

RecDiff: Diffusion Model for Social Recommendation

Zongwei Li
University of Hong Kong
Hong Kong, China
zongwei9888@gmail.com

Lianghao Xia
University of Hong Kong
Hong Kong, China
aka_xia@foxmail.com

Chao Huang*
University of Hong Kong
Hong Kong, China
chaohuang75@gmail.com

ABSTRACT

Social recommendation has emerged as a powerful approach to enhance personalized recommendations by leveraging the social connections among users, such as following and friend relations observed in online social platforms. The fundamental assumption of social recommendation is that socially-connected users exhibit homophily in their preference patterns. This means that users connected by social ties tend to have similar tastes in user-item activities, such as rating and purchasing. **However, this assumption is not always valid due to the presence of irrelevant and false social ties, which can contaminate user embeddings and adversely affect recommendation accuracy.** To address this challenge, we propose a novel diffusion-based social denoising framework for recommendation (RecDiff). Our approach utilizes a simple yet effective hidden-space diffusion paradigm to alleviate the noisy effect in the compressed and dense representation space. By performing multi-step noise diffusion and removal, RecDiff possesses a robust ability to identify and eliminate noise from the encoded user representations, even when the noise levels vary. The diffusion module is optimized in a downstream task-aware manner, thereby maximizing its ability to enhance the recommendation process. We conducted extensive experiments to evaluate the efficacy of our framework, and the results demonstrate its superiority in terms of recommendation accuracy, training efficiency, and denoising effectiveness. The source code for the model implementation is publicly available at: <https://github.com/HKUDS/RecDiff>.

ACM Reference Format:

Zongwei Li, Lianghao Xia, and Chao Huang. 2024. RecDiff: Diffusion Model for Social Recommendation. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/nnnnnnn>.

1 INTRODUCTION

Personalized recommender systems are pivotal in inferring users' preferences for items and are extensively utilized in various online services, including e-commerce systems (e.g., Alibaba and Amazon) [43, 47] and content sharing platforms (e.g., Tiktok and Netflix) [17]. To overcome the challenge posed by limited user-item interactions, recent research has incorporated social relation data among users as an additional source of information. Known as social recommendation, these approaches aim to uncover shared preference patterns among socially-connected users, thereby enhancing user representations and improving recommendation performance [6].

Various social recommendation models have emerged, utilizing different technologies such as matrix factorization [42], attention

mechanism [2], and graph neural network [37]. These models encode user-item interactions and user-wise social connections in a latent space, collaboratively fusing representations to enhance preference modeling. Recent studies, including SAMN [2] and EATNN [3], incorporate attention mechanisms for capturing user relationships through weighted aggregation. Additionally, graph neural networks have also gained attention, with proposed encoders like DiffNet [37], DGRec [24], GraphRec [7], and DANSER [38] emphasizing the extraction of high-order social connectivities.

Despite significant progress in social recommendation, a persistent challenge is the presence of inherent noise in social information. While auxiliary social data captures user relationships reflecting shared interests, it can also include irrelevant or false social connections that contradict users' similarity in preferences. Real-world social networks often have social links such as acquaintances and colleagues, which are unrelated to shared interests, resulting in socially-connected users exhibiting diverse preference patterns. This noisy social information can misguide recommenders by overemphasizing similarities between socially-connected users with limited shared interests. State-of-the-art GNN-based methods are particularly susceptible to this misleading effect due to their information propagation process on noisy social edges.

While there are limited works that attempt to address denoising social information for recommendation, handcrafted self-supervised learning techniques may offer limited help in reducing ambiguity in social links. For instance, MHCN [45] and KCGN [10] utilize local-global mutual information maximization to constrain the extracted semantics of specific social edges using social community features at the subgraph level. SDCRec [6] and DcRec [35] employ self-discrimination tasks, leveraging the contrastive learning paradigm for social recommenders. However, their pre-defined self-supervised learning (SSL) tasks may not effectively align with the objective of denoising social recommendation in two aspects.

(i) These approaches heavily depend on handcrafted targets that may contain noise. In particular, the local-global infomax paradigm assumes less noise in global information, but the global social context can undergo semantic shifts. Contrastive learning introduces noise through random permutations.

(ii) These self-supervised learning tasks fail to consider the wide array of social noise types and intensities. Infomax and contrastive approaches solely concentrate on specific deviations in social-aware user preferences without distinguishing between varying levels of noisy social dependencies. For instance, users with different social behaviors may exhibit varying degrees of irrelevant or false social links when it comes to modeling similar interests.

Based on the above discussions, two fundamental questions arise regarding our social information denoiser for recommendation:

*Chao Huang is the Corresponding Author.

- How can we design suitable training targets that directly align with the denoising task for social recommendation?
- How can we effectively capture social dependencies in user preferences while accounting for varying levels of noisy connections?

To address the aforementioned challenges, we present a social diffusion model, called RecDiff, which combines the distinguishing capabilities of generative diffusion models with effective denoising-based training objectives. Our RecDiff model excels in capturing data distributions and filtering noise in social graph structures, enabling accurate modeling of user similarities based on their preferences. Additionally, we employ a refined noise diffusion and removal process, allowing our social recommender to effectively handle various types of connection noise.

Our RecDiff addresses the challenge of eliminating noisy social information by leveraging a diffusion model within a joint modeling framework that encompasses social relationships among users and interaction between users and items. To begin, RecDiff encodes the structural features from both the social and interaction graphs into low-dimensional embeddings. These embeddings serve as the foundation for subsequent refinement through our diffusion-based denoiser. The denoiser is trained using a multi-step noise diffusion and elimination process within the representation space. By sampling different diffusion steps, our approach gains exposure to a wide range of noise scales, enhancing its ability to handle various types of social noise. Compared to diffusion models operating directly on the original graph domain, our hidden-space diffusion paradigm exhibits higher efficiency and a more compact solution space. Finally, the revised social embeddings are merged with user-item interaction modeling, resulting in improved recommendations that incorporate denoised social information.

The key contributions of this paper are summarized as follows:

- We **present the RecDiff framework, a novel approach that enhances social recommender systems by effectively denoising social connections among users with a diffusion model.**
- Within our RecDiff framework, we **introduce an effective and efficient hidden-space diffusion paradigm.** Through multi-step noise propagation and removal training, our model acquires robust denoising capabilities, enabling it to effectively handle diverse social connections among users. As a result, our model produces accurate user preference representations.
- We conduct a comprehensive empirical study of our RecDiff framework, demonstrating its superiority in terms of recommendation performance and denoising capability.

2 PRELIMINARIES

2.1 Social-aware Collaborative Graph

We denote the users as $\mathcal{U} = \{u\}$ and the items as $\mathcal{V} = \{v\}$. The user-item interactions are represented by $\mathbf{R} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}|}$, while the user-user social relations are represented by $\mathbf{S} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{U}|}$. The user-item interactions capture historical user behaviors on items, e.g., clicks and views. Each element $r_{u,v} \in \mathbf{R}$ is set as 1 if an interaction is observed between user u and item v , and $r_{u,v} = 0$ otherwise. Similarly, the social relation between a pair of users (u, u') is represented by the binary element $s_{u,u'} \in \mathbf{S}$. Here, $s_{u,u'} =$

1 denotes social connections like friendship or following, while $s_{u,u'} = 0$ indicates no such observation between user u and u' .

To capture high-order connections in user-item interactions and user-user social relations, we transform them into graph-structured forms. The collaborative graph for user-item interactions, denoted as $\mathcal{G}_r = (\mathcal{U}, \mathcal{V}, \mathcal{E}_r)$, represents the relationships between users and items, where $\mathcal{E}_r = (u, v) | r_{u,v} = 1$ represents the edge set. Similarly, the social graph based on user connections, denoted as $\mathcal{G}_s = (\mathcal{U}, \mathcal{E}_s)$, captures the social relations between users, with $\mathcal{E}_s = (u, u') | s_{u,u'} = 1$ representing the edge set.

2.2 Graph-based Social Recommender

State-of-the-art social recommendation methods typically utilize GNN-based encoding functions on collaborative graphs and social graphs to learn representations for users and items. These representations capture preference patterns and enable accurate prediction of user-item interactions. The paradigm can be formulated as:

$$\hat{r}_{u,v} = \text{Pred}(\mathbf{e}_u, \mathbf{e}_v), \mathbf{E} = \text{Agg}(\mathbf{E}^r, \mathbf{E}^s), \mathbf{E}^* = \text{Enc}(\mathcal{G}_*) \quad (1)$$

where \mathbf{E}^* denotes \mathbf{E}^r or \mathbf{E}^s and \mathcal{G}_* denotes \mathcal{G}_r or \mathcal{G}_s for simplicity. The collaborative graph \mathcal{G}_r and the social graph \mathcal{G}_s are firstly encoded by the GNN-based encoding function $\text{Enc}(\cdot)$ to produce node representations $\mathbf{E}^r \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{V}|) \times d}$ and $\mathbf{E}^s \in \mathbb{R}^{|\mathcal{U}| \times d}$, respectively. The dual-view embeddings are then aggregated by a pooling method $\text{Agg}(\cdot)$ to yield the final node embeddings $\mathbf{E} \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{V}|) \times d}$. The embeddings are finally used by a predicting function $\text{Pred}(\cdot)$, such as the dot-product operation, to output predictions $\hat{r}_{u,v}$ for user-item interactions.

3 METHODOLOGY

3.1 Graph-based Collaborative Pattern Encoder

Taking inspiration from the success of simplified Graph Neural Networks (GNNs) [8], we utilize a lightweight Graph Convolutional Network (GCN) as the graph encoder in our social denoising framework. This encoder operates on the user-item interaction graph \mathcal{G}_r and follows an iterative message passing process described as:

$$\mathbf{E}^r = \sum_{l=0}^L \mathbf{E}_l^r, \mathbf{E}_l^r = h(\mathbf{D}^{-1/2} \mathbf{A}^r \mathbf{D}^{-1/2} \mathbf{E}_{l-1}^r \mathbf{W}) \quad (2)$$

Here, $\mathbf{A}^r \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{V}|) \times (|\mathcal{U}|+|\mathcal{V}|)}$ denotes the adjacency matrix of the collaborative graph \mathcal{G}_r , and \mathbf{D} is the corresponding diagonal degree matrix. The embedding matrix $\mathbf{E}_l^r \in \mathbb{R}^{(|\mathcal{U}|+|\mathcal{V}|) \times d}$ captures the embeddings in the l -th iteration of the GCN. The initial embeddings, \mathbf{E}_0^r , are randomly generated learnable parameters. In each iteration, the current embeddings are propagated to their neighboring nodes using the Laplacian-normalized adjacency matrix, and the l_2 embedding normalization function $h(\cdot)$ is applied. After L iterations, the final user and item embeddings, \mathbf{E}^r , are obtained by element-wise summation of the multi-order node embeddings.

3.2 Diffusion-based Social Relation Denoising

3.2.1 Hidden-Space Social Diffusion. Drawing inspiration from the success of diffusion models in generating noise-free data across various domains, such as images [23] and text [13], our RecDiff framework introduces a diffusion model to generate denoised social

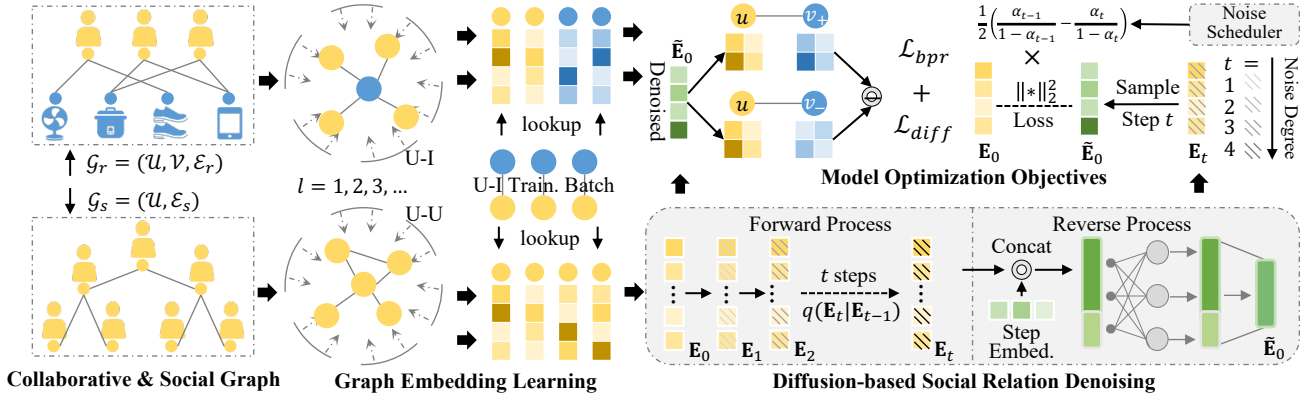


Figure 1: Overall architecture of the proposed RecDiff framework.

relation data. Considering the inherent sparsity of social graph data, we propose an approach that enables efficient and effective social diffusion by conducting forward and reverse diffusion processes in the latent space, rather than the graph data space. Figure 2 provides an illustration of this hidden-space social diffusion mechanism, and its process can be summarized by the following formula:

$$\mathcal{G}_s \xrightarrow{\phi} \mathbf{E}^s \xrightarrow{\psi} \tilde{\mathbf{E}}^s \xrightarrow{\psi'} \hat{\mathbf{E}}^s \xrightarrow{\phi'} \hat{\mathcal{G}}_s \quad (3)$$

Here, ϕ and ϕ' represent bidirectional projections between the graph data domain and the hidden embedding domain. ψ and ψ' refer to the forward and reverse processes of the diffusion model, where ψ introduces noise to \mathbf{E}^s , and ψ' is optimized to remove this noise in the hidden space.

In graph-based recommendation, the encoding function $\text{Enc}(\cdot)$ and the predictive function $\text{Pred}(\cdot)$ are crucial for accurately compressing and reconstructing graph structures. By regarding them as the aforementioned projections ϕ and ϕ' , respectively, these paired projections become invertible to each other. This property implies that by learning the hidden-space denoiser ψ' , our hidden-space diffusion model can effectively filter out noise present in the graph-structured data \mathcal{G} . Furthermore, due to the low-dimensional nature of the hidden space, training the hidden-space denoiser ψ' is significantly easier compared to directly denoising in the graph data space. Building on the aforementioned considerations, we utilize a similar GCN model as the learnable ϕ projection, following the formulation in Eq 2. Subsequently, we design our social denoiser based on the acquired embeddings $\mathbf{E}^s \in \mathbb{R}^{|\mathcal{U}| \times d}$.

3.2.2 Forward and Reverse Diffusion. Based on the social relation data \mathbf{E}^s in the latent embedding space, we design the forward and reverse processes for our hidden-space diffusion. In the forward process, Gaussian noise is incrementally added to the original data \mathbf{E}^s until it eventually transforms into complete Gaussian noise. Conversely, the backward process utilizes trainable neural networks to eliminate noise, enabling the generation of noise-less social data.

Forward Process. In the forward process, our RecDiff iteratively applies T noise steps, where T is a hyperparameter. The data at the t -th step is denoted as \mathbf{E}_t (omitting the superscript s for simplicity), and the 0-step data is the original data, i.e., $\mathbf{E}_0 = \mathbf{E}^s$. The t -step data

is calculated from the $(t-1)$ -step data as follows:

$$q(\mathbf{E}_t | \mathbf{E}_{t-1}) = \mathcal{N}(\mathbf{E}_t; \sqrt{1 - \beta_t} \mathbf{E}_{t-1}, \beta_t \mathbf{I}) \quad (4)$$

Here, \mathcal{N} represents the Gaussian distribution. The parameter β_t controls the magnitude of the noise. It has been shown that by gradually increasing β_t for $t = 1, \dots, T$, the noised data \mathbf{E}_t converges to complete Gaussian noise as t increases [9]. This property allows our noise diffusion process to cover a wide range of noise levels in the data. To generate β_t , we use a linear interpolation sequence s between two hyperparameters \bar{s}_{\max} and \bar{s}_{\min} :

$$\beta_t = 1 - s_t / s_{t-1}, \quad s = (1, \bar{s}_{\max}, \dots, \bar{s}_{\min}) \quad (5)$$

Due to the additivity property of Gaussian distributions, the t -step data can be directly calculated using only \mathbf{E}_0 and pre-computed values related to the β_t sequence. This significantly speeds up the forward process by avoiding iterative calculations. Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{t'=1}^t \alpha_{t'}$, the iterative formulation of \mathbf{E}_t can be simplified to an equation that depends only on \mathbf{E}_0 and $\bar{\alpha}_t$:

$$\begin{aligned} \mathbf{E}_t &= \sqrt{\alpha_t} \mathbf{E}_{t-1} + \sqrt{\beta_t} \mathbf{N}_1 \\ &= \sqrt{\alpha_t} (\sqrt{\alpha_{t-1}} \mathbf{E}_{t-2} + \sqrt{\beta_{t-1}} \mathbf{N}_2) + \sqrt{\beta_t} \mathbf{N}_1 \\ &= \sqrt{\alpha_t \alpha_{t-1}} \mathbf{E}_{t-2} + \sqrt{1 - \alpha_t \alpha_{t-1}} \mathbf{N}'_2 \\ &= \sqrt{\bar{\alpha}_t} \mathbf{E}_0 + \sqrt{1 - \bar{\alpha}_t} \mathbf{N}'_t \end{aligned} \quad (6)$$

Here, \mathbf{N}, \mathbf{N}' denote independent Gaussian distributions following $\mathcal{N}(0, \mathbf{I})$. Due to the addition rule of Gaussian distributions, the term $\sqrt{\alpha_t \alpha_{t-1}} \mathbf{N}_2 + \sqrt{1 - \alpha_t \alpha_{t-1}} \mathbf{N}_1$ follows $\mathcal{N}(0, \sqrt{1 - \alpha_t \alpha_{t-1}})$. By pre-calculating $\bar{\alpha}_t$ for $1 \leq t \leq T$, \mathbf{E}_t can be efficiently obtained without recursion, facilitating random sampling for the diffusion step t .

Reverse Process. In the reverse process, our objective is to restore the social relation data in the hidden space from the noisy data \mathbf{E}_t , where $t = 1, \dots, T$. Specifically, this process aims to estimate the following conditional probability using learnable neural networks:

$$p(\mathbf{E}_{t-1} | \mathbf{E}_t) = \mathcal{N}(\mathbf{E}_{t-1}; \mu_\theta(\mathbf{E}_t, t), \Sigma_\theta(\mathbf{E}_t, t)) \quad (7)$$

Here, $\mu_\theta(\cdot)$ and $\Sigma_\theta(\cdot)$ represent neural networks with learnable parameters θ used to estimate the Gaussian distribution. We concatenate the t -step data vector with a time step-specific embedding as the input. The network consists of two fully-connected layers:

$$\mu_\theta(\mathbf{e}_t, t) = \text{FC}^2(\mathbf{e}_t \| \mathbf{h}_t), \quad \text{FC}(\mathbf{x}) = \delta(\mathbf{W}\mathbf{x} + \mathbf{b}) \quad (8)$$

In the above equations, $\text{FC}^2(\cdot)$ represents two consecutive fully-connected layers. $\mathbf{e}_t \in \mathbb{R}^d$ denotes a node embedding vector in the t -th diffusion step, and $\mathbf{h}_t \in \mathbb{R}^{d'}$ represents the embedding for the t -th time step. $\delta(\cdot)$, \mathbf{W} , and \mathbf{b} refer to the activation function, linear transformation, and bias vectors for the fully-connected layer.

3.2.3 Diffusion Loss Function. The learnable denoising process of our hidden-space social diffusion is optimized by maximizing the evidence lower bound (ELBO) of the input social embeddings \mathbf{E}_0 . This ELBO term can be decomposed as follows:

$$\begin{aligned} \log p(\mathbf{e}_0) &= \log \int p(\mathbf{e}_{0:T}) d\mathbf{e}_{1:T} = \log \mathbb{E}_{q(\mathbf{e}_{1:T}|\mathbf{e}_0)} \frac{p(\mathbf{e}_{0:T})}{q(\mathbf{e}_{1:T}|\mathbf{e}_0)} \\ &\geq \underbrace{\mathbb{E}_{q(\mathbf{e}_1|\mathbf{e}_0)} [\log p_\theta(\mathbf{e}_0|\mathbf{e}_1)]}_{\text{(reconstruction term)}} - \underbrace{D_{\text{KL}}(q(\mathbf{e}_T|\mathbf{e}_0) \| p(\mathbf{e}_T))}_{\text{(prior matching term)}} \\ &\quad - \sum_{t=2}^T \underbrace{\mathbb{E}_{q(\mathbf{e}_t|\mathbf{e}_0)} [D_{\text{KL}}(q(\mathbf{e}_{t-1}|\mathbf{e}_t, \mathbf{e}_0) \| p_\theta(\mathbf{e}_{t-1}|\mathbf{e}_t))]}_{\text{(denoising matching term)}} \end{aligned} \quad (9)$$

Since the prior matching term is a constant, it can be omitted in the loss function for optimization. The denoising matching term aims to minimize the KL divergence between the true distribution $q(\mathbf{e}_{t-1}|\mathbf{e}_t, \mathbf{e}_0)$ and our denoiser $p_\theta(\mathbf{e}_{t-1}|\mathbf{e}_t)$. Following previous works (Ho et al., 2020; Wang et al., 2023), we simplify the learning process by omitting the learning of the standard deviation network and assume that $\Sigma_\theta(\mathbf{e}_t, t) = \sigma^2(t)\mathbf{I}$. The denoising matching term for the t -th time step can be defined as follows:

$$\begin{aligned} \mathcal{L}_t &= \mathbb{E}_{q(\mathbf{e}_t|\mathbf{e}_0)} [D_{\text{KL}}(q(\mathbf{e}_{t-1}|\mathbf{e}_t, \mathbf{e}_0) \| p_\theta(\mathbf{e}_{t-1}|\mathbf{e}_t))] \\ &= \mathbb{E}_{q(\mathbf{e}_t|\mathbf{e}_0)} \left[\frac{1}{2\sigma^2(t)} \left[\|\mu_\theta(\mathbf{e}_t, t) - \mu(\mathbf{e}_t, \mathbf{e}_0, t)\|_2^2 \right] \right] \end{aligned} \quad (10)$$

$\mu_\theta(\mathbf{e}_t, t)$ represents the output of our mean value predictor for an embedding vector \mathbf{e}_t , and $\mu(\mathbf{e}_t, \mathbf{e}_0, t)$ denotes the mean value for the true probability. These mean value terms can be decomposed into components related to \mathbf{e}_t , \mathbf{e}_0 , and the output of a prediction network $\hat{\mathbf{e}}_\theta(\mathbf{e}_t, t)$ for the real data \mathbf{e}_0 . As a result, \mathcal{L}_t can be expressed as:

$$\mathcal{L}_t = \mathbb{E}_{q(\mathbf{e}_t|\mathbf{e}_0)} \left[\frac{1}{2} \left(\frac{\bar{\alpha}_{t-1}}{1 - \bar{\alpha}_{t-1}} - \frac{\bar{\alpha}_t}{1 - \bar{\alpha}_t} \right) \|\hat{\mathbf{e}}_\theta(\mathbf{e}_t, t) - \mathbf{e}_0\|_2^2 \right] \quad (11)$$

The reconstruction term can be represented as the squared loss between the predicted value $\hat{\mathbf{e}}_\theta(\mathbf{e}_1, 1)$ and the real embedding vector \mathbf{e}_0 . This term is denoted as \mathcal{L}'_t and defined as follows:

$$\mathcal{L}'_t = \mathbb{E}_{q(\mathbf{e}_t|\mathbf{e}_0)} [\|\hat{\mathbf{e}}_\theta(\mathbf{e}_1, 1) - \mathbf{e}_0\|_2^2] \quad (12)$$

3.2.4 Inference Process. During the inference process, we focus on removing noise from the observed social data and utilize the denoised embeddings for making predictions. The social denoiser takes the encoded social embeddings \mathbf{E}^s as input, skipping the forward noise diffusion process. In the denoising step, we iteratively remove noise by updating $\hat{\mathbf{e}}_{t-1} = \hat{\mu}_\theta(\hat{\mathbf{e}}_t, t)$, where $\hat{\mu}_\theta(\hat{\mathbf{e}}_t, t)$ is defined as follows:

$$\hat{\mu}_\theta(\hat{\mathbf{e}}_t, t) = \frac{\sqrt{\alpha_t(1 - \bar{\alpha}_{t-1})}}{1 - \bar{\alpha}_t} \hat{\mathbf{e}}_t + \frac{\sqrt{\bar{\alpha}_{t-1}(1 - \alpha_t)}}{1 - \bar{\alpha}_t} \hat{\mathbf{e}}_\theta(\hat{\mathbf{e}}_t, t) \quad (13)$$

We predict $\hat{\mathbf{e}}_0$ based on $\hat{\mathbf{e}}_t$ and t , denoted as $\hat{\mathbf{e}}_\theta$. We then use the derived $\hat{\mathbf{e}}_0$ for consecutive predictions. The inference procedure is outlined in Algorithm 1.

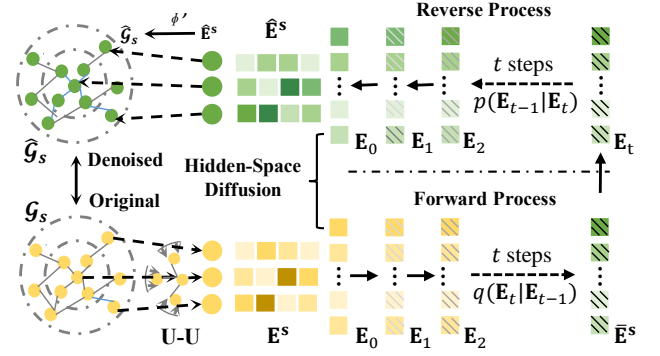


Figure 2: Illustration for the hidden-space social diffusion.

Algorithm 1: Inference of our RecDiff framework.

Input: The social tie embedding \mathbf{E}^s .
Output: the denoising embedding $\hat{\mathbf{e}}_0$.

- 1 Set $\hat{\mathbf{E}}_t = \mathbf{E}_0$;
- 2 **for** $t = T, \dots, 1$ **do**
- 3 $\hat{\mathbf{e}}_{t-1} = \hat{\mu}_\theta(\hat{\mathbf{e}}_t, t)$ obtained from $\hat{\mathbf{e}}_t$ and $\hat{\mathbf{e}}_\theta(\cdot)$ via Eq 13;
- 4 **end**
- 5 return the denoising $\hat{\mathbf{e}}_0$;

3.3 Prediction and Optimization

Using the hidden-space social diffusion module, we combine the denoised social relation with the encoded interaction patterns to obtain final embeddings for predictions. This is done as follows:

$$\hat{\mathbf{r}}_{u,v} = \tilde{\mathbf{e}}_u^\top \mathbf{e}_v^r, \quad \tilde{\mathbf{e}}_u = \mathbf{e}_u^r + \hat{\mathbf{e}}_\theta(\mathbf{e}_{u,t}, t) \quad (14)$$

where t represents a sampled diffusion step for user u . We optimize our RecDiff using the predictions $\hat{\mathbf{r}}_{u,v}$ and a combination of recommendation loss and diffusion loss functions:

$$\mathcal{L} = \sum_{u,v^+,v^-} -\log \sigma(\hat{\mathbf{r}}_{u,v^+} - \hat{\mathbf{r}}_{u,v^-}) + \lambda_1 \sum_t \mathcal{L}'_t \quad (15)$$

损失函数

(u, v^+, v^-) is a triplet sample for pairwise recommendation training [22]. The diffusion loss is calculated on sampled diffusion steps t for embeddings. We also apply weight-decay regularization, with weight λ_2 , to all trainable parameters Θ . The learning process of RecDiff, including graph encoding, multi-step forward and back diffusion, and loss calculations.

3.4 Model Complexity Analysis

This section provides a comprehensive analysis of the time and space complexity of our RecDiff with the social diffusion module.

Time Complexity: Initially, RecDiff performs graph-level information propagation on both the holistic collaborative graph \mathcal{G}_r and the social graph \mathcal{G}_s . This process requires $O((|\mathcal{E}_r| + |\mathcal{E}_s|) \times d)$ calculations for message passing and $O((|\mathcal{U}| + |\mathcal{V}|) \times d^2)$ for embedding transformation. However, our social diffusion model operates exclusively on the current batch during each training step. Let B be the number of user-item interaction pairs in each batch. The forward diffusion process costs $O(B \times d)$ computations, while the reverse process costs $O(B \times (d^2 + dd'))$. Theoretical analysis suggests that our RecDiff achieves comparable time costs to common

Table 1: Statistics of experimental datasets.

Dataset	Ciao	Yelp	Epinions
# of Users	1925	99262	14680
# of Items	15053	105142	233261
# of User-Item Interactions	23223	672513	447312
# of Social Interactions	65084	1298522	632144

social recommendation methods based on GNNs.

Memory Complexity: The graph encoding process of our RecDiff model requires a similar number of parameters as conventional graph-based social recommenders. The hidden-space diffusion network employs $O(d^2 + dd')$ parameters for the denoiser. In comparison, diffusion models operating on the original graph data typically require $O(|\mathcal{U}| \times d)$ parameters for the trained denoising process. This distinction results in the denoiser of our RecDiff having a smaller solution space, thereby alleviating optimization challenges.

4 EVALUATION

In this section, we analyze the performance of our RecDiff framework by exploring the following research questions (RQs):

- **RQ1:** How does the performance of our RecDiff model compare to various recommendation baseline methods?
- **RQ2:** What are the effects of different designed modules in our RecDiff framework on recommendation performance?
- **RQ3:** How do different settings of key hyperparameters impact the recommendation accuracy of our RecDiff method?
- **RQ4:** What impact does the noise scale in the noise diffusion process of our RecDiff have on the model’s performance?
- **RQ5:** How does the efficiency of RecDiff compare to baselines?
- **RQ6:** How effectively can our RecDiff with the social diffusion model handle noisy user connections?
- **RQ7:** Can our social relation denoiser provide explainability?

数据集

4.1 Experimental Settings

4.1.1 Experimental Datasets. We conducted experiments on three publicly available datasets from real-world commercial platforms: **Yelp**, **Ciao**, and **Epinions**. The user-item interaction data was based on users’ review records, where an interaction (u, v) exists if user u reviewed item v . For user-wise social relationships, we established connections between users (u, u') if user u' is in the trust list of user u . Table 1 presents the statistics of these datasets. Here is a summary of the dataset information:

- **Yelp:** This data originates from Yelp and records user feedback on venues. It includes social relationships between users, emphasizing networks formed by individuals with shared interests.
- **Ciao:** The Ciao dataset is sourced from the Ciao platform and captures user reviews and ratings across a wide range of products and services. It provides detailed information on social interactions among users, highlighting networks formed through shared preferences and engagements.
- **Epinions:** This data collects user feedback on various items from the social network-based review platform Epinions [7]. It categorizes ratings from 1 to 5 into distinct interaction categories: negative, below average, neutral, above average, and positive.

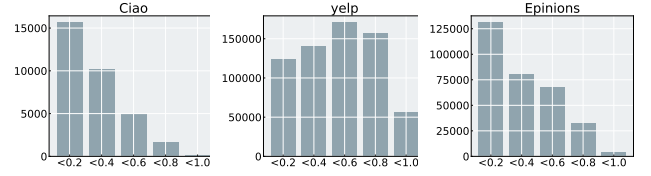


Figure 3: The distribution of social relation pairs across different datasets based on embedding similarity levels.

We conducted an analysis to reveal the inconsistency between users’ social relationships and their interaction patterns, indicating the presence of noise. By computing cosine similarity of social relation pairs’ embeddings in the three datasets, we observed that a certain percentage of user pairs exhibit low-level similarity (cosine similarity < 0.2), suggesting the existence of noise in social relationships. Detailed results can be found in Table 3.

4.1.2 Evaluation Protocols. In our experiments, we used a 7:1:2 ratio to create training, validation, and test sets for each dataset, following standard data partitioning criteria in graph-based recommender systems [8, 30]. To mitigate sampling bias, we employed an all-ranking protocol [15] to evaluate prediction accuracy for all items. The evaluation metrics used were **Recall@N** and **NDCG@N**, widely adopted in Top-N recommendations. 评价指标

4.1.3 Compared Baselines. We compared our proposed model with 11 baseline methods representing diverse research approaches. The baseline methods used for comparison include:

(i) Conventional and Attention-based Social Recommenders:

- **TrustMF** [42]: Joint matrix factorization for user-item interaction and user-user trust matrices to enhance collaborative recommendations.
- **SAMN** [2]: Two-stage attention mechanism modeling social-aware relations between users and their social neighbors.

(ii) Graph Collaborative Filtering Models:

- **NGCF** [30]: A representative graph-enhanced collaborative filtering model that captures collaborative signals by propagating embeddings on GCN-layers using a user-item bipartite graph.

(iii) GNN-based Social Recommender Systems:

- **GraphRec** [7]: Introduces a graph attention network for attentive information propagation on the social network, merging social connection and user-item interaction information to enhance user representations.
- **DiffNet** [37]: Utilizes a layer-wise diffusion architecture to represent social relations through graph information propagation, capturing recursive social influences.
- **GDMSR** [21]: Presents a robust graph-based denoising framework that effectively filters out noise, improving recommendation quality through preference guidance and relational modeling.

(iv) Temporal-aware Social Recommendation Frameworks:

- **DGRec** [24]: Combines RNNs with graph attention layers to capture dynamic user interests and social connections.

(v) Knowledge-enhanced Social Recommender Systems:

- **KCGN** [10]: Integrates interdependent knowledge among items and social influences among users within a multi-task learning framework for social recommendation.

(vi) Self-Supervised Learning Social Recommenders:

- **MHCN** [45]: Self-supervised learning with multi-channel hypergraph neural networks to enhance model performance.
- **SMIN**

Table 2: Recommendation performance of all methods in terms of Recall@20 and NDCG@20.

Dataset	Metrics	TrustMF	SAMN	DiffNet	GraphRec	DGRec	NGCF	MHCN	KCGN	SMIN	GDMSR	DSL	RecDiff	p-val
Ciao	Recall	0.0539	0.0604	0.0528	0.0540	0.0517	0.0559	0.0621	0.0602	0.0588	0.0560	0.0606	0.0712	4.047-12
	Imprv	32.10%	17.88%	34.85%	31.85%	37.72%	27.37%	14.65%	18.27%	21.09%	27.14%	17.49%	-	-
	NDCG	0.0343	0.0384	0.0328	0.0335	0.0319	0.0363	0.0378	0.0350	0.0354	0.0355	0.0389	0.0419	4.675-4
	Imprv	22.16%	9.11%	27.74%	25.07%	31.35%	15.43%	10.85%	19.71%	18.36%	18.03%	7.71%	-	-
Yelp	Recall	0.0371	0.0403	0.0557	0.0419	0.0410	0.0450	0.0567	0.0460	0.0485	0.0513	0.0504	0.0597	4.288-10
	Imprv	60.92%	48.14%	7.18%	42.48%	45.61%	32.67%	5.29%	29.78%	23.09%	16.37%	18.45%	-	-
	NDCG	0.0193	0.0208	0.0292	0.0201	0.0209	0.0230	0.0292	0.0234	0.0251	0.0246	0.0259	0.0308	5.064-08
	Imprv	59.59%	48.08%	5.48%	53.23%	47.37%	33.91%	5.48%	31.62%	22.71%	25.20%	18.92%	-	-
Epinions	Recall	0.0265	0.0329	0.0384	0.0334	0.0326	0.0353	0.0438	0.0370	0.0333	0.0368	0.0365	0.0460	1.117-08
	Imprv	73.58%	39.82%	19.79%	37.72%	41.10%	30.31%	5.02%	24.32%	38.14%	25.00%	26.03%	-	-
	NDCG	0.0195	0.0226	0.0273	0.0246	0.0236	0.0243	0.0321	0.0264	0.0228	0.0241	0.0267	0.0336	4.200-08
	Imprv	72.31%	48.67%	23.08%	36.59%	42.37%	38.27%	4.67%	27.27%	47.37%	39.42%	25.84%	-	-

[18]: Proposes a meta-path-guided heterogeneous graph learning approach with self-supervised signals based on mutual information maximization to enhance training in social recommendation.

- **DSL**[27]: Introduces an adaptive self-supervision task for personalized social information denoising, preserving valuable social relationships for user preference modeling.

4.1.4 Hyperparameter Settings. We implement RecDiff using PyTorch and optimize it with the Adam algorithm. The embeddings’ dimensionality is tuned from 8, 16, 32, 64. The learning rate ranges from 0.001, 0.005, 0.01, and the batch size varies between 512 and 4096. The number of diffusion steps ranges from 10 to 200. Timestep embedding size is selected from 4, 8, 16, 32. Other hyperparameter details can be found in our release code. For baselines, we use released code or implement them based on the original paper. Hyperparameters are optimized through grid search, with a standardized embedding dimension of 64. The batch size for Ciao is 2048, while for Yelp and Epinions, it is 4096. GNN models use 1 to 3 propagation layers for optimal performance across baselines.

4.2 Overall Performance Comparison (RQ1)

We compare the overall recommendation performance of our RecDiff with baselines. The comparison results are presented in Table 2 for top-20 evaluation and Table 3 for varying top-N evaluation. Based on these results, we draw the following conclusions.

- **Superiority of RecDiff.** Our RecDiff consistently outperforms state-of-the-art baselines, demonstrating superior recommendation accuracy. T-tests confirm the statistical significance of our results across all datasets and evaluation metrics. The performance advantage of RecDiff remains consistent across different top-N settings (Table 3). Our diffusion-based social relation denoising module removes irrelevant and false information, allowing RecDiff to effectively mine valuable social ties for enhanced recommendations.

- **Negative Impact of Noisy Social Information.** Some social recommendation methods, such as DGRec, DiffNet, and GraphRec, perform worse than the social information-agnostic method NGCF. This suggests that social connections can have a negative influence on user-item relation modeling due to false or irrelevant components. Our RecDiff framework addresses this issue by denoising social information and consistently outperforms the baseline model GDMSR. It effectively filters out noise from social connections and

identifies meaningful and influential social ties, accurately encoding user preferences for precise recommendations.

- **Advantages of diffusion-based supervision augmentation.** Baseline methods incorporating self-supervised learning (SSL) consistently outperform other approaches in recommendation performance. Methods like MHCN, KCGN, and SMIN utilize variations of the local-global infomax technique, while DSL employs a prediction alignment SSL task. This highlights the positive impact of auxiliary supervision signals in addressing data deficiency challenges in social recommendation, such as noise and sparsity. In contrast, our RecDiff introduces a multi-step denoising method based on the diffusion model, generating a larger number of supervision signals at different noise levels. This robust denoising capability leads to superior recommendation performance, surpassing the baselines.

4.3 Model Ablation Study (RQ2)

In this section, we investigate the influence of different sub-modules in our RecDiff framework through an ablation study. We evaluate the performance of several variants obtained by removing or replacing essential modules. The following ablated models are compared:

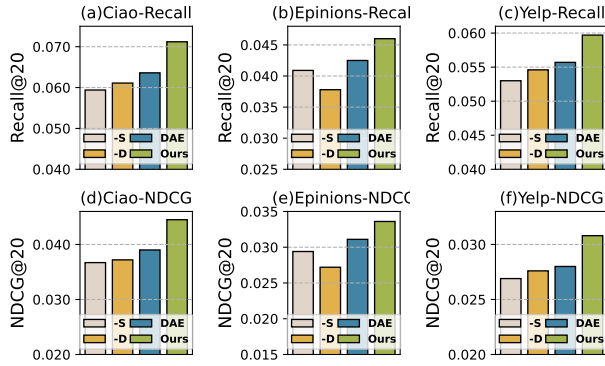
- **-D:** Removes the holistic diffusion module, retaining only the social and user-item relation learning GNN.
- **-S:** Does not utilize social information. Instead, it solely relies on the user-item interaction graph to make recommendations.
- **DAE:** Replaces the diffusion-based denoising module of RecDiff with a denoising autoencoder. This DAE-based denoising module is trained to reconstruct randomly masked user representations.

We evaluate these variants on the Ciao and Yelp datasets using the top-20 recommendation setting. The results, depicted in Figure 4, consistently demonstrate that RecDiff outperforms all four variants. These findings strongly support the effectiveness of the social relation learning and diffusion-based denoising components in our model. Notable observations from our analysis include:

- (1) Removing the diffusion module (-D) leads to significant performance degradation, highlighting the denoising function’s effectiveness provided by our latent feature-level diffusion model.
- (2) Comparing the -S variant to RecDiff highlights the significant improvement obtained by incorporating users’ social context information in user preference learning. However, subgraphs (b) and (e) suggest that in the presence of noisy social information in the epinions dataset, the -S variant may outperform the -D variant.

Table 3: Recommendation performance with varying top-N settings, in terms of Recall@N and NDCG@N.

Model	Ciao@ 10		Ciao@ 40		Yelp@ 10		Yelp@ 40		Epinions @ 10		Epinions @ 40	
	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG	Recall	NDCG
TrustMF	0.0341	0.0289	0.0796	0.0416	0.0224	0.0149	0.0606	0.0254	0.0165	0.0163	0.0394	0.0236
SAMN	0.0345	0.0289	0.0801	0.0429	0.0289	0.0195	0.0700	0.0308	0.0193	0.0181	0.0496	0.0280
DiffNet	0.0328	0.0271	0.0780	0.0397	0.0381	0.0247	0.0739	0.0312	0.0238	0.0227	0.0587	0.0335
GraphRec	0.0322	0.0266	0.0838	0.0420	0.0233	0.0144	0.0711	0.0277	0.0207	0.0206	0.0521	0.0304
DGRec	0.0296	0.0254	0.0733	0.0381	0.0245	0.0158	0.0656	0.0272	0.0197	0.0194	0.0517	0.0293
NGCF	0.0366	0.0301	0.0804	0.0343	0.0276	0.0177	0.0711	0.0297	0.0217	0.0206	0.0550	0.0304
MHCN	0.0343	0.0286	0.0929	0.0473	0.0342	0.0225	0.0890	0.0377	0.0272	0.0272	0.0674	0.0321
KCGN	0.0360	0.0263	0.0926	0.0448	0.0284	0.0182	0.0702	0.0295	0.0221	0.0219	0.056	0.0325
SMIN	0.0326	0.0275	0.0813	0.0453	0.0316	0.0198	0.0768	0.0312	0.0203	0.0186	0.0531	0.0289
GDMSR	0.0340	0.276	0.0804	0.0409	0.0369	0.0196	0.0744	0.0293	0.0226	0.0206	0.0536	0.0291
DSL	0.0412	0.0329	0.0873	0.0473	0.0315	0.0203	0.0786	0.0332	0.0229	0.0226	0.0594	0.0338
RecDiff	0.0457	0.0361	0.1023	0.0535	0.0391	0.0249	0.0941	0.0394	0.0282	0.0275	0.0696	0.0343


Figure 4: Ablation studies on Yelp, Ciao, and Epinions datasets for different sub-modules in our proposed RecDiff framework, measuring Recall@20 and NDCG@20.

- (3) The suboptimal performance of the DAE variant showcases the superior denoising ability of our designed diffusion module compared to vanilla denoising techniques. RecDiff effectively models complex representation distributions by gradually learning each denoising transition step from t to $t - 1$ through shared neural networks, enhancing noise reduction in latent features.

4.4 Impact of Hyperparameters (RQ3)

This section investigates the impact of crucial hyperparameters on model performance: the dimensionality of hidden representations (d), the dimensionality of time step embeddings (d'), and the maximum number of diffusion steps (T). Evaluation is conducted on all three experimental datasets, measuring Recall@20 and NDCG@20 metrics. The results, shown in Figure 6, are analyzed as follows:

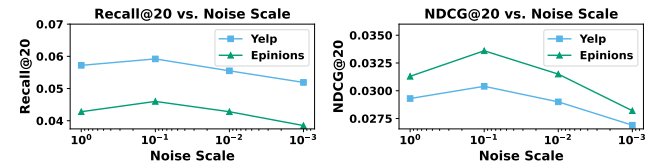
- **Embedding dimensionality d :** Increasing d improves performance, except for the Ciao and Epinions datasets where larger values lead to slight degradation due to overfitting.
- **Time step embedding size d' :** Larger dimensions enhance the positive impact of diffusion steps on denoising, improving performance. However, excessively high dimensions hinder the model’s diffusion ability, resulting in decreased performance.
- **Maximum diffusion steps T :** Increasing T enhances performance with more diffusion steps. However, extremely large values damage social information and reduce denoising effectiveness.

4.5 Impact of Noise Scale (RQ4)

This section investigates the impact of the noise scale factor (τ) on our noising process. By adjusting the minimum noise (\bar{s}_{min}) and maximum noise (\bar{s}_{max}) in the noise scheduler to $\tau \cdot \bar{s}_{min}$ and $\tau \cdot \bar{s}_{max}$, respectively, we examine the model’s performance with different noise scale values of 1, 0.1, 0.01, 0.001. The results, depicted in Figure 5, reveal the following insights:

- Increasing the noise scale effectively improves model performance, showcasing the effectiveness of our denoising mechanism within our proposed RecDiff framework.
- Further increasing the noise scale beyond a certain threshold leads to a decline in performance. This effect is particularly pronounced in sparsely populated datasets like Yelp and Epinion.

We attribute this decline to excessive noise obscuring the intrinsic personalized data, thereby hindering the model’s ability to retain and process essential individualized information.


Figure 5: Impact of noise scale over model performance.

4.6 Training Efficiency Study (RQ5)

This section optimizes the efficiency of our RecDiff compared to baseline models (MHCN, SMIN, and KCGN) on the Ciao and Yelp datasets. Using an A40 graphics card with 48GB of GPU memory, we compared the time costs of these baselines (Table 4). Our RecDiff demonstrates significant efficiency advantages in both training and testing. For each training epoch, we evaluated and recorded the test set performance to analyze improvements (Figure 8).

- **Training efficiency of RecDiff:** Our RecDiff consistently outperforms the baselines in training efficiency, benefiting from effective denoising diffusion for accelerated optimization.
- **Limitations of baselines:** SMIN shows overfitting effects, potentially due to reliance on metagraphs, limiting generalization. MHCN achieves high final performance but converges slower due

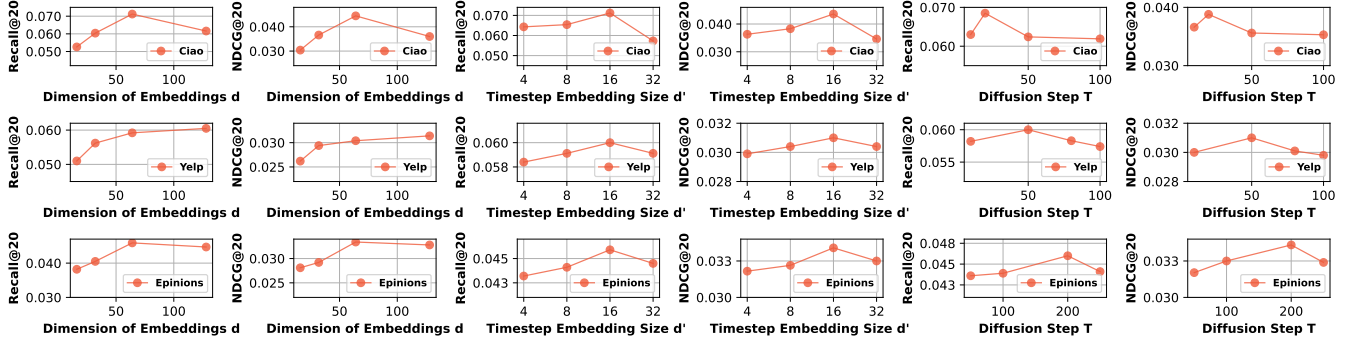


Figure 6: Hyperparameter study on important parametric configurations of RecDiff, in terms of Recall@20 and NDCG@20.

Table 4: The running time (in seconds) per epoch for different models is compared on diverse evaluation datasets.

Model	Training			Testing		
	ciao	yelp	epinions	ciao	yelp	epinions
MHCN	0.46	28.51	9.98	0.33	44.39	22.74
KCGN	0.33	75.84	63.11	0.25	31.23	13.61
SMIN	0.95	60.86	66.15	1.25	35.54	27.42
Ours	0.19	4.23	2.78	0.31	27.91	13.54

to its complex hypergraph structure. In contrast, our RecDiff benefits from a compact neural architecture without handcrafted priors, enabling faster optimization with auxiliary signals.

• **Fluctuations on Ciao data:** In comparing convergence curves, significant recommendation performance fluctuations are observed on the smaller Ciao dataset, indicating training instability.

4.7 Further Exploration of Anti-Noise Capacity with our RecDiff Framework (RQ6)

We evaluate the robustness of RecDiff in the presence of data noise by introducing random fake edges to replace varying percentages of genuine social connections in the user-user graph. The model is then retrained using the corrupted graph and evaluated on the true test set. Specifically, we analyze the effects of replacing 0%, 20%, and 50% of the social relations with noise signals. Comparing the performance of RecDiff with MHCN and DiffNet, the results in Figure 9 (a) and (b) show the original evaluation outcomes, while (c) illustrates the relative performance change in NDCG. Based on these results, the following observations are made:

• **Advantageous robustness of RecDiff:** Our RecDiff model outperforms the baselines with a smaller performance drop, showcasing its superior denoising capabilities in social recommendation.

• **Denoising effect of vanilla SSL:** The MHCN model shows promise in denoising, but it falls short compared to our RecDiff model. This highlights that general-purpose self-supervised learning tasks may not effectively address the specific denoising requirements of social recommendation.

• **Higher noise ratio in Ciao dataset:** The Ciao dataset demonstrates a larger performance drop, suggesting a higher noise ratio in comparison to other datasets.

4.8 Case Study (RQ7)

This section explores the denoising effect of RecDiff on specific user/item cases. Four subgraph cases are illustrated in Figure 7,

highlighting the need for denoising. The baseline methods, KCGN and MHCN, fail to identify false social connections, resulting in high cosine scores for the incorrect social neighbors. In contrast, our proposed RecDiff effectively recognizes these noise instances, yielding significantly lower similarity scores and producing distinct embeddings for falsely-connected users. These findings demonstrate the superior noise elimination capability of RecDiff across different noise situations.

Two additional cases are presented, involving user pairs sharing interactions with items that significantly differ in category from other items the users interact with. These isolated interactions are likely to be noisy items, rendering the associated social links noisy as well. Once again, RecDiff successfully identifies and eliminates the noise, assigning lower similarity scores and generating more distinct embeddings for false social neighbors. These cases further exemplify the denoising effectiveness of our RecDiff approach.

5 RELATED WORK

• **Social-aware Recommender Systems.** Deep learning-based social recommender systems, such as DiffNet [37], RecoGCN [40], and KCGN [10], leverage Graph Neural Networks to effectively model the connections between users and items. Approaches like SAMN [2] and GraphRec [7] further enhance social recommendation by incorporating attention mechanisms to differentially differentiate the influence levels among users. More recent self-supervised learning (SSL) based methods, including MHCN [45], SMIN [18], SDCRec [6], and DcRec [35], have shown promising and encouraging results in bolstering social recommenders through innovative SSL-based data augmentation techniques. In contrast to these existing established approaches, our proposed method, RecDiff, takes a unique and distinctive stance by concentrating on denoising relation learning in social recommender systems with the power of diffusion models.

• **Recommendation with Graph Neural Networks.** Graph neural networks have achieved state-of-the-art performance in recommendation scenarios [25, 41, 48]. Earlier works like NGCF [30] and Pinsage [44] introduced higher-order connectivity extraction using graph convolutional networks (GCNs). LightGCN [8] simplified training by removing non-linear activations, while DGCF [31] focused on intent-aware modeling through graph disentangling. Studies have also incorporated auxiliary information, such as multi-modal [33] and knowledge [29] data, to further enhance recommendation performance. Recent advancements in graph self-supervised learning include SGL [36], AutoCF [39], and NCL [15],

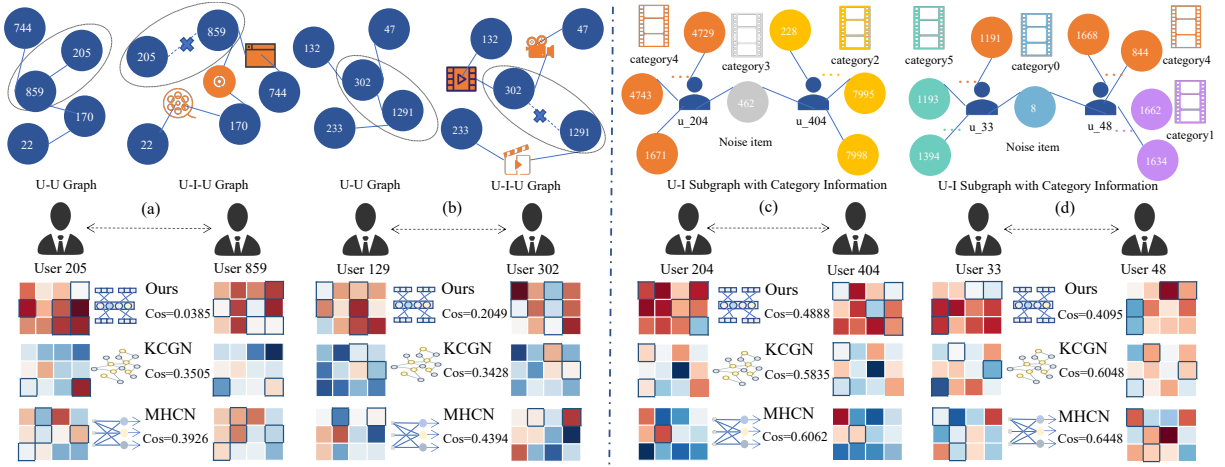


Figure 7: Case study for the user relation recalibration effect of our social denoiser RecDiff.

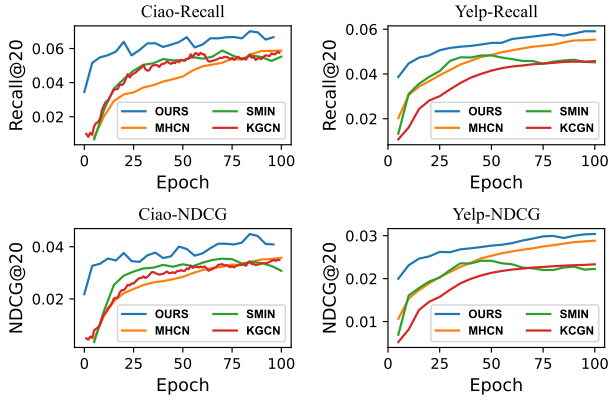


Figure 8: Model performance by training epochs on the Ciao and Yelp test sets, measured by Recall@20 and NDCG@20.

which have demonstrated promising results. In contrast, our work takes a unique approach by enhancing the denoising capabilities of recommenders through the development of an auxiliary task for social recommendation, leveraging a diffusion-based paradigm.

• **Generative Models for Recommendation.** Generative models, such as GANs [26] and VAEs [46], have gained prominence in recommender systems for data generation to enhance preference modeling [16, 34]. Some approaches focus on generating synthetic user-item interaction data to address data sparsity and cold start issues. GAN-based methods [4, 12, 32] use adversarial training to mimic real user behaviors. VAE-based generative recommendation models [14, 19] employ variational autoencoders to make accurate predictions. More recently, studies have explored the use of diffusion models [1, 5, 20] for improved data generation. For example, DiffRec [28] and DiffKG [11] use diffusion models to denoise the interaction data and knowledge graph, respectively, leading to enhanced recommendation performance.

In contrast, our work takes a distinct approach. RecDiff leverages diffusion models’ denoising paradigm to refine social representations for recommendation. It encodes structural features of the social graph into low-dimensional embeddings and performs denoising with a hidden-space diffusion paradigm.

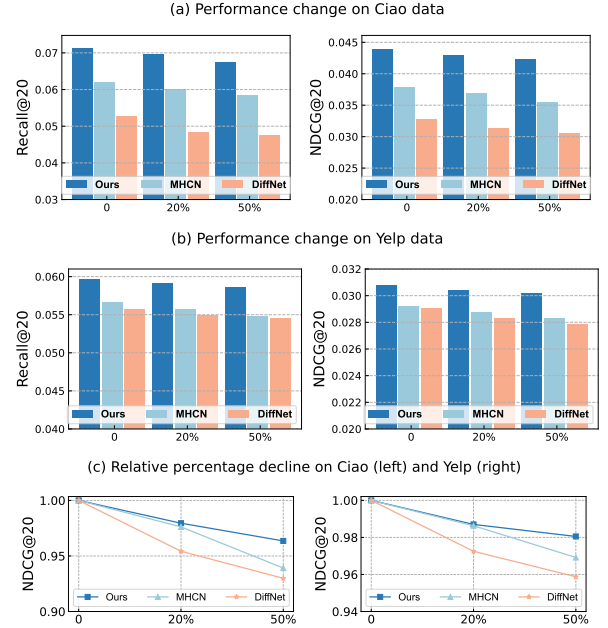


Figure 9: Investigating the Influence of Different Noise Ratios on Performance Degradation.

6 CONCLUSION

This paper aims to enhance social-aware recommender systems by eliminating false or irrelevant user-wise social links. To achieve this goal, we propose RecDiff, a novel diffusion model that trains a social denoiser through a multi-step noise propagation and elimination task. This diffusion process operates in the hidden space, utilizing encoded user representations for both effectiveness and simplicity. By training the model with varying diffusion steps, RecDiff demonstrates exceptional capabilities in handling diverse noisy effects. We evaluate the effectiveness of our model through experiments on real-world datasets, showing significant improvements in recommendation accuracy compared to existing methods. **In the future, we plan to explore the potential of our model in diverse recommendation scenarios, incorporating multi-modal information.**

REFERENCES

- [1] J. Austin, D. D. Johnson, J. Ho, D. Tarlow, and R. Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *NeurIPS*, 34:17981–17993, 2021.
- [2] C. Chen, M. Zhang, Y. Liu, and S. Ma. Social attentional memory network: Modeling aspect-and friend-level differences in recommendation. In *WSDM*, pages 177–185, 2019.
- [3] C. Chen, M. Zhang, C. Wang, W. Ma, M. Li, Y. Liu, and S. Ma. An efficient adaptive transfer neural network for social-aware recommendation. In *SIGIR*, pages 225–234, 2019.
- [4] H. Chen, Z. Wang, F. Huang, X. Huang, Y. Xu, Y. Lin, P. He, and Z. Li. Generative adversarial framework for cold-start item recommendation. In *SIGIR*, pages 2565–2571, 2022.
- [5] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023.
- [6] J. Du, Z. Ye, L. Yao, B. Guo, and Z. Yu. Socially-aware dual contrastive learning for cold-start recommendation. In *SIGIR*, pages 1927–1932, 2022.
- [7] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin. Graph neural networks for social recommendation. In *WWW*, pages 417–426, 2019.
- [8] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *SIGIR*, pages 639–648, 2020.
- [9] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020.
- [10] C. Huang, H. Xu, Y. Xu, P. Dai, L. Xia, M. Lu, L. Bo, H. Xing, X. Lai, and Y. Ye. Knowledge-aware coupled graph neural network for social recommendation. In *AAAI*, volume 35, pages 4115–4122, 2021.
- [11] Y. Jiang, Y. Yang, L. Xia, and C. Huang. Diffkg: Knowledge graph diffusion model for recommendation. In *WSDM*, 2024.
- [12] B. Jin, D. Lian, Z. Liu, Q. Liu, J. Ma, X. Xie, and E. Chen. Sampling-decomposable generative adversarial recommender. *NeurIPS*, 33:22629–22639, 2020.
- [13] X. Li, J. Thickstun, I. Gulrajani, P. S. Liang, and T. B. Hashimoto. Diffusion-lm improves controllable text generation. *NeurIPS*, 35:4328–4343, 2022.
- [14] D. Liang, R. G. Krishnan, M. D. Hoffman, and T. Jebara. Variational autoencoders for collaborative filtering. In *WWW*, pages 689–698, 2018.
- [15] Z. Lin, C. Tian, Y. Hou, and W. X. Zhao. Improving graph collaborative filtering with neighborhood-enriched contrastive learning. In *WWW*, pages 2320–2329, 2022.
- [16] F. Liu, Z. Cheng, L. Zhu, Z. Gao, and L. Nie. Interest-aware message-passing gcn for recommendation. In *WWW*, pages 1296–1305, 2021.
- [17] Y. Liu, Q. Liu, Y. Tian, C. Wang, Y. Niu, Y. Song, and C. Li. Concept-aware denoising graph neural network for micro-video recommendation. In *CIKM*, pages 1099–1108, 2021.
- [18] X. Long, C. Huang, Y. Xu, H. Xu, P. Dai, L. Xia, and L. Bo. Social recommendation with self-supervised metagraph informax network. In *CIKM*, pages 1160–1169, 2021.
- [19] J. Ma, C. Zhou, P. Cui, H. Yang, and W. Zhu. Learning disentangled representations for recommendation. *NeurIPS*, Jan 2019.
- [20] V. Popov, I. Vovk, V. Gogoryan, T. Sadekova, and M. Kudinov. Grad-tts: A diffusion probabilistic model for text-to-speech. In *ICML*, pages 8599–8608. PMLR, 2021.
- [21] Y. Quan, J. Ding, C. Gao, L. Yi, D. Jin, and Y. Li. Robust preference-guided denoising for graph based social recommendation. In *WWW*, pages 1097–1108, 2023.
- [22] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *UAI*, pages 452–461, 2009.
- [23] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi. Palette: Image-to-image diffusion models. In *SIGGRAPH*, pages 1–10, 2022.
- [24] W. Song, Z. Xiao, Y. Wang, L. Charlin, M. Zhang, and J. Tang. Session-based social recommendation via dynamic graph attention networks. In *WSDM*, pages 555–563, 2019.
- [25] J. Tang, Y. Yang, W. Wei, L. Shi, L. Xia, D. Yin, and C. Huang. Higtpt: Heterogeneous graph language model. *arXiv preprint arXiv:2402.16024*, 2024.
- [26] J. Wang, L. Yu, W. Zhang, Y. Gong, Y. Xu, B. Wang, P. Zhang, and D. Zhang. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *SIGIR*, pages 515–524, 2017.
- [27] T. Wang, L. Xia, and C. Huang. Denoised self-augmented learning for social recommendation. In *IJCAI*, 2023.
- [28] W. Wang, Y. Xu, F. Feng, X. Lin, X. He, and T.-S. Chua. Diffusion recommender model. In *SIGIR*, pages 832–841, 2023.
- [29] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua. Kgat: Knowledge graph attention network for recommendation. In *KDD*, pages 950–958, 2019.
- [30] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua. Neural graph collaborative filtering. In *SIGIR*, pages 165–174, 2019.
- [31] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, and T.-S. Chua. Disentangled graph collaborative filtering. In *SIGIR*, pages 1001–1010, 2020.
- [32] Z. Wang, W. Ye, X. Chen, W. Zhang, Z. Wang, L. Zou, and W. Liu. Generative session-based recommendation. In *WWW*, pages 2227–2235, 2022.
- [33] Y. Wei, X. Wang, L. Nie, X. He, R. Hong, and T.-S. Chua. Mmgcn: Multi-modal graph convolution network for personalized recommendation of micro-video. In *ACM MM*, pages 1437–1445, 2019.
- [34] Y. Wei, X. Wang, L. Nie, S. Li, D. Wang, and T.-S. Chua. Causal inference for knowledge graph based recommendation. *TKDE*, 2022.
- [35] J. Wu, W. Fan, J. Chen, S. Liu, Q. Li, and K. Tang. Disentangled contrastive learning for social recommendation. In *CIKM*, pages 4570–4574, 2022.
- [36] J. Wu, X. Wang, F. Feng, X. He, L. Chen, J. Lian, and X. Xie. Self-supervised graph learning for recommendation. In *SIGIR*, pages 726–735, 2021.
- [37] L. Wu, P. Sun, Y. Fu, R. Hong, X. Wang, and M. Wang. A neural influence diffusion model for social recommendation. In *SIGIR*, pages 235–244, 2019.
- [38] Q. Wu, H. Zhang, X. Gao, P. He, P. Weng, H. Gao, and G. Chen. Dual graph attention networks for deep latent representation of multifaceted social effects in recommender systems. In *WWW*, pages 2091–2102, 2019.
- [39] L. Xia, C. Huang, C. Huang, K. Lin, T. Yu, and B. Kao. Automated self-supervised learning for recommendation. In *WWW*, pages 992–1002, 2023.
- [40] F. Xu, J. Lian, Z. Han, Y. Li, Y. Xu, and X. Xie. Relation-aware graph convolutional networks for agent-initiated social e-commerce recommendation. In *CIKM*, pages 529–538, 2019.
- [41] H. Yan, S. Wang, Y. Yang, B. Guo, T. He, and D. Zhang. Siterrec: Store site recommendation under the o2o model via multi-graph attention networks. In *ICDE*, pages 525–538. IEEE, 2022.
- [42] B. Yang, Y. Lei, J. Liu, and W. Li. Social collaborative filtering by trust. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 39(8):1633–1647, 2016.
- [43] C.-Y. Yeh, H.-W. Chen, D.-N. Yang, W.-C. Lee, S. Y. Philip, and M.-S. Chen. Planning data poisoning attacks on heterogeneous recommender systems in a multiplayer setting. In *ICDE*, pages 2510–2523. IEEE, 2023.
- [44] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *KDD*, pages 974–983, 2018.
- [45] J. Yu, H. Yin, J. Li, Q. Wang, N. Q. V. Hung, and X. Zhang. Self-supervised multi-channel hypergraph convolutional network for social recommendation. In *WWW*, pages 413–424, 2021.
- [46] X. Yu, X. Zhang, Y. Cao, and M. Xia. Vaegan: A collaborative filtering framework based on adversarial variational autoencoders. In *IJCAI*, pages 4206–4212, 2019.
- [47] S. Zhang, L. Yao, A. Sun, and Y. Tay. Deep learning based recommender system: A survey and new perspectives. *ACM computing surveys (CSUR)*, 52(1):1–38, 2019.
- [48] X. Zhou, D. Lin, Y. Liu, and C. Miao. Layer-refined graph convolutional networks for recommendation. In *ICDE*, pages 1247–1259. IEEE, 2023.