that could enhance recommendation tasks.

Another issue is the inherent popularity bias in the social networks. Although existing approaches have made some progresses to address the sparse connections in social networks, they neglect the underlying popularity bias in the social networks and exhibit notable limitations on the improvements for recommendation. Specifically, as we have illustrated in the Table 1, those social users (either popular or cold users) interacts with at least 64% popular items and only at most 6% niche items (indicated by the underlines). In other words, users in the social networks have the similar popularity bias towards items with those ones in the interaction graphs. In this way, directly incorporating such social information into the recommendation tasks will inevitably reinforce the recommendation system's bias towards popular users, hindering the discovery of personalized social and interaction preferences. Although existing works have made some progresses on the social network sparsity issue, they mainly focus on increasing the connections between users and neglecting the inherent bias information. Hence, their methods not only fails to effectively increase recommendations for niche items (N-Item) but also exacerbates the system's bias towards popular items (P-Item).

To address these challenges, we propose a **Ca**usal **D**isentanglement Enhanced **Di**ffusion Denoising for **S**ocial **R**ecommendation (**CaDDiSR**) framework. To tackle the first challenge (inadequate denoising techniques), we novelly design a causal disentanglement-enhanced diffusion process to efficiently distill the beneficial social relations and remove the redundant or noise information for recommendations. Concretely, in this process, we adopt the multi-layer perceptron (MLP) to generate the initial representations and employ an adaptive causal graph to capture the causal relationships and further generate users' causal social representations. Note that our learned representations could not only uncover the preference-driven user connections, but also encode the higher-order neighbor influences, which breaks through the limitations of simple similarity measures. Those generated causal representations will be used to guide the diffusion process to effectively remove noise from social relationships while retaining useful information for recommendations. For the second challenge (inherent popularity bias in social networks), we propose an multi-task fusion mechanism to significantly incorporate the social information into recommendation tasks. In detail, we integrate the bidirectional distillation method and cross-domain contrastive learning to mutually transfer information between two scenarios and address the data-sparsity issue. In this way, the refined social networks will be further reshaped by the augmented interaction information, which effectively we utilize alleviate the popularity bias and improve the recommendation performance.

In brief, the salient contributions of our work include:

- we propose a causal disentanglement enhanced diffusion denoising framework for social recommendation. Our work could efficiently distill the beneficial information from social networks and alleviate the inherent popularity bias issue in the social networks, significantly improving the performance of recommendation systems.
- We design a causal graph to adaptively disentangle the complex factors that influences the social connections and extract the beneficial preference-driven relationships. The generated causal representations are further

used to guide the diffusion process to remove the irrelevant or noise information, obtaining a more refined social network.

- To tackle the problem of popularity bias in social networks, we propose a multi-task fusion mechanism. Through bidirectional knowledge distillation and cross-domain contrastive learning, information is effectively transferred between social and interaction scenarios, mitigating the recommendation system's bias toward popular items and enhancing personalized recommendation performance.
- Extensive experiments on multiple real-world datasets demonstrate that the CaDDiSR framework achieves significant improvements over existing baseline models. Comparative experiments, ablation studies, and robustness tests further validate the effectiveness and superiority of the various components of our model.

2 PROBLEM FORMULATION

In this section, we introduce the notation and definitions used in SR tasks. We assume that it is composed of interaction history and social history.

Interaction History. Let $\mathcal{U} = \{u_1, u_2, \ldots, u_m\}$ denote the set of m users that share both interaction and social records, and let $\mathcal{I} = \{i_1, i_2, \ldots, i_n\}$ denote the set of n items. The interaction history can be represented as a user-item interaction matrix $\mathcal{R} \in \mathbb{R}^{m \times n}$, where the element $r_{u,i} = 1$ if user u has interacted with item i, and $r_{u,i} = 0$ otherwise.

Social History. Let $\mathcal{S} \in \mathbb{R}^{m \times m}$ denote the social relationship matrix, The element $s_{i,j}$ in the matrix denotes the existence of a social connection between users u_i and u_j from the set of users $\mathcal{U}$. Specifically, $s_{i,j} = 1$ implies that there is a social relationship between the two users, and $s_{i,j} = 0$ otherwise.

Problem Statement. The goal of social recommendation is to learn a function $\mathcal{F}(u, i \mid \mathcal{R}, \mathcal{S}, \Theta)$ that predicts the set of items that a user $u \in \mathcal{U}$ would like to interact with. This function leverages both the interaction history $\mathcal{R}$ and the social history $\mathcal{S}$, where Θ represents the model parameters.

3 METHODOLOGY

The CaDDiSR structure proposed in this section is shown in Figure 1. The overall model can be divided into three components: i) Representations Generation In Different Scenarios. It introduces interaction encoders and social encoders to generate representations of users and items within interaction and social scenarios, respectively. ii) Causal Disentanglement-Enhanced Diffusion Process. It utilizes a causal graph to extract users' additional causal-related social representations, which are used to enhance the initial input of the Diffusion model, guiding the denoising of user social representations. iii) Multi-task Contrastive Fusion. It introduces cross-domain contrastive learning, which aligns the user representations from the interaction and social views, while mitigating the Matthew effect on both interaction and social cold-start users.

3.1 Representations Generation In Different Scenarios

In interaction scenarios, users typically possess explicit collaborative preference information that is directly relevant to the recommendation task, while most of their social relationships are redundant for recommendation purposes. If these social relationships are directly incorporated into the recommendation system, they may adversely affect its performance. To effectively distinguish and extract useful information, our approach divides the initial user representation into two channels, which are then encoded separately for the interaction bipartite graph and the social graph using graph neural networks. This allows us to effectively purify the social graph by removing redundant information. Specifically,

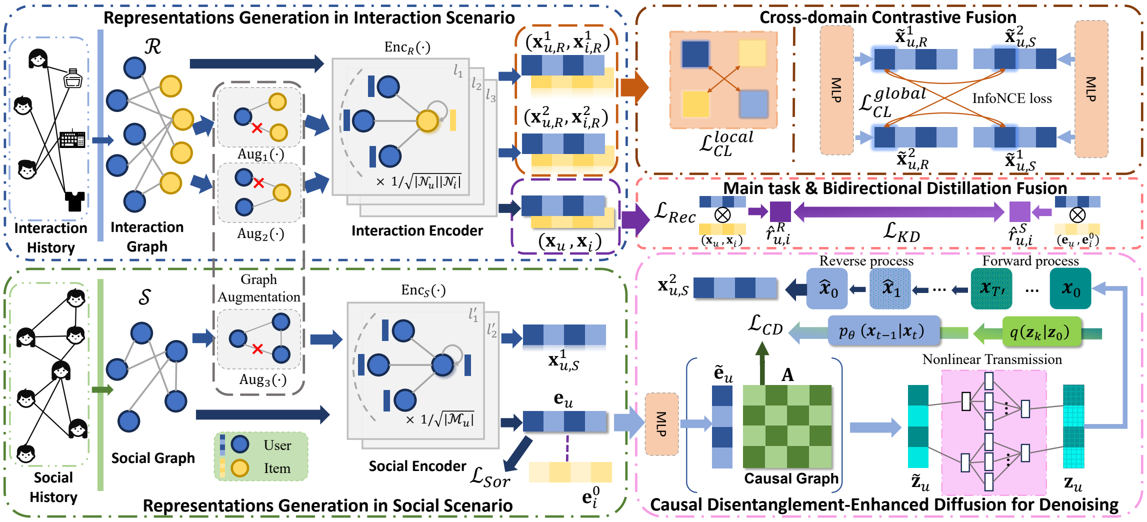


Fig. 1. The overall architecture of CaDDiSR. The representations of user in the interaction and social scenarios are generated separately, and then the representations in social scenario are denoised and augmented through the causal disentanglement-enhanced diffusion process. The user representations in the two scenarios are fused through bidirectional distillation and cross-domain contrastive learning to improve recommendation performance.

we designed separate encoders based on LightGCN [11] for the interaction bipartite graph and the social network, capturing high-order user representations in these two distinct scenarios.

3.1.1 Representations Generation of Interaction Scenario. For interaction scenarios, we assume that the initial input of users and items is $(\mathbf{e}_u^0, \mathbf{e}_i^0)$. The interaction encoder $\mathrm{Enc}_R(\cdot)$ is formulated as follows:

$$\mathbf{e}_u^{(l+1)} = \sum_{i \in \mathcal{N}_u} \frac{1}{\sqrt{|\mathcal{N}_u||\mathcal{N}_i|}} \mathbf{e}_i^{(l)}, \ \mathbf{e}_i^{(l+1)} = \sum_{u \in \mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_i||\mathcal{N}_u|}} \mathbf{e}_u^{(l)}, \tag{1}$$

where $\mathbf{e}_u^{(l)}$ and $\mathbf{e}_i^{(l)}$ is the encoded representations of user u and item i at the l-th graph propagation layer. $\mathcal{N}_u$ and $\mathcal{N}_i$ is the set of users that connected to item i and the items that user u has interacted with, respectively. After L-layer aggregation, we use average pooling to fuse user/item representations from each layer, resulting in $\mathbf{x}_u$ and $\mathbf{x}_i$.

3.1.2 Representations Generation of Social Scenario. For the sake of simplicity in the model, we refer to previous work [28, 31, 46] and design a social encoder $\mathrm{Enc}_S(\cdot)$ within social scenario as follows:

$$\mathbf{e}_u^{(l')} = \sum_{u \in \mathcal{M}_u} \frac{1}{|\mathcal{M}_u|} \mathbf{e}_u^{(l'-1)}, \tag{2}$$

where $\mathcal{M}_u$ represents the set of users who have social relationships with user u. $\mathbf{e}_u^{(l')}$ is the user embedding representation of l'-th graph propagation layer. We ultimately obtains the final output $\mathbf{e}_u$ by averaging pooling the outputs of each layer in the L' layer network.

3.2 Causal Disentanglement-Enhanced Diffusion Process

Considering the complexity of factors involved in social relationship generation and their low correlation with item preferences, we first explicitly incorporate causal graph into the causal modeling of users' social relationships. Subsequently, the causal disentanglement-enhanced user representation output is used as a guiding factor and applied to the diffusion model as a generation method. Then, the user representation containing specific structural information is reconstructed by forward and reverse diffusion process to further achieve the effect of high-quality data augmentation.

3.2.1 Causal Disentanglement for Social Networks. Our causal model is integrated into the Diffusion framework. Inspired by the relevant causal disentanglement works [42], we assume c potential causal factor and design a specific causal graph to generate a linear representation. Specifically, it is represented as follows:

$$\begin{aligned} \tilde{\mathbf{z}} &= \mathbf{A}^T\mathbf{z} + \tilde{\mathbf{e}}_\mathbf{u} = (\mathbf{I} - \mathbf{A}^T)^{-1}\tilde{\mathbf{e}}_\mathbf{u}, \\ \tilde{\mathbf{e}}_\mathbf{u} &= \mathrm{MLP}(\mathbf{e}_u; \theta_C), \end{aligned} \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{c\times c}$ is a learnable matrix in the form of an upper triangle, the user social representation $\tilde{e}_u$ is treated as an exogenous factor after passing through the MLP layer with learnable parameters θ_C.

To accommodate the directionality of causal effects, we design a K-layer smooth nonlinear transmission module. This module ensures that the causal representation components at each layer exert a unidirectional influence on one another, allowing information to propagate exclusively from the upper layer to the lower layer with little reverse influence. The corresponding formula is presented as follows:

$$\mathbf{z} = \|_{k=1}^{K} g_k(\tilde{\mathbf{z}}_k; \eta_k). \tag{4}$$

where g_k is the k-th layer non-linear function in the module, and η_k is its corresponding learnable parameters. Finally, we concatenate these unidirectionally propagated components to form the ultimate user causal representation $\mathbf{z}$.

3.2.2 Causal Guided Diffusion Generation. The user representations encoded from the original social network contain excessive redundant social information that are unrelated to the interaction scenario. Inspired by recent work on diffusion models [32, 45], we argue that diffusion models can assist in generating user social representations with less noise and more collaborative preference information. We merge $\mathbf{z}$ with the current timestamp and set it as the initial state $\boldsymbol{x}_0$.

Analogous to general diffusion processes [12], our forward noise injection procedure can be directly formulated as follows, propagating from the initial representation $\boldsymbol{x}_0$ to the t-th step representation $\boldsymbol{x}_t$:

$$\begin{aligned} q(\boldsymbol{x}_t|\boldsymbol{x}_0) &= \mathcal{N}(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_{t-1}, (1-\bar{\alpha}_t)\boldsymbol{I}); \\ \boldsymbol{x}_t &= \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}), \end{aligned} \tag{5}$$

where β_t governs the degree of Gaussian noise at the t-th step, we set $\alpha_t = 1 - \beta_t$. Furthermore, $\bar{\alpha}_t = \prod_{t'=1}^{t} \alpha_{t'}$ denotes the cumulative scaling factor. The proportion of noise introduced exhibits a linear dependence on the step index t, such that $1 - \bar{\alpha}_t \propto t$.

The reverse process is controlled by a neural network, which commences from $\boldsymbol{x}_t$ and removes the noise through each step t to restore the user's social representation.

$$p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t), \Sigma_\theta(\boldsymbol{x}_t, t)), \tag{6}$$

where the terms $\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t)$ and $\Sigma_\theta(\boldsymbol{x}_t, t)$ represent Gaussian parameters generated by neural networks.

3.2.3 Optimization. Diffusion models seek to optimize the parameter θ in the neural network by primarily aiming to maximize the ELBO linked to the initial state $\boldsymbol{x}_0$. Utilizing the KL divergence, the main component of the process approximates the fitted distribution $p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ to the manageable distribution $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$. Aligning with the approaches [12, 14], we optimize the second term at step t using $\mathcal{L}_t$ as an optimization objective:

$$\begin{aligned}\mathcal{L}_t &= \mathbb{E}_{q(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[D_{KL}(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)) \parallel p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t))\right] \\ &= \mathbb{E}_{q(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[\frac{1}{2}\left(\frac{\bar{\alpha}_{t-1}}{1-\bar{\alpha}_{t-1}} - \frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}\right)\|\hat{\boldsymbol{x}}_\theta(\boldsymbol{x}_t, t) - \boldsymbol{x}_0\|_2^2\right] + C,\end{aligned} \tag{7}$$

where $\hat{\boldsymbol{x}}_\theta(\boldsymbol{x}_t, t)$ is computed by $\boldsymbol{x}_t$ and the time embedding of step t into a MLP, which outputs an estimate of $\boldsymbol{x}_0$. The constant C depends only on the first $\boldsymbol{x}_1$ and the initial states $\boldsymbol{x}_0$.

In practical implementation, we uniformly sample step $t \sim \mathcal{U}(1, T)$ to optimize $\mathcal{L}(\boldsymbol{x}_0, \theta)$, where we utilize $\sum_{t=1}^{T} \mathcal{L}_t$ to optimize the ELBO, formalized as follows:

$$\mathcal{L}(\boldsymbol{x}_0, \theta) = \mathbb{E}_{t\sim\mathcal{U}(1,t)}\mathcal{L}_t. \tag{8}$$

Due to the acyclicity inherent in causal representations, we adopt a continuously differentiable constraint function to maintain $\boldsymbol{A}$ as a Directed Acyclic Graph (DAG). We then perform a joint optimization of the ELBO loss associated with the initial state and the overall denoising process loss $\mathcal{L}_{CD}$, which can be formalized as follows:

$$\mathcal{L}_{CD} = \mathcal{L}(\boldsymbol{x}_0, \theta) + \mathrm{tr}((\boldsymbol{I} + \omega\mathbf{A} \circ \mathbf{A})^c) - c \tag{9}$$

where ω is an empirical constant, and $\circ$ is the element-wise multiplication.

Finally, in our denoising strategy, we first gradually corrupt the representations in the forward process, obtaining $\boldsymbol{x}_{T'}$. Then, we set $\hat{\boldsymbol{x}}_T = \boldsymbol{x}_{T'}$ and perform the reverse denoising process using $\hat{\boldsymbol{x}}_{t-1} = \mu_\theta(\boldsymbol{x}_t, t)$. Subsequently, we leverage $\hat{\boldsymbol{x}}_0$ as the final user social representation $\hat{\mathbf{e}}_u$.

3.3 Multi-task Fusion

In the preceding section, we extract useful information from the social scenario. To further integrate this information effectively into the interaction scenario and alleviate the popularity bias, we design a multi task fusion mechanism to align and sift user information in both interaction and social scenarios. Unlike directly merging heterogeneous user representations, we adopt a bidirectional distillation method to balance the user-item prediction rankings between social and interactive scenarios, while utilizing the cross-domain contrastive learning to tackle the data sparsity in the two scenarios.

3.3.1 Main task optimization. To maintain simplicity, we compute the inner product between the user and item representations $(\mathbf{x}_u, \mathbf{x}_i)$ in the interaction scenario, and between the user social representations $\mathbf{e}_u$ and the initialized item representations $\mathbf{e}_i^0$ in the social scenario, to obtain the predicted rankings, which can be formalized as follows:

$$\hat{r}_{u,i}^R = \mathbf{x}_u^\top \cdot \mathbf{x}_i, \quad \hat{r}_{u,i}^S = \mathbf{e}_u^\top \cdot \mathbf{e}_i^0. \tag{10}$$

Subsequently, we employ the BPR loss as the objective function for both our primary recommendation task and our social relationship optimization process, which are defined as follows:

$$\mathcal{L}_{Rec} = \sum_{u\in U}\sum_{i\in\mathcal{N}_u}\sum_{i'\notin\mathcal{N}_u} -\log\ \sigma(\hat{r}_{u,i}^R - \hat{r}_{u,i'}^R), \tag{11}$$

$$\mathcal{L}_{Sor} = \sum_{u \in U} \sum_{v \in \mathcal{M}_u} \sum_{v' \notin \mathcal{M}_u} -\log\ \sigma(\hat{r}^S_{u,v} - \hat{r}^S_{u,v'}), \tag{12}$$

where $\sigma(\cdot)$ indicates the sigmoid activation function. For the recommendation task, $i' \notin \mathcal{N}_u$ refers to any item that user u has not engaged with, obtained via random sampling; $\hat{r}^R_{u,i}$ denotes the predicted score for item $i \in \mathcal{N}_u$, and $\hat{r}^R_{u,i'}$ corresponds to the model's predicted rating for any item that user u has not interacted with. For the social relationship optimization, $v' \notin \mathcal{M}_u$ refers to any user that user u has no relationship with, obtained via random sampling; $\hat{r}^S_{u,v} = \mathbf{e}_u \cdot \mathbf{e}_v$ denotes the predicted score for user $v \in \mathcal{M}_u$; and $\hat{r}^S_{u,v'}$ corresponds to the model's predicted rating for any user that user u has no relationship with.

3.3.2 Bidirectional Distillation Fusion. We define our KD loss using binary cross-entropy concept, which serves to extract knowledge from each other's scenarios. It is formalized as follows:

$$\begin{aligned}
\mathcal{L}^{S\to R}_{KD} &= -\sum_{(u,i)} \sigma(\hat{r}^R_{u,i}) \ln \sigma(\hat{r}^S_{u,i}) - (1-\hat{r}^R_{u,i}) \ln(1-\sigma(\hat{r}^S_{u,i})); \\
\mathcal{L}^{R\to S}_{KD} &= -\sum_{(u,i)} \sigma(\hat{r}^S_{u,i}) \ln \sigma(\hat{r}^R_{u,i}) - (1-\hat{r}^S_{u,i}) \ln(1-\sigma(\hat{r}^R_{u,i})); \\
\mathcal{L}_{KD} &= \gamma_1 \mathcal{L}^{S\to R}_{KD} + \gamma_2 \mathcal{L}^{R\to S}_{KD},
\end{aligned} \tag{13}$$

where taking $R \longrightarrow S$ as an example, R represents the teacher model in the KD framework, while S represents the student model. When updating the gradient of student model, the teacher's gradient will be frozen. γ_1 and γ_2 donates two hyperparameters to adjust the regularization weights of bidirectional distillation.

3.3.3 Cross-domain Contrastive Fusion. The role of bidirectional knowledge distillation is mutually transfer information between two scenarios, which could align users representations in user-item relationships and user-user relationships to make them more consistent. To further address the data sparsity issue, we conduct contrastive learning on both the two scenarios and Moreover, we also utilize the cross-domain contrastive fusion to further distinguish the positive samples with those hard-negative samples. This could endow the recommendation system to accurately discern those truly beneficial relational patterns, thereby improving the accuracy of the recommendations.

we argue that if during this process, hard negative samples—those that are difficult to be distinguished with the positive samples in the recommendation task—are also brought closer, the recommendation system may become confused. This could hinder its ability to accurately discern truly useful relational patterns, thereby affecting the accuracy of the recommendations.

Inspired by the operations in recent works [36], we adopt forms of graph data augmentation, and then leverage the encoder formulation shown in Eq. 1 to generate the augmented user/item representations as follows:

$$\begin{aligned}
\mathbf{x}^1_{u,R}, \mathbf{x}^1_{i,R} &= \mathrm{Enc}_R\left((\mathbf{e}^0_u, \mathbf{e}^0_i); \mathrm{Aug}_1(\mathcal{R})\right); \\
\mathbf{x}^2_{u,R}, \mathbf{x}^2_{i,R} &= \mathrm{Enc}_R\left((\mathbf{e}^0_u, \mathbf{e}^0_i); \mathrm{Aug}_2(\mathcal{R})\right); \\
\mathbf{x}^1_{u,S} &= \mathrm{Enc}_S\left((\mathbf{e}^0_u); \mathrm{Aug}_3(\mathcal{S})\right),
\end{aligned} \tag{14}$$

where Aug_1, Aug_2, and Aug_3 denotes any data augmentation operator, i.e., *Node Drop*, *Edge Drop*, or *Random Walk*. We also augment the original user social representations to obtain $\mathbf{x}^1_{u,S}$, which serves as a companion reference group while treating the diffusion-processed representations $\hat{\mathbf{e}}_u$ as augmented representations $\mathbf{x}^2_{u,S}$. This provides an additional supervisory signal to guide the subsequent contrastive learning fusion process.

Subsequently, we utilize a disentangled cross-domain InfoNCE loss to maximize the mutual information. This loss function comprises a local term focusing on the interaction scenario representations, and a global term optimize between the interaction and social scenario representations. The local term is defined as follows:

$$\begin{aligned}\mathcal{L}_{CL}^{lcoal} = &\sum_{u\in\mathcal{U}} -\log \frac{\exp(\mathrm{sim}(\mathbf{x}_u^1,\ \mathbf{x}_u^2)/\tau)}{\sum_{u'\in\mathcal{U},u'\neq u}\exp(\mathrm{sim}(\mathbf{x}_u^1,\ \mathbf{x}_{u'}^2)/\tau)} \\ &+ \sum_{i\in\mathcal{N}_u} -\log \frac{\exp(\mathrm{sim}(\mathbf{x}_i^1,\ \mathbf{x}_i^2)/\tau)}{\sum_{i'\in\mathcal{I},i'\neq i}\exp(\mathrm{sim}(\mathbf{x}_i^1, \mathbf{x}_{i'}^2)/\tau)}\end{aligned} \tag{15}$$

where u' signifies a negative user sample, indicating any user in the training set excluding the target user u; i' denotes a negative item sample, corresponding to any item the user u has not interacted with in the training data; τ represents the temperature hyperparameter; and $\mathrm{sim}(\cdot)$ is the cosine function employed to quantify the similarity between the sampled pairs.

The global term aligns user representations across the scenarios in a shared mapping space, separating representations of different users. Specifically, we employ dedicated MLPs as the mapping functions for each set of representations, defined as follows:

$$\tilde{\mathbf{x}}^s = \mathrm{MLP}(\mathbf{x}^s; \theta_F^s) \tag{16}$$

where θ_F^s represents the trainable parameter of the MLP corresponding to the s-th representation $\mathbf{x}^s$, and $s \in \{1, 2\}$.

Then, we calculate the global term in the form of cross-domain InfoNCE, which is defined as follows:

$$\mathcal{L}_{CL}^{global} = \sum_{(u,v)\in\mathcal{S}} -\log \frac{\exp(\mathrm{sim}(\tilde{\mathbf{x}}_{u,R}^j,\ \tilde{\mathbf{x}}_{v,S}^k)/\tau)}{\sum_{v'\in\mathcal{U},v'\neq v}\exp(\mathrm{sim}(\tilde{\mathbf{x}}_{u,R}^j,\ \tilde{\mathbf{x}}_{v',S}^k)/\tau)}, \tag{17}$$

where v' denotes a user who has not established a relationship with the target user u in the social history $\mathcal{S}$. $j \in \{1, 2\}$ and $k \in \{1, 2\}$ represent the indices of the corresponding augmented representations for the user across the two distinct scenarios.

Finally, we fuse the local and global contrastive losses to obtain our final cross-domain contrastive fusion loss:

$$\mathcal{L}_{CL} = \lambda_1 \mathcal{L}_{CL}^{local} + \lambda_2 \mathcal{L}_{CL}^{global} \tag{18}$$

where λ_1 and λ_2 donates two hyperparameters to adjust the regularization weights of two terms of the contrastive loss.

3.3.4 Joint Optimization. According to the above all loss functions, the comprehensive optimization loss of this model is:

$$\mathcal{L} = \mathcal{L}_{Rec} + \xi\mathcal{L}_{Sor} + \mathcal{L}_{KD} + \mathcal{L}_{CL} + \zeta||\Theta||_2^2, \tag{19}$$

where ξ is a hyperparameter for adjusting the weight of social relationship optimization loss. Θ denotes the parameter set of the model, while the L_2 regularization hyperparameter ζ can be manipulated to fine-tune the weight value, thereby averting overfitting.

3.4 Discussion

This paper identifies issues related to the low overlap between social and recommendation scenario and their common sparsity, resulting in sparse supervision signals and integration difficulties. As shown in Table 2, most previous methods have addressed only one of these problems. In contrast, our approach not only proposes strategies to tackle both issues but also considers the inherent latent factors in social relationships that are independent of collaborative information. By employing causal disentanglement, we more comprehensively capture users' interaction-related social preferences.

Table 2. Comparison of key modules used between baseline and CaDDiSR.

Methods	GNN	SSL	Denoising	Distanglement	Causal
DiffNet [38]	✓	✗	✗	✗	✗
SEPT [46]	✓	✓	✗	✗	✗
MHCN [48]	✓	✓	✗	✗	✗
DSL [31]	✓	✓	✓	✗	✗
DISGCN [17]	✓	✓	✗	✓	✗
DESIGN [28]	✓	✗	✗	✗	✗
DcRec [35]	✓	✓	✗	✓	✗
GDMSR [27]	✓	✗	✓	✗	✗
CaDDiSR	✓	✓	✓	✓	✓

4 EXPERIMENTS

In this section, we conduct extensive experiments to thoroughly evaluate the performance of our proposed CaDDiSR model. Specifically, we aim to address the following research questions:

- **RQ1:** How does the performance of CaDDiSR compare to that of different types of recommendation methods?
- **RQ2:** What is the contribution of the various key modules within the CaDDiSR framework to the overall performance?
- **RQ3:** How do different hyperparameter settings affect the performance of CaDDiSR?
- **RQ4:** How does the robustness of CaDDiSR to data with different level of sparsity?
- **RQ5:** How does each module optimize user representation in social scenarios from a more intuitive perspective?

4.1 Experiment Settings

4.1.1 Datasets. Our experiments are conducted on three public datasets gathered from diverse real-world platforms: Douban-book for book recommendations, Epinoins for product review recommendation, and Yelp2018 for commercial venue recommendations. As shown in Table 3, these experimental datasets possess varying interaction density and social network characteris.

Table 3. Analysis of the statistical properties of the dataset employed in the experiment.

Datasets	Douban-book	Epinions	Yelp2018
# User	13, 025	18, 203	45, 919
# Item	22, 348	47, 450	45, 538
# Interaction	598, 420	338, 400	1, 183, 610
# Relation	169, 150	595, 049	709, 459
U-I Density	2.06×10^{-3}	3.20×10^{-4}	5.66×10^{-4}
U-U Density	1.04×10^{-3}	1.80×10^{-3}	8.01×10^{-4}

4.1.2 Baselines. We select 9 recommender models as the baselines for comparative experiments, in order to explore the performance improvement of our model relative to other methods.

- **DiffNet** [38]: The proposal is a deep influence propagation model to simulate how users are influenced by recursive social diffusion, resulting in better performance.
- **NGCF** [33]: It exploits the user-item graph structure and employs conventional GCN to explicitly inject collaborative signals into the embedding process for learning user and item representations.
- **LightGCN** [11]: It improves NGCF by simplifying the GCN of collaborative filtering to only retain symmetric sqrt normalization, ensuring the efficiency and simplicity of collaborative filtering.
- **SEPT** [46]: It leverages tri-training to enhance SR task by augmenting user representations with social information and iteratively improves multiple encoders using self-supervision signals.
- **MHCN** [48]: It proposes a multi-channel hypergraph network with self-supervised learning for SR, leveraging high-order social relations to model complex user relations.
- **DSL** [31]: It introduces a self-augmented learning paradigm that retains valuable social relations and enables personalized cross-view knowledge sharing.
- **DESIGN** [28]: It introduces an integrated approach that can more effectively encode both the U-I and U-U graphs, while incorporating knowledge distillation between auxiliary models.
- **DcRec** [35]: It learns disentangled user representations from both interaction and social domains, and employs contrastive learning to facilitate knowledge transfer between the learned representations, enhancing social recommendations.
- **LightGCL** [3]: It employs singular value decomposition for contrastive augmentation of the U-I graph, preserving semantic structures and improving robustness.

Table 4. Overall performance comparison on three datasets. The best performance is **bolded**, and the second-best performance is underlined. Improv. indicates the percentage improvement of our CaDDiSR method compared to other baselines.

Dataset	Metric	NGCF	LightGCN	DiffNet	DESIGN	SEPT	MHCN	DSL	DcRec	LightGCL	**CaDDiSR**	Improv.
Douban-book	R@10	0.0882	0.0989	0.0973	0.1046	0.1030	0.1043	0.0960	0.1140	0.1089	**0.1223**	7.32%
	N@10	0.1012	0.1163	0.1123	0.1270	0.1221	0.1237	0.1111	0.1329	0.1354	**0.1452**	7.25%
	R@20	0.1347	0.1499	0.1312	0.1532	0.1547	0.1513	0.1413	0.1610	0.1568	**0.1746**	8.44%
	N@20	0.1106	0.1274	0.1046	0.1315	0.1314	0.1326	0.1168	0.1392	0.1442	**0.1525**	5.77%
Epinions	R@10	0.0381	0.0452	0.0440	0.0455	0.0443	0.0467	0.0462	0.0482	0.0489	**0.0541**	10.65%
	N@10	0.0270	0.0319	0.0302	0.0331	0.0326	0.0337	0.0334	0.0349	0.0353	**0.0392**	10.82%
	R@20	0.0608	0.0702	0.0682	0.0711	0.0701	0.0733	0.0679	0.0755	0.0750	**0.0833**	10.37%
	N@20	0.0337	0.0393	0.0389	0.0401	0.0395	0.0414	0.0378	0.0426	0.0434	**0.0473**	9.09%
Yelp2018	R@10	0.0345	0.0412	0.0369	0.0413	0.0409	0.0370	0.0385	0.0472	0.0476	**0.0492**	3.41%
	N@10	0.0289	0.0345	0.0302	0.0345	0.0340	0.0311	0.0326	0.0400	0.0403	**0.0414**	2.73%
	R@20	0.0584	0.0673	0.0609	0.0720	0.0696	0.0658	0.0632	0.0776	0.0782	**0.0809**	3.37%
	N@20	0.0375	0.0436	0.0383	0.0461	0.0448	0.0426	0.0413	0.0507	0.0513	**0.0528**	2.96%

4.1.3 Evaluation Metrics. The user-item interaction datasets are randomly partitioned, with 80% allocated for training and the remaining 20% reserved for testing. The evaluation metrics employed are the widely used Recall@K and NDCG@K (abbreviated to R@K and N@K below), which are standard metrics for top-K recommendation tasks, with K set to 10 and 20.

4.1.4 Parameter settings. Our CaDDiSR is implemented through PyTorch and the experiment is built on NVIDIA GeForce RTX 3090. With the aim of providing a fair assessment, we align the experimental settings with those used by the baseline methods. Specifically, the dimensionality of the user/item representations is set to 64, the learning rate is fixed at $1e^{-3}$, and the batch size is configured as 2048.

In our experiment, the fixed parameters are set as follows: The layer L of the interaction encoder shown in Eq. 1 is set to 3, while the layer L' of the social encoder shown in Eq. 2 is set to 2. The empirical constant ω in Eq. 9 is set to $5e^{-3}$. In order to unify with relevant methods [14, 32], the total step T' of the forward process in our diffusion is set to 0, and the total step T of the reverse process is set to 5. Temperature parameter τ used in both local and global terms for contrastive learning is set to 0.2.

4.2 Comparison of Performance (RQ1)

In this section, we perform a comprehensive evaluation of the performance of CaDDiSR and the baseline methods mentioned previously. The results of the experiments conducted on the three datasets are presented in Table 4, from which the following insights can be derived:

- Our CaDDiSR consistently outperform all baseline methods across the three evaluation metrics, validating the effectiveness of our causal disentanglement enhanced diffusion in mining potential factors of user social relationships; It also verifies the indivisibility of user representation fusion and differentiation in dual scenarios through the combination of bidirectional knowledge distillation and cross-domain contrastive learning. Despite the diversity of datasets and evaluation scenarios, the consistently superior results highlight the broad applicability and multifunctionality of our approach.
- Compared to the traditional bipartite graph-based methods, the SN-based approaches, such as Diffnet and DESIGN, are able to capture additional social information. However, CaDDiSR is able to outperform these conventional SN-based methods because it can effectively encode and extract the causal social information, which provides further performance improvements.
- Compared to the SSL-based SR methods, such as SEPT, MHCN, DSL, and DcRec, the traditional SR approaches are more susceptible to biases and disturbances due to the lack of an augmented perception process for sparse information. Our CaDDiSR model employs a dual combination of knowledge distillation and cross-domain contrastive learning, which not only effectively extracts the supervisory signals in different scenarios but also mitigates the impact of data sparsity. Notably, in contrast to DcRec, which emphasizes the disentanglement of user's social and interaction character, our disentanglement focuses on the causal relationships formed by social connections, which is a fundamental factor underlying user-item interactions.
- It is worth noting that LightGCL, without utilizing social information, is able to achieve state-of-the-art performance across many metrics by solely adopting a novel graph encoding representation and integrating it with self-supervised learning techniques. This suggests that the information contained in the bipartite graphs can be more deeply exploited, while social information as an auxiliary feature often faces the problem of low overlap with CF-based recommendations. Our approach has better adapted to the social and interaction scenarios in the self-supervised learning context, and therefore achieved greater performance improvements compared to LightGCL.

4.3 Ablation Experiment (RQ2)

To investigate the individual contribution of each module within the CaDDiSR method towards the overall model performance, the ablation study section primarily conducts comparative experiments on five variants of our proposed model:

- "w/o diff": a variant that replaces the causal-enhanced diffusion process with random augmentation. This modification results in a model that not only lacks a causal representation of social relationships, but also lacks an optimizable denoising mechanism.
- "w/o global": a variant that removes the global cross-domain contrastive module results in a model that can only learn the self-supervised signals within the interaction scenario. This modification leads to a lack of differentiation and alignment of user representations across two different scenarios.
- "w/o sor": a variant that removes the social relationship optimization module from the primary task optimization, retaining only the collaborative filtering BPR. This modification results in a model that lacks the capacity to perceive the rich and salient information regarding users' social preference patterns.
- "w/o kd": a variant that eliminate the bidirectional knowledge distillation module compromises the model's ability to effectively integrate ground truth from interaction and social domains. This leads the model to overlook items valuable for user interactions in social scenario and interactions that would enable more robust social signal learning.
- "w/o ssl": a variant that removes the self-supervised learning methodology. This architectural modification renders the model's learning process highly susceptible to the effects of sparsity and popularity biases. Furthermore, the model is also subject to the additional influence of social information biases.

Table 5. Ablation study of CaDDiSR across different datasets and metrics.

Dataset	Douban-book				Epinions				Yelp2018			
Metrics	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20
ours	**0.1223**	**0.1452**	0.1739	0.1524	**0.0541**	**0.0392**	**0.0833**	**0.0473**	**0.0492**	**0.0414**	**0.0809**	**0.0528**
w/o diff	0.1161	0.1405	0.1674	0.1480	0.0527	0.0381	0.0806	0.0464	0.0487	0.0410	0.0802	0.0522
w/o global	0.1213	0.1442	**0.1746**	**0.1525**	0.0517	0.0374	0.0789	0.0453	0.0484	0.0409	0.0802	0.0522
w/o sor	0.1210	0.1430	0.1725	0.1512	0.0504	0.0364	0.0784	0.0447	0.0479	0.0402	0.0788	0.0515
w/o kd	0.1187	0.1400	0.1714	0.1490	0.0517	0.0375	0.0798	0.0459	0.0474	0.0402	0.0783	0.0513
w/o ssl	0.0703	0.0830	0.1114	0.0915	0.0396	0.0286	0.0629	0.0355	0.0348	0.0292	0.0576	0.0374

Table 5 presents a performance comparison of the original CaDDiSR and its five ablation variants across two evaluation metrics. The ablation study findings enable the following observations:

- From the perspective of whether to utilize self-supervised learning or not, the variant "w/o ssl" that removes the entire self-supervised module exhibits a significant performance degradation, while removing the global component of cross-domain contrastive learning, i.e., "w/o global" also leads to suboptimal model performance on the Epinions and Yelp2018 datasets. This suggests that simply incorporating limited social information without proper differentiation and augmentation can result in collaborative filtering being susceptible to the interference of user social biases. It is worth noting that the performance of "w/o global" on the Douban-book dataset is actually better, indicating that in the context of user social interaction based on platform diversified interests, comparing and distinguishing the user group with a single interest in books can on the contrary cause interference.

- The variant that excludes the causal-enhanced diffusion denoising process for social information, denoted as "w/o diff", exhibits more pronounced performance improvements on the Douban-book and Epinions datasets. We hypothesize that this is because the user social relationships in these two datasets are more complex, where connections formed solely based on shared item preferences have limited impact on preference learning. Conversely, the user relationship patterns in the Yelp2018 dataset are relatively simpler, leading to less significant performance changes similar to providing random augmentation.
- From the perspective of jointly encoding, optimizing, and integrating both of social and interaction scenarios, the three aspects are tightly coupled. Optimizing the interaction preference representation alone (the variant "w/o sor") leads to the user's social information being inadequately differentiated. Conversely, the lack of effective cross-scenario knowledge distillation (the variant "w/o kd") results in the feature learning for the two scenarios becoming disconnected or mutually interfering.

4.4 Parameter Analysis (RQ3)

4.4.1 Impact of potential causal factor. To investigate the impact of potential causal factor c on causal learning performance, we conduct parameter tuning experiments on the Douban-book, Epinions, and Yelp2018 datasets, setting the range to $\{10, 32, 64, 128, 256\}$. As observed from Figure 2, there is a significant variation in the optimal number of causal factor across different datasets. Generally, a higher number of causal factor indicates more complex implicit reasons for forming social relationships. For instance, in the Yelp2018 dataset, performance peaks at $c = 64$. However, for the Douban-book dataset, the optimal points for Recall and NDCG lie between 64 and 128. Interestingly, in the Epinions dataset, we discover that Recall values are relatively high when the number of causal factor is particularly low, such as 10, while NDCG remains relatively stable. This suggests that an excessive number of decoupled social factor can lead to overfitting, causing information confusion and negatively impacting recommendation performance.

4.4.2 Impact of social relationship optimization loss. To investigate the impact of social relationship optimization loss, we conduct experiments on the Douban-book, Epinions, and Yelp2018 datasets by adjusting the corresponding weight parameter ξ within the range $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1\}$. As shown in Figure 2, the optimal performance on the Douban-book dataset is achieved at $\xi = 0.01$, whereas for the Epinions and Yelp2018 datasets, the best results are observed at $\xi = 0.1$. When the weight parameter is insufficiently large, the encoding of users' social information becomes constrained, rendering it challenging for the model to explicitly capture users' latent social preferences in the absence of labeled data, thereby hindering the identification of meaningful social relationship patterns that could benefit the recommendation task. Conversely, when the weight parameter is excessively large, the representations of numerous niche users become analogous to those of prominent users within social scenario, consequently causing the model hard to disentangle personalized features across the user population.

4.4.3 Impact of parameters pair on bidirectional distillation. To investigate the optimal balance between the two scenarios in bidirectional knowledge distillation, we employ the automated hyperparameter optimization method [1], and conduct experiments on the Douban-book and Epinions datasets. The star and circles in Figure 3(a) represent the optimal parameter and its convergence region, respectively. It is observed that for Douban-book, the weights for distillation in both scenarios are relatively low, indicating a low overlap between social and interaction scenarios. This is consistent with the broad range of interests of Douban-book users (such as movies, music, or other entertainment projects, not just books). For Epinions, the proportion of knowledge distilled from the social scenario to the interaction

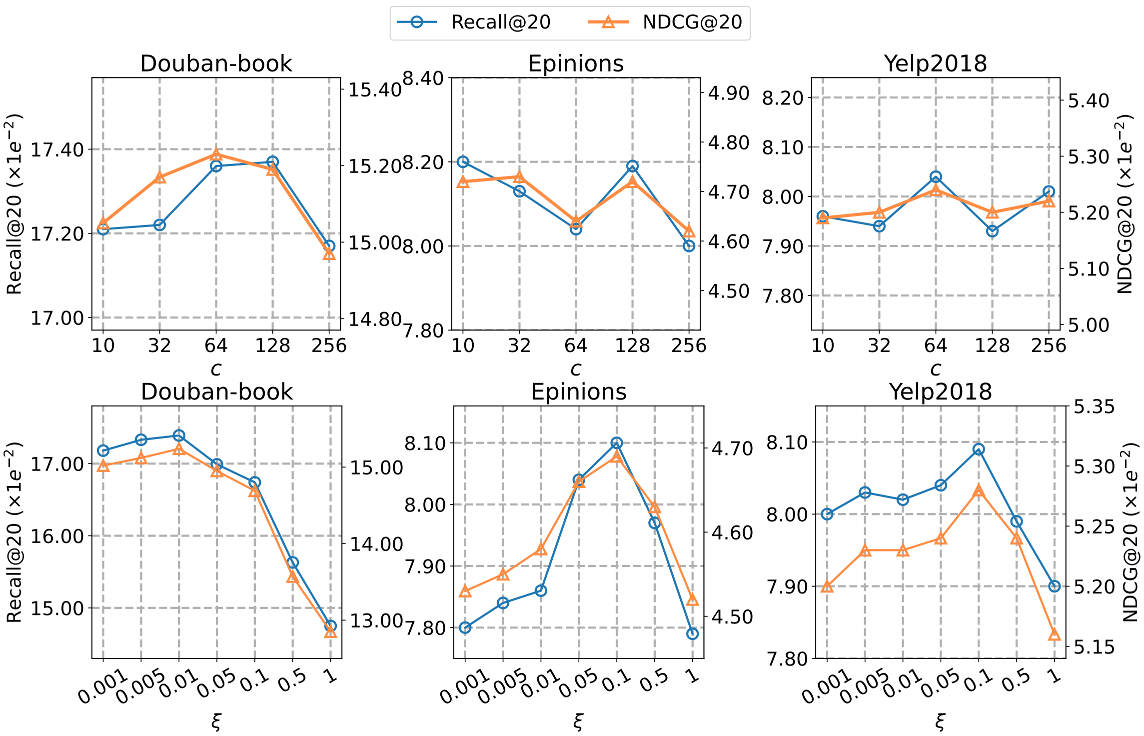


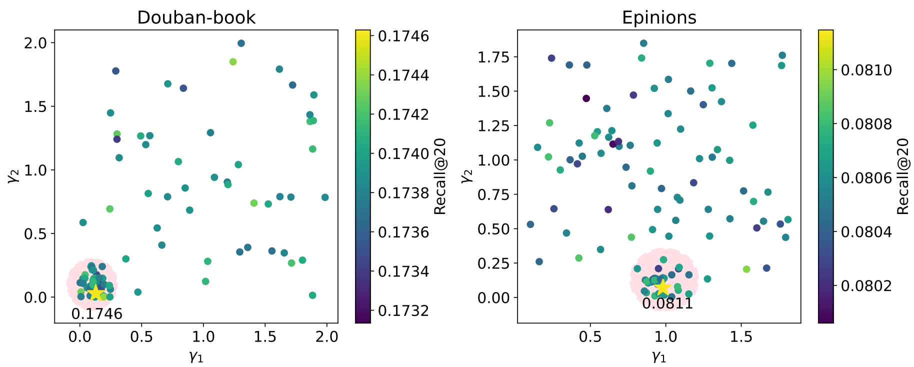
Fig. 2. Impact of potential causal factor c and weight of social relationship optimization loss ξ.

scenario is higher, indicating a strong correlation between users' social preferences and interaction item preferences on this dataset. Therefore, leveraging the social network can provide more collaborative information.

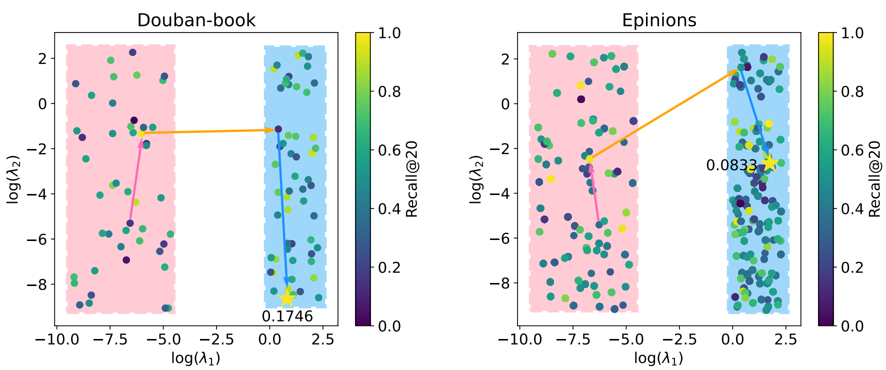
4.4.4 Impact of parameters pair on contrastive learning. To explore the impact of balancing local and global cross-domain contrastive learning on our overall performance, we also employ an automated hyperparameter optimization method on the Douban-book and Epinions datasets. Given the wide range of parameter values, we perform two rounds of local optimization, as illustrated by the differently colored rectangles in Figure 3(b). After optimizing these local tuning processes, we compare the optimal parameters, with the pentagon representing the best parameters. It can be observed that on both datasets, local contrastive learning holds a significant weight, indicating that the sparsity of the supervision signal in the interaction graph greatly affects recommendation performance. However, due to the low task overlap in the Douban-book dataset, the weight for global cross-domain part is minimal (less than $1e-8$). In contrast, global cross-domain contrastive learning exhibits a substantial impact on the Epinions dataset, where even when the weight for local part is low, a higher weight for global cross-domain part is beneficial for performance enhancement.

4.5 Model Robustness Study (RQ4)

To evaluate the robustness of CaDDiSR across user groups with varying social capabilities, we filter the original social history to generate three distinct social graph variants based on user social degrees: above 150, between 10 and 150, and



(a) Parameters pair on bidirectional distillation

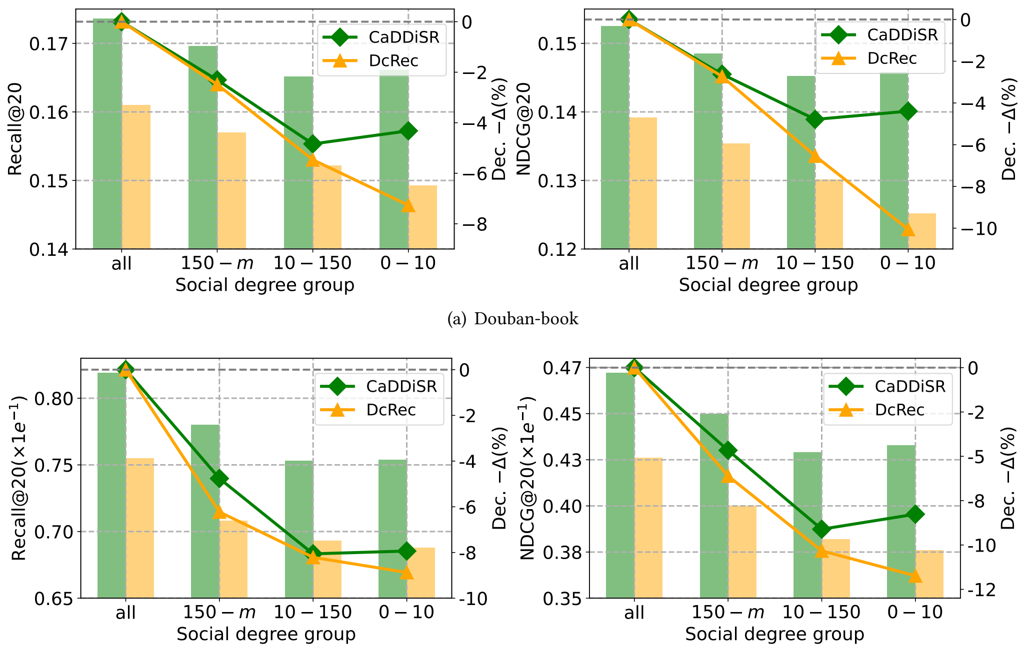


(b) Parameters pair on contrastive learning

Fig. 3. Impact of parameters pairs on bidirectional distillation and on contrastive learning over Douban-book and Epinions datasets.

below 10 (i.e., $150-m$, $10-150$, $0-10$). Subsequently, we assess the performance of our CaDDiSR model and another CL-based model, DcRec, on various variants of the Douban-book and Epinions datasets.

Figure 4 demonstrates that our CaDDiSR method consistently exhibits the smallest performance degradation across all groups. This observation can be attributed to the following reasons: (i) SR tasks rely on high-quality social graphs, and users with stronger social capabilities provide more useful and stable collaborative information. Consequently, as the degrees decrease, indicating an increase in the sparsity of the social graph, then the performance of all models deteriorates. However, due to the knowledge distillation fusion mechanism, our model is less affected by unstable information, resulting in slower degradation compared to DcRec. (ii) Among cold-start users in social scenarios, our model's performance actually improves. This is because cold-start users in social scenario have sparse social relationships and are more sensitive to denoising. Our causal disentanglement-enhanced diffusion denoising process effectively captures recommendation-specific relationship patterns, preserving the critical relationships for cold-start users. In contrast, DcRec's random augmentation of social relationships disrupts key relationship patterns and retains redundant noise unrelated to the primary task.



(a) Douban-book

(b) Epinions

Fig. 4. Performance on Douban-book and Epinions datasets with different user degree groups.

4.6 Visualization Analysis (RQ5)

To further investigate the role of each module in our model within social scenario, we select all user representations of yelp2018 dataset at initial state, post-social encoder state, and post-diffusion denoising state (named as *Pure*, *Social*, and *Diffusion*, respectively) at the 10th and 20th epochs. We visualize user representations at different stages and epochsusing t-SNE [30] and K-means clustering [23], as shown in Figure 5. It can be observed that both *Pure* and *Social* user representations exhibit a large, overly uniform distribution with a few isolated small clusters. This indicates that while self-supervised learning can mitigate the popularity effect, it still struggles to bring closer the representations of cold-start users who deviate significantly from the majority, resulting in an "island effect." Additionally, the clustering effect of *Social* is better than that of *Pure*, suggesting that encoding social information can uncover features such as naturally formed social circles, beneficial for mining social homogeneity information crucial for the main recommendation task. Due to the presence of causal disentanglement and denoising processes, the user representations in the *Diffusion* state are mostly coherent yet not overly uniform, exhibiting distinct clustering characteristics.

5 RELATED WORK

5.1 Social Recommendation

The proliferation of online social networks has highlighted the value of users' social relationships. As SR tasks delve into social homogeneity and influence, recommendation algorithms become more effective. The key challenge is to accurately describe interactions between acquaintances and integrate interpersonal communication with social attributes.

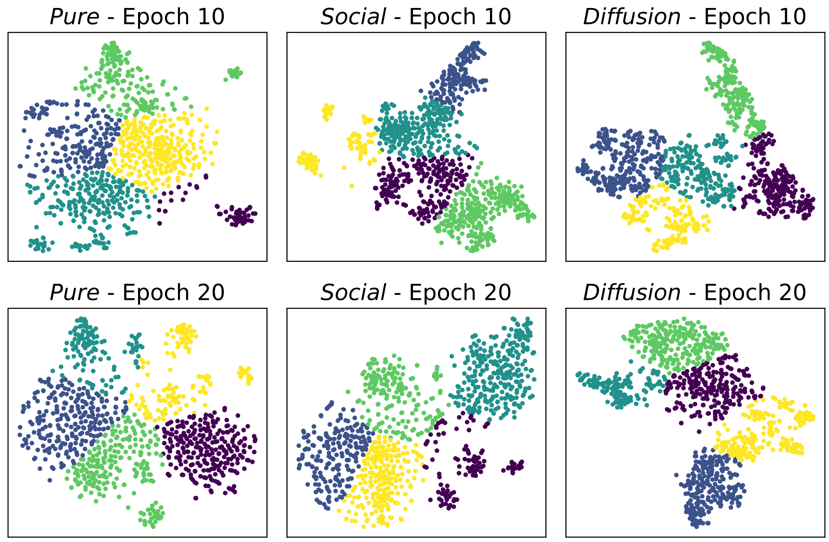


Fig. 5. The visualization results of tSNE for user representations of Yelp2018 dataset in three stages at different epoch. Respectively, *Pure*: user representations at the beginning stage; *Social*: user representations through social encoder; *Diffusion*: user representations after diffusion denoising.

Early research on social recommendation (SR) [9, 13, 21, 22] primarily involved combining collaborative filtering (CF) [16] techniques. These studies demonstrated that integrating social data with traditional CF could enhance recommendation accuracy [2], although challenges related to data sparsity and noise persisted. Recently, there has been rapid development in GNN-based recommender systems [8, 10, 11, 29, 34], propelling recommendation algorithms into the deep learning era. Considering that social networks (SNs) can be modeled as graph structures, the application of GNNs to encode SNs has become increasingly prevalent in the SR domain. Graph neural networks and structure-aware models enhance the modeling of complex relationships and resilience against noise in social recommendations [5, 6]. Existing studies [19, 38, 40, 47] have shown that GNN-based social network encoding can lead to substantial improvements in recommendation performance. GraphRec [7] captures interactions and opinions in the user-item graph, considering the heterogeneous strength of social relationships. SocialGCN [39] uses a GCN-based model to capture the propagation of user preferences in social networks via a hierarchical diffusion mechanism. DESIGN [28] enhances predictive performance by combining user-item interaction graphs and user-user social graphs using distillation-enhanced social graph networks and knowledge distillation techniques. However, these methods lack handling of data sparsity and popularity bias in recommendation tasks.

SSL-based social recommendation [35, 46–48] adopt augmented signals for capturing users' potential interests and social pattern, addressing data sparsity and bias. SEPT [46] combines social relations and self-supervised signals in a triple-training framework. PerFedRec [20] leverages personalized federated learning to address the sparsity of supervision signals in social networks, allowing for more tailored and effective social recommendations.MHCN [48] employs a multi-channel hypergraph convolutional network to maximize hierarchical mutual information but lacks dedicated denoising. DcRec [35] uses decoupled comparison learning for cross-domain knowledge transfer, also without explicit denoising. DISGCN [17] further refines this approach by separating user representations across different domains for more precise feature extraction.

Social denoising models debias for SR-based collaborative filtering by identifying and removing irrelevant social network information. GDMSR [27] uses a preference-guided denoising approach to retain information-rich social connections, with self-correcting learning modules and adaptive strategies. DSL [31] employs a dual-view graph neural network to capture latent relationships and filters unreliable user connections to improve preference modeling. GBSR [44] applies the information bottleneck principle to eliminate redundant social relationships by maximizing mutual information between the denoised social graph and recommendation labels, while minimizing it with the original social graph.

5.2 Contrastive Learning for Recommendation

Contrastive learning for recommendation is a self-supervised approach that maximizes mutual information via data augmentation and positive/negative sample optimization, compensating for sparse data to learn more robust user-item representations and improve recommender performance.

In the direction of collaborative filtering recommendation, several studies [4, 18, 36, 49] have leveraged user-item interaction graphs and self-supervised learning to enhance node representation learning. SGL [36] generates multiple views of a node and maximizes the consistency between them using a self-supervised task, effectively learning robust node representations for long-tail item recommendation in graph convolutional networks. NCL [18] constructs comparison pairs by exploiting both structural and semantic neighbor relationships in the graph. In contrast, LightGCL [4] generates comparison views through Singular Value Decomposition (SVD), which preserves the global collaborative signals of user-item interactions, overcoming the potential information loss and noise interference associated with the random augmentation strategies used in SGL.

Some SSL-based recommendation methods also utilize knowledge graphs (KGs) to supplement item side supervision signals by providing semantic and contextual information, enhancing understanding of the items and users. MCCLK [51] enhances knowledge-aware recommendation through multi-level cross-view comparison that captures global, local, and semantic graph features and structural information. KGCL [43] adopts augmented KGs to guide self-supervised signal generation of interaction graphs.

In the direction of sequence recommendation, S3-Rec [50] uses self-supervised pre-training with multi-level comparisons by capturing attribute, item, and segment-level features of user behavior. CL4Rec [41] exploits the self-supervised signals in user behavior sequences to infer accurate user representations and constructs different views of user sequences using three data augmentation methods. DuoRec [26] combat representation degradation through contrast regularization, Dropout-based enhancement, and using sequences with same target items as hard positives for contrastive learning.

6 CONCLUSION

In this work, we propose a Causal Disentanglement-Enhanced Diffusion Denoising framework (CaDDiSR) for social recommendation, addressing the limitations of existing methods in handling noise in social relationships and popularity bias. By employing causal disentanglement to extract meaningful social signals and integrating them into the diffusion denoising process, the method significantly enhances recommendation performance. Additionally, the combination of bidirectional knowledge distillation and cross-domain contrastive learning enables the model to effectively balance and transfer knowledge between social and recommendation scenarios, mitigating data sparsity issues and ensuring robust user representations.

Extensive experiments on multiple public datasets, including Douban-book, Epinions, and Yelp2018, demonstrate that CaDDiSR consistently outperforms existing baseline methods across various evaluation metrics, highlighting its effectiveness and broad applicability in different use cases. The key contributions of this work include enhancing the robustness of the model to noisy social data through the combination of causal disentanglement and diffusion denoising, mitigating the impact of data sparsity on recommendation accuracy via cross-domain contrastive learning, and validating the generalizability and stability of the framework through comprehensive evaluations across multiple datasets.

REFERENCES

[1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2623–2631.

[2] Alejandro Bellogin, Iván Cantador, Fernando Díez, Pablo Castells, and Enrique Chavarriaga. 2013. An empirical comparison of social, collaborative filtering, and hybrid recommenders. *ACM Transactions on Intelligent Systems and Technology (TIST)* 4, 1 (2013), 1–29.

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

[4] 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 (ICLR '23)*.

[5] Xu Chen. 2024. Robust Structure-Aware Graph-based Semi-Supervised Learning: Batch and Recursive Processing. *ACM Transactions on Intelligent Systems and Technology (TIST)* 15, 4 (2024), 66:1–66:25.

[6] Ehsan Elahi, Sajid Anwar, Babar Shah, Zahid Halim, Abrar Ullah, Imad Rida, and Muhammad Waqas. 2024. Knowledge Graph Enhanced Contextualized Attention-Based Network for Responsible User-Specific Recommendation. *ACM Transactions on Intelligent Systems and Technology (TIST)* 15, 4 (2024), 83:1–83:24.

[7] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. 2019. Graph Neural Networks for Social Recommendation. In *The World Wide Web Conference (WWW '19)*. 417–426.

[8] Chenyuan Feng, Zuozhu Liu, Shaowei Lin, and Tony Q.S. Quek. 2019. Attention-based Graph Convolutional Network for Recommendation System. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 7560–7564.

[9] Guibing Guo, Jie Zhang, and Neil Yorke-Smith. 2015. TrustSVD: Collaborative Filtering with Both the Explicit and Implicit Influence of User Trust and of Item Ratings. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '15, Vol. 29)*.

[10] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30.

[11] 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)*. 639–648.

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.

[13] Mohsen Jamali and Martin Ester. 2010. A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks. In *Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys '10)*. 135–142.

[14] Yangqin Jiang, Yuhao Yang, Lianghao Xia, and Chao Huang. 2024. DiffKG: Knowledge Graph Diffusion Model for Recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 313–321.

[15] Huan Liu Jiliang Tang, Xia Hu. 2013. Social recommendation: a review. *Social Network Analysis and Mining* 3, 1113–1133.

[16] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42, 8, 30–37.

[17] Nian Li, Chen Gao, Depeng Jin, and Qingmin Liao. 2022. Disentangled modeling of social homophily and influence for social recommendation. *IEEE Transactions on Knowledge and Data Engineering* 35, 6 (2022), 5738–5751.

[18] Zihan Lin, Changxin Tian, Yupeng Hou, and Wayne Xin Zhao. 2022. Improving Graph Collaborative Filtering with Neighborhood-enriched Contrastive Learning. In *Proceedings of the ACM Web Conference 2022 (WWW '22)*. 2320–2329.

[19] Zhiwei Liu, Liangwei Yang, Ziwei Fan, Hao Peng, and Philip S. Yu. 2022. Federated Social Recommendation with Graph Neural Network. *ACM Transactions on Intelligent Systems and Technology (TIST)* 13, 5 (2022), 55:1–55:25.

[20] Sichun Luo, Yuanzhang Xiao, Xinyi Zhang, Yang Liu, Wenbo Ding, and Long Song. 2024. PerFedRec++: Enhancing Personalized Federated Recommendation with Self-Supervised Pre-Training. *ACM Transactions on Intelligent Systems and Technology (TIST)* (2024).

[21] Hao Ma, Haixuan Yang, Michael R. Lyu, and Irwin King. 2008. SoRec: Social Recommendation Using Probabilistic Matrix Factorization. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM '08)*. 931–940.

[22] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. 2011. Recommender Systems with Social Regularization. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM '11)*. 287–296.

[23] James MacQueen et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1. Oakland, CA, USA, 281–297.

[24] Peter V. Marsden and Noah E. Friedkin. 1993. Network Studies of Social Influence. *Sociological Methods & Research* 22, 127–151.

[25] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Review of Sociology* 27, 415–444.

[26] Ruihong Qiu, Zi Huang, Hongzhi Yin, and Zijian Wang. 2022. Contrastive Learning for Representation Degeneration Problem in Sequential Recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 813–823.

[27] Yuhan Quan, Jingtao Ding, Chen Gao, Lingling Yi, Depeng Jin, and Yong Li. 2023. Robust Preference-Guided Denoising for Graph based Social Recommendation. In *Proceedings of the ACM Web Conference 2023 (WWW '23)*. 1097–1108.

[28] Ye Tao, Ying Li, Su Zhang, Zhirong Hou, and Zhonghai Wu. 2022. Revisiting Graph based Social Recommendation: A Distillation Enhanced Social Graph Network. In *Proceedings of the ACM Web Conference 2022 (WWW '22)*. 2830–2838.

[29] Rianne van den Berg, Thomas N. Kipf, and Max Welling. 2018. Graph Convolutional Matrix Completion. In *KDD Workshop on Deep Learning Day*.

[30] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[31] Tianle Wang, Lianghao Xia, and Chao Huang. 2023. Denoised Self-Augmented Learning for Social Recommendation. In *International Joint Conference on Artificial Intelligence (IJCAI '23)*.

[32] Wenjie Wang, Yiyan Xu, Fuli Feng, Xinyu Lin, Xiangnan He, and Tat-Seng Chua. 2023. Diffusion recommender model. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 832–841.

[33] 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.

[34] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering *(SIGIR '19)*. 165–174.

[35] Jiahao Wu, Wenqi Fan, Jingfan Chen, Shengcai Liu, Qing Li, and Ke Tang. 2022. Disentangled contrastive learning for social recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4570–4574.

[36] 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 (SIGIR '21)*. 726–735.

[37] Le Wu, Junwei Li, Peijie Sun, Richang Hong, Yong Ge, and Meng Wang. 2020. DiffNet++: A Neural Influence and Interest Diffusion Network for Social Recommendation. *IEEE Transactions on Knowledge and Data Engineering* 34, 4753–4766.

[38] Le Wu, Peijie Sun, Yanjie Fu, Richang Hong, Xiting Wang, and Meng Wang. 2019. A Neural Influence Diffusion Model for Social Recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*. 235–244.

[39] Le Wu, Peijie Sun, Richang Hong, Yanjie Fu, Xiting Wang, and Meng Wang. 2018. SocialGCN: An Efficient Graph Convolutional Network based Model for Social Recommendation. *arXiv preprint arXiv:1811.02815* (2018).

[40] Qitian Wu, Hengrui Zhang, Xiaofeng Gao, Peng He, Paul Weng, Han Gao, and Guihai Chen. 2019. Dual Graph Attention Networks for Deep Latent Representation of Multifaceted Social Effects in Recommender Systems. In *The World Wide Web Conference (WWW '19)*. 2091–2102.

[41] Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. 2022. Contrastive Learning for Sequential Recommendation. *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 1259–1273.

[42] Mengyue Yang, Furui Liu, Zhitang Chen, Xinwei Shen, Jianye Hao, and Jun Wang. 2020. Causalvae: Structured causal disentanglement in variational autoencoder. *arXiv preprint arXiv:2004.08697* (2020).

[43] Yuhao Yang, Chao Huang, Lianghao Xia, and Chenliang Li. 2022. Knowledge Graph Contrastive Learning for Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*. 1434–1443.

[44] Yonghui Yang, Le Wu, Zihan Wang, Zhuangzhuang He, Richang Hong, and Meng Wang. 2024. Graph Bottlenecked Social Recommendation. *arXiv preprint arXiv:2406.08214* (2024). Accepted by KDD 2024.

[45] Zhengyi Yang, Jiancan Wu, Zhicai Wang, Xiang Wang, Yancheng Yuan, and Xiangnan He. 2024. Generate What You Prefer: Reshaping Sequential Recommendation via Guided Diffusion. *Advances in Neural Information Processing Systems* 36 (2024).

[46] Junliang Yu, Hongzhi Yin, Min Gao, Xin Xia, Xiangliang Zhang, and Nguyen Quoc Viet Hung. 2021. Socially-Aware Self-Supervised Tri-Training for Recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '21)*. 2084–2092.

[47] Junliang Yu, Hongzhi Yin, Jundong Li, Min Gao, Zi Huang, and Lizhen Cui. 2022. Enhancing Social Recommendation With Adversarial Graph Convolutional Networks. *IEEE Transactions on Knowledge and Data Engineering* 34, 8, 3727–3739.

[48] Junliang Yu, Hongzhi Yin, Jundong Li, Qinyong Wang, Nguyen Quoc Viet Hung, and Xiangliang Zhang. 2021. Self-Supervised Multi-Channel Hypergraph Convolutional Network for Social Recommendation. In *Proceedings of the Web Conference 2021 (WWW '21)*. 413–424.

[49] 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.

[50] Kun Zhou, Hui Wang, Wayne Xin Zhao, Yutao Zhu, Sirui Wang, Fuzheng Zhang, Zhongyuan Wang, and Ji-Rong Wen. 2020. S3-Rec: Self-Supervised Learning for Sequential Recommendation with Mutual Information Maximization. In *Proceedings of the 29th ACM International Conference on*

Information & Knowledge Management. 1893–1902.

[51] Ding Zou, Wei Wei, Xian-Ling Mao, Ziyang Wang, Minghui Qiu, Feida Zhu, and Xin Cao. 2022. Multi-level Cross-view Contrastive Learning for Knowledge-aware Recommender System. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1358–1368.