Neurocomputing 618 (2025) 129133

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Causal disentanglement for regulating social influence bias in social recommendation

Li Wang [a], Min Xu [a],[*], Quangui Zhang [b], Yunxiao Shi [a], Qiang Wu [a]

[a] School of Electrical and Data Engineering, University of Technology Sydney, 15, Broadway, Ultimo, Sydney, 2000, NSW, Australia

[b] School of Artificial Intelligence, Chongqing University of Arts and Sciences, 319, Honghe Avenue, Yongchuan District, Chongqing, 402160, China

ARTICLE INFO

Communicated by L.L. Minku

Keywords:
Graph Neural Network (GNN)
Social recommendation
Causal inference
Disentanglement
Mutual information

ABSTRACT

Social recommendation systems face the problem of social influence bias, which can lead to an overemphasis on recommending items that friends have interacted with. Addressing this problem is crucial, and existing methods often rely on techniques such as weight adjustment or leveraging unbiased data to eliminate this bias. However, we argue that not all biases are detrimental, i.e., some items recommended by friends may align with the user's interests. Blindly eliminating such biases could undermine these positive effects, potentially diminishing recommendation accuracy. In this paper, we propose a **C**ausal **D**isentanglement-based framework for **R**egulating **S**ocial influence **B**ias in social recommendation, named CDRSB, to improve recommendation performance. From the perspective of causal inference, we find that the user social network could be regarded as a confounder between the user and item embeddings (treatment) and ratings (outcome). Due to the presence of this social network confounder, two paths exist from user and item embeddings to ratings: a non-causal social influence path and a causal interest path. Building upon this insight, we propose a disentangled encoder that focuses on disentangling user and item embeddings into interest and social influence embeddings. Mutual information-based objectives are designed to enhance the distinctiveness of these disentangled embeddings, eliminating redundant information. Additionally, a regulatory decoder that employs a weight calculation module to dynamically learn the weights of social influence embeddings for effectively regulating social influence bias has been designed. Experimental results on four large-scale real-world datasets Ciao, Epinions, Dianping, and Douban book demonstrate the effectiveness of CDRSB compared to state-of-the-art baselines. We release our code at https://github.com/Lili1013/CDRSB.

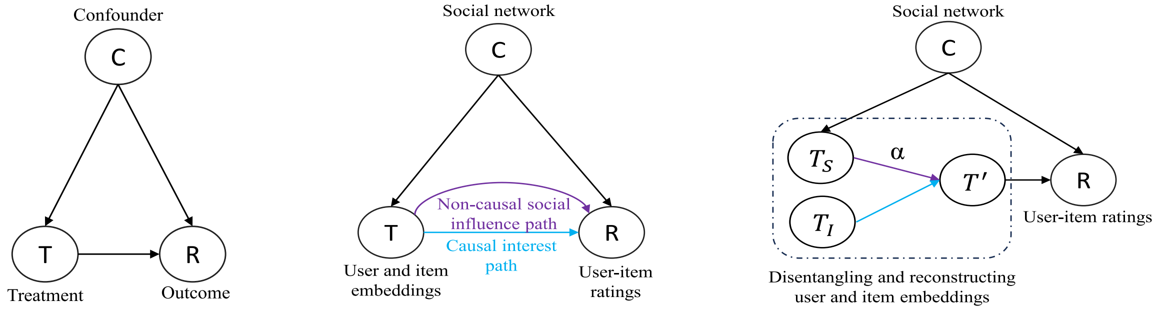
1. Introduction

Social recommendation systems (SR) play a crucial role in addressing the data-sparsity challenge and enhancing recommendation performance by incorporating social network information [1–3]. These systems leverage data from users' social interactions, such as friendships, shared content, and comments, to learn comprehensive and personalized user and item representations. Notable methods like SocialMF [4], GraphRec [5], and DiffNet [2] have proposed innovative models, such as integrating social network information into traditional matrix factorization or utilizing Graph Neural Networks (GNN) to model high-order social relationships.

Although these methods have achieved advancements, they face a critical challenge that the acquired user and item embeddings contain social influence bias, which may not accurately reflect users' genuine interests. This bias originates from the phenomenon where individuals, under the influence of their social circles, might make choices that deviate from their personal preferences. For example, a user who loves classical music but belongs to a social circle where most friends prefer pop music might receive pop music recommendations, which are inconsistent with the user's genuine interests. Recently, some existing methods [6–8] have focused on mitigating social influence bias generated from network information. For instance, SIDR [6] disentangles user and item representations into three latent factors: user interest, item popularity, and user social influence, thereby mitigating the social influence bias. DENC [7] proposes an exposure model and a deconfounding model to effectively control and eliminate social influence bias. Conversely, D2Rec [8] utilizes network information to disentangle user and item representations into exposure, confounder, and prediction factors. It designs a reweighting function to mitigate social influence bias. While SIDR [6] employs causal disentanglement

[*] Corresponding author.
E-mail addresses: li.wang-13@student.uts.edu.au (L. Wang), Min.Xu@uts.edu.au (M. Xu), zhqgui@cqwu.edu.cn (Q. Zhang), Yunxiao.Shi@student.uts.edu.au (Y. Shi), Qiang.Wu@uts.edu.au (Q. Wu).

https://doi.org/10.1016/j.neucom.2024.129133
Received 23 December 2023; Received in revised form 26 November 2024; Accepted 2 December 2024
Available online 9 December 2024



(a) Fork structure of causal graph.

(b) Causal graph of social recommendations.

(c) Causal graph of CDRSB.

Fig. 1. (a) Fork structure of causal graph: confounder affects both the treatment and the outcome. (b) In social recommendations, we treat the social network as a confounder, with user and item embeddings as the treatment and user–item ratings as the outcome. Due to the existence of the social network confounder, there are two paths between user and item embeddings and user–item ratings: a non-causal social influence path and a causal interest path. (c) CDRSB disentangles the user and item embeddings into social influence and interest components and learns dynamic weights of social influence embeddings to fuse them, thereby effectively regulating social influence bias. C: social network, T: user and item embeddings, R: user–item ratings, T_S: social influence embeddings, T_I: interest embeddings, T': reconstructed user and item representations, α: the weight of social influence embeddings.

to separate social influence, it focuses on mitigating social influence bias rather than strategically leveraging it.

Rather than indiscriminately mitigating social influence bias, it is crucial to recognize that not all biases are harmful. Some users perceive recommendations from friends as thoughtful selections, indicating high quality and alignment with their interests. In such cases, the influence exerted by friends has a positive impact on users, and these products are worth recommending. Blindly mitigating this bias may lead to the loss of essential information, preventing the recommendations that align with users' interests. Thus, a dilemma arises: eliminating social influence bias sacrifices meaningful recommendations, while preserving it may lead to undesirable social conformity. Therefore, it is crucial and urgent to propose a method that can reasonably regulate social influence bias to enhance recommendation performance. Such a method should preserve positive social influence bias while mitigating its negative counterpart.

To gain a deeper understanding of how social influence affects recommendations, we introduce the causal graph to analyze this process. The causal graph is a directed acyclic graph (DAG), where nodes represent variables, and edges indicate causal influences from one variable to another. It is used to depict and analyze causal relationships between variables. Motivated by the fork structure of the causal model proposed by Pearl [9], as shown in Fig. 1(a), the confounder is a variable that is related to both the treatment variable (the cause) and the outcome variable (the effect). The presence of a confounder can lead to a spurious correlation between the treatment and outcome variables, making it challenging to establish the causal relationship. In the context of social recommendations, as illustrated in Fig. 1(b), we can regard user and item embeddings as the treatment and user–item ratings as the outcome. The social network emerges as a potential confounder, called social network confounder, since it exerts simultaneous influence on both the user and item embeddings as well as user–item ratings. Due to the existence of the social network confounder, there are two paths from the user and item embeddings (treatment) T and the user–item ratings (outcome) R, including the non-causal social influence path ($C \rightarrow T \rightarrow R$) introduced by the confounder and the causal interest ($T \rightarrow R$) path which represents the reason why a user likes an item. We refer to the bias introduced by the non-causal social influence path as social influence bias. This bias has the undesirable effect of bias amplification because it increases the exposure probability of items that friends interact with, even if these items do not match the user's interests. However, some of the items recommended by friends may align with the user's interests and deserve to be recommended. Therefore, it is crucial to design a method that can effectively regulate social influence bias, preserving its positive effects while mitigating the negative ones.

Based on the above analysis, we propose a causal disentanglement-based framework for regulating social influence bias in social recommendations, named CDRSB, as illustrated in Fig. 1(c). We assume that the treatment (user and item embeddings) can be causally decomposed into two independent embeddings: interest embedding which represents the user's real preferences, and social influence embedding which indicates the social influence bias. We design a disentangled encoder aimed at separating interest and social influence embeddings, along with a regulatory decoder that designs a weight calculation module to reasonably regulate social influence bias. Specifically, for the disentangled encoder, we first learn user and item embeddings via GNN-based learning networks with the user social network and user–item interaction network. Then, we design a causal disentanglement component to separate the user and item embeddings into interest and social influence embeddings. To make these two components independent of each other and contain more semantic information, we introduce mutual information-based objectives. Regarding the regulatory decoder, we first introduce a weight calculation module to learn varying weights of social influence embeddings, which could regulate the social influence bias. These weights are then utilized to fuse interest and social influence embeddings to learn more accurate user and item representations. The main contributions of this work are as follows:

- Based on the non-causal social influence path and the causal interest path introduced by the social network confounder, we propose a disentangled encoder. This encoder disentangles user and item embeddings into interest and social influence embeddings. Mutual information-based objectives are designed to ensure the separation of these disentangled embeddings.
- We propose a regulatory decoder that introduces a weight calculation module to regulate social influence bias and learn more accurate user and item representations, enhancing the model's performance.
- Extensive experiments are conducted on four large-scale real-world datasets Ciao, Epinions, Dianping, and Douban book. The comprehensive results demonstrate the effectiveness of CDRSB compared to state-of-the-art baselines.

The remainder of this work is organized as follows: In Section 2, we provide an overview of the relevant work. Section 3 presents the detailed methods of the CDRSB. We then conduct experiments on four public datasets and compare the results with baselines in Section 4. Finally, in Section 5, we provide a conclusion summarizing our study and outline future work.

2. Related work

In this section, we present several relevant studies related to social recommendation, disentangled representation learning in recommendation, and causal recommendation.

Social Recommendation. In recent years, social recommendation systems have gained significant attention due to the widespread adoption of social media platforms and the increasing number of users. The social network provides rich social relationships and interaction information among users, which can solve the long-standing data-sparsity problem [10–12]. Existing social recommendations could be categorized as matrix factorization (MF)-based [4,13] and graph neural network (GNN)-based approaches [2,5,14,15]. MF-based methods usually jointly factorize the user–item interaction matrix and user-social relationship matrix or add a regularizer to restrain user and item embeddings. SocialMF [4] incorporates the user's social network into traditional collaborative filtering models for recommendation. In [13], authors introduce a social regularization term to incorporate users' social relationships into the recommendation model. In contrast, GNN-based methods utilize the connectivity of graphs to directly model user and item embeddings. GraphRec [5] introduce a graph attention mechanism with both the user social network and the user–item interaction graph to learn user and item embeddings. ConsisRec [14] is an improved version of GraphRec that addresses the social inconsistency problem. To improve the performance of social recommendation, DiffNet [2] and DiffNet++ [15] learn the social influence diffusion process. However, these methods ignore the problem of social influence bias, which may degrade the recommendation performance. Among these approaches, ConsisRec is the most closely related model to ours, as it also aims to aggregate the positive effects of friends. However, our method analyzes the positive and negative influences from the perspective of causal inference, making it more explainable.

Disentangled Representation Learning in Recommendation. Disentangled representation learning has been recognized as an effective way to enhance the robustness and interpretability of models [16,17]. Disentangled representation learning has been applied in generative recommendations [18–20] and graph recommendations [21–23]. Authors in [18] propose a model called MacridVAE, which is a disentangled variational auto-encoder capable of learning representations from user behavior. This model achieves both macro-disentanglement of high-level concepts and micro-disentanglement of isolated low-level factors. DGCF [21] learns disentangled representations that capture fine-grained user intents from the user–item interaction graph. DisenHAN [22] learns disentangled user/item representations from various aspects in a heterogeneous information network, utilizing meta relations to decompose high-order connectivity between node pairs. Recently, disentangled representation learning has been applied to causal recommendation systems [24–26]. DICE [24] constructs cause-specific data according to causal effect and disentangles user and item embeddings into interest and conformity components. DIB [25] mitigates confounding bias by decomposing user and item embeddings into unbiased and biased components via information bottleneck. Authors in [26] utilize user search data to decouple corresponding actual preferences, providing a model-agnostic approach to causal embedding learning in recommendation systems. Nevertheless, these methods may not be suitable for social recommendations where social influence plays an important role in modeling user preferences.

Causal Recommendation. In contrast to traditional recommendation systems that mainly emphasize correlational patterns, causal recommendations delve into the realm of causality. Their objective is to discern and address the causal relationships between user actions and the recommended outcomes [27]. These systems leverage approaches rooted in causal inference, such as inverse propensity weighting [7,28], backdoor adjustment [29,30], frontdoor adjustment [31,32], and counterfactual inference [33,34], to understand and mitigate various biases like confounding bias, selection bias or spurious correlations. For example, AutoDebias [35] leverages uniform data collected by a random logging policy and meta-learning technique to mitigate various biases. Zhang et al. [28] introduce the Multi-IPW model, employing a multi-task learning approach to estimate Inverse Propensity Scores (IPS) and simultaneously mitigate selection bias. DCR [29] introduces the notion of item confounding features and employs backdoor adjustment combined with a mixture-of-experts (MoE) strategy to alleviate spurious correlations arising from them. DCCF [31] utilizes frontdoor adjustment to alleviate confounding bias. Wei et al. [33] focus specifically on countering popularity bias through the application of counterfactual inference. Recently, some causal disentanglement-based methods have been proposed to achieve unbiased recommendations [7,8]. For example, DENC [7] and D2Rec [8] both focus on disentangling three causes: inherent, confounder, and exposure factors, and mitigating bias produced by network information. Nevertheless, these approaches mainly eliminate biases, it is worth noting that certain biases, such as popularity bias and social influence bias, can occasionally be beneficial for learning user preferences [36].

3. Methodology

In this part, we first present the definitions and notations used in this paper and then provide a concise overview of the overall framework. Finally, we introduce each component in detail.

3.1. Definitions and notations

Let $U = \{u_1, u_2, \ldots, u_n\}$ and $V = \{v_1, v_2, \ldots, v_m\}$ denote the user and item sets, where n and m are the number of users and items, respectively. $\mathbf{p}_i \in \mathbb{R}^d$ represents the ID embedding of user u_i and $\mathbf{q}_j \in \mathbb{R}^d$ denotes the ID embedding of item v_j, d is the embedding size. $\mathbf{Y} \in \mathbb{R}^{m\times n}$ represents the user–item rating matrix, where $y_{ij} \in \mathbf{Y}$ denotes the rating given by user u_i to item v_j. Let use $\mathbf{T} \in \mathbb{R}^{n\times n}$ to denote the user social network, where $T_{ij} = 1$ if there is a relation between u_i and u_j. $\mathbf{e}^u_{ij}$ represents the rating embedding of user u_i on item v_j that u_i has interacted with, and $\mathbf{e}^v_{ji}$ denotes the rating embedding of item v_j rated by user u_i. We use $\mathbf{z}^u_i$, $\mathbf{c}^u_i$, $\mathbf{z}^v_j$, and $\mathbf{c}^v_j$ to denote the interest embedding and social influence embedding of user u_i and item v_j, respectively. $\mathbf{x}^u_i$ and $\mathbf{x}^v_j$ represent the embeddings learned by a GNN-based network of user u_i and item v_j. In addition, $\mathbf{h}^u_i$ and $\mathbf{h}^v_j$ denote the reconstructed user and item representations. The mathematical notations used in this paper are summarized in Table 1.

3.2. An overview of the proposed model

We propose a causal disentanglement-based framework for regulating social influence bias in social recommendation, named CDRSB. Fig. 2 shows the overall model architecture.

This framework mainly includes two modules: (1) **Disentangled Encoder**. We aim to disentangle user and item embeddings learned from network information into interest and social influence embeddings. (2) **Regulatory Decoder**. We introduce a weight calculation module to regulate the social influence bias, which could improve the model performance.

3.3. Disentangled encoder

In this section, based on the non-causal social influence path and causal interest path of the causal graph for social recommendations, our focus is on disentangling user and item embeddings into interest and social influence embeddings. Initially, we employ a GNN-based learning network with the user–item interaction graph and user-social graph to learn user and item embeddings. Subsequently, we introduce independent MLP layers to disentangle interest and social influence embeddings.

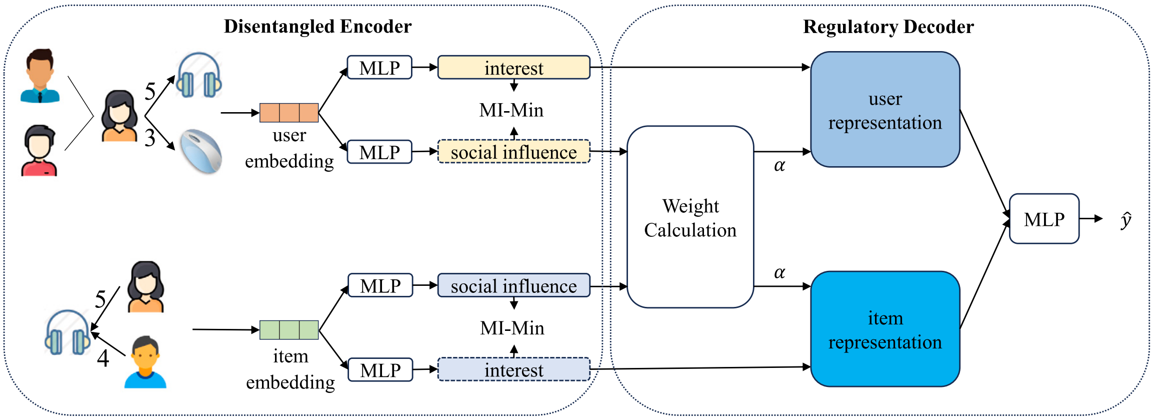


Fig. 2. The overall framework of CDRSB. It contains two modules: (1) a disentangled encoder that disentangles user and item embeddings learned from a GNN-based network into interest and social influence embeddings. We minimize mutual information-based objectives to reduce redundancy and ensure the separation of these disentangled embeddings. (2) a regulatory decoder that learns dynamic weights to combine interest and social influence embeddings into final user and item representations, achieving reasonable utilization of social influence bias.

Table 1
Notations.

Symbols	Definitions and Notations
Y	user–item rating matrix
U	user set
V	item set
T	user social network
$C(i)$	the set of items which user u_i have interacted with
$N(i)$	the set of neighboring users connected to user u_i
$B(j)$	the set of users who interact with item v_j
$\mathbf{p}_i$	ID embedding of user u_i
$\mathbf{q}_j$	ID embedding of item v_j
$\mathbf{x}_i^u$	embedding learned from a GNN-based network of user u_i
$\mathbf{x}_j^v$	embedding learned from a GNN-based network of item v_j
$\mathbf{z}_i^u$	interest embedding of user u_i
$\mathbf{c}_i^u$	social influence embedding of user u_i
$\mathbf{z}_j^v$	interest embedding of item v_j
$\mathbf{c}_j^v$	social influence embedding of item v_j
$\mathbf{e}_{ij}^u$	rating embedding of user u_i for item v_j that u_i has interacted with
$\mathbf{e}_{ji}^v$	rating embedding of item v_j by user u_i
$\mathbf{r}_{ij}^v$	item embedding from item set $C(i)$ of user u_i
$\mathbf{r}_{ji}^u$	user embedding from user set $B(j)$ of item v_j
$\mathbf{t}_{ij}^u$	user embedding from user set $N(i)$ of user u_i
$\mathbf{h}_i^u$	reconstructed representation of user u_i
$\mathbf{h}_j^v$	reconstructed representation of item v_j

3.3.1. *GNN-based learning network*

Motivated by the approach introduced in [5], we design a similar GNN-based method to learn the initial user and item embeddings, which contains two layers: the embedding layer and the concatenation layer.

Embedding layer: We input the user social network and user–item interaction network into the embedding layer to learn the ID embedding, neighboring item embedding, and neighboring user embedding of user u_i. In addition, we also consider each user–item rating in the user–item interaction network, which represents the user's preference level.

$$\begin{aligned}
&\mathbf{p}_i = \delta(\mathbf{W} \cdot \mathbf{o}_i + \mathbf{b}), \\
&\mathbf{r}^v_{ij,\forall j \in C(i)} = \delta(\mathbf{W} \cdot \mathbf{o}^r_{ij} + \mathbf{b}), \\
&\mathbf{e}^u_{ij,\forall j \in C(i)} = \delta(\mathbf{W} \cdot \mathbf{o}^e_{ij} + \mathbf{b}), \\
&\mathbf{t}^u_{ij,\forall j \in N(i)} = \delta(\mathbf{W} \cdot \mathbf{o}^t_{ij} + \mathbf{b}),
\end{aligned} \tag{1}$$

where $\mathbf{p}_i$ and $\mathbf{o}_i$ represent the ID embedding and one-hot vectors of user u_i. $\mathbf{r}^v_{ij}$ and $\mathbf{o}^r_{ij}$ are the embedding and one-hot vectors of item v_j that user u_i has interacted with. $\mathbf{e}^u_{ij}$ and $\mathbf{o}^e_{ij}$ are the rating embedding and one-hot vectors of user u_i for the item v_j that user u_i has interacted with. $\mathbf{t}^u_{ij}$ and $\mathbf{o}^t_{ij}$ are the embedding and one-hot vectors of user u_j that user u_i has connected with. δ denotes the non-linear activation function, $\mathbf{W}$ and $\mathbf{b}$ represent the weight matrix and bias vector, respectively.

Concatenation layer: In this part, we concatenate user ID embedding $\mathbf{p}_i$, neighboring item embedding $\mathbf{r}^v_{ij}$, and rating embedding $\mathbf{e}^u_{ij}$ from the user–item interaction network, as well as the neighboring user embedding $\mathbf{t}^u_{ij}$ from the user-social network to learn the user embedding $\mathbf{x}^u_i$.

$$\begin{aligned}
\mathbf{x}^u_i = \delta(\mathbf{W} \cdot (\mathbf{p}_i &\oplus Agg_{items}(\mathbf{r}^v_{ij}, \forall j \in C(i)) \\
&\oplus Agg_{ratings}(\mathbf{e}^u_{ij}, \forall j \in C(i)) \\
&\oplus Agg_{users}(\mathbf{t}^u_{ij}, \forall j \in N(i))) + \mathbf{b}),
\end{aligned} \tag{2}$$

where Agg_{items} represents the aggregation operation for items, $Agg_{ratings}$ represents the aggregation operation for user–item ratings, and Agg_{users} represents the user aggregation function. We have tried various aggregation methods, including sum aggregation, mean aggregation, and neural network aggregation. Among them, the mean aggregation obtained the best results. $C(i)$ is a set of items that user u_i has interacted with. $N(i)$ is the set of neighboring users trusted by user u_i. $\oplus$ represents the concatenation operation.

Similarly, we utilize the user–item interaction network to learn the item embedding $\mathbf{x}^v_j$.

$$\begin{aligned}
&\mathbf{q}_j = \delta(\mathbf{W} \cdot \mathbf{o}_j + \mathbf{b}), \\
&\mathbf{r}^u_{ji,\forall i \in B(j)} = \delta(\mathbf{W} \cdot \mathbf{o}^r_{ji} + \mathbf{b}), \\
&\mathbf{e}^v_{ji,\forall i \in B(j)} = \delta(\mathbf{W} \cdot \mathbf{o}^e_{ji} + \mathbf{b}),
\end{aligned} \tag{3}$$

$$\begin{aligned}
\mathbf{x}^v_j = \delta(\mathbf{W} \cdot (\mathbf{q}_j &\oplus Agg_{users}(\mathbf{r}^u_{ji}, \forall i \in B(j)) \oplus \\
&Agg_{ratings}(\mathbf{e}^v_{ji}, \forall i \in B(j))) + \mathbf{b}),
\end{aligned} \tag{4}$$

where $B(j)$ is the set of users who interacted with item v_j.

3.3.2. *Causal disentanglement*

In this subsection, we design four independent MLP networks to decompose user and item embeddings into interest and social influence embeddings, minimizing parameter sharing and reducing information redundancy. Specifically, for the user embedding $\mathbf{x}^u_i$, we extract two independent interest and social influence components. We feed $\mathbf{x}^u_i$ into two separate MLP layers,

$$\mathbf{z}^u_i = MLP(\mathbf{x}^u_i; \Theta_0); \qquad \mathbf{c}^u_i = MLP(\mathbf{x}^u_i; \Theta_1). \tag{5}$$

Similarly, the decomposing process of item embedding $\mathbf{x}^v_j$ is as follows,

$$\mathbf{z}^v_j = MLP(\mathbf{x}^v_j; \Theta_2); \qquad \mathbf{c}^v_j = MLP(\mathbf{x}^v_j; \Theta_3), \tag{6}$$

where $\Theta_0, \Theta_1, \Theta_2$ and Θ_3 are parameters for MLP layers.

Despite using separate networks to learn interest and social influence embeddings, we cannot guarantee the absence of redundant information. Therefore, we aim to minimize the mutual information between them to ensure their independence. Mutual information is a statistical measure that quantifies the amount of information one random variable contains about another, indicating the level of dependence between these two variables. High mutual information indicates a strong relationship or dependency between the variables, while low mutual information suggests independence or little shared information.

Traditionally, mutual information is calculated based on examples a_i and b_i sampled from two distributions A and B. However, in our case, the true distributions of A and B are unknown. Inspired by recent advancements in contrastive learning and sample-based mutual information estimation, such as the CLUB framework [37], we estimate mutual information using the difference in conditional probabilities between positive and negative sample pairs. We regard a sample pair with the same index as a positive pair, such as (a_i, b_i), and consider a sample pair with a different index as a negative pair, such as (a_i, b_j).

Since we cannot directly compute the conditional distribution $P(\mathbf{c}_i^u|\mathbf{z}_i^u)$ and $P(\mathbf{c}_j^v|\mathbf{z}_j^v)$, we use variational distributions $Q_{\theta_u}(\mathbf{c}_i^u|\mathbf{z}_i^u)$ and $Q_{\theta_v}(\mathbf{c}_j^v|\mathbf{z}_j^v)$ with parameters θ_u and θ_v, which could be implemented by neural networks, to approximate $P(\mathbf{c}_i^u|\mathbf{z}_i^u)$ and $P(\mathbf{c}_j^v|\mathbf{z}_j^v)$. The output of $Q_{\theta_u}(\mathbf{c}_i^u|\mathbf{z}_i^u)/Q_{\theta_v}(\mathbf{c}_j^v|\mathbf{z}_j^v)$ are mean μ_i^u/μ_j^v and log variance $log(\delta_i^{u2})/log(\delta_j^{v2})$. To update the parameter θ_u and θ_v, we maximize the corresponding conditional log-likelihood loss function,

$$\begin{aligned}
L_u^{LLD} &= \frac{1}{N}\sum_{i=1}^{N} logQ_{\theta_u}(\mathbf{c}_i^u|\mathbf{z}_i^u) \\
&= \frac{1}{N}\sum_{i=1}^{N} -\frac{(\mu_i^u - \mathbf{c}_i^u)^2}{e^{log(\delta_i^{u2})}}; \\
L_v^{LLD} &= \frac{1}{N}\sum_{j=1}^{N} logQ_{\theta_v}(\mathbf{c}_j^v|\mathbf{z}_j^v) \\
&= \frac{1}{N}\sum_{i=1}^{N} -\frac{(\mu_j^v - \mathbf{c}_j^v)^2}{e^{log(\delta_j^{v2})}},
\end{aligned} \tag{7}$$

where N is the number of samples. $\mathbf{z}_i^u$, $\mathbf{c}_i^u$, $\mathbf{z}_j^v$ and $\mathbf{c}_j^v$ represent the interest embedding and social influence embedding of user u_i and item v_j, respectively.

To ensure independence between the decomposed interest and social influence embeddings, we introduce the mutual information loss function and minimize it,

$$\begin{aligned}
L_u^{MI} &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1,k\neq i}^{N}[logQ_{\theta_u}(\mathbf{c}_i^u|\mathbf{z}_i^u) - logQ_{\theta_u}(\mathbf{c}_k^u|\mathbf{z}_i^u))] \\
&= \frac{1}{N}\sum_{i=1}^{N}[logQ_{\theta_u}(\mathbf{c}_i^u|\mathbf{z}_i^u) - \frac{1}{N}\sum_{k=1,k\neq i}^{N} logQ_{\theta_u}(\mathbf{c}_k^u|\mathbf{z}_i^u)] \\
&= \frac{1}{N}\sum_{i=1}^{N}[-\frac{(\mu_i^u - \mathbf{c}_i^u)^2}{e^{log(\delta_i^{u2})}} + \frac{1}{N}\sum_{k=1,k\neq i}^{N}\frac{(\mu_i^u - \mathbf{c}_k^u)^2}{e^{log(\delta_i^{u2})}}]; \\
L_v^{MI} &= \frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1,k\neq j}^{N}[logQ_{\theta_v}(\mathbf{c}_j^v|\mathbf{z}_j^v) - logQ_{\theta_v}(\mathbf{c}_k^v|\mathbf{z}_j^v)] \\
&= \frac{1}{N}\sum_{j=1}^{N}[logQ_{\theta_v}(\mathbf{c}_j^v|\mathbf{z}_j^v) - \frac{1}{N}\sum_{k=1,k\neq j}^{N} logQ_{\theta_v}(\mathbf{c}_k^v|\mathbf{z}_j^v)] \\
&= \frac{1}{N}\sum_{j=1}^{N}[-\frac{(\mu_j^v - \mathbf{c}_j^v)^2}{e^{log(\delta_j^{v2})}} + \frac{1}{N}\sum_{k=1,k\neq j}^{N}\frac{(\mu_j^v - \mathbf{c}_k^v)^2}{e^{log(\delta_j^{v2})}}],
\end{aligned} \tag{8}$$

where $(\mathbf{c}_i^u, \mathbf{z}_i^u)/(\mathbf{c}_j^v, \mathbf{z}_j^v)$ is the positive pair and $(\mathbf{c}_k^u, \mathbf{z}_i^u)/(\mathbf{c}_k^v, \mathbf{z}_j^v)$ is the negative pair.

The total mutual information loss function and variational approximation loss are as follows,

$$\begin{aligned}
L^{MI} &= L_u^{MI} + L_v^{MI}, \\
L^{LLD} &= -L_u^{LLD} - L_v^{LLD}.
\end{aligned} \tag{9}$$

3.4. Regulatory decoder

In traditional causal recommendation methods, social influence bias is often discarded. However, in our method, we argue that social influence bias is individual-specific. The impact of social influence bias can be detrimental when a user interacts with an item recommended by friends, deviating from their genuine interests. Conversely, it can be beneficial when a user engages with an item recommended by friends that is of high quality and aligns with his interests. Blindly eliminating this bias may lead to suboptimal recommendation performance. Therefore, we present a regulatory decoder designed to regulate social influence bias by reasonably incorporating positive social influence embeddings and mitigating the negative ones. This decoder encompasses a dynamic weight calculation module and a fusion module. The weight calculation module dynamically computes the weight of social influence embeddings, while the fusion module integrates interest and social influence embeddings into the final user and item representations.

User-interacted items generally reflect a user's preferences. When users purchase a product recommended by their friends that aligns with their historical preferences, it indicates that the primary reason for the purchase is personal interest. In this case, the influence of friends is beneficial. Conversely, if the purchased item differs significantly from the user's historical preferences, the decision is likely driven by herd mentality rather than genuine interest. Specifically, for a user–item interaction pair (u_i, v_j), where v_j is recommended by friends of user u_i, we hypothesize that the primary reason for u_i interacting with v_j is personal interests if the social influence embedding of v_j is similar to interest embeddings of most items previously interacted with by u_i. In such cases, the impact of friends is beneficial, and we should incorporate positive social influence embeddings. Conversely, it might be attributed to the influence of herd mentality rather than the personal preference, we need to mitigate the negative social influence embeddings. For the user–item interaction pair (u_i, v_j), if user u_i's friends have interacted with the item v_j, we consider v_j to be recommended by u_i's friends.

First, we calculate the cosine similarity, a widely used metric in recommendation methods [38,39], between v_j's social influence embedding $\mathbf{c}_j^v$ and the interest embedding $\mathbf{z}_k^v$ of the kth item v_k previously interacted with by u_i,

$$s_{jk} = f(\mathbf{c}_j^v, \mathbf{z}_k^v) \qquad = \frac{\mathbf{c}_j^v \cdot \mathbf{z}_k^v}{\|\mathbf{c}_j^v\| \| \mathbf{z}_k^v\|}, \tag{10}$$

where f is the cosine similarity function, $\|\mathbf{z}\|$ denotes the vector norm. We can obtain the similarity set $S_j = \{s_{j1}, \dots, s_{jm'}\}$, where m' is the number of interactions.

We then calculate the average similarity score for this set to determine the weight of the social influence embedding,

$$\alpha = \frac{\sum_{k=1,k\neq j}^{m'} s_{jk}}{|S_j|}. \tag{11}$$

The value of α reflects the alignment between the social influence of v_j and u_i's preferences. A higher value indicates that social influence aligns well with u_i's preferences, suggesting that the interaction is driven by personal interest. Conversely, a small value represents that the user dislikes the item v_j, indicating that the interaction is influenced by herd mentality.

It is worth noting that within the user–item interaction pair (u_i, v_j), if the item v_j is not recommended by the friends of user u_i, we assume that u_i clicked on item v_j based on personal preferences, entirely independent of any influence from the friends of user u_i. Consequently, we set the weight α to 0. By leveraging α, the model can effectively balance the incorporation of positive social influence and mitigate the negative impact of herd mentality.

Subsequently, a more accurate user representation $\mathbf{h}_i^u$ and item representation $\mathbf{h}_j^v$ can be learned as follows,

$$\mathbf{h}_i^u = \mathbf{z}_i^u + \alpha \mathbf{c}_i^u, \\ \mathbf{h}_j^v = \mathbf{z}_j^v + \alpha \mathbf{c}_j^v. \tag{12}$$

After reconstructing the user representation $\mathbf{h}_i^u$ and item representation $\mathbf{h}_j^v$, we concatenate them and put them into the final prediction layers,

$$\hat{y}_{ij} = MLP(\mathbf{h}_i^u \oplus \mathbf{h}_j^v). \tag{13}$$

We aim to minimize the following loss function,

$$L^O = \frac{1}{N} \sum_{i,j=1}^{N} l(y_{ij} - \hat{y}_{ij}), \tag{14}$$

where N is the batch size, l denotes the mean squared error in the rating prediction task and the cross-entropy loss function in the ranking task.

Finally, we optimize all parameters by minimizing the final loss function L,

$$L = L^O + \lambda(L^{MI} + L^{LLD}), \tag{15}$$

where λ is the weight parameter.

4. Experiments

In this section, we conduct a series of comprehensive experiments on four publicly available datasets to evaluate the performance of our proposed model, CDRSB, in the rating prediction and ranking tasks. We aim to answer the following questions.

- RQ1: Does our model achieve superior performance compared to other state-of-the-art baseline methods?
- RQ2: How do different components, such as mutual information-based objectives and social influence embeddings, affect the outcomes of our model?
- RQ3: Are the interest and social influence embeddings we have acquired genuinely disentangled?
- RQ4: How does our model's performance vary with different hyper-parameters?
- RQ5: What is the primary reason for user interaction with an item?

4.1. Experimental settings

4.1.1. Datasets

We conduct experiments on four large-scale datasets: Ciao,[1] Epinions[1], Dianping,[2] and Douban book.[3] These datasets comprise user–item ratings along with user trust relationships. The Ciao and Epinions datasets are collected from popular social websites Ciao,[4] and Epinions[5] where users have the ability to rate items and establish social connections by adding friends. The Dianping dataset is collected from a leading local restaurant search and review platform in China.[6] This dataset is crawled by authors in [40]. The Douban book dataset is extracted from a Chinese book forum.[7] The rating scale in all datasets ranges from 1 to 5. For the ranking task, we discretize the ratings into binary values of 0 and 1 to indicate whether the user has interacted with the item or not. To ensure data quality and address sparsity issues, we apply filtering criteria to remove records with insufficient interactions. For the Epinions dataset, we delete records with fewer than five interactions. For the Ciao dataset, we remove samples with less than three interactions between users and items. For the Dianping and Douban book datasets, samples involving less than ten interactions between users and items are omitted. Given the substantial data volume in both Dianping and Douban book datasets, we employ random sampling to enhance training efficiency. Table 2 provides an overview of the basic statistics of these datasets.

Table 2
Statistic of the datasets: Ciao, Epinions, Dianping and Douban book.

Datasets	Ciao	Epinions	Dianping	Douban book
# Users	7108	20 461	20 000	7000
# Items	21 978	31 678	9511	16 421
# Ratings	184 960	545 861	725 637	443 334
Rating Density	0.119%	0.084%	0.380%	0.386%
# Social Connections	53 019	311 235	77 146	11 267
Social Connection Density	0.105%	0.074%	0.019%	0.023%

Table 3
Summary of baselines for rating prediction and ranking tasks.

	Rating Prediction	Ranking
Traditional recommendations	NeuMF PMF	NeuMF LightGCN
Social recommendations	SocialMF GraphRec ConsisRec	DiffNet DiffNet++
Causal recommendations	CausE D2Rec IPS-MF	DICE D2Rec SIDR

4.1.2. Evaluation metrics

To assess the recommendation performance of the CDRSB model and baselines, we depend on the following metrics:

Rating Prediction Metrics. We utilize two commonly used evaluation metrics: root mean squared error (RMSE) and mean absolute error (MAE). These metrics are widely employed in collaborative prediction algorithms [41]. RMSE measures the square root of the average squared difference between the predicted ratings $\hat{y}_i$ and the true ratings y_i for a set of N instances. A lower RMSE indicates better accuracy in predicting ratings. MAE calculates the average absolute difference between the predicted ratings $\hat{y}_i$ and the true ratings y_i.

Ranking Metrics. We utilize two widely adopted metrics, Hit Rate (HR) and Normalized Discounted Cumulative Gain (NDCG), to assess the ranking performance. HR measures the proportion of samples where a user-interacted item appears within the top-K recommended items. NDCG considers both the relevance and the position of the recommended items, assigning higher weights to items ranked higher in the list.

4.1.3. Baseline methods

To verify the effectiveness of our model, we compare the performance of CDRSB with three sets of baselines: traditional recommendation systems, social recommendation systems, and causal recommendation systems. We carefully selected several of the most representative methods from each category. This comprehensive comparison allows us to assess the relative performance of CDRSB against different types of recommendation approaches and gain insights into its strengths and advantages.

Furthermore, distinct sets of baselines are employed for the rating prediction and ranking tasks, as outlined in Table 3.

[1] http://www.cse.msu.edu/~tangjili/trust.html
[2] https://lihui.info/data/dianping/
[3] https://www.dropbox.com/s/u2ejjezjk08lz1o/Douban.tar.gz?e=1&dl=0
[4] http://www.ciao.co.uk
[5] http://www.epinions.com
[6] https://www.dianping.com
[7] https://book.douban.com/

- **Traditional recommendation systems**
  - **PMF** [42] is a traditional recommendation method that decomposes the user–item rating matrix into low-dimensional latent feature matrices. It learns the relationships among these latent features for rating prediction.
  - **NeuMF** [43] combines collaborative filtering and neural networks to capture complex user–item interaction relationships. For the rating prediction task, we modify its loss function to the square loss.
  - **LightGCN** [44] is a simple GCN model that directly propagates user and item embeddings through the user–item interaction graph without introducing complex operations or auxiliary information.
- **Social recommendation systems**
  - **SocialMF** [4] incorporates user trust information into a matrix factorization model, leveraging user social relationships to infer user ratings for items.
  - **GraphRec** [5] proposes a framework for social recommendation that leverages Graph Neural Networks to learn user and item representations with user–item interactions and user social networks.
  - **ConsisRec** [14] is an enhanced method based on GraphRec that addresses the issue of social inconsistency by leveraging consistent neighbor aggregation.
  - **DiffNet** [2] proposes a deep influence propagation model to simulate how users are influenced by the recursive social diffusion process for SR.
  - **DiffNet++** [15] is an improved model based on DiffNet, incorporating the modeling of interest diffusion with a user–item graph.
- **Causal recommendation systems**
  - **IPS-MF** [45] utilizes the inverse propensity score (IPS) to mitigate the selection bias.
  - **CausE** [46] proposes a domain adaptation method that trains the model by utilizing biased data and predicts results based on random exposure.
  - **DICE** [24] first disentangles user and item embeddings into interest and conformity with cause-specific data and then eliminates the confounding effect of popularity bias.
  - **D2Rec** [8] disentangles user and item representations into inherent, confounder, and exposure factors, and then mitigates social influence bias by a reweighting function.
  - **SIDR** [6] causally disentangles the user and item latent features to mitigate social influence bias in implicit feedback for social recommendation.

4.1.4. Parameter settings

We implement the CDRSB model using Python with the Pytorch framework, all baseline methods are conducted based on their GitHub source code and carefully adjusted the hyperparameters. We randomly split datasets into training, test, and validation sets according to an 8:1:1 ratio. The optimal hyperparameters are obtained by optimizing the loss function (15) using the RMSprop optimizer. Based on the validation set, we evaluate the performance of the model using different parameter combinations. The embedding size of original embedding, decomposing embedding, and batch size are searched within the range of [8, 16, 32, 64, 128, 256]. Ultimately, we set them to 64, 64, and 128, respectively. We conduct tests with different learning rates [0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1] and λ values [0.0001, 0.001, 0.01, 0.1]. We determine that the optimal learning rate is 0.0001, and we set λ to 0.001. To prevent overfitting, we apply batch normalization, dropout, and early stopping techniques where the training is stopped when the test evaluation metrics increase for 5 epochs.

4.2. Results and analysis (RQ1)

We evaluate the performance of CDRSB and the baselines using commonly used evaluation metrics for rating prediction task w.r.t. RMSE and MAE and ranking task w.r.t. HR@10 and NDCG@10 on four datasets. The results for each task are shown in Table 4 and Table 5.

- Our model, CDRSB, outperforms other baselines in terms of metrics for rating prediction (RMSE and MAE) and ranking (HR@10 and NDCG@10) on four datasets. CDRSB demonstrates superior performance, achieving average improvements of up to 33.11% and 22.48% in terms of HR@10 and NDCG@10 over the best baseline model LightGCN in traditional recommendation methods. Furthermore, it outperforms the best baseline model DiffNet++ by an average of 29.04% and 17.69% in terms of HR@10 and NDCG@10 in social recommendation methods. Additionally, CDRSB exhibits better performance than the best baseline model D2Rec, with improvements of 7.18% and 7.64% in terms of RMSE and MAE in causal recommendation methods. The results indicate the effectiveness and rationality of CDRSB.
- Compared to D2Rec and SIDR, which mitigate social influence bias, CDRSB achieves the highest performance. This demonstrates that properly regulating social influence bias could improve the model's overall performance.
- Social recommendation methods based on causal debiasing, such as D2Rec and SIDR, outperform traditional social recommendation methods across all evaluation metrics. This emphasizes the significance of mitigating social influence bias in enhancing recommendation performance.
- SocialMF, GraphRec, and ConsisRec, which incorporate social network information, outperform traditional recommendation models NeuMF and PMF. This is attributed to the fact that social networks can complement user preferences, particularly when user features are sparse.
- Both NeuMF and PMF utilize rating information for recommendations. However, NeuMF outperforms PMF, indicating the superior learning capability of deep learning models in recommendation systems.
- GraphRec outperforms SocialMF by an average of 8.63% and 12.12% in terms of RMSE and MAE across four datasets. These results highlight the advantages of GNN and the incorporation of rating information into the learning process of user and item embeddings.

4.3. Ablation studies (RQ2)

To evaluate the effectiveness of each component in CDRSB, we conduct ablation experiments on four datasets. We create three variants of CDRSB, denoted as w/o wt, w/o sl, and w/o mi by removing specific components.

- w/o wt: It deletes the weight calculation module, thereby fixing the weights of social influence embeddings at 1.
- w/o sl: It removes user and item social influence embeddings and only uses user and item interest embeddings for recommendation.
- w/o mi: It eliminates the mutual information minimization objective from the joint loss.

The results of the ablation studies for rating prediction and ranking tasks are presented in Table 6 and Table 7.

Based on the above results, we can observe that each component of the overall model plays a crucial role. The model w/o mi, which eliminates mutual information restraints, experiences a significant drop in performance. This outcome might be attributed to the fact that mutual information minimization ensures that the decoupled interest embedding and social influence embedding have non-redundant information, leading to enhanced performance. Comparing CDRSB with

Table 4
Overall performance comparison for the rating prediction task. The optimal performance is highlighted using bold fonts, and the second-best performance is denoted by underlines.

Method type	Method	Epinions		Ciao		Dianping		Douban book	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Traditional recommendations	PMF	1.2905	1.0203	1.1309	0.9107	1.0260	0.8530	1.0176	0.8425
	NeuMF	1.1290	0.9040	1.0713	0.8145	0.9606	0.7887	0.9513	0.7802
Social recommendations	SociaMF	1.0934	0.8442	1.0534	0.8223	0.9512	0.7945	0.9428	0.7845
	GraphRec	1.0657	0.8134	0.9978	0.7523	0.8205	0.6023	0.8116	0.5928
	ConsisRec	1.0467	0.8096	0.9823	0.7436	0.8073	0.5796	0.8025	0.5768
Causal recommendations	IPS-MF	1.1023	0.8813	1.0423	0.7904	0.9649	0.7514	0.9584	0.7461
	CausE	1.1145	0.8956	1.0378	0.7810	0.8546	0.6481	0.8475	0.6362
	D2Rec	1.0223	0.8545	0.9394	0.7103	0.7256	0.5478	0.7124	0.5405
	CDRSB	**0.9396**	**0.7109**	**0.8286**	**0.5964**	**0.6754**	**0.5117**	**0.6688**	**0.5284**
	Imp.%	↑ 8.27%	↑ 9.87%	↑ 11.08%	↑ 11.39%	↑ 5.02%	↑ 3.61%	↑ 4.36%	↑ 1.21%

Table 5
Overall performance comparison for ranking task. The optimal performance is highlighted using bold fonts, and the second-best performance is denoted by underlines.

Method type	Method	Epinions		Ciao		Dianping		Douban book	
		HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
Traditional recommendations	NeuMF	0.3815	0.2389	0.3393	0.2231	0.4123	0.2701	0.4163	0.2572
	LightGCN	0.4013	0.2527	0.3456	0.2367	0.4234	0.2825	0.4389	0.2826
Social recommendations	DiffNet	0.4264	0.2756	0.3589	0.2501	0.4495	0.3026	0.4527	0.3009
	DiffNet++	0.4479	0.3016	0.3670	0.2704	0.4726	0.3313	0.4848	0.3428
Causal recommendations	DICE	0.6735	0.3902	0.5572	0.3301	0.6972	0.4231	0.7026	0.4237
	D2Rec	0.6703	0.3876	0.5674	0.3356	0.7004	0.4294	0.7128	0.4162
	SIDR	0.6824	0.4036	0.5863	0.3528	0.7183	0.4368	0.7345	0.4596
	CDRSB	**0.7721**	**0.5204**	**0.6547**	**0.4152**	**0.7187**	**0.4675**	**0.7884**	**0.5504**
	Imp.%	↑ 8.97%	↑ 11.68%	↑ 6.84%	↑ 6.24%	↑ 0.04%	↑ 3.07%	↑ 5.39%	↑ 9.08%

Table 6
Results of the ablation studies on the rating prediction task.

Method	Epinions		Ciao		Dianping		Douban book	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
w/o wt	0.9529	0.7285	0.8487	0.616	0.6929	0.5315	0.6829	0.5536
w/o sl	0.9495	0.7228	0.8474	0.6087	0.6878	0.5267	0.6789	0.5408
w/o mi	0.9458	0.7165	0.8429	0.6042	0.6820	0.5223	0.6747	0.5364
CDRSB	**0.9396**	**0.7109**	**0.8286**	**0.5964**	**0.6754**	**0.5117**	**0.6688**	**0.5284**

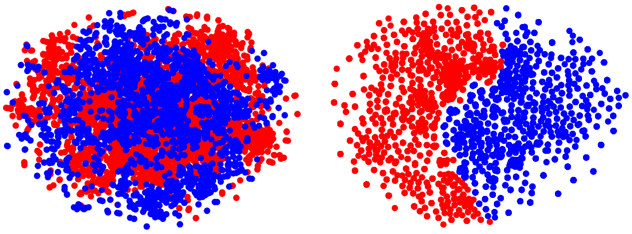

(a) Initialization (b) Convergence

Fig. 3. Visualization of user's interest embedding (red points) and social influence embedding (blue points) for different stages: (a) initialization, (b) convergence on the dataset Epinions.

the model w/o wt, where negative social influence embeddings are incorporated, it becomes evident that the weight calculation module plays a pivotal role in effectively controlling social influence bias. Moreover, the inferior performance of the model w/o sl emphasizes the significant contributions of the positive social influence embeddings to the final outcome.

4.4. Visualization of disentangled embeddings (RQ3)

In this section, we aim to gain deeper insights into the impact of the disentangled encoder on the representative learning process in CDRSB. Specifically, we investigate whether the interest embedding and social influence embedding become independent of each other during the training process. We employ t-SNE [47], a data visualization technique that projects high-dimensional data into a lower-dimensional space, to visualize this phenomenon. The disentangled interest embedding and social influence embedding for the rating prediction task on the Epinions dataset are visualized in Fig. 3 and Fig. 4.

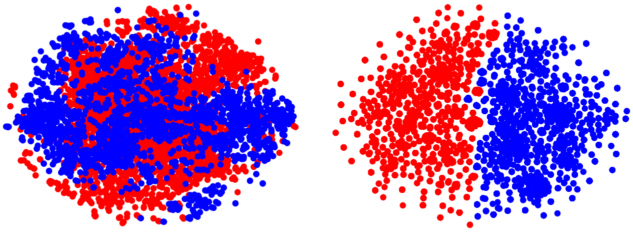

(a) Initialization (b) Convergence

Fig. 4. Visualization of item's interest embedding (red points) and social influence embedding (blue points) for different stages: (a) initialization, (b) convergence on the dataset Epinions.

We observe that with an increasing number of model training iterations, a distinction emerges between the interest embedding represented in red and the social influence embedding shown in blue. This clear separation validates the effectiveness of the disentangled encoder in disentangling and distinguishing the interest and social influence within the embeddings.

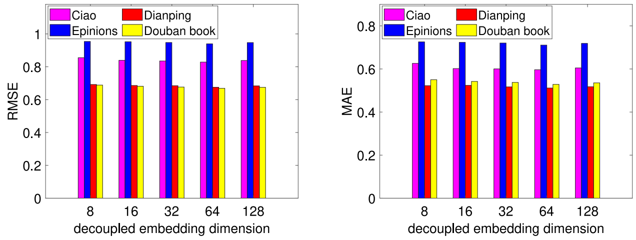
4.5. Parameter sensitivity (RQ4)

In this section, we evaluate the performance of CDRSB under various settings of two crucial parameters: the decoupling embedding dimension and the weight parameter λ.

- **The impact of decoupling embedding dimension.** For the rating prediction and ranking tasks, we present the comparative results in Fig. 5 and Fig. 6, respectively. It is observed that optimal performance is achieved when the dimension is set to

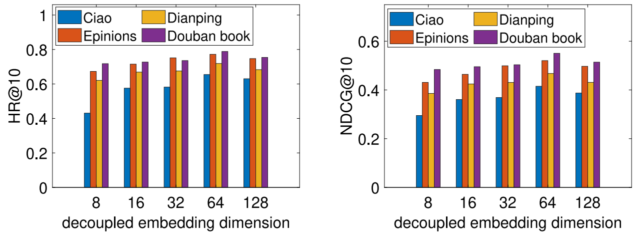
Table 7
Results of the ablation studies on the ranking task.

Method	Epinions		Ciao		Dianping		Douban book	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
w/o wt	0.7424	0.4933	0.6063	0.3656	0.6729	0.4238	0.7526	0.5026
w/o sl	0.7390	0.4871	0.6068	0.3685	0.6697	0.4154	0.7434	0.4861
w/o mi	0.7447	0.4935	0.6128	0.3733	0.6765	0.4268	0.7553	0.5106
CDRSB	**0.7721**	**0.5204**	**0.6547**	**0.4152**	**0.7187**	**0.4675**	**0.7884**	**0.5504**



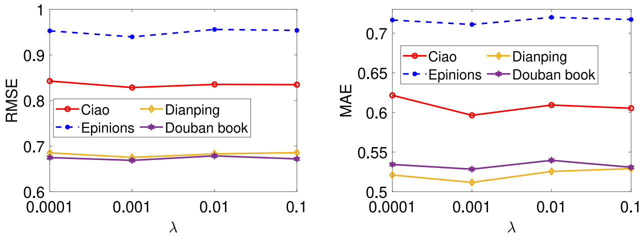
(a) (b)

Fig. 5. RMSE and MAE on different decoupled embedding dimensions: (a) RMSE (b) MAE.



(a) (b)

Fig. 6. HR@10 and NDCG@10 on different decoupled embedding dimensions: (a) HR@10 (b) NDCG@10.



(a) (b)

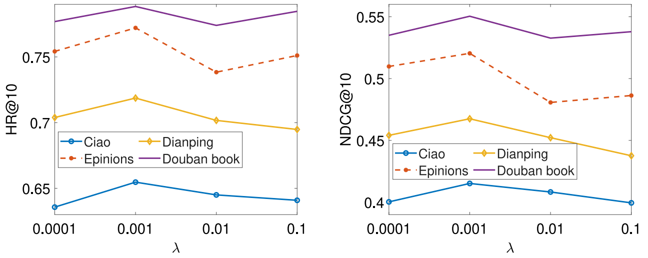
Fig. 7. RMSE and MAE on different λ: (a) RMSE (b) MAE.

64. Increasing the dimension improves the model's effectiveness. However, excessively large dimensions may lead to overfitting. Thus, choosing an appropriate decoupled embedding dimension is crucial for balancing model complexity and performance.

- **The impact of weight parameter** λ**.** The comparative results for rating prediction and ranking tasks are shown in Fig. 7 and Fig. 8. It is evident that CDRSB obtains the optimal result when λ is set to 0.001. This indicates that a moderate weight parameter strikes a balance between incorporating the mutual information minimization objective and preserving the overall model performance.

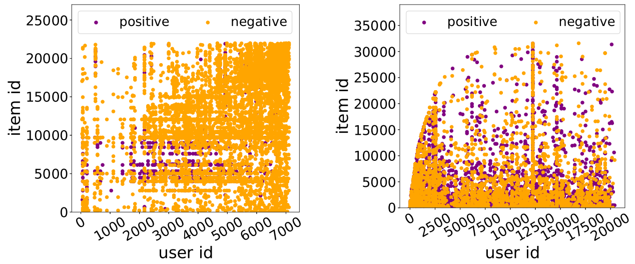
4.6. Case study (RQ5)

In this section, we delve into the motivations behind user–item interactions, exploring whether they primarily stem from individual preferences or are influenced by herd mentality. We randomly sample 1000 users who received recommendations from friends across all datasets and subsequently visualize the different effects of social influence embeddings on each user–item interaction pair for the rating prediction task, as shown in Figs. 9 and 10.



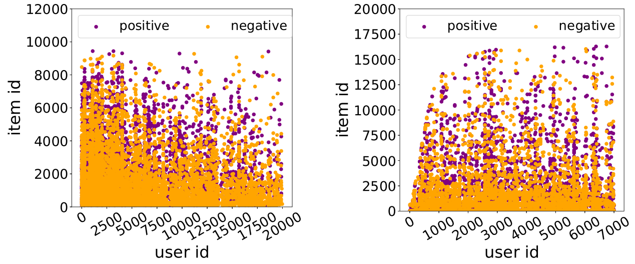
(a) (b)

Fig. 8. HR@10 and NDCG@10 on different λ: (a) HR@10 (b) NDCG@10.



(a) (b)

Fig. 9. The main reason for user–item interactions: (a) Ciao (b) Epinions.



(a) (b)

Fig. 10. The main reason for user–item interactions: (a) Dianping (b) Douban book.

In Figs. 9 and 10, the purple points denote that the social embeddings have a positive effect, implying that recommendations from friends align with users' interests, showcasing interactions driven by personal preferences. Conversely, the yellow points signify a negative effect of social influence embeddings, indicating recommendations from friends that diverge from users' interests and reflect interactions influenced by herd mentality.

5. Conclusion and future work

In this paper, we propose a causal disentanglement-based framework for regulating social influence bias in social recommendation, named CDRSB. First, due to the existence of the social network confounder, there are two paths between user and item embeddings and user–item ratings in SR: a non-causal social influence path and a causal interest path. Therefore, we propose a disentangled encoder that decomposes user and item embeddings into interest and social influence embeddings. We leverage mutual information minimization techniques to ensure the independence of these two components from each other. Next, we present a regulatory decoder to regulate social influence bias and integrate interest and social influence embeddings into more accurate user and item representations. Experimental results on four datasets Ciao, Epinions, Dianping, and Douban book demonstrate the effectiveness of our proposed model CDRSB and its various components.

In the future, we plan to explore the following three directions. Firstly, we will investigate other effective disentanglement methods that can enhance the separation of user preferences and social influence bias. Secondly, we aim to incorporate additional side information, such as user search data or review texts, to assist in decomposing true preferences and social influence more accurately. Additionally, we intend to further disentangle users' interests into finer-grained factors, such as clothing color, style, and price.

CRediT authorship contribution statement

Li Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Min Xu:** Writing – review & editing. **Quangui Zhang:** Writing – review & editing. **Yunxiao Shi:** Writing – review & editing, Data curation. **Qiang Wu:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] W. Fan, Q. Li, M. Cheng, Deep modeling of social relations for recommendation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, No. 1, 2018.

[2] L. Wu, P. Sun, Y. Fu, R. Hong, X. Wang, M. Wang, A neural influence diffusion model for social recommendation, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019, pp. 235–244.

[3] P. Zhu, D. Cheng, S. Luo, F. Yang, Y. Luo, W. Qian, A. Zhou, SI-News: Integrating social information for news recommendation with attention-based graph convolutional network, Neurocomputing 494 (2022) 33–42.

[4] M. Jamali, M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks, in: Proceedings of the Fourth ACM Conference on Recommender Systems, 2010, pp. 135–142.

[5] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, D. Yin, Graph neural networks for social recommendation, in: The World Wide Web Conference, 2019, pp. 417–426.

[6] P. Sheth, R. Guo, L. Cheng, H. Liu, K.S. Candan, Causal disentanglement for implicit recommendations with network information, ACM Trans. Knowl. Discov. Data 17 (7) (2023) 1–18.

[7] Q. Li, X. Wang, Z. Wang, G. Xu, Be causal: De-biasing social network confounding in recommendation, ACM Trans. Knowl. Discov. Data 17 (1) (2023) 1–23.

[8] P. Sheth, R. Guo, K. Ding, L. Cheng, K.S. Candan, H. Liu, Causal disentanglement with network information for debiased recommendations, in: International Conference on Similarity Search and Applications, Springer, 2022, pp. 265–273.

[9] J. Pearl, et al., Models, Reasoning and Inference, vol. 19, (2) CambridgeUniversityPress, Cambridge, UK, 2000, p. 3.

[10] H. Ma, H. Yang, M.R. Lyu, I. King, Sorec: social recommendation using probabilistic matrix factorization, in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, 2008, pp. 931–940.

[11] X. Wang, W. Pan, C. Xu, Hgmf: Hierarchical group matrix factorization for collaborative recommendation, in: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, 2014, pp. 769–778.

[12] J. Chen, C. Wang, S. Zhou, Q. Shi, Y. Feng, C. Chen, Samwalker: Social recommendation with informative sampling strategy, in: The World Wide Web Conference, 2019, pp. 228–239.

[13] H. Ma, D. Zhou, C. Liu, M.R. Lyu, I. King, Recommender systems with social regularization, in: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 2011, pp. 287–296.

[14] L. Yang, Z. Liu, Y. Dou, J. Ma, P.S. Yu, Consisrec: Enhancing gnn for social recommendation via consistent neighbor aggregation, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 2141–2145.

[15] L. Wu, J. Li, P. Sun, R. Hong, Y. Ge, M. Wang, Diffnet++: A neural influence and interest diffusion network for social recommendation, IEEE Trans. Knowl. Data Eng. 34 (10) (2020) 4753–4766.

[16] X. Wang, H. Chen, S. Tang, Z. Wu, W. Zhu, Disentangled representation learning, 2022, arXiv preprint arXiv:2211.11695.

[17] J. Wu, X. Li, X. Ao, Y. Meng, F. Wu, J. Li, Improving robustness and generality of nlp models using disentangled representations, 2020, arXiv preprint arXiv:2009.09587.

[18] J. Ma, C. Zhou, P. Cui, H. Yang, W. Zhu, Learning disentangled representations for recommendation, Adv. Neural Inf. Process. Syst. 32 (2019).

[19] C.P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, A. Lerchner, Understanding disentangling in β-VAE, 2018, arXiv:1804.03599.

[20] D. Bouchacourt, R. Tomioka, S. Nowozin, Multi-level variational autoencoder: Learning disentangled representations from grouped observations, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, No. 1, 2018.

[21] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, T.-S. Chua, Disentangled graph collaborative filtering, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 1001–1010.

[22] Y. Wang, S. Tang, Y. Lei, W. Song, S. Wang, M. Zhang, Disenhan: Disentangled heterogeneous graph attention network for recommendation, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 1605–1614.

[23] A. Li, Z. Cheng, F. Liu, Z. Gao, W. Guan, Y. Peng, Disentangled graph neural networks for session-based recommendation, IEEE Trans. Knowl. Data Eng. (2022).

[24] Y. Zheng, C. Gao, X. Li, X. He, Y. Li, D. Jin, Disentangling user interest and conformity for recommendation with causal embedding, in: Proceedings of the Web Conference 2021, 2021, pp. 2980–2991.

[25] D. Liu, P. Cheng, H. Zhu, Z. Dong, X. He, W. Pan, Z. Ming, Mitigating confounding bias in recommendation via information bottleneck, in: Proceedings of the 15th ACM Conference on Recommender Systems, 2021, pp. 351–360.

[26] Z. Si, X. Han, X. Zhang, J. Xu, Y. Yin, Y. Song, J.-R. Wen, A model-agnostic causal learning framework for recommendation using search data, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 224–233.

[27] C. Gao, Y. Zheng, W. Wang, F. Feng, X. He, Y. Li, Causal inference in recommender systems: A survey and future directions, 2022, arXiv preprint arXiv:2208.12397.

[28] W. Zhang, W. Bao, X.-Y. Liu, K. Yang, Q. Lin, H. Wen, R. Ramezani, Large-scale causal approaches to debiasing post-click conversion rate estimation with multi-task learning, in: Proceedings of the Web Conference 2020, 2020, pp. 2775–2781.

[29] X. He, Y. Zhang, F. Feng, C. Song, L. Yi, G. Ling, Y. Zhang, Addressing confounding feature issue for causal recommendation, ACM Trans. Inf. Syst. 41 (3) (2023) 1–23.

[30] W. Wang, F. Feng, X. He, X. Wang, T.-S. Chua, Deconfounded recommendation for alleviating bias amplification, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 1717–1725.

[31] S. Xu, J. Tan, S. Heinecke, V.J. Li, Y. Zhang, Deconfounded causal collaborative filtering, ACM Trans. Recomm. Syst. 1 (4) (2023) 1–25.

[32] X. Zhu, Y. Zhang, F. Feng, X. Yang, D. Wang, X. He, Mitigating hidden confounding effects for causal recommendation, 2022, arXiv preprint arXiv:2205.07499.

[33] T. Wei, F. Feng, J. Chen, Z. Wu, J. Yi, X. He, Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 1791–1800.

[34] M. He, C. Li, X. Hu, X. Chen, J. Wang, Mitigating popularity bias in recommendation via counterfactual inference, in: International Conference on Database Systems for Advanced Applications, Springer, 2022, pp. 377–388.

[35] J. Chen, H. Dong, Y. Qiu, X. He, X. Xin, L. Chen, G. Lin, K. Yang, AutoDebias: Learning to debias for recommendation, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 21–30.

[36] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, X. He, Bias and debias in recommender system: A survey and future directions, ACM Trans. Inf. Syst. 41 (3) (2023) 1–39.

[37] P. Cheng, W. Hao, S. Dai, J. Liu, Z. Gan, L. Carin, Club: A contrastive log-ratio upper bound of mutual information, in: International Conference on Machine Learning, PMLR, 2020, pp. 1779–1788.

[38] J. Lu, G. Sun, X. Fang, J. Yang, W. He, A contrastive learning framework for dual-target cross-domain recommendation, in: Proceedings of the 31st ACM International Conference on Multimedia, ACM, Ottawa ON Canada, 2023, pp. 6332–6339, http://dx.doi.org/10.1145/3581783.3612250.

[39] C. Zhao, H. Zhao, M. He, J. Zhang, J. Fan, Cross-domain recommendation via user interest alignment, in: Proceedings of the ACM Web Conference 2023, 2023, pp. 887–896.

[40] H. Li, D. Wu, W. Tang, N. Mamoulis, Overlapping community regularization for rating prediction in social recommender systems, in: Proceedings of the 9th ACM Conference on Recommender Systems, 2015, pp. 27–34.

[41] H. Wang, N. Wang, D.-Y. Yeung, Collaborative deep learning for recommender systems, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1235–1244.

[42] A. Mnih, R.R. Salakhutdinov, Probabilistic matrix factorization, Adv. Neural Inf. Process. Syst. 20 (2007).

[43] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural collaborative filtering, in: Proceedings of the 26th International Conference on World Wide Web, 2017, pp. 173–182.

[44] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, M. Wang, Lightgcn: Simplifying and powering graph convolution network for recommendation, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 639–648.

[45] D. Liang, L. Charlin, D.M. Blei, Causal inference for recommendation, in: Causation: Foundation to Application, Workshop At UAI. AUAI, 2016.

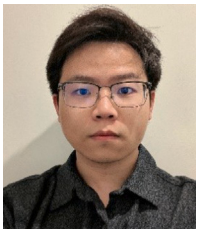
[46] S. Bonner, F. Vasile, Causal embeddings for recommendation, in: Proceedings of the 12th ACM Conference on Recommender Systems, 2018, pp. 104–112.

[47] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).

Li Wang is currently a Ph.D. in University of Technology Sydney. Her research interests include machine learning, data mining and recommendation systems.

Min Xu received the B.Eng. from University of Science and Technology of China(USTC), the M.Comp. from National University of Singapore (NUS) and the Ph.D. degree from the University of Newcastle, Australia. She is a Professor at the School of Electrical and Data Engineering, University of Technology Sydney (UTS) and currently the Leader of Visual and Aural Intelligence Laboratory within the Global Big Data Technologies Center at UTS. Min is a researcher in the fields of multimedia, computer vision and machine learning. She has published 200+ research papers in prestigious international journals and conference proceedings, including IEEE T-PAMI, IEEE T-NNLS, IEEE T-MM, IEEE T-MC, PR, ICLR, ICML, CVPR, ICCV, ACM MM, AAAI and so on. She is an editorial board member for Elsevier Journal of Neurocomputing and Journal of Ambient Intelligence and Humanised Computing; and has served in various chair roles for many major conferences.

Quangui Zhang received the Ph.D. degree in Beijing University of Technology in China. He is an associate professor at Chongqing University of Arts and Sciences, China. His research interests include Machine Learning and Recommender Systems.



Yunxiao Shi is an enthusiastic researcher specializing in recommender systems. His work primarily focuses on streaming recommendation, incremental learning, and the integration of large language models into recommender systems. Yunxiao is also interested in addressing the challenges of learning with noise labels, aiming to enhance the data quality of recommendation model's training.

Qiang Wu received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998, respectively, and the Ph.D. degree from the University of Technology Sydney, Australia, in 2004. He is currently an Associate Professor and a Core Member of the Global Big Data Technologies Centre, University of Technology Sydney. His research interests include computer vision, image processing, pattern recognition, machine learning, and multimedia processing. The application fields where the research outcomes are applied span over video security surveillance, biometrics, video data analysis, and human– computer interaction. His research outcomes have been published in many premier international conferences, including ECCV, CVPR, ICCV, ICIP, and ICPR and the major international journals, such as the IEEE TIP, IEEE TSMC-B, IEEE TCSVT, IEEE TIFS, PR, PRL, and Signal Processing.