# DisenCDR: Learning Disentangled Representations for Cross-Domain Recommendation

Jiangxia Cao
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
caojiangxia@iie.ac.cn

Xixun Lin
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
linxixun@iie.ac.cn

Xin Cong
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
congxin@iie.ac.cn

Jing Ya
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
yajing@iie.ac.cn

Tingwen Liu*
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
liutingwen@iie.ac.cn

Bin Wang
Xiaomi AI Lab,
Xiaomi Inc
wangbin11@xiaomi.com

## ABSTRACT

Data sparsity is a long-standing problem in recommender systems. To alleviate it, Cross-Domain Recommendation (CDR) has attracted a surge of interests, which utilizes the rich user-item interaction information from the related source domain to improve the performance on the sparse target domain. Recent CDR approaches pay attention to aggregating the source domain information to generate better user representations for the target domain. However, they focus on designing more powerful interaction encoders to learn both domains simultaneously, but fail to model different user preferences of different domains. Particularly, domain-specific preferences of the source domain usually provide useless information to enhance the performance in the target domain, and directly aggregating the domain-shared and domain-specific information together maybe hurts target domain performance. This work considers a key challenge of CDR: How do we transfer shared information across domains? Grounded in the information theory, we propose **DisenCDR**, a novel model to disentangle the domain-shared and domain-specific information. To reach our goal, we propose two mutual-information-based disentanglement regularizers. Specifically, an exclusive regularizer aims to enforce the user domain-shared representations and domain-specific representations encoding exclusive information. An information regularizer is to encourage the user domain-shared representations encoding predictive information for both domains. Based on them, we further derive a tractable bound of our disentanglement objective to learn desirable disentangled representations. Extensive experiments show that DisenCDR achieves significant improvements over state-of-the-art baselines on four real-world datasets.

*Tingwen Liu is the corresponding author.



*SIGIR '22, July 11–15, 2022, Madrid, Spain.*

ACM ISBN 978-1-4503-8732-3/22/07.
https://doi.org/10.1145/3477495.3531967

## CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Computing methodologies** → **Neural networks**.

## KEYWORDS

Cross-Domain Recommendation; Variational Autoencoder; Disentangled Representation Learning

**ACM Reference Format:**
Jiangxia Cao, Xixun Lin, Xin Cong, Jing Ya, Tingwen Liu, Bin Wang. 2022. DisenCDR: Learning Disentangled Representations for Cross-Domain Recommendation. In *Proceedings of the 45th Int'l ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3477495.3531967

## 1 INTRODUCTION

As an important tool to understand user intentions, recommender systems (RS) have been successfully applied in many real-world applications, such as Amazon and Youtube. Most current RS models [11, 12, 55] follow a collaborative filtering (CF) paradigm to make recommendations by learning the representations of users and items. However, when user-item interactions are very sparse, those CF-based methods suffer from a strong decrease in recommendation performance [61]. Recently, cross-domain recommendation (CDR) [26, 30] becomes a promising way to alleviate this data sparsity problem by utilizing the rich information from the source domain to improve recommendation performance on the target domain. For instance (see Figure 1(a)), CDR can leverage the interest of a user's taste about the `Book` domain as an important support to make recommendations for the `Film` domain.

With the great success of neural network techniques in RS, several deep neural network approaches of CDR have been proposed by designing more powerful interaction encoders to model both domains simultaneously. As shown in Figure 1(b), to generate robust representations, most of existing works first apply two base encoders to model each domain interactions, and then introduce different transferring layers to fuse the learned representations of base encoders in a symmetrical way. For example, CoNet [14]

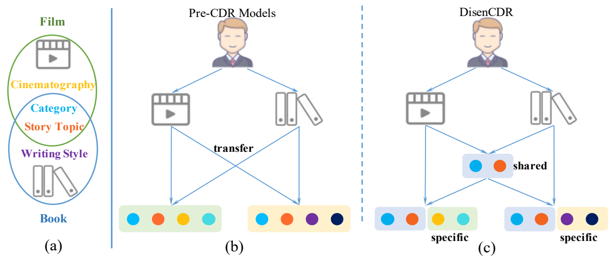


**Figure 1: (a): An example of user preferences in `Film` domain and `Book` domain. (b-c): A model comparison: in contrast with previous CDR models, DisenCDR disentangles the domain-shared and domain-specific representations to transfer knowledge across domains.**

leverages MLPs as the base encoder for each domain and devises a cross-connections network to transfer information. DDTCDR [26] extends the CoNet by learning a latent orthogonal mapping function to transfer users' similarity across domains. PPGN [58] learns user/item representations by stacking several GCNs [23] to directly aggregate the interaction information of both domains. BiTGCF [30] utilizes the LightGCN [11] as the encoder to aggregate interaction information for each domain and further introduces a feature transfer layer to enhance the two base graph encoders.

Although above methods achieve promising results to some extent, most of them still neglect to disentangle the domain-shared information and the domain-specific information, which limits the model transferring effectiveness. A concrete example is depicted in Figure 1(a), there are two domains: `Film` and `Book`. The domain-shared information, such as the 'Story Topic' and 'Category', can provide valuable information for both domains. But the domain-specific information, e.g., the 'Writing Style' in `Book` domain may provide useless information to making recommendation for `Film` domain, even causing the negative transfer problem for CDR [57]. Unfortunately, existing CDR methods ignore this problem and directly aggregate the domain-shared and domain-specific information of both domains. As a result, the learned user representations entangle both domain preferences together, which would lead to sub-optimal recommendation results.

To disentangle the domain-shared and domain-specific information, we propose a novel deep generation model termed DisenCDR to learn Disentangled Representations for CDR, which follows the VAE framework [20] to model the data distribution. As shown in Figure 1(c), our DisenCDR learns three separate representations for each user: one domain-shared representation and two domain-specific representations, for disentangling the user preferences. To disentangle the three representations, we introduce a variational bipartite graph encoder (VBGE) and two mutual-information-based regularizers: the exclusive regularizer and the informative regularizer, to constrain them. The exclusive regularizer aims at enforcing the user domain-shared representations and domain-specific representations encoding exclusive information. The information regularizer aims at encouraging the user domain-shared representations encoding predictive information for both domains. Based on the two mutual information regularizers, we derive a tractable bound of our disentanglement objective for CDR. Leveraging disentangled representations learned by DisenCDR, we show that only transferring the domain-shared information is a more powerful transfer strategy than other CDR methods.

Our main contributions are summarized as follows:

- We introduce a fresh perspective to solve CDR by disentangling the domain-shared information and domain-specific information and only transferring the domain-shared information to enhance model recommendation performance. To the best of our knowledge, this paper is the first work to learn disentangled representations for CDR.
- We propose a novel model named DisenCDR which contains two well-designed mutual-information-based regularizers, and derive a tractable disentanglement objective to learn meaningful domain-shared and domain-specific user representations.
- We conduct experiments on four real-world CDR datasets to evaluate model performance. Extensive results demonstrate that our DisenCDR achieves consistent and significant improvements over state-of-the-art baselines. Besides, we also conduct comprehensive ablation studies and detailed analyses to investigate the effectiveness of our model components. Our source codes and datasets are available at github[1] for further comparisons.

## 2 PROBLEM DEFINITION

This work considers a general CDR scenario that two domains have a shared user set. Let $\mathcal{D}^X = (\mathcal{U}, \mathcal{V}^X, \mathcal{E}^X)$ and $\mathcal{D}^Y = (\mathcal{U}, \mathcal{V}^Y, \mathcal{E}^Y)$ denote the interaction data of domain $X$ and $Y$, where $\mathcal{U}$ denotes the shared user set in both domains, $\mathcal{V}$ denotes the item set and $\mathcal{E}$ denotes edge set in each domain. Additionally, there are two binary interaction matrices $\boldsymbol{A}^X \in \{0,1\}^{|\mathcal{U}|\times|\mathcal{V}^X|}$ and $\boldsymbol{A}^Y \in \{0,1\}^{|\mathcal{U}|\times|\mathcal{V}^Y|}$ for $X$ and $Y$ respectively, where each element $A_{ij}$ describes whether user $u_i \in \mathcal{U}$ has interacted with item $v_j \in \mathcal{V}$ in the edge set $\mathcal{E}$.

Given the observed interactions of both domains, DisenCDR aims to learn disentangled user/item representations, i.e., $Z_u^S$, $Z_u^X$, $Z_u^Y$, $Z_v^X$ and $Z_v^Y$, and transfers the domain-shared representations $Z_u^S$ to enhance recommendation performance in both domains. Here $Z_u^S$, $Z_u^X$, $Z_u^Y$ represent the domain-shared, the domain-$X$-specific and the domain-$Y$-specific user representations; $Z_v^X$ and $Z_v^Y$ represent items domain-$X$-specific and the domain-$Y$-specific representations.

## 3 METHODOLOGY

Figure 2 gives a high-level overview of DisenCDR, including the variational bipartite graph encoders (VBGE), the generation and inference procedures. This section first describes the embedding layer which provides the initialized representations for users and items in both domains. Then, VBGE formulates the user-item interactions as a bipartite graph, and generates the user/item approximate posterior distribution by aggregating their homogeneous information. Afterward, following the VAE framework, we give the generation and inference procedures of DisenCDR. The inference procedure aims to **encode** the disentangled representations from the observed

[1] https://github.com/cjx96/DisenCDR

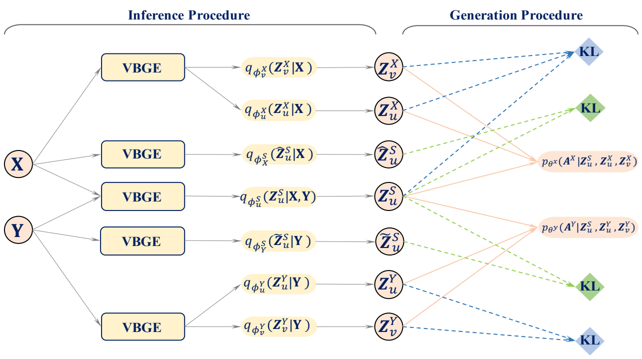


**Figure 2: An overview of DisenCDR. The VBGE means the variational bipartite graph encoder. The $Z_v^X, Z_u^X, Z_u^Y$ and $Z_v^Y$ are user/item domain-specific representations, and $Z_u^S$ is the user domain-shared representations. The blue KL means calculating KL divergence with prior distribution $\mathcal{N}(0, I)$. The green KL means calculating KL divergence between inputs. The latent variables $\widetilde{Z}_u^S$, $\widetilde{Z}_u^S$ are used to calculate our disentanglement objective.**

interactions and the generation procedure aims to **decode** the observed interactions by the learned disentanglement representations while constrained by our exclusive and informative regularizers.

## 3.1 Embedding Layer

The embedding layer embeds user and item into a low-dimensional vector space. To guarantee that the domain-shared and domain-specific representations are independent, we introduce multiple initialization embedding matrices for users. Formally, for the shared users in $\mathcal{U}$, we embed them with three $F$-dimensional embedding matrices $\boldsymbol{U}^S \in \mathbb{R}^{|\mathcal{U}| \times F}$, $\boldsymbol{U}^X \in \mathbb{R}^{|\mathcal{U}| \times F}$, and $\boldsymbol{U}^Y \in \mathbb{R}^{|\mathcal{U}| \times F}$ as domain-shared, domain-$X$-specific and domain-$Y$-specific initialized embedding matrices, respectively. Meanwhile, we also use $\boldsymbol{V}^X \in \mathbb{R}^{|\mathcal{V}^X| \times F}$ and $\boldsymbol{V}^Y \in \mathbb{R}^{|\mathcal{V}^Y| \times F}$ for two item sets in domain $X$ and domain $Y$ respectively.

## 3.2 Variational Bipartite Graph Encoder

The user-item interactions are naturally formed as a heterogeneous bipartite graph (i.e., the interaction graph contains two types of nodes: the user-type nodes and the item-type nodes), where **users are indirectly connected with an even-number-hop**, e.g., 2-hop, 4-hop, etc. Nevertheless, existing graph encoders [22, 23, 37, 47, 49, 53] always focus on aggregating features from 1-hop neighboring information, which leads to the improper information propagation from its heterogeneous neighbors. Therefore, our VBGE utilizes a two-step information propagation procedure. It first generates intermediate representations with homogeneous neighbors, and then generates latent variables (a.k.a. representations) by intermediate representations (as shown in Figure 3).

For simplicity, we take the information propagation procedure of users in the domain $X$ as an example. We first generate the

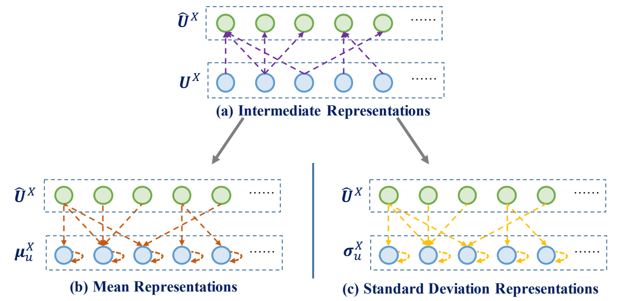


**Figure 3: A simple illustration of VBGE. Blue and green circles denote users and items. Purple, orange and yellow lines describe the two-step information propagation procedure. The $\widehat{U}^X$ is learned intermediate representations. The $\mu_u^X$ and $\sigma_u^X$ are used to sample final representations.**

intermediate representations $\widehat{\boldsymbol{U}}^X$ as[2]:

$$\widehat{\boldsymbol{U}}^X = \delta\Big(\mathtt{Norm}\big((\boldsymbol{A}^X)^\top\big)\boldsymbol{U}^X \boldsymbol{W}_u^X\Big), \tag{1}$$

where $\mathtt{Norm}(\cdot)$ denotes the row normalized function, $(\boldsymbol{A}^X)^\top$ denotes the transpose interaction matrix, $\delta(\cdot)$ is the LeakyReLU function and the $\boldsymbol{W}_u^X \in \mathbb{R}^{F \times F}$ is a parameter matrix. **Note that the $\widehat{\boldsymbol{U}}^X \in \mathbb{R}^{|\mathcal{V}^X| \times F}$ can be considered as user-type representations** to build a bridge between users with their 2-hop neighbors. In this way, VBGE generates the domain-$X$-specific latent variables (a.k.a. domain-$X$-specific representations) $\boldsymbol{Z}_u^X$ for users in domain $X$ as follows:

$$\begin{aligned}
\boldsymbol{\mu}_u^X &= \delta\Big(\big[\delta(\mathtt{Norm}(\boldsymbol{A}^X)\widehat{\boldsymbol{U}}^X \widehat{\boldsymbol{W}}_{u,\mu}^X) \oplus \boldsymbol{U}^X\big]\boldsymbol{W}_{u,\mu}^X\Big),\\
\boldsymbol{\sigma}_u^X &= \varphi\Big(\big[\delta(\mathtt{Norm}(\boldsymbol{A}^X)\widehat{\boldsymbol{U}}^X \widehat{\boldsymbol{W}}_{u,\sigma}^X) \oplus \boldsymbol{U}^X\big]\boldsymbol{W}_{u,\sigma}^X\Big),\\
\boldsymbol{Z}_u^X &\sim \mathcal{N}\Big(\boldsymbol{\mu}_u^X, [\mathrm{diag}\{\boldsymbol{\sigma}_u^X\}]^2\Big),
\end{aligned} \tag{2}$$

where $\oplus$ is the concatenation operation, $\varphi(\cdot)$ is the Softplus function, $\widehat{\boldsymbol{W}}_{u,\mu}^X \in \mathbb{R}^{F \times F}$, $\widehat{\boldsymbol{W}}_{u,\sigma}^X \in \mathbb{R}^{F \times F}$, $\boldsymbol{W}_{u,\mu}^X \in \mathbb{R}^{2F \times F}$ and $\boldsymbol{W}_{u,\sigma}^X \in \mathbb{R}^{2F \times F}$ are parameter matrices. $\boldsymbol{\mu}_u^X$ and $\boldsymbol{\sigma}_u^X$ are mean and standard deviation of Gaussian distribution, which is used to sample the latent variables $\boldsymbol{Z}_u^X \in \mathbb{R}^{|\mathcal{U}^X| \times F}$. Concretely, we apply reparameterization trick [20] to sample the latent variable $\boldsymbol{z}_{u_i}^X$ for user $u_i$:

$$\boldsymbol{z}_{u_i}^X = \boldsymbol{\mu}_{u_i}^X + \boldsymbol{\sigma}_{u_i}^X \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \tag{3}$$

where $\odot$ is the element-wise product, $\boldsymbol{\epsilon} \in \mathbb{R}^F$ is a normal Gaussian noise vector, $\boldsymbol{\mu}_{u_i}^X \in \mathbb{R}^F$ is the mean vector and $\boldsymbol{\sigma}_{u_i}^X \in \mathbb{R}^F$ is the standard deviation vector of user $u_i$.

According to Eq.(1-3), the information propagation procedure of generated users latent variables $\boldsymbol{Z}_u^X$ only relies on users representations $\boldsymbol{U}^X$, which achieves proper information propagation from its homogeneous neighbors. Similarly, the latent variables of item $\boldsymbol{Z}_v^X$ can be also obtained via analogous calculations. All user/item domain-specific latent variables for the domain $X$ and the domain $Y$ can be summarized as $\boldsymbol{Z}_u^X$, $\boldsymbol{Z}_v^X$, $\boldsymbol{Z}_u^Y$ and $\boldsymbol{Z}_v^Y$, respectively.

[2] Although recent works [11, 52] empirically show that the nonlinear functions in GNNs are harmful to capture the collaborative filtering signal, but we still find that the activation function works in our work.

Besides the domain-specific representations, VBGE can also utilize the interactions from both domains to generate domain-shared latent variables $Z_u^S$ for users. To do so, we first use the initialized domain-shared user representation $\boldsymbol{U}^S$ as input to generate $\overline{\boldsymbol{\mu}}_u^X$, $\overline{\boldsymbol{\sigma}}_u^X$, $\overline{\boldsymbol{\mu}}_u^Y$ and $\overline{\boldsymbol{\sigma}}_u^Y$ for $X$ and $Y$. Afterwards, we use a gate operator [6] to combine them as:

$$\begin{aligned} \boldsymbol{\mu}_u^S &= \boldsymbol{\lambda}_u \odot \overline{\boldsymbol{\mu}}_u^X + (1 - \boldsymbol{\lambda}_u) \odot \overline{\boldsymbol{\mu}}_u^Y, \\ \boldsymbol{\sigma}_u^S &= \boldsymbol{\lambda}_u \odot \overline{\boldsymbol{\sigma}}_u^X + (1 - \boldsymbol{\lambda}_u) \odot \overline{\boldsymbol{\sigma}}_u^Y, \\ \lambda_{u_i} &= \frac{N_{u_i}^X}{N_{u_i}^X + N_{u_i}^Y}, \quad Z_u^S \sim \mathcal{N}(\boldsymbol{\mu}_u^S, [\mathrm{diag}\{\boldsymbol{\sigma}_u^S\}]^2), \end{aligned} \tag{4}$$

where the $N_{u_i}^X$ and $N_{u_i}^Y$ denote the number of 1-hop neighbors of user $u_i$ in domain $X$ and $Y$, and the fixed factor $\lambda_{u_i}$ controls the contribution ratio of different domains.

Up to now, we have introduced the learning process to generate domain-shared and domain-specific representations. In the following section, we will dive into the VAE framework and explain how we devise the mutual-information-based tractable objective to constrain these representations to achieve desirable disentanglement.

## 3.3 Generation and Inference

Following the VAE framework [20], we assume that the observed interactions, $\mathcal{D}^X$ and $\mathcal{D}^Y$, are sampled from an joint interaction distribution $p_{\mathcal{D}}(u, v^X, v^Y)$. Each triple $(u_i, v_j, v_k) \sim p_{\mathcal{D}}(u, v^X, v^Y)$ describes that the user $u_i$ has interacted with the item $v_j \in \mathcal{V}^X$ and with the the item $v_k \in \mathcal{V}^Y$. Given observed interaction data, the goal of DisenCDR is to learn the disentangled domain-shared and domain-specific representations. To do so, we adopt the generation procedure to decode the joint distribution by the inference procedure learned domain-shared (i.e., $Z_u^S$) and domain-specific (i.e., $Z_u^X$, $Z_v^X$, $Z_u^Y$ and $Z_v^Y$) representations.

*3.3.1 Generation Procedure.* Based on the structural assumption in Figure 4(a), we maximize the likelihood of the joint distribution:

$$\begin{aligned} p_\theta(u, v^X, v^Y) = \int & p_{\theta^X}(\boldsymbol{A}^X | Z_u^S, Z_u^X, Z_v^X) p_{\theta^Y}(\boldsymbol{A}^Y | Z_u^S, Z_u^Y, Z_v^Y) \\ & p(Z_u^S) p(Z_u^X) p(Z_u^Y) p(Z_v^X) p(Z_v^Y) \mathrm{d}Z_u^S \mathrm{d}Z_u^X \mathrm{d}Z_u^Y \mathrm{d}Z_v^X \mathrm{d}Z_v^Y. \end{aligned} \tag{5}$$

Noticeably, $p_\theta(u, v^X, v^Y)$ can be divided as two parts:

- Prior distributions: $p(Z_u^S), p(Z_u^X), p(Z_u^Y), p(Z_v^X), p(Z_v^Y)$.
- Decoders: $p_{\theta^X}(\boldsymbol{A}^X | Z_u^S, Z_u^X, Z_v^X), p_{\theta^Y}(\boldsymbol{A}^Y | Z_u^S, Z_u^Y, Z_v^Y)$.

For the prior distributions, we set all of them as normal Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ as suggested by various variational models [5, 13, 28, 29]. The decoders aim to reconstruct observed interactions. Given a user-item triple $(u_i, v_j, v_k)$, we have:

$$\begin{aligned} p_{\theta^X}(A_{ij}^X | z_{u_i}^S, z_{u_i}^X, z_{v_j}^X) &= \sigma\big(\mathrm{S}_{\theta_S^X}(z_{v_j}^X, z_{u_i}^S) + \mathrm{S}_{\theta_X^X}(z_{v_j}^X, z_{u_i}^X)\big), \\ p_{\theta^Y}(A_{ik}^Y | z_{u_i}^S, z_{u_i}^Y, z_{v_k}^Y) &= \sigma\big(\mathrm{S}_{\theta_S^Y}(z_{v_k}^Y, z_{u_i}^S) + \mathrm{S}_{\theta_Y^Y}(z_{v_k}^Y, z_{u_i}^Y)\big), \end{aligned} \tag{6}$$

where $A_{ij}^X / A_{ik}^Y$ is an element of $\boldsymbol{A}^X / \boldsymbol{A}^Y$, $\sigma(\cdot)$ is the sigmoid function, $\mathrm{S}_{\theta_S^X}(\cdot), \mathrm{S}_{\theta_S^Y}(\cdot), \mathrm{S}_{\theta_X^X}(\cdot), \mathrm{S}_{\theta_Y^Y}(\cdot)$ are score functions to estimate the different contributions of those representations. Besides, there are many alternative methods that can be used to implement the score functions such as MLPs. In this work, we choose the inner product operation for fast training speed.

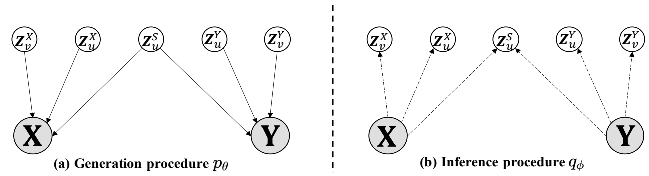


**Figure 4: A graphical illustration of the generation and inference procedures in DisenCDR. The $Z_u^S$, $Z_u^X$, $Z_u^Y$, $Z_v^X$ and $Z_v^Y$ are disentangled representations of users and items.**

*3.3.2 Inference Procedure.* Directly maximizing the joint distribution in Eq.(5) is intractable, since the true posterior distribution $p_\theta(Z_u^X, Z_u^Y, Z_u^S, Z_v^X, Z_v^Y | \mathrm{X}, \mathrm{Y})$ is unknown. Therefore, we employ the amortized inference [7] to approximate the true posterior distribution by using the generated approximated posterior distribution of our VBGEs. From structural assumption of DisenCDR in Figure 4(b), we factorize the approximated posterior distribution as:

$$\begin{aligned} q_\phi(Z_u^X, Z_u^Y, Z_u^S, Z_v^X, Z_v^Y | \mathrm{X}, \mathrm{Y}) &= q_{\phi_u^X}(Z_u^X | \mathrm{X}) q_{\phi_u^Y}(Z_u^Y | \mathrm{Y}) \\ & q_{\phi_v^X}(Z_v^X | \mathrm{X}) q_{\phi_v^Y}(Z_v^Y | \mathrm{Y}) q_{\phi_u^S}(Z_u^S | \mathrm{X}, \mathrm{Y}), \end{aligned} \tag{7}$$

where $\phi = \{\phi_u^X, \phi_u^Y, \phi_v^X, \phi_v^Y, \phi_u^S\}$ is parameter set of all VBGEs, and $\mathrm{X}, \mathrm{Y}$ denote the interaction information of domain $X$ and $Y$. To achieve ideal disentanglement, the first four terms should encode the domain-specific information and the last term needs to encode the domain-shared information.

Nevertheless, Eq.(7) does not guarantee that the domain-shared and domain-specific representations can be disentangled correctly. The main reason lies in that we have no control over the encoding information by three separate users latent variables $Z_u^X$, $Z_u^Y$ and $Z_u^S$. This observation motivates us to devise the regularizers of them to achieve effective disentanglement.

## 3.4 Disentanglement Objective

In this section, we analyze the entanglement issue from the information theory perspective and further derive a novel disentanglement objective for DisenCDR. Generally, the encoding information of ideal domain-shared and domain-specific representations should be mutually exclusive. Thus we devise regularizers to constrain them to meet the disentanglement goal in the following section.

Before going on, it is necessary to introduce two mutual information definitions for better understanding:

*Definition 3.1.* **Conditional Mutual Information** $I(\mathrm{X}; \mathrm{Y} | \mathrm{Z})$. For three random variables $\mathrm{X}, \mathrm{Y}, \mathrm{Z}$, the conditional mutual information is defined as follows:

$$\begin{aligned} I(\mathrm{X}; \mathrm{Y} | \mathrm{Z}) &= I(\mathrm{X}; \mathrm{Y}, \mathrm{Z}) - I(\mathrm{X}; \mathrm{Z}) \\ &= H(\mathrm{X} | \mathrm{Z}) - H(\mathrm{X} | \mathrm{Y}, \mathrm{Z}) \end{aligned} \tag{8}$$

where $H(\cdot)$ refers to the information entropy.

*Definition 3.2.* **Interaction Information** $I(\mathrm{X}; \mathrm{Y}; \mathrm{Z})$. For three random variables $\mathrm{X}, \mathrm{Y}, \mathrm{Z}$, the interaction information is defined as:

$$\begin{aligned} I(\mathrm{X}; \mathrm{Y}; \mathrm{Z}) &= I(\mathrm{X}; \mathrm{Y}) - I(\mathrm{X}; \mathrm{Y} | \mathrm{Z}) \\ &= I(\mathrm{X}; \mathrm{Y}) + I(\mathrm{X}; \mathrm{Z}) - I(\mathrm{X}; \mathrm{Y}, \mathrm{Z}) \\ &= H(\mathrm{Y}) - H(\mathrm{Y} | \mathrm{X}) - H(\mathrm{X} | \mathrm{Z}) + H(\mathrm{X} | \mathrm{Y}, \mathrm{Z}) \end{aligned} \tag{9}$$

**Algorithm 1** The training procedure of DisenCDR.

**Input:** the interaction data $\mathcal{D}^X$ and $\mathcal{D}^Y$; the interaction matrix $\boldsymbol{A}^X$ and $\boldsymbol{A}^Y$; the dimension size $F$; the batch size $B$; the depth of VBGE; the disentanglement factor $\beta$

**Output:** the users' domain-shared representations $Z_u^S$; the users' domain-specific representations $Z_u^X$ and $Z_u^Y$; the items' domain-specific representations $Z_v^X$ and $Z_v^Y$

Initialize parameters of Embedding Layer and VBGEs
**while** not convergence **do**
  Encode user approximate posterior distribution $q_{\phi_u^X}(Z_u^X|\mathrm{X})$, $q_{\phi_u^Y}(Z_u^Y|\mathrm{Y})$ and $q_{\phi_u^S}(Z_u^S|\mathrm{X},\mathrm{Y})$ in Eq.(7)
  Encode item approximate posterior distribution $q_{\phi_v^X}(Z_v^X|\mathrm{X})$, and $q_{\phi_v^Y}(Z_v^Y|\mathrm{Y})$ in Eq.(7)
  Sample a batch of interaction triples $\mathcal{B}$ from $\mathcal{D}^X$ and $\mathcal{D}^Y$
  **for** each triple $(u_i, v_j, v_k)$ in $\mathcal{B}$ **do**
    Make prediction via Eq.(6)
    Calculate ELBO in Eq.(12)
  **end for**
  Encode variational distribution $q_{\phi_Y^S}(\widetilde{Z}_u^S|\mathrm{Y})$ and $q_{\phi_X^S}(\widehat{Z}_u^S|\mathrm{X})$ to estimate $I(Z_u^S;\mathrm{X}|\mathrm{Y})$ and $I(Z_u^S;\mathrm{Y}|\mathrm{X})$ in Eq.(12)
  Calculate our disentanglement objective $\mathcal{L}$ in Eq.(13)
  Update model parameters
**end while**
**return** $Z_u^S, Z_u^X, Z_u^Y, Z_v^X, Z_v^Y$

*3.4.1 Exclusive regularizer.* For encouraging the domain-shared and domain-specific latent variables to encode with the exclusive information, we first introduce the exclusive regularizer to **minimize mutual information** between domain-shared and domain-specific latent variables of users, i.e., $I(Z_u^X; Z_u^S)$ and $I(Z_u^Y; Z_u^S)$. To better analyze the influence of minimizing this regularizer for learning disentanglement representations, we rewrite $I(Z_u^X; Z_u^S)$ as follows:

$$\begin{aligned} I(Z_u^X; Z_u^S) &= I(Z_u^X; Z_u^S) - \big(H(Z_u^X|\mathrm{X}) - H(Z_u^X|Z_u^S,\mathrm{X})\big) \\ &= I(Z_u^X; Z_u^S) - I(Z_u^X; Z_u^S|\mathrm{X}) \\ &= I(Z_u^X; Z_u^S; \mathrm{X}) \\ &= I(\mathrm{X}; Z_u^X) + I(\mathrm{X}; Z_u^S) - I(\mathrm{X}; Z_u^X, Z_u^S). \end{aligned} \tag{10}$$

Here we first use the structural assumption of DisenCDR in Figure 4(b), i.e., $q(Z_u^X|\mathrm{X}) = q(Z_u^X|\mathrm{X}, Z_u^S)$, and then utilize the Definition 3.1 and Definition 3.2. In general, minimizing the above regularizer can be interpreted heuristically in the following way: (1) The last term is to maximize $I(\mathrm{X}; Z_u^X, Z_u^S)$, which encourages $Z_u^X$,$Z_u^S$ jointly to be correlated with interaction information $\mathbf{X}$. (2) The first two terms are minimized to penalize the separate correlation with $Z_u^X$ and $Z_u^S$. Thus, minimizing Eq.(10) is forced to preserve exclusive information between $Z_u^X$ and $Z_u^S$.

Nevertheless, the above regularizer is still insufficient for learning the ideal disentangled representations. The main reason is that any arbitrary mutually exclusive information decomposition can satisfy the regularizer, even if the $Z_u^X$ encodes total information and $Z_u^S$ is non-informative for both domains. This observation motivates us to devise another regularizer to achieve a better disentanglement on domain-shared and domain-specific representations.

*3.4.2 Informative regularizer.* To enable the domain-shared representation $Z_u^S$ to be informative for both domains, we **maximize** the mutual information $I(Z_u^S;\mathrm{X};\mathrm{Y})$ for encouraging $Z_u^S$ encoding the shared information across domains. Specifically, for the domain $X$, we have:

$$\begin{aligned} I(Z_u^S;\mathrm{X};\mathrm{Y}) &= I(Z_u^S;\mathrm{X}) - I(Z_u^S;\mathrm{X}|\mathrm{Y}) \\ &= I(Z_u^S;\mathrm{X}) - \big(I(Z_u^S;\mathrm{X},\mathrm{Y}) - I(Z_u^S;\mathrm{Y})\big). \end{aligned} \tag{11}$$

In practice, maximizing the above regularizer also can be interpreted heuristically in the following way: (1) The former term is maximized $I(Z_u^S;\mathrm{X})$ to encourage $Z_u^S$ to be correlated with domain $X$. (2) The latter term is minimized, which means the domain-shared representation $Z_u^S$ not only can be inferred from joint information of domains $X$ and $Y$, but also can be directly inferred from separate domain $Y$. Thus, maximizing Eq.(11) will naturally encourage $Z_u^S$ to encode the shared information across domains.

*3.4.3 Objective Function.* Based on the two mutual-information-based regularizers, we derive our final disentanglement objective by **combining the regularizers on both domains together**:

$$\begin{aligned} \mathcal{L} =& I(Z_u^X; Z_u^S) + I(Z_u^Y; Z_u^S) - 2I(Z_u^S;\mathrm{X};\mathrm{Y}) \\ =& I(\mathrm{X}; Z_u^X) + I(Z_u^S;\mathrm{X}|\mathrm{Y}) - I(\mathrm{X}; Z_u^X, Z_u^S) \\ &+ I(\mathrm{Y}; Z_u^Y) + I(Z_u^S;\mathrm{Y}|\mathrm{X}) - I(\mathrm{Y}; Z_u^Y, Z_u^S) \\ \leq& I(\mathrm{X}; Z_u^X) + I(\mathrm{X}; Z_v^X) + I(\mathrm{Y}; Z_u^Y) + I(\mathrm{Y}; Z_v^Y) \\ &+ I(\mathrm{X},\mathrm{Y}; Z_u^S) + I(Z_u^S;\mathrm{X}|\mathrm{Y}) + I(Z_u^S;\mathrm{Y}|\mathrm{X}) \\ &- I(\mathrm{X}; Z_u^X, Z_u^S, Z_v^X) - I(\mathrm{Y}; Z_u^Y, Z_u^S, Z_v^Y) \\ \leq& \mathrm{ELBO} + I(Z_u^S;\mathrm{X}|\mathrm{Y}) + I(Z_u^S;\mathrm{Y}|\mathrm{X}). \end{aligned} \tag{12}$$

Fortunately, many terms of our disentanglement objective are already covered in the standard Evidence Lower Bound (ELBO) [29] according to the relationship between the variation bound and the mutual information [39]. Therefore, the tractable objective of our objective can derive:

- For the encoding terms in ELBO (e.g., $I(\mathrm{X}; Z_u^X)$), as most variational models adopted [20, 28], those terms encourage that the generated posterior distributions (e.g., $q_{\phi_u^X}(Z_u^X|\mathrm{X})$) close to the prior distributions (e.g., $p(Z_u^X)$). We measure it by the Kullback-Leibler (KL) divergence;
- For the tractable objective of the decoding terms in ELBO (e.g., $I(\mathrm{X}; Z_u^X, Z_u^S, Z_v^X)$), as discussed in BiVAE [46], those aims to reconstruct the observed interactions. We estimate them by maximizing the log-likelihood of our decoders;
- For the rest two terms, i.e., $I(Z_u^S;\mathrm{X}|\mathrm{Y})$ and $I(Z_u^S;\mathrm{Y}|\mathrm{X})$, as discussed above, those encourage the representations $Z_u^S$ encoding meaningful domain-shared information. To estimate them, we first additionally introduce two VBGEs to generate variational distributions $q_{\phi_Y^S}(\widetilde{Z}_u^S|\mathrm{Y})$, $q_{\phi_X^S}(\widehat{Z}_u^S|\mathrm{X})$, and then calculate KL divergence with $q_{\phi_u^S}(Z_u^S|\mathrm{X},\mathrm{Y})$. Note that the encoding procedure of the two additionally distributions are similar with $q_{\phi_u^X}(Z_u^X|\mathrm{X})$ and $q_{\phi_u^Y}(Z_u^Y|\mathrm{Y})$.

Thus, the variational lower bound of our disentanglement objective is defined as[3]:

$$
\begin{aligned}
\mathcal{L} \leq & \mathbb{D}_{KL}(q(Z_u^X|\mathrm{X})||p(Z_u^X)) + \mathbb{D}_{KL}(q(Z_v^X|\mathrm{X})||p(Z_v^X)) \\
& + \mathbb{D}_{KL}(q(Z_u^Y|\mathrm{Y})||p(Z_u^Y)) + \mathbb{D}_{KL}(q(Z_u^S|\mathrm{X},\mathrm{Y})||p(Z_u^S)) \\
& + \mathbb{D}_{KL}(q(Z_v^Y|\mathrm{Y})||p(Z_v^Y)) \\
& - \mathbb{E}_{q(Z_u^X, Z_v^X|\mathrm{X})q(Z_u^S|\mathrm{X},\mathrm{Y})}[\log p(A^X|Z_u^S, Z_u^X, Z_v^X)] \\
& - \mathbb{E}_{q(Z_u^Y, Z_v^Y|\mathrm{Y})q(Z_u^S|\mathrm{X},\mathrm{Y})}[\log p(A^Y|Z_u^S, Z_u^Y, Z_v^Y)] \\
& + \beta\mathbb{D}_{KL}(q(Z_u^S|\mathrm{X},\mathrm{Y})||q(\widetilde{Z}_u^S|\mathrm{Y})) + \beta\mathbb{D}_{KL}(q(Z_u^S|\mathrm{X},\mathrm{Y})||q(\widehat{Z}_u^S|\mathrm{X})) \\
= & \mathbb{E}_{p_{\mathcal{D}}(u,v^X,v^Y)}[\mathrm{ELBO}(p,q) + \beta\mathbb{D}_{KL}(q(Z_u^S|\mathrm{X},\mathrm{Y})||q(\widetilde{Z}_u^S|\mathrm{Y})) \\
& + \beta\mathbb{D}_{KL}(q(Z_u^S|\mathrm{X},\mathrm{Y})||q(\widehat{Z}_u^S|\mathrm{X}))],
\end{aligned}
\tag{13}
$$

where $\beta > 0$ is a disentanglement factor to control the disentangling capacity of users representations. The pseudo-code of the training procedure is shown in Algorithm 1.

*3.4.4 Time Complexity.* DisenCDR can be optimized with mini-batch manner, which keeps promising time complexity. In particular, the computational complexity of the generation procedure is $\mathcal{O}(BF)$, and the inference procedure is $\mathcal{O}(|\mathcal{E}^X| + |\mathcal{E}^Y|)F^2)$, where $B$ is the batch size. Empirically, in the same running environment, DisenCDR, PPGN [58] and BiTGCF [30] would cost around 0.69s/0.121ms, 0.50s/0.108ms and 0.53s/0.112ms pre 1000 training/inference samples on Elec&Cloth dataset, respectively.

## 4 EXPERIMENTS

In this section, we give detailed analyses to answer the following major research questions (RQs):

- **RQ1**: Does our method achieve the significant performances in comparison with other state-of-the-art methods?
- **RQ2**: When we change the basic graph encoder, is our proposed disentanglement objective still able to achieve improvements?
- **RQ3**: Besides of the experiment results, is our proposed disentanglement objective reaches the desirable disentanglement? Further, what impact does our disentanglement objective function achieve?
- **RQ4**: How does different hyperparameter settings influence the performance of our method?

### 4.1 Datasets

For a fair comparison with previous methods, we evaluate our model on four real-world benchmark datasets from Amazon[4] [10]: Elec, Phone, Sport and Cloth. As used in BiTGCF [30], they preprocess those datasets and combine them into four CDR scenarios: Elec&Phone, Sport&Cloth, Sport&Phone and Elec&Cloth. However, in their preprocessed datasets, we found that there are many **cold-start items acting as ground truths**[5] in the preprocessed test sets, which may lead to improper evaluations. Thus, in this work, we use the same training set with BiTGCF, but we filter out the cold-start item entry in the test set. The concrete statistics of our preprocessed data are listed in Table 1.

[3]For notation brevity, we omit the subscripts $\theta$ and $\phi$.
[4]http://jmcauley.ucsd.edu/data/amazon/index_2014.html
[5]We find about 20% ground-truth items are cold-start items in BiTGCF released datasets.

**Table 1: Statistics of four CDR scenarios.**

| Datasets | $\lvert\mathcal{U}\rvert$ | $\lvert\mathcal{V}\rvert$ | Training | Test | Density |
|---|---|---|---|---|---|
| Elec | 3,325 | 17,709 | 50,407 | 2,559 | 0.089% |
| Phone | 3,325 | 38,706 | 115,554 | 2,560 | 0.091% |
| Sport | 9,928 | 30,796 | 92,612 | 8,326 | 0.033% |
| Cloth | 9,928 | 39,008 | 87,829 | 7,540 | 0.024% |
| Sport | 4,998 | 20,845 | 50,558 | 3,698 | 0.052% |
| Phone | 4,998 | 13,655 | 42,446 | 3,999 | 0.068% |
| Elec | 15,761 | 51,447 | 210,865 | 13,824 | 0.027% |
| Cloth | 15,761 | 48,781 | 121,083 | 12,526 | 0.017% |

### 4.2 Experiment Setting

*4.2.1 Evaluation Protocol.* Following BiTGCF [30], the widely used *leave-one-out* method is adapted to show the performance. To ensure unbiased discoveries for every method, we follow Rendle's suggestion [24] to calculate 1000 records. Concretely, for a ground truth user-item pair $(u_i, v_j)$ in domain $X$, DisenCDR predicts 1000 candidates score (including 999 negative items $\tilde{v}_j$ and 1 ground truth item $v_j$) by the learned representation $z_{u_i}^S, z_{u_i}^X, z_{v_j}^X$ and $z_{\tilde{v}_j}^X$. Afterward, we adopt two widely-used metrics HR (Hit Ratio) and NDCG (Normalized Discounted Cumulative Gain [17]) to evaluate performance on the *top-10* ranking result.

*4.2.2 Compared Methods.* We compare DisenCDR with the following strong single-domain and cross-domain baselines.

To verify the effectiveness of CDR methods, we mix all interactions of both domains, and then apply the following CF-based single domain methods: (1) **BPRMF** [41] is a famous method which learns representations with pairwise ranking loss. (2) **NeuMF** [12] is also a well-known approach which learns representations with several MLP layers. (3) **NGCF** [49] stacks three GNN layers to aggregate the high-order neighboring information to learn representations. (4) **LightGCN** [11] is an extension of NGCF, which devises a linear propagating encoder to learn representations.

We also compare the state-of-the-art cross-domain methods: (1) **CDFM** [32] focuses on transferring knowledge from source domain to target domain. In our setting, we run two times for each CDR scenario to get the source and target evaluation results. (2) **CoNet** [14] first models interactions of two domains by two base networks, and then transfer knowledge by a cross-connection network between the two base networks. (3) **DDTCDR** [26] focuses on transfer users' similarity across domains, it first encodes users/items representations for each domain, then introduces a latent orthogonal mapping function to transfer users' similarity across domains. (4) **PPGN** [58] learns user/item representations by two separate GCNs with a shared initialized user embedding layer. (5) **BiTGCF** [30] extends the LightGCN to CDR task. It first uses two linear graph encoders to generate users/items representations for each domain, then utilizes a feature transfer layer to fuse users representations.

Additionally, we further implement our introduced VBGE as a cross-domain model, which utilizes the standard ELBO as the objective function. We rename this variant as **VBGE***.

*4.2.3 Implementation Details.* In our experiments, the side information of users/items is not exploited. For BPRMF and CDFM, we

**Table 2: Performance comparison (%) of different methods on four CDR scenarios ($\beta = 0.7$, 2-layer VBGEs).**

| Datasets | Metrics@10 | Single-Domain Methods | | | | Cross-Domain Methods | | | | | Ours |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BPRMF | NeuMF | NGCF | LightGCN | CDFM | CoNet | DDTCDR | PPGN | BiTGCF | DisenCDR |
| Elec | HR | 15.71 | 16.17 | 18.55 | 19.17 | 18.24 | 17.22 | 18.47 | 21.68 | 22.14 | **24.57* (+2.43)** |
| | NDCG | 9.19 | 9.24 | 10.87 | 10.28 | 10.92 | 9.86 | 11.08 | 11.63 | 12.20 | **14.51* (+2.31)** |
| Phone | HR | 16.32 | 15.84 | 22.79 | 23.25 | 17.97 | 17.66 | 17.23 | 24.54 | 25.71 | **28.76* (+3.05)** |
| | NDCG | 8.53 | 8.02 | 12.38 | 12.72 | 9.72 | 9.30 | 8.58 | 13.34 | 13.93 | **16.13* (+2.20)** |
| Sport | HR | 10.43 | 10.74 | 13.13 | 13.19 | 11.61 | 12.09 | 11.86 | 15.10 | 14.83 | **17.55* (+2.45)** |
| | NDCG | 5.41 | 5.46 | 6.87 | 6.94 | 6.33 | 6.41 | 6.37 | 8.03 | 7.95 | **9.46* (+1.51)** |
| Cloth | HR | 11.53 | 11.18 | 13.22 | 13.58 | 12.32 | 12.40 | 12.54 | 14.23 | 14.68 | **16.31* (+1.63)** |
| | NDCG | 6.25 | 6.02 | 6.97 | 7.29 | 7.05 | 6.62 | 7.13 | 7.68 | 7.93 | **9.03* (+1.10)** |
| Sport | HR | 9.89 | 10.11 | 16.06 | 16.33 | 11.97 | 12.88 | 12.14 | 18.00 | 18.63 | **20.17* (+1.54)** |
| | NDCG | 5.16 | 5.19 | 8.53 | 9.16 | 6.55 | 6.91 | 6.47 | 10.54 | 10.11 | **11.80* (+1.26)** |
| Phone | HR | 13.60 | 14.67 | 17.07 | 16.47 | 16.32 | 16.60 | 16.17 | 20.40 | 21.10 | **23.55* (+2.45)** |
| | NDCG | 7.27 | 7.80 | 9.22 | 8.95 | 9.01 | 9.15 | 8.98 | 11.09 | 11.25 | **12.97* (+1.72)** |
| Elec | HR | 20.65 | 20.08 | 20.20 | 19.97 | 21.09 | 21.26 | 21.70 | 21.85 | 21.61 | **23.71* (+1.86)** |
| | NDCG | 11.66 | 11.79 | 11.74 | 10.73 | 11.89 | 12.61 | 13.10 | 12.36 | 12.25 | **13.56* (+1.20)** |
| Cloth | HR | 9.47 | 10.84 | 10.86 | 11.24 | 10.37 | 11.35 | 11.47 | 12.98 | 13.11 | **15.13* (+2.02)** |
| | NDCG | 5.07 | 5.80 | 5.92 | 6.11 | 5.63 | 6.19 | 6.38 | 6.88 | 6.80 | **8.37* (+1.49)** |

* indicates that the improvements are statistically significant for $p < 0.05$ judged with the runner-up result in each case by paired t-test.

**Table 3: Performance comparison (%) of different variants.**

| Datasets | Metrics@10 | VPPGN variants | | VBGE variants | |
|---|---|---|---|---|---|
| | | VPPGN* | VPPGN# | VBGE* | DisenCDR |
| Elec | HR | 22.34 | 23.40 | 23.36 | 24.57 |
| | NDCG | 12.16 | 13.86 | 13.71 | 14.51 |
| Phone | HR | 26.37 | 27.43 | 27.35 | 28.76 |
| | NDCG | 14.34 | 15.43 | 15.25 | 16.13 |
| Sport | HR | 16.32 | 17.21 | 16.74 | 17.55 |
| | NDCG | 8.84 | 9.23 | 9.08 | 9.46 |
| Cloth | HR | 14.88 | 15.92 | 15.76 | 16.31 |
| | NDCG | 8.60 | 9.08 | 8.85 | 9.03 |
| Sport | HR | 18.93 | 19.84 | 19.36 | 20.17 |
| | NDCG | 10.73 | 11.43 | 10.92 | 11.80 |
| Phone | HR | 20.95 | 22.98 | 21.83 | 23.55 |
| | NDCG | 11.16 | 12.39 | 11.75 | 12.97 |
| Elec | HR | 22.15 | 22.78 | 22.19 | 23.71 |
| | NDCG | 12.53 | 12.79 | 12.66 | 13.56 |
| Cloth | HR | 13.95 | 14.87 | 14.33 | 15.13 |
| | NDCG | 7.37 | 8.15 | 7.80 | 8.37 |

implement them by ourselves. Except them, we directly use official implementations of other baseline models in our experiment setting. For all methods, the common hyper-parameters are listed as follows: the initializing embedding dimension $F$ is fixed as 128, the mini-batch size $B$ is fixed as 1024, the learning rate is fixed as 0.001, the L2 regularization coefficient is fixed as 0.0005, the dropout rate is fixed as 0.3, the number of graph encoder is selected from 1 to 4, the negative sampling number is fixed as 1, the cross entropy is used as reconstruction loss and the Adam [21] optimizer is used to update all parameters. Besides, the slope of LeakyReLU is fixed as 0.1, the $\beta$ of our disentanglement objective is chosen from {0.1, 0.3, 0.5, 0.7, 0.9}. We train all models with 100 epochs for convergence, and evaluate the model prediction scores every 10 epochs. In the following section, we conduct experiments under the $\beta$ is 0.7 and the 2-layer VBGEs by default.

**Table 4: KL divergence between domain-shared&specific Rep.**

| Variants | Elec&Phone | Sport&Cloth | Sport&Phone | Elec&Cloth |
|---|---|---|---|---|
| VBGE* | 50.4 | 24.7 | 37.6 | 58.2 |
| DisenCDR | 270.9 | 273.5 | 238.1 | 300.6 |

## 4.3 Performance Comparisons (RQ1)

Table 2 shows the comparison results on the four CDR scenarios according to HR@10 and NDCG@10. From the experimental results, we have the following observations: (1) Compared with BPRMF and NeuMF, the GNN-based methods NGCF and LightGCN achieve significant improvement. It demonstrates the effectiveness of capturing the high-order neighboring information to learn users/items representations. Meanwhile, NGCF and LightGCN are comparable in most cases, it might be that the key idea of their information propagating layers are similar. (2) The cross-domain methods are superior to corresponding single-domain methods (e.g., CoNet with NeuMF, BiTGCF with LightGCN), which demonstrates that designing different transferring strategiesfor CDR is better than using one single network to model the mixed dataset. (3) Compared with BPRMF and NeuMF, CoNet gains better performance over them, which demonstrates that transferring knowledge between two base networks is a promising transferring strategy. (4) DDTCDR reaches better performance than CoNet, which indicates that transferring the similarity between users could be a better transfer strategy than directly transferring user-item interaction features. (5) Moreover, PPGN and BiTGCF achieve significant improvement than other methods, which demonstrates that transferring knowledge

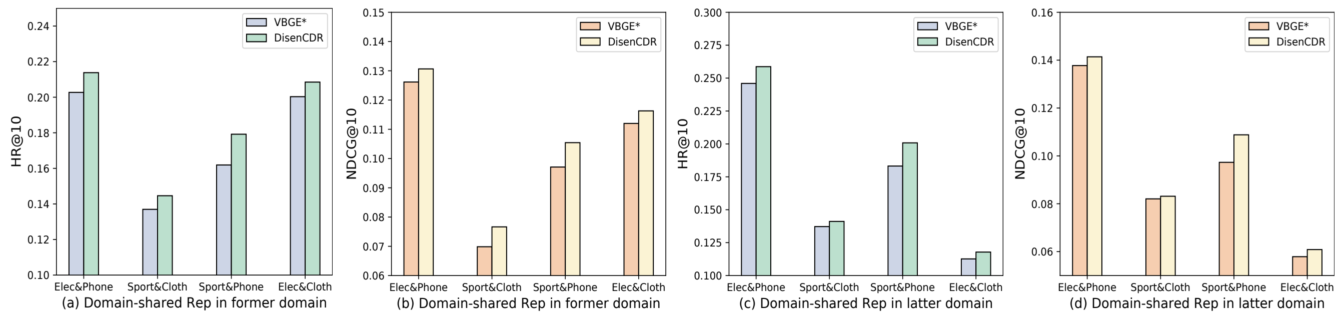


**Figure 5: The predictive ability of domain-shared representations with different objective. The (a) and (b) show the former domain results (e.g., Elec of Elec&Phone), (c) and (d) show the latter domain results (e.g., Cloth of Sport&Cloth).**

**Table 5: The impact of observed interaction in both domains.**

| #Elec_Inter | #Cloth_Inter | VBGE* | | DisenCDR | |
|---|---|---|---|---|---|
| | | HR | NDCG | HR | NDCG |
| 1-10 | 1-10 | 22.93 | 13.02 | 23.46 | 13.74 |
| | 11-20 | 24.88 | 14.34 | 25.49 | 14.82 |
| | 21-30 | 30.06 | 20.23 | 31.28 | 21.02 |
| | >30 | 30.88 | 19.79 | 32.35 | 20.45 |
| 11-20 | 1-10 | 21.35 | 12.01 | 21.88 | 12.16 |
| | 11-20 | 26.42 | 14.73 | 27.46 | 15.32 |
| | 21-30 | 22.33 | 13.45 | 24.27 | 15.15 |
| | >30 | 24.32 | 14.24 | 27.02 | 18.58 |
| 21-30 | 1-10 | 19.88 | 11.31 | 20.72 | 11.53 |
| | 11-20 | 30.81 | 15.71 | 31.35 | 16.82 |
| | 21-30 | 35.71 | 17.84 | 35.71 | 18.13 |
| | >30 | 16.66 | 6.93 | 16.66 | 9.25 |
| >30 | 1-10 | 19.73 | 10.20 | 20.26 | 11.00 |
| | 11-20 | 21.03 | 11.78 | 22.68 | 12.06 |
| | 21-30 | 15.49 | 8.14 | 16.90 | 8.75 |
| | >30 | 16.32 | 12.40 | 20.40 | 14.82 |

between GNNs is a strong transferring strategy for CDR. (6) Compared with state-of-the-art baselines, our DisenCDR consistently yields the best performances on these datasets for all metrics, which reveals that learning disentangled representations and transferring the user domain-shared representations is a more powerful transfer strategy.

## 4.4 Discussion of Model Variants (RQ2)

In this section, we construct several variants and conduct experiments on four CDR scenarios. Specifically, to validate our transferring framework the extensibility and effectiveness. We adapt the PPGN into our framework, named VPPGN, and then we employ VPPGN to replace VBGE and train them with two different objectives. **We use '*' and '#' to denote the variant using standard ELBO or our disentanglement objective, respectively.** The experimental results are reported in Table 3, and we can draw the following observations. (1) For VPPGN variants, it is obvious that VPPGN# outperforms BiTGCF, which demonstrates our transfer strategy is effective for other graph encoders. Further, VPPGN# gains stability improvement from VPPGN*, indicating our objective is helpful to learn user domain-shared representations to transfer knowledge. (2) The VBGE variants show more robust performance than VPPGN variants. We suppose the reason lies in that VBGE only aggregates user-level homogenous information which can learn more fine-grained domain-shared and domain-specific representations. Moreover, comparing VBGE*, our DisenCDR further yields satisfactory improvement. It also verifies the effectiveness of our disentanglement objective.

## 4.5 Analysis of Disentanglement (RQ3)

In this section, to validate that our disentanglement objective is able to learn the domain-shared and domain-specific representations for users, we conduct an analysis between DisenCDR and VBGE* (The VBGE* only uses the standard ELBO as the objective function). To be specific, we directly calculate average KL divergence to measure the mutual information in Eq.(10) after the training procedure of both models is finished (higher KL divergence means lower mutual information). According to Table 4, it is obvious that the KL divergence score of DisenCDR is much higher than the VBGE* variant, which reflects our DisenCDR achieves better disentanglement between domain-shared and domain-specific representations.

Besides, to investigate the impact of our disentanglement objective, we conduct another comparison between VBGE* and our DisenCDR. As shown in Figure 5, this ablation study tests the predictive ability of the trained domain-shared representation which plays the central role to transfer knowledge across domains. Specifically, for VBGE* and our DisenCDR, **we only leverage the domain-shared representations to predict the target items** for each domain in the test stage. From Figure 5, we have the following observations: (1) The learned domain-shared representations from our framework always show robust recommendation performance, which reflects that transferring the domain-shared information is a powerful strategy in the CDR task. (2) The domain-shared representations of DisenCDR show consistent improvements than the domain-shared representations of VBGE*. It demonstrates that our disentanglement objective could learn enhanced domain-shared representation to transfer knowledge across domains.

Moreover, to analyze our objective effectiveness regarding the number of observed interactions, we further conduct experiments on groups with different amounts of source interactions. In Table 5, we report the Elec domain results from Elec&Cloth scenario in

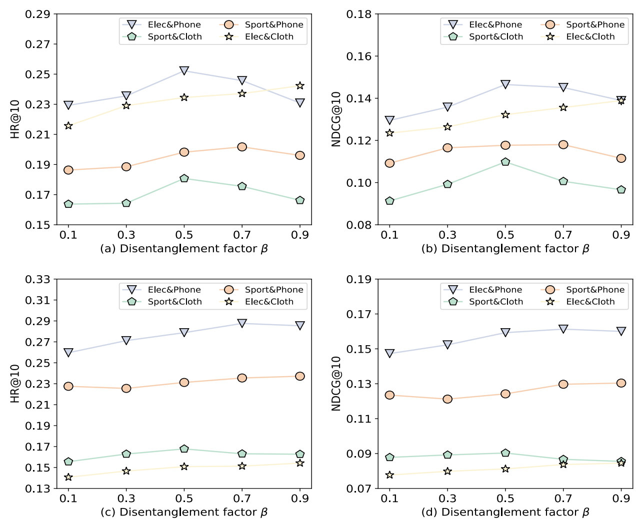


**Figure 6: Results of disentanglement factor $\beta$.**

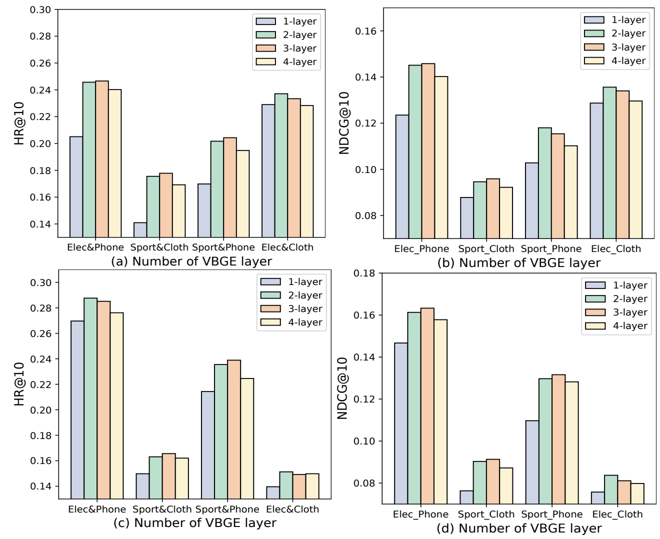


**Figure 7: Impact of VBGE layer number.**

terms of the four different interaction number groups. For each Elec group, we also divide four sub-groups in terms of interaction number in Cloth domain. From it, we can observe that: (1) For each Elec group, more interactions shown in Cloth domain are helpful for better recommendation, especially in the data sparsity group, e.g., 1-10 in Elec. (2) DisenCDR shows statistic improvements than VBGE*, and especially in the cases that Cloth domain contains diverse information, i.e., >30 interactions in each sub-groups. This fact indicates that our disentanglement objective has the ability to recognize the domain-shared information from diverse user preferences.

## 4.6 Parameter Sensitivity (RQ4)

In this section, we investigate two hyper-parameter sensitivity: the disentanglement factor $\beta$ and the number of VBGE layers on the recommendation performance. The Figure 6(a)(b) and 7(a)(b) show results of former domains (e.g., Elec of Elec&Phone), and the (c)(d) for latter domain. For the disentanglement factor $\beta$, according to Figure 6, we find $\beta$ can be set as a larger value for larger interaction scenarios, such as Elec&Cloth. The reason might be that the larger interaction scenario contains more domain-specific information which should increase the factor to reach better disentanglement. For the number of VBGE layers, according to Figure 7, the 1-layer variant always shows the lowest performance, which demonstrates aggregating higher-order neighboring information helpful to learn robust representations. Further, 2-layer and 3-layer variants achieve best results than the 4-layer variant. The reason might be that the deep graph neural networks easily cause the over-smoothing issue. In addition, the experiment results are stable which verifies that our method is robust to the change of the two parameters.

# 5 RELATED WORK

## 5.1 Cross-Domain Recommendation

Following the collaborative filtering (CF) paradigm, a significant amount of studies [3, 11, 12, 49, 55] have been proposed to build an effective yet efficient recommender system, to capture the dynamic user interests from massive user-item interaction networks [2]. Although existing CF-based methods achieve promising $top-k$ results to some extent, but they severely suffer from the cold-start and data sparsity issues. To alleviate such issues, the progress of CDR [27, 31, 43, 48, 54, 60, 62, 63] can be roughly divided into the following two branches: (1) recommendation items for cold-start users and (2) recommendation items for shared users. Since the total different research purposes for the two group users, therefore, the previous works learning idea of them are completely in distinct ways.

**To recommend items for cold-start users**, several CDR approaches have been proposed to model different correlations between source and target domains, such as CBMF [38], DSN [18], EMCDR [36], SSCDR [19], CATN [59] and CDRIB [4]. To be specific, CBMF learns the cluster-level information of users and items across domains. DSN learns source representations to predict in the target domain by the domain adaption technique. EMCDR and SSCDR propose a general CDR framework to a learn user feature mapping function between source and target domain. CATN develops a review-based model to match user’s aspect-level preferences across domains. CDRIB is the most related work, which attempts capturing the domain-shared information via information bottleneck principle [45]. Nevertheless, these methods aim to make recommendation for the non-overlapped (cold-start) users in target domain, having fundamental different task definition with this work.

**To recommend items for shared users**, most typical CDR approaches, including shallow models and deep models, focus on transferring knowledge from the source domain to the target domain. Specifically, shallow models, such as CMF [44], CDFM [32] and CDCF [15, 25], are based on Matrix Factorization [12], Factorization Machines [9, 40] or CF paradigm, and try to effectively transfer knowledge by utilizing source domain interaction as additional features. Additionally, several neural network models have been proposed, such as CoNet [14], DDTCDR [26], PPGN [58] and BiTGCF [30]. CoNet transfers knowledge across domains by using the cross-connections network between feed-forward MLPs. DDTCDR learns a latent orthogonal mapping function of shared users to transfer user preferences across domains. Besides, the DARec [56] transfers knowledge across domain by the domain adaptive technique, but fail to learn the disentanglement representations. The PPGN and BiTGCF are two state-of-the-art graph neural network [53] based models, PPGN simply stacking several graph convolution networks [23] on two domains and transfer knowledge by sharing common user hidden features. BiTGCF also aims to transfer the domain-shared information. But it devises a feature transfer layer between domain-specific representations, which is hard to disentangle the domain-specific and domain-shared information.

Our DisenCDR belongs to the latter branch and aims to transfer the domain-shared information across domains, which reaches state-of-the-art performance and learns disentangled representations.

## 5.2 Disentangled Representation Learning

Disentangled representation learning focuses on factorizing the unobservable structural factors from data, which achieves great success on many downstream tasks [8, 34, 42, 50]. Following variational inference [7] based on VAE [20], the well-known method $\beta$-VAE [13] demonstrates that penalizing the KL divergence term in the ELBO can easily derive the disentanglement representations. Based on $\beta$-VAE, extensive researches of learning disentangled representation are proposed by introducing extra regularization terms to ELBO, such as total correlation term [51] and information bottleneck term [1, 45]. Although existing learning disentangled representation approaches made majority contributions in the computer vision field [5, 8, 16], it is not studied over graph-structured data. In recent years, some works try to learn disentangled representation of users and items for different recommendation tasks, such as Top-k [33, 34, 50] and sequential [35] recommendation. Different from these works, DisenCDR focus on learning disentangled representation for cross-domain recommendation task, which is a challenging and unexplored area.

## 6 CONCLUSION

In this paper, we present a novel deep generation method named DisenCDR. Different from previous works which entangle the domain-shared information and domain-specific information, our model utilizes two mutual-information-based regularizers to disentangle these two types of information. Specifically, the exclusive regularizer aims to enforce the user domain-shared representations and domain-specific representations encoding exclusive information. The information regularizer is to encourage the user domain-shared representations encoding predictive information for both domains. Empirically experimental results on four real-world datasets demonstrate the effectiveness of DisenCDR on cross-domain recommendation against several state-of-the-art baseline models. Besides, detailed analyses in various CDR datasets show the effectiveness of our transferring strategy and robustness of our model components. In the future, we will investigate learning disentanglement representations for multi-domain recommendation.

## ACKNOWLEDGEMENT

We would like to thank Shiyao Cui, Jiawei Sheng, Yanzeng Li, Xinghua Zhang, Bowen Yu, Zhenyu Zhang and Shicheng Wang for their helpful discussions and feedback of this work. We would also like to thank anonymous reviewers for their constructive comments. This work was supported by the National Key Research and Development Program of China under Grant No.2021YFB3100600, the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No.XDC02040400, and the Youth Innovation Promotion Association of CAS under Grant No.2021153.

## REFERENCES


[1] Alexander A Alemi, Ian Fischer, Joshua V Dillon, and Kevin Murphy. 2017. Deep Variational Information Bottleneck. In *International Conference on Learning Representations (ICLR)*.

[2] Jiangxia Cao, Xixun Lin, Xin Cong, Shu Guo, Hengzhu Tang, Tingwen Liu, and Bin Wang. 2021. Deep Structural Point Process for Learning Temporal Interaction Networks. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*.

[3] Jiangxia Cao, Xixun Lin, Shu Guo, Luchen Liu, Tingwen Liu, and Bin Wang. 2021. Bipartite Graph Embedding via Mutual Information Maximization. In *ACM International Conference on Web Search and Data Mining (WSDM)*.

[4] Jiangxia Cao, Jiawei Sheng, Xin Cong, Tingwen Liu, and Bin Wang. 2022. Cross-Domain Recommendation to Cold-Start Users via Variational Information Bottleneck. In *IEEE International Conference on Data Engineering (ICDE)*.

[5] Ricky TQ Chen, Xuechen Li, Roger Grosse, and David Duvenaud. 2018. Isolating Sources of Disentanglement in Variational Autoencoders. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*.

[6] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *ArXiv* (2014).

[7] Samuel Gershman and Noah Goodman. 2014. Amortized Inference in Probabilistic Reasoning. In *Proceedings of the Annual Meeting of The Cognitive Science Society*.

[8] Abel Gonzalez-Garcia, Joost Van De Weijer, and Yoshua Bengio. 2018. Image-to-Image Translation for Cross-Domain Disentanglement. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*.

[9] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, Xiuqiang He, and Zhenhua Dong. 2017. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

[10] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *The World Wide Web Conference (WWW)*.

[11] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and Powering Graph Convolution Network for Recommendation. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.

[12] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *The World Wide Web Conference (WWW)*.

[13] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. 2017. beta-vae: Learning Basic Visual Concepts with a Constrained Variational Framework. In *International Conference on Learning Representations (ICLR)*.

[14] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. Conet: Collaborative Cross Networks for Cross-Domain Recommendation. In *ACM International Conference on Information and Knowledge Management (CIKM)*.

[15] Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Can Zhu. 2013. Personalized Recommendation via Cross-Domain Triadic Factorization. In *The World Wide Web Conference (WWW)*.

[16] HyeongJoo Hwang, Geon-Hyeong Kim, Seunghoon Hong, and Kee-Eung Kim. 2020. Variational Interaction Information Maximization for Cross-domain Disentanglement. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*.

[17] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* (2002).

[18] Heishiro Kanagawa, Hayato Kobayashi, N. Shimizu, Yukihiro Tagami, and Taiji Suzuki. 2019. Cross-domain Recommendation via Deep Domain Adaptation. In *European Conference on IR Research (ECIR)*.

[19] SeongKu Kang, Junyoung Hwang, Dongha Lee, and Hwanjo Yu. 2019. Semi-Supervised Learning for Cross-Domain Recommendation to Cold-Start Users. In *ACM International Conference on Information and Knowledge Management (CIKM)*.

[20] Diederik P Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. In *International Conference on Learning Representations (ICLR)*.

[21] P. Diederik Kingma and Lei Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations (ICLR)*.

[22] Thomas N Kipf and Max Welling. 2016. Variational Graph Auto-Encoders. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*.

[23] Thomas N Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.

[24] Walid Krichene and Steffen Rendle. 2020. On sampled metrics for item recommendation. In *ACM Knowledge Discovery and Data Mining (KDD)*.

[25] Bin Li, Qiang Yang, and Xiangyang Xue. 2009. Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

[26] Pan Li and Alexander Tuzhilin. 2020. DDTCDR: Deep Dual Transfer Cross Domain Recommendation. In *ACM International Conference on Web Search and Data Mining (WSDM)*.

[27] Siqing Li, Liuyi Yao, Shanlei Mu, Wayne Xin Zhao, Yaliang Li, Tonglei Guo, Bolin Ding, and Ji-Rong Wen. 2021. Debiasing Learning based Cross-domain Recommendation. In *ACM Knowledge Discovery and Data Mining (KDD)*.

[28] Xiaopeng Li and James She. 2017. Collaborative variational autoencoder for recommender systems. In *ACM Knowledge Discovery and Data Mining (KDD)*.

[29] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. 2018. Variational autoencoders for collaborative filtering. In *The World Wide Web Conference (WWW)*.

[30] Meng Liu, Jianjun Li, Guohui Li, and Peng Pan. 2020. Cross Domain Recommendation via Bi-directional Transfer Graph Collaborative Filtering Networks. In *ACM International Conference on Information and Knowledge Management (CIKM)*.

[31] Ziqi Liu, Yue Shen, Xiaocheng Cheng, Qiang Li, Jianping Wei, Zhiqiang Zhang, Dong Wang, Xiaodong Zeng, Jinjie Gu, and Jun Zhou. 2021. Learning Representations of Inactive Users: A Cross Domain Approach with Graph Neural Networks. In *ACM International Conference on Information and Knowledge Management (CIKM)*.

[32] Babak Loni, Yue Shi, Martha Larson, and Alan Hanjalic. 2014. Cross-Domain Collaborative Filtering with Factorization Machines. In *European Conference on IR Research (ECIR)*.

[33] Jianxin Ma, Peng Cui, Kun Kuang, Xin Wang, and Wenwu Zhu. 2019. Disentangled Graph Convolutional Networks. In *International Conference on Machine Learning (ICML)*.

[34] Jianxin Ma, Chang Zhou, Peng Cui, Hongxia Yang, and Wenwu Zhu. 2019. Learning Disentangled Representations for Recommendation. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*.

[35] Jianxin Ma, Chang Zhou, Hongxia Yang, Peng Cui, Xin Wang, and Wenwu Zhu. 2020. Disentangled Self-Supervision in Sequential Recommenders. In *ACM Knowledge Discovery and Data Mining (KDD)*.

[36] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach.. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

[37] Kelong Mao, Jieming Zhu, Jinpeng Wang, Quanyu Dai, Zhenhua Dong, Xi Xiao, and Xiuqiang He. 2021. SimpleX: A Simple and Strong Baseline for Collaborative Filtering. In *ACM International Conference on Information and Knowledge Management (CIKM)*.

[38] Nima Mirbakhsh and Charles X Ling. 2015. Improving Top-n Recommendation for Cold-Start Users via Cross-Domain Information. *ACM Transactions on Knowledge Discovery from Data (TKDD)* (2015).

[39] Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. 2019. On Variational Bounds of Mutual Information. In *International Conference on Machine Learning (ICML)*.

[40] Steffen Rendle. 2010. Factorization Machines. In *IEEE International Conference on Data Mining (ICDM)*.

[41] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. *International Conference on Uncertainty in Artificial Intelligence (UAI)* (2009).

[42] Eduardo Hugo Sanchez, Mathieu Serrurier, and Mathias Ortner. 2020. Learning Disentangled Representations via Mutual Information Estimation. In *European Conference on Computer Vision (ECCV)*.

[43] Xiang-Rong Sheng, Liqin Zhao, Guorui Zhou, Xinyao Ding, Binding Dai, Qiang Luo, Siran Yang, Jingshan Lv, Chi Zhang, Hongbo Deng, et al. 2021. One Model to Serve All: Star Topology Adaptive Recommender for Multi-Domain CTR Prediction. In *ACM International Conference on Information and Knowledge Management (CIKM)*.

[44] Ajit P Singh and Geoffrey J Gordon. 2008. Relational Learning via Collective Matrix Factorization. In *ACM Knowledge Discovery and Data Mining (KDD)*.

[45] Naftali Tishby and Noga Zaslavsky. 2015. Deep Learning and The Information Bottleneck Principle. In *2015 IEEE Information Theory Workshop (ITW)*.

[46] Quoc-Tuan Truong, Aghiles Salah, and Hady W Lauw. 2021. Bilateral Variational Autoencoder for Collaborative Filtering. In *ACM International Conference on Web Search and Data Mining (WSDM)*.

[47] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations (ICLR)*.

[48] Chen Wang, Yueqing Liang, Zhiwei Liu, Tao Zhang, and Philip S Yu. 2021. Pre-training Graph Neural Network for Cross Domain Recommendation. *ArXiv* (2021).

[49] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.

[50] Xiang Wang, Hongye Jin, An Zhang, Xiangnan He, Tong Xu, and Tat-Seng Chua. 2020. Disentangled Graph Collaborative Filtering. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.

[51] Satosi Watanabe. 1960. Information Theoretical Analysis of Multivariate Correlation. *IBM Journal of Research and Development* (1960).

[52] Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. 2019. Simplifying graph convolutional networks. In *International Conference on Machine Learning (ICML)*.

[53] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. 2020. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)* (2020).

[54] Kun Xu, Yuanzhen Xie, Liang Chen, and Zibin Zheng. 2021. Expanding Relationship for Cross Domain Recommendation. In *ACM International Conference on Information and Knowledge Management (CIKM)*.

[55] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph Convolutional Neural Networks for Web-Scale Recommender Systems. In *ACM Knowledge Discovery and Data Mining (KDD)*.

[56] Feng Yuan, Lina Yao, and Boualem Benatallah. 2019. DARec: Deep Domain Adaptation for Cross-Domain Recommendation via Transferring Rating Patterns. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

[57] Tianzi Zang, Yanmin Zhu, Haobing Liu, Ruohan Zhang, and Jiadi Yu. 2021. A Survey on Cross-domain Recommendation: Taxonomies, Methods, and Future Directions. *ArXiv* (2021).

[58] Cheng Zhao, Chenliang Li, and Cong Fu. 2019. Cross-Domain Recommendation via Preference Propagation GraphNet. In *ACM International Conference on Information and Knowledge Management (CIKM)*.

[59] Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, and Aixin Sun. 2020. CATN: Cross-Domain Recommendation for Cold-Start Users via Aspect Transfer Network. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.

[60] Feng Zhu, Yan Wang, Chaochao Chen, Guanfeng Liu, and Xiaolin Zheng. 2020. A Graphical and Attentional Framework for Dual-Target Cross-Domain Recommendation.. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

[61] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-Domain Recommendation: Challenges, Progress, and Prospects. *ArXiv* (2021).

[62] Yongchun Zhu, Kaikai Ge, Fuzhen Zhuang, Ruobing Xie, Dongbo Xi, Xu Zhang, Leyu Lin, and Qing He. 2021. Transfer-Meta Framework for Cross-domain Recommendation to Cold-Start Users. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.

[63] Yongchun Zhu, Zhenwei Tang, Yudan Liu, Fuzhen Zhuang, Ruobing Xie, Xu Zhang, Leyu Lin, and Qing He. 2022. Personalized Transfer of User Preferences for Cross-domain Recommendation. In *ACM International Conference on Web Search and Data Mining (WSDM)*.