Disentangled Contrastive Bundle Recommendation with Conditional Diffusion

Jiuqiang Li[1,2]

[1]School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China
[2]Engineering Research Center of Sustainable Urban Intelligent Transportation, Ministry of Education, China
jiuqiangli@outlook.com

Abstract

Bundle recommendation aims to improve user experience by suggesting complementary items that users are likely to purchase together. Although recent advances in recommendation systems have shown promise, there are still significant challenges: i) The dynamic nature of user preferences and interactions introduces noise that can distort the effectiveness of recommendations. ii) Existing methods frequently exhibit limited robustness when addressing the sparsity of user interactions with bundles in real-world scenarios. To tackle these issues, we introduce a disentangled contrastive bundle recommendation (DCBR) framework with conditional diffusion. First, we propose a conditional bundle diffusion model for denoising the user-bundle interaction graph, introducing a bundle latent consistency constraint during the optimization process to mitigate the degradation of original interaction information. Subsequently, we design a triple-view denoised graph learning module to obtain effective representations from multiple views. Furthermore, we present a dual-level disentangled contrastive learning paradigm, which addresses the latent relationships at two levels: between views (inter-view) and within each view (intra-view). By maximizing the consistency between positive samples in these contrastive views, we generate disentangled contrastive signals, overcoming interaction sparsity and alleviating noise issues. Our experimental evaluations on three benchmark datasets reveal that DCBR significantly outperforms state-of-the-art methods.

Code — https://github.com/recomall/DCBR

Introduction

In recent years, the field of recommendation systems has evolved significantly, with a particular focus on improving user experience through enhanced item suggestions. Bundle recommendation, which aims to recommend complementary items that users are likely to purchase together, has emerged as a promising area of research (Chen et al. 2019a; Chang et al. 2020; Ma et al. 2022).

Early research on bundle recommendation (Rendle, Freudenthaler, and Schmidt-Thieme 2010) has been viewed as a special form of user-item recommendation, with conventional solutions typically employing Collaborative Filtering (CF) methods to analyze user interactions with bundles. However, the bundle recommendation context includes information such as user-bundle interactions, user-item interactions, and bundle-item affiliations, while these approaches primarily focus on the user-bundle interaction space, neglecting other views. Based on this, some research methods have considered user-item interactions and bundle-item associations to fully utilize the valuable information provided by the scenario (Sun et al. 2024). Factorization models (Cao et al. 2017; Chen et al. 2019a) and Graph Neural Networks (GNNs) (Deng et al. 2020; Chang et al. 2020, 2021) are some commonly used techniques that have proven to be effective in handling complex high-order relationships between users, bundles, and items. However, the highly sparse interaction space in real-world contexts restricts the ability of GNNs to effectively model intricate user preferences in a fully supervised manner. To address data sparsity in recommendations, an approach is to leverage Self-Supervised Learning (SSL) to extract features from unlabeled user behavior data (Wu et al. 2021; Yu et al. 2022a, 2023; Jeon et al. 2024), which constructs self-supervised contrastive views by methods such as randomly dropping nodes, dropping edges, random walks, and adding random noise. Recent studies have explored the integration of SSL to improve bundle recommendation, with examples such as MIDGN (Zhao et al. 2022), CrossCBR (Ma et al. 2022), EBRec (Du et al. 2023), and MultiCBR (Ma et al. 2024). SSL-based bundle recommendation methods primarily aim to leverage relationships between multiple views to construct self-supervised tasks, alleviating the challenges posed by insufficient supervision due to sparse interactions.

Despite the widespread application of SSL in bundle recommendation, several limitations persist: i) In bundle recommendation, factors such as user behavior uncertainty, including erroneous clicks on bundles, inevitably introduce noise that can significantly mislead the model's learning process, thereby affecting the quality of the final recommendations. ii) In real-world scenarios, existing methods often demonstrate poor robustness in the face of sparse user-bundle interactions. Although previous work has attempted to integrate SSL into bundle recommendation, some studies typically rely on directly using the final fused user (bundle) representations for contrastive learning. This approach inevitably introduces confounding noise due to semantic gaps between different views, resulting in suboptimal contrastive signals. To overcome the aforementioned issues, we propose a novel disentangled contrastive bundle recommendation (**DCBR**) framework with conditional diffusion. Specifically, inspired by the outstanding performance of diffusion models (Ho, Jain, and Abbeel 2020; Wang et al. 2023) in data denoising, we aim to introduce diffusion models into the bundle recommendation domain to effectively remove or mitigate the inevitable latent noise in the user-bundle interaction graph. However, conventional diffusion models applied directly for denoising graph data may mistakenly filter out genuine interaction information as noise. To address this issue, we propose a conditional bundle diffusion model specifically designed to denoise user-bundle interaction data. To alleviate the degradation of the original interaction information during the denoising process, we introduce a bundle latent consistency constraint to maximize the consistency of the latent bundle representations between the generated denoised views and their original counterparts. To capture effective higher-order collaborative relationships among user-bundle, user-item, and bundle-item interactions, we design the triple-view denoised graph learning module. This module adaptively fuses the representations of users and bundles across the multiple views using tunable parameters for subsequent preference score calculations of users towards bundles. Moreover, to mitigate the introduction of confounding noise due to semantic discrepancies between different views during the adaptive fusion process, we propose a dual-level disentangled contrastive learning paradigm. The dual levels represent the exploration of potential relationships between multiple views (inter-view) and within a single view (intra-view) in the recommendation task. At the inter-view level, we enhance the shared node features between views by learning the disentangled relationships among different views. At the intra-view level, we rely on the comparative analysis of local features within a single view to increase each view's sensitivity to local interactions. "Disentangled" can be understood as constructing contrastive views utilizing the original features before fusion of features. By maximizing the similarity between positive samples in these contrastive views and minimizing the similarity between negative samples, we generate disentangled contrastive signals for users (bundles) to address interaction sparsity while alleviating confounding noise.

To summarize, the key contributions of our research are outlined as follows:

- We design a conditional bundle diffusion model for denoising the core user-bundle interaction graph in recommendation tasks. The bundle latent consistency constraint effectively balances the reduction of useful interaction information and the denoising learning capability.
- We propose a dual-level disentangled contrastive learning paradigm, which effectively avoids the formation of semantic noise during the multi-view feature fusion process and provides robust auxiliary contrastive signals for recommendation tasks.
- The experimental results on three public datasets validate the performance improvement and effectiveness of our proposed **DCBR** in bundle recommendation.

Related Work

Self-Supervised Learning for Recommendation. Self-supervised learning (SSL) has proven to be an effective solution to address the issue of scarce labels in recommendation systems (Wu et al. 2021). Popular approaches utilize unlabeled data from user-item interactions to generate additional self-supervised signals, enhancing the original supervised learning tasks. For example, SimGCL (Yu et al. 2022a), XSimGCL (Yu et al. 2023), NCL (Lin et al. 2022) and LightGCL (Cai et al. 2023) employ various graph augmentation techniques, such as random edge dropout, node dropout, and semantic neighbor identification, to generate self-supervised signals by contrasting positive node pairs. In the bundle recommendation, MultiCBR (Ma et al. 2024) advocates self-contrastive learning on the fused multi-view representations. Conversely, we propose a novel dual-level disentangled contrastive learning paradigm that combines global information between views with local representations within views to achieve robust user preference learning.

Recommendation with Diffusion Models. Diffusion Models (DMs) (Ho, Jain, and Abbeel 2020; Sohl-Dickstein et al. 2015) have excelled in various fields, such as image generation (Epstein et al. 2023) and inpainting (Lugmayr et al. 2022) in the visual domain, as well as text generation (Austin et al. 2021) in natural language processing. Recently, DMs have been extensively utilized in recommender systems, exemplified by approaches such as DiffRec (Wang et al. 2023), GDSSL (Li and Wang 2024), DiffKG (Jiang et al. 2024), and DDRM (Zhao et al. 2024). DiffKG employs generative diffusion models as a data augmentation technique to enhance representation learning in knowledge graphs, while DDRM enhances the robustness of user and item embeddings through a multi-step denoising process to address noisy implicit feedback. In contrast, our conditional bundle diffusion model introduces a bundle latent consistency constraint designed to preserve the original user-bundle interaction information during the denoising process.

Bundle Recommender Systems. Bundle recommendation aims to model user preferences for bundled items and accordingly recommend predefined bundles to potentially interested users. Inspired by the outstanding performance of Graph Convolutional Networks (GCNs) (Kipf and Welling 2017) in representation learning, BGCN (Chang et al. 2020) utilizes GCNs to capture user preferences at the bundle and item levels through a dual view approach, focusing on the user-bundle interaction graph and the bundle-item association graph. With the introduction of contrastive learning in recommendation systems, MIDGN (Zhao et al. 2022) separates user-bundle preferences into local and global views, applying contrastive loss between these two views. CrossCBR (Ma et al. 2022) utilizes contrastive learning in cross views to improve the similarity of representations for the same node. BundleGT (Wei et al. 2023) designs a hierarchical graph transformer to model strategy-based representations for bundles and users. MultiCBR (Ma et al. 2024) performs self-supervised contrastive learning after fusion of multi-view representations. In contrast, our **DCBR** leverages DMs to integrate contrastive learning, enhancing the denoising of relation learning in bundle recommendation.

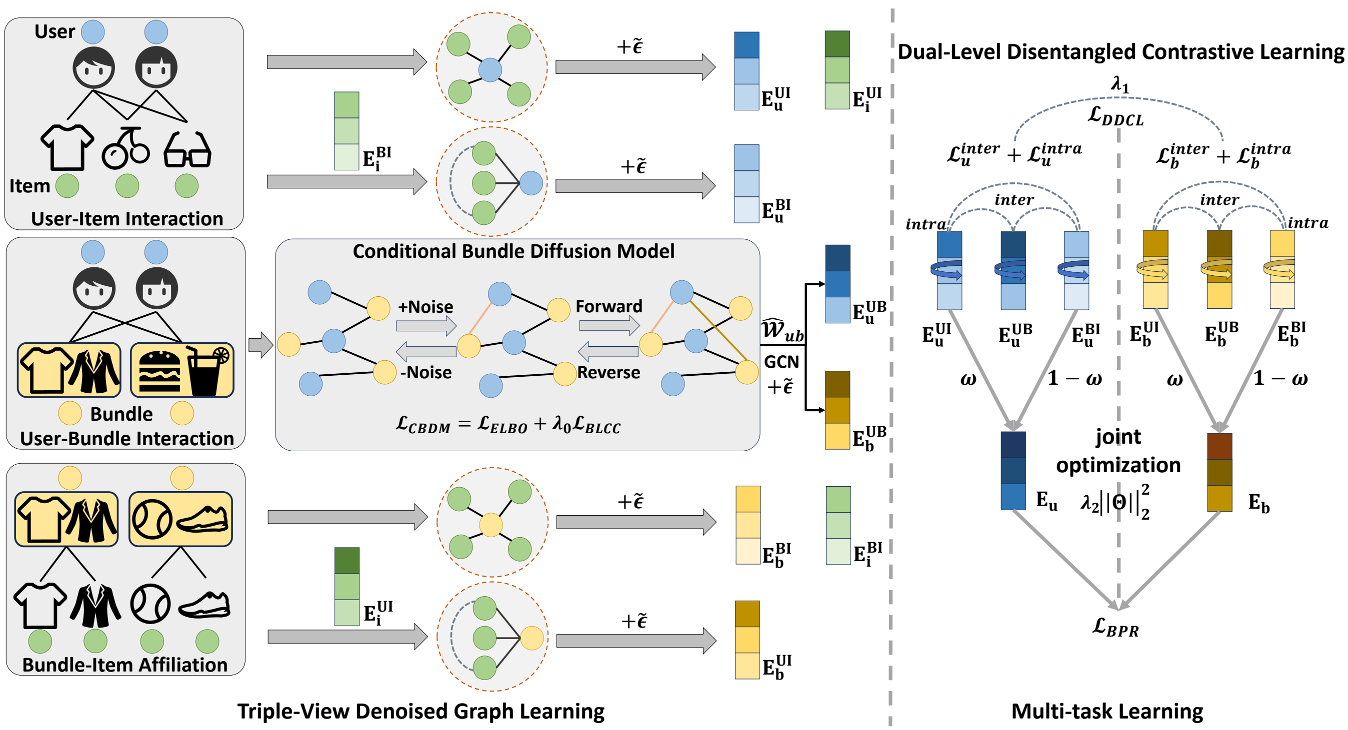


Figure 1: Architecture of our proposed disentangled contrastive bundle recommendation with conditional diffusion.

Preliminary

In this section, we recapitulate the fundamental concepts of the groundbreaking DDPM in establishing the diffusion model. The primary objective of the DDPM parameterized by ϕ is to characterize the data-generating distribution of the target data x_0, denoted as $p_\phi(x_0)$. In the forward process, DDPM progressively introduces Gaussian noise to x_0 with a variance schedule of $[\beta_1, \cdots, \beta_t, \cdots, \beta_T]$:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t} x_0, (1-\bar{\alpha}_t)\mathbf{I}), \quad \bar{\alpha}_t = \prod_{t'=1}^{t}(1-\beta_{t'}). \quad (1)$$

In the reverse process, the denoised data $\hat{x}_\phi$ are generated through the learned parameter ϕ. Formularily,

$$p_\phi(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\phi(x_t,t), \Sigma_\phi(x_t,t)). \quad (2)$$

Moreover, to cater to the domain of recommendation systems, DiffRec builds upon DDPM by designing denoising optimization objectives:

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{t\sim U(1,T)}\left(\mathbb{E}_{q(x_t|x_0)}\left(\hat{\alpha}_t \left\|\hat{x}_\phi(x_t,t) - x_0\right\|_2^2\right)\right),$$
$$\hat{\alpha}_t = \frac{\bar{\alpha}_{t-1} - \bar{\alpha}_t}{2\left(1-\bar{\alpha}_{t-1}\right)\left(1-\bar{\alpha}_t\right)}. \quad (3)$$

This work mainly explores how to effectively adapt the diffusion model to bundle recommendation.

Methodology

Task Formulation

In the bundle recommendation scenario, we define the user set as $\mathcal{U} = \{u_1, u_2, \cdots, u_M\}$, the bundle set as $\mathcal{B} = \{b_1, b_2, \cdots, b_K\}$, and the item set as $\mathcal{I} = \{i_1, i_2, \cdots, i_N\}$, where M, K, and N represent the sizes of the corresponding sets. Given the user-bundle interaction graph $\mathcal{G}_{ub} = \{(u,b)|u \in \mathcal{U}, b \in \mathcal{B}\}$, the user-item interaction graph $\mathcal{G}_{ui} = \{(u,i)|u \in \mathcal{U}, i \in \mathcal{I}\}$, and the bundle-item affiliation graph $\mathcal{G}_{bi} = \{(b,i)|b \in \mathcal{B}, i \in \mathcal{I}\}$, represented by their adjacency matrices $\mathcal{W}_{ub} \in \mathbb{R}^{M\times K}$, $\mathcal{W}_{ui} \in \mathbb{R}^{M\times N}$, and $\mathcal{W}_{bi} \in \mathbb{R}^{K\times N}$, the goal is to learn a function $\mathcal{F}((u,b)|\Theta)$ to predict the likelihood $\hat{y}_{ub}$ of a user u adopting an unseen bundle b where Θ denotes the learnable parameters of $\mathcal{F}$.

The architecture diagram of our proposed **DCBR** is shown in Figure 1, consisting mainly of learnable embedding parameters $\mathbf{E}_u^{(0)} \in \mathbb{R}^{M\times d}$ for users, $\mathbf{E}_b^{(0)} \in \mathbb{R}^{K\times d}$ for bundles, and $\mathbf{E}_i^{(0)} \in \mathbb{R}^{N\times d}$ for items, all of which are initialized randomly. Here, d represents the embedding size.

Conditional Bundle Diffusion Model

In order to learn denoising for user-bundle interaction graph $\mathcal{G}_{ub}$, our proposed Conditional Bundle Diffusion Model (CBDM) also includes forward and reverse processes as shown in Eq. (1) and Eq. (2), with the target data x_0 being the adjacency matrix $\mathcal{W}_{ub}$ of $\mathcal{G}_{ub}$. Based on prior knowledge, the basic optimization objective of the model is $\mathcal{L}_{\text{ELBO}}$, and the denoised output of the model is $\hat{\mathcal{W}}_{ub}$.

General diffusion models directly applied to the user-bundle interaction graph $\mathcal{G}_{ub}$, without alignment with the recommendation task, fail to achieve substantial denoising effects. Therefore, we propose the Bundle Latent Consistency Constraint (BLCC) to ensure that the denoised graph $\hat{\mathcal{G}}_{ub}$ more accurately captures the true interaction scenario.

Specifically, CBDM samples all bundles for batched users, with the embeddings of the batched users represented as $\tilde{\mathbf{E}}_u^{(0)} \in \mathbb{R}^{B\times d}$, where B denotes the batch size. BLCC aims to maximize the consistency between the latent bundle representations of the generated denoised view and its original view, alleviating the degradation of original interaction information during the denoising process. Formally,

$$\mathcal{L}_{\texttt{BLCC}} = \left\| \tilde{\mathcal{W}}_\phi^\top \cdot \tilde{\mathbf{E}}_u^{(0)} - \mathbf{E}_b^{(0)} \right\|_2^2, \tag{4}$$

where $\tilde{\mathcal{W}}_\phi \in \mathbb{R}^{B\times K}$ represents the denoised interactive adjacency matrix between a batch of users and all bundles. Based on batch training and inference, $\hat{\mathcal{W}}_{ub}$ can be obtained by concatenating $\tilde{\mathcal{W}}_\phi$ from all batches. Ultimately, the optimization objective for CBDM can be represented as follows:

$$\arg\min_\phi \mathcal{L}_{\texttt{CBDM}} = \mathcal{L}_{\texttt{ELBO}} + \lambda_0 \mathcal{L}_{\texttt{BLCC}}. \tag{5}$$

Here, λ_0 is used to control the contribution of the BLCC loss relative to the ELBO loss.

Triple-View Denoised Graph Learning

To effectively capture higher-order collaborative relationships across user-bundle (UB), user-item (UI), and bundle-item (BI) interactions, we design the triple-view denoised graph learning module utilizing graph neural networks. For simplicity, we define $\mathcal{W}_x = \{\hat{\mathcal{W}}_{ub}, \mathcal{W}_{ui}, \mathcal{W}_{bi}\}$ to represent the corresponding graph structures. Inspired by previous work (Yu et al. 2022a; Ma et al. 2024), we incorporate random noise perturbation $\tilde{\epsilon}_x^{(l)}$ at each propagation layer l to strengthen robustness against triple-view noise. Specifically,

$$\tilde{\epsilon}_x^{(l)} = \upsilon_x \cdot \mathrm{sign}(\mathbf{E}_x^{(l)}) \odot \frac{\epsilon_x^{(l)}}{\left\|\epsilon_x^{(l)}\right\|_2}, \;\; \epsilon_x^{(l)} \in \mathbb{R}^d \sim U(0,1), \tag{6}$$

$$\mathbf{E}_x^{(l+1)} = \left(\mathbf{D}_x^{-\frac{1}{2}} \begin{bmatrix} \mathbf{0} & \mathcal{W}_x \\ \mathcal{W}_x^\top & \mathbf{0} \end{bmatrix} \mathbf{D}_x^{-\frac{1}{2}} \right) \cdot \mathbf{E}_x^{(l)} + \tilde{\epsilon}_x^{(l)}, \tag{7}$$

where $\mathbf{E}_x^{(l)}$ represents the embedding of the corresponding interaction graph structure $\mathcal{W}_x$ after l iterations of graph message passing. The initial embeddings, $\mathbf{E}_{\texttt{UB}}^{(0)}$, $\mathbf{E}_{\texttt{UI}}^{(0)}$, and $\mathbf{E}_{\texttt{BI}}^{(0)}$, are constructed by stacking $\mathbf{E}_u^{(0)}$ with $\mathbf{E}_b^{(0)}$, $\mathbf{E}_u^{(0)}$ with $\mathbf{E}_i^{(0)}$, and $\mathbf{E}_b^{(0)}$ with $\mathbf{E}_i^{(0)}$, respectively. $\mathbf{D}_x$ denotes the diagonal degree matrix of the bidirectional adjacency matrix from $\mathcal{W}_x$, essential for normalization. υ_x is the coefficient that controls the intensity of noise in the corresponding view.

To effectively aggregate embeddings from various layers for complex and diverse scenarios, our **DCBR** uses weighted average pooling to derive the final representations $\mathbf{E}_{\texttt{UB}}$, $\mathbf{E}_{\texttt{UI}}$ and $\mathbf{E}_{\texttt{BI}}$ for the three views, which is expressed as: $\mathbf{E}_x = \sum_{l=0}^{L} \xi_x^{(l)} \mathbf{E}_x^{(l)}$. Here, L represents the number of propagation layers, and $\xi_x^{(l)}$ is the fusion weight associated with the view. Subsequently, $\mathbf{E}_u^{\texttt{UB}} \in \mathbb{R}^{M\times d}$ and $\mathbf{E}_b^{\texttt{UB}} \in \mathbb{R}^{K\times d}$, $\mathbf{E}_u^{\texttt{UI}} \in \mathbb{R}^{M\times d}$ and $\mathbf{E}_i^{\texttt{UI}} \in \mathbb{R}^{N\times d}$, $\mathbf{E}_b^{\texttt{BI}} \in \mathbb{R}^{K\times d}$ and $\mathbf{E}_i^{\texttt{BI}} \in \mathbb{R}^{N\times d}$ are split from $\mathbf{E}_{\texttt{UB}}$, $\mathbf{E}_{\texttt{UI}}$, $\mathbf{E}_{\texttt{BI}}$, respectively.

Furthermore, we extract latent bundle representations $\mathbf{E}_b^{\texttt{UI}}$ in user-item interactions and latent user representations $\mathbf{E}_u^{\texttt{BI}}$ in bundle-item affiliations to enhance the embeddings of relevant nodes. The specific learning process is as follows:

$$\bar{\mathbf{E}}_b^{\texttt{UI}} = \bar{\mathbf{D}}_{\texttt{BI}}^{-1} \mathcal{W}_{bi} \cdot \mathbf{E}_i^{\texttt{UI}}, \quad \bar{\mathbf{E}}_u^{\texttt{BI}} = \bar{\mathbf{D}}_{\texttt{UI}}^{-1} \mathcal{W}_{ui} \cdot \mathbf{E}_i^{\texttt{BI}}, \tag{8}$$

$$\mathbf{E}_b^{\texttt{UI}} = \bar{\mathbf{E}}_b^{\texttt{UI}} + \upsilon_{bi} \cdot \mathrm{sign}(\bar{\mathbf{E}}_b^{\texttt{UI}}) \odot \frac{\epsilon_{\texttt{BI}}}{\|\epsilon_{\texttt{BI}}\|_2}, \tag{9}$$

$$\mathbf{E}_u^{\texttt{BI}} = \bar{\mathbf{E}}_u^{\texttt{BI}} + \upsilon_{ui} \cdot \mathrm{sign}(\bar{\mathbf{E}}_u^{\texttt{BI}}) \odot \frac{\epsilon_{\texttt{UI}}}{\|\epsilon_{\texttt{UI}}\|_2}, \tag{10}$$

where $\bar{\mathbf{D}}_{\texttt{BI}} \in \mathbb{R}^{K\times K}$ and $\bar{\mathbf{D}}_{\texttt{UI}} \in \mathbb{R}^{M\times M}$ are diagonal matrices associated with the views $\mathcal{W}_{bi}$ and $\mathcal{W}_{ui}$, respectively. $\bar{\mathbf{E}}_b^{\texttt{UI}}$ and $\bar{\mathbf{E}}_u^{\texttt{BI}}$ denote original representations obtained by graph convolution. On this basis, we introduce the adaptive noise $\epsilon_{\texttt{BI}}, \epsilon_{\texttt{UI}} \in \mathbb{R}^d \sim U(0,1)$ as in Eq. (6) to further enhance the noise resistance ability of our model.

Finally, the node representation ($\mathbf{E}_u \in \mathbb{R}^{M\times d}, \mathbf{E}_b \in \mathbb{R}^{K\times d}$) learned by our **DCBR** are obtained through the adaptive fusion from the dual views, which are represented as:

$$\mathbf{E}_u = \omega \mathbf{E}_u^{\texttt{UI}} + (1-\omega)\mathbf{E}_u^{\texttt{BI}}, \;\; \mathbf{E}_b = \omega \mathbf{E}_b^{\texttt{UI}} + (1-\omega)\mathbf{E}_b^{\texttt{BI}}. \tag{11}$$

Here, ω is used to control the weight of the two views. $\mathbf{E}_u$ and $\mathbf{E}_b$ represent the final learned representations of users and bundles, respectively.

Dual-Level Disentangled Contrastive Learning

In recent years, self-supervised learning, especially contrastive learning, has gradually become an important technique to overcome the sparsity of interaction behaviors in bundle recommendations. However, previous methods (Ma et al. 2024) have attempted to directly use the final fused user (bundle) representations for contrastive learning, which inevitably introduces entangled noise due to semantic gaps between different views, leading to suboptimal contrastive signals. To address this challenge, we propose Dual-level Disentangled Contrastive Learning (DDCL), considering the latent relations both between views (inter-view) and within each view (intra-view). Motivated by InfoNCE (Oord, Li, and Vinyals 2018), we maximize the similarity between positive user (bundle) samples across the two contrastive views while pushing away negative samples, generating disentangled contrastive signals to mitigate interaction sparsity and alleviate entangled noise, which is as follows:

$$\mathcal{C}(\mathbf{E}_o^*, \mathbf{E}_o^{**}) = \frac{1}{|\mathcal{O}|} \sum_{o\in\mathcal{O}} -\log \frac{\exp(s(\mathbf{E}_o^*, \mathbf{E}_o^{**})/\tau)}{\sum_{o'\in\mathcal{O}} \exp(s(\mathbf{E}_o^*, \mathbf{E}_{o'}^{**})/\tau)}, \tag{12}$$

where $\mathcal{C}(,)$ represents the disentangled contrastive signal generator, $\mathbf{E}_o^*$ and $\mathbf{E}_o^{**}$ denote the original and augmented views of node $o \in \{u, b\}$, respectively. $\mathcal{O}$ represents the set of nodes o, with $|\mathcal{O}|$ denoting the size of the corresponding set. τ controls the sensitivity of our model to the similarity difference between positive and negative samples. $s(,)$ denotes the cosine similarity between the two views, expressed as $s(\mathbf{E}_o^*, \mathbf{E}_o^{**}) = \mathbf{E}_o^{*\top} \cdot \mathbf{E}_o^{**} / (\|\mathbf{E}_o^*\|_2 \cdot \|\mathbf{E}_o^{**}\|_2)$.

In the inter-view level, we focus on discovering disentangled relationships between various views. To adequately learn shared features across views, the contrastive learning task is designed to capture global cross-view information by maximizing the consistency between different views:

$$\mathcal{L}_o^{inter} = \mathcal{C}(\mathbf{E}_o^{\texttt{UB}}, \mathbf{E}_o^{\texttt{UI}}) + \mathcal{C}(\mathbf{E}_o^{\texttt{UB}}, \mathbf{E}_o^{\texttt{BI}}) + \mathcal{C}(\mathbf{E}_o^{\texttt{UI}}, \mathbf{E}_o^{\texttt{BI}}). \tag{13}$$

In the intra-view level, we emphasize exploring the fine-grained relationships within each view. We enhance our model's sensitivity to local interactions by leveraging the complementarity of local features within the same view:

$$\mathcal{L}_o^{intra} = \mathcal{C}(\mathbf{E}_o^{\mathtt{UB}}, \mathbf{E}_o^{\mathtt{UB}}) + \mathcal{C}(\mathbf{E}_o^{\mathtt{UI}}, \mathbf{E}_o^{\mathtt{UI}}) + \mathcal{C}(\mathbf{E}_o^{\mathtt{BI}}, \mathbf{E}_o^{\mathtt{BI}}). \quad (14)$$

Ultimately, our dual-level disentangled contrastive loss (DDCL) can be expressed as:

$$\mathcal{L}_{\mathtt{DDCL}} = \gamma_1(\mathcal{L}_u^{inter} + \mathcal{L}_b^{inter}) + \gamma_2(\mathcal{L}_u^{intra} + \mathcal{L}_b^{intra}). \quad (15)$$

Here, γ_1 and γ_2 are two tunable weights used to control the relative strength of the inter-view and intra-view levels.

Multi-task Learning

The training of our **DCBR** consists mainly of two parts: the CBDM (as defined in Eq. (5)) and the recommendation model. For the recommendation task, we define a triplet that includes a user, a bundle b^+ that the user u has interacted with and one b^- the user has not:

$$\mathcal{R} = \{(u, b^+, b^-) | (u, b^+) \in \mathcal{G}_{ub}, (u, b^-) \notin \mathcal{G}_{ub}\}, \quad (16)$$

where $\mathcal{R}$ represents the set of triplets used for training. We apply Bayesian Personalized Ranking (BPR) (Rendle et al. 2009) to optimize the recommendation model:

$$\mathcal{L}_{\mathtt{BPR}} = \frac{1}{|\mathcal{R}|} \sum_{(u,b^+,b^-)\in\mathcal{R}} -\ln\sigma(\hat{y}_{u,b^+} - \hat{y}_{u,b^-}), \quad (17)$$

$$\hat{y}_{u,b^+} = \mathbf{E}_u^\top \cdot \mathbf{E}_{b^+}, \quad \hat{y}_{u,b^-} = \mathbf{E}_u^\top \cdot \mathbf{E}_{b^-}. \quad (18)$$

Here, σ denotes the Sigmoid activation function. $\hat{y}_{u,b^+}$ and $\hat{y}_{u,b^-}$ represent the preference scores of user u for bundles b^+ and b^-, calculated through the inner product, respectively. Finally, integrating the proposed DDCL loss into the BPR loss constitutes the optimization objective $\mathcal{L}_{\mathtt{BRec}}$ of our bundle recommendation model, which can be expressed as:

$$\arg\min_{\Theta} \mathcal{L}_{\mathtt{BRec}} = \mathcal{L}_{\mathtt{BPR}} + \lambda_1 \mathcal{L}_{\mathtt{DDCL}} + \lambda_2 \|\Theta\|_2^2, \quad (19)$$

where λ_1 controls the strength of our dual-level disentangled contrastive loss, λ_2 represents the $L2$ regularization term to prevent overfitting, and $\Theta = \{\mathbf{E}_u^{(0)}, \mathbf{E}_b^{(0)}, \mathbf{E}_i^{(0)}\}$.

Computational Complexity Analysis

The parameters of **DCBR** consist of embeddings for users, bundles, and items: $\mathbf{E}_u^{(0)}, \mathbf{E}_b^{(0)}, \mathbf{E}_i^{(0)}$. Therefore, the total space complexity of **DCBR** is $\mathcal{O}((M+K+N)d)$. For time complexity, the triple-view denoised graph learning module employs graph convolutional networks to extract representations from multiple graphs, with a time complexity of $\mathcal{O}((2L|\hat{\mathcal{G}}_{ub}| + (2L+1)(|\mathcal{G}_{ui}| + |\mathcal{G}_{bi}|))d)$, where $|\hat{\mathcal{G}}_{ub}|$, $|\mathcal{G}_{ui}|$, and $|\mathcal{G}_{bi}|$ represent the number of edges in the corresponding graphs. The BPR loss has a time complexity of $\mathcal{O}(Bd)$, while the dual-level disentangled contrastive learning process requires $\mathcal{O}(B^2 d)$ time complexity.

Dataset	MealRec$_H^+$	MealRec$_L^+$	iFashion
# User (U)	1,575	1,928	53,897
# Bundle (B)	3,817	3,578	27,694
# Item (I)	7,280	10,589	42,563
# U-B Interaction	46,767	11,807	1,679,708
U-B Sparsity	99.2221%	99.8288%	99.8875%
# U-I Interaction	151,148	181,087	2,290,645
U-I Sparsity	98.6818%	99.1130%	99.9001%
# B-I Affiliation	11,451	10,734	106,916
B-I Sparsity	99.9588%	99.9717%	99.9909%

Table 1: Statistics of experimental datasets.

Experiments

Experimental Settings

Datasets The datasets utilized in the evaluation include MealRec$_H^+$, MealRec$_L^+$ (Li et al. 2024), and iFashion (Chen et al. 2019b), corresponding to meal and fashion outfit recommendation scenarios, respectively. MealRec$_H^+$ and MealRec$_L^+$ represent MealRec$^+$ datasets pre-processed with 5 cores and 2 cores, respectively. The statistical properties of the data are summarized in Table 1, and the data partitioning follows previous work (Ma et al. 2022; Li et al. 2024).

Evaluation Metrics and Protocols Following previous work (Ma et al. 2022), we evaluate the performance of bundle recommendation methods using two widely adopted metrics: Recall@K (R@K) and NDCG@K (N@K), where K = {10, 20}. All experimental results are based on the model that achieves the highest Top-20 metrics in the validation set. We adopt the all-ranking evaluation protocol (He et al. 2020; Wu et al. 2021) to calculate the metrics.

Baselines We compare our **DCBR** with various baselines: i) Collaborative Filtering: **Pop** (Cremonesi, Koren, and Turrin 2010), **MF-BPR** (Rendle et al. 2009), **NGCF** (Wang et al. 2019), **LightGCN** (He et al. 2020), **SGL** (Wu et al. 2021), **SimGCL** (Yu et al. 2022a), **XSimGCL** (Yu et al. 2023), **BIGCF** (Zhang, Sang, and Zhang 2024); and ii) Bundle Recommendation: **BGCN** (Chang et al. 2020), **UHBR** (Yu et al. 2022b), **CrossCBR** (Ma et al. 2022), **DSCBR** (Wu et al. 2023), **EBRec** (Du et al. 2023), **BundleGT** (Wei et al. 2023), **MultiCBR** (Ma et al. 2024).

Implementation Details To ensure fair experimental comparisons, our proposed **DCBR** and all comparative baselines are implemented using PyTorch (Paszke et al. 2019), optimized with Adam optimizer (Kingma and Ba 2015) at a learning rate of $1e^{-3}$, and evaluated on an NVIDIA RTX 3090 GPU with 24GB of memory. All models use Xavier initialization (Glorot and Bengio 2010) for their embeddings, with the embedding size fixed at 64 and the mini-batch size set at 2048. The number of negative samples and the test interval are fixed at 1 and 5, respectively. For our **DCBR**, the number of graph propagation layers L is fixed at 2, λ_2 is selected in $\{1e^{-5}, 1e^{-6}, 1e^{-7}\}$, and the υ_x, $\xi_x^{(l)}$, ω, τ, $\gamma_i \in [0; 1]$ are optimized through grid search. λ_0 and λ_1 are tuned from the ranges of $\{1e^0, 1e^1, 1e^2, 1e^3, 1e^4\}$ and $\{0.01, 0.02, 0.03, 0.04, 0.05, 0.2, 0.3, 0.4\}$, respectively.

Model	Reference	MealRec^+_H				MealRec^+_L				iFashion			
		R@10	**N@10**	**R@20**	**N@20**	**R@10**	**N@10**	**R@20**	**N@20**	**R@10**	**N@10**	**R@20**	**N@20**
Pop	RecSys'10	0.0163	0.0101	0.0339	0.0168	0.0142	0.0059	0.0481	0.0166	0.0126	0.0113	0.0220	0.0152
MF-BPR	UAI'09	0.1094	0.0757	0.1632	0.0917	0.0257	0.0157	0.0378	0.0190	0.0398	0.0359	0.0648	0.0463
NGCF	SIGIR'19	0.1189	0.0843	0.1704	0.0992	0.0291	0.0160	0.0418	0.0193	0.0420	0.0376	0.0676	0.0481
LightGCN	SIGIR'20	0.1397	0.0957	0.1963	0.1123	0.0447	0.0277	0.0525	0.0300	0.0519	0.0477	0.0824	0.0602
SGL	SIGIR'21	0.1543	0.1099	0.2114	0.1259	0.0465	0.0279	0.0510	0.0293	0.0582	0.0535	0.0911	0.0670
SimGCL	SIGIR'22	0.1433	0.1059	0.2038	0.1233	0.0454	0.0265	0.0627	0.0303	0.0659	0.0611	0.1023	0.0759
XSimGCL	TKDE'23	0.1483	0.1072	0.2061	0.1241	0.0483	0.0274	0.0689	0.0324	0.0661	0.0616	0.1022	0.0763
BIGCF	SIGIR'24	0.1488	0.1085	0.2110	0.1267	0.0453	0.0257	0.0597	0.0296	0.0660	0.0612	0.1022	0.0760
BGCN	SIGIR'20	0.1800	0.1323	0.2440	0.1501	0.0736	0.0439	0.1069	0.0529	0.0526	0.0483	0.0834	0.0609
UHBR	KBS'22	0.1417	0.0992	0.2032	0.1167	0.0451	0.0226	0.0789	0.0316	0.0654	0.0608	0.1013	0.0755
CrossCBR	KDD'22	0.2727	0.2137	0.3670	0.2400	0.1252	0.0807	0.1678	0.0921	0.0760	0.0717	0.1132	0.0868
DSCBR	TCSS'23	0.2564	0.1975	0.3385	0.2208	0.1336	0.0822	0.1670	0.0915	0.0748	0.0691	0.1133	0.0849
EBRec	TORS'23	0.2481	0.1969	0.3303	0.2200	0.1311	0.0839	0.1744	0.0957	0.0765	0.0724	0.1154	0.0883
BundleGT	SIGIR'23	0.2596	0.2085	0.3617	0.2358	0.1278	0.0724	0.1694	0.0841	0.0806	0.0759	0.1214	0.0926
MultiCBR	TOIS'24	0.3196	0.2408	0.4211	0.2693	0.2666	0.1678	0.3369	0.1871	0.1058	0.1027	0.1497	0.1203
DCBR	-	**0.4113**	**0.3159**	**0.5261**	**0.3483**	**0.2761**	**0.1916**	**0.3611**	**0.2144**	**0.1189**	**0.1191**	**0.1633**	**0.1370**
#Improv.	-	28.69%	31.19%	24.93%	29.34%	3.56%	14.18%	7.18%	14.59%	12.38%	15.97%	9.08%	13.88%

Table 2: Overall performance of **DCBR** and compared baselines. The best result is **bold** and the second best is underlined.

Overall Performance

In this section, we compare the overall recommendation performance of our **DCBR** framework with several baseline methods. The results of our evaluations are summarized in Table 2 for the top-K recommendations, which are observed: (1) **Performance superiority of DCBR.** Our **DCBR** demonstrates consistent superiority over state-of-the-art (SOTA) baselines across all datasets and evaluation metrics. We attribute the significant improvement to: i) CBDM effectively eliminates irrelevant and erroneous information from user-bundle interactions; ii) DDCL captures the latent relationships among multiple views to compensate for insufficient supervision. (2) **Effectiveness of triple-view learning.** The introduction of semantically rich interactions between users and items, as well as the affiliation information between bundles and items, effectively enhances bundle recommendations. By effectively modeling the information across three views, bundle recommendation systems generally achieve better results than general recommendation methods. (3) **Significant advantages of disentangled contrastive learning.** Experimental results demonstrate that contrastive learning-based methods significantly outperform other approaches. Furthermore, the superiority of **DCBR** over MultiCBR highlights that the dual-level disentangled contrastive learning paradigm not only enhances the robustness of feature representations but also enables effective mitigation of noise signals caused by semantic discrepancies between views during the feature fusion process.

Ablation Study

In this section, we analyze the impact of different core components in our **DCBR**. We conduct performance evaluation by comparing **DCBR** with multiple variants obtained by removing key modules. The following are the variants used for comparison: "*w/o BLCC*": only discards the proposed BLCC loss, optimizing our conditional bundle diffusion model with the ELBO loss. "*w/o CBDM*": removes the CBDM and directly uses the original user-bundle interaction graph instead of the denoised version. "*w/o inter*", "*w/o intra*", and "*w/o DDCL*": respectively eliminate the auxiliary contrastive signal of inter-view level, intra-view level, and dual-level disentangled contrastive learning.

Data	Metrics	DCBR	*w/o BLCC*	*w/o CBDM*	*w/o inter*	*w/o intra*	*w/o DDCL*
MealRec^+_H	R@10	**0.4113**	0.3891	0.3711	0.2108	0.3957	0.0174
	N@10	**0.3159**	0.3068	0.2893	0.1454	0.3077	0.0123
	R@20	**0.5261**	0.5025	0.4731	0.2856	0.5070	0.0410
	N@20	**0.3483**	0.3393	0.3183	0.1661	0.3400	0.0192
MealRec^+_L	R@10	0.2761	0.2741	**0.2798**	0.1923	0.2714	0.0455
	N@10	**0.1916**	0.1858	0.1834	0.1222	0.1792	0.0221
	R@20	**0.3611**	0.3597	0.3464	0.2616	0.3375	0.0814
	N@20	**0.2144**	0.2085	0.2016	0.1407	0.1971	0.0317
iFashion	R@10	**0.1189**	0.1172	0.1070	0.1023	0.1154	0.0245
	N@10	**0.1191**	0.1184	0.1064	0.1003	0.1145	0.0219
	R@20	**0.1633**	0.1612	0.1495	0.1462	0.1611	0.0404
	N@20	**0.1370**	0.1360	0.1236	0.1182	0.1330	0.0284

Table 3: Ablation study on different components of **DCBR**.

We evaluate the results for all the experimental data, as illustrated in Table 3, which demonstrates that **DCBR** consistently outperforms the five variants. Specifically, the variant without BLCC shows a certain degree of performance decline, validating that the BLCC loss effectively enhances the denoising effect by mitigating the degradation of original user-bundle interaction information during the denoising process. Removal of CBDM leads to a notable decline in recommendation performance, demonstrating the effectiveness of our CBDM in denoising user-bundle interaction information. This variant directly utilizes the original graph as the encoding object during the learning process, which may lead to potential noise interference in the learned representations. Abolishing inter-view, intra-view, or DDCL leads to a significant decrease in performance, illustrating that our designed contrastive loss effectively avoids entangled noise caused by semantic gaps between different views and compensates for the lack of supervision due to data sparsity.

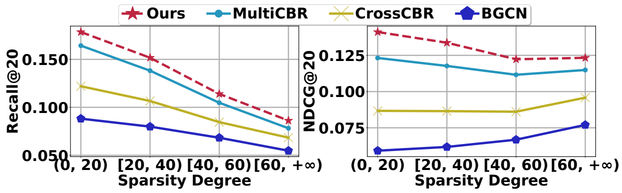


Figure 2: Performance w.r.t. different user interaction sparsity degree on iFashion datasets.

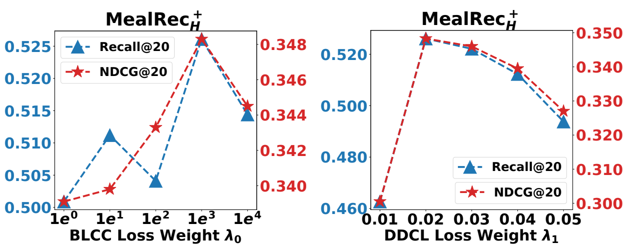


Figure 3: Hyperparameter analysis on the loss weights of BLCC and DDCL of **DCBR** on MealRec$_H^+$ dataset.

Robustness Investigation against Sparsity

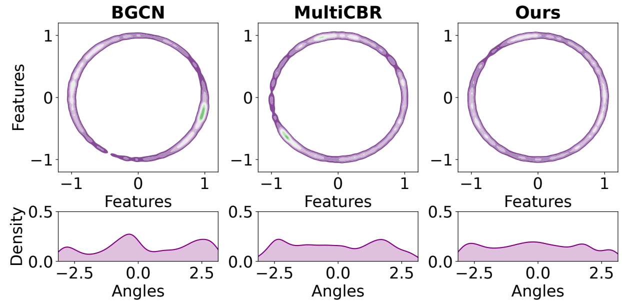
We further investigate the robustness of the model in addressing sparse user interactions by conducting experiments with **DCBR** alongside three representative bundle recommendation baselines: BGCN, CrossCBR, and MultiCBR. Specifically, we partition the user set into four groups based on the degree of user nodes in the user-bundle training interaction graph of iFashion dataset, defined as $(0, 20)$, $[20, 40)$, $[40, 60)$, and $[60, \infty)$. From the results presented in Figure 2, it is evident that **DCBR** outperforms all comparative baselines across user groups with varying levels of sparsity. This further validates that denoising augmentation provided by the conditional bundle diffusion model enables **DCBR** to effectively compensate for the lack of supervision in scenarios characterized by scarce interaction labels through the generation of disentangled contrastive self-supervised signals.

Hyperparameter Analysis

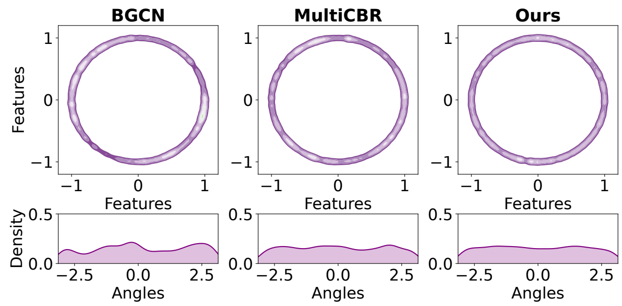
We explore the impact of key hyperparameters, the BLCC loss weight λ_0 and the DDCL loss weight λ_1, on the recommendation performance of **DCBR**. Figure 3 shows the results of the R@20 and N@20 on the MealRec$_H^+$ datasets. Based on the results, it can be concluded that increasing λ_0 to a certain extent can improve the performance of **DCBR**, but higher values can lead to a slight decline in performance due to overly strong latent consistency constraints, resulting in less significant denoising effects. Increasing λ_1 can enhance performance by more effectively removing entangled noise. However, excessively large values can misguide the supervision task, leading to a decreased performance.

Case Study

In this section, we delve into a case study to qualitatively investigate the effectiveness of our disentangled contrastive learning framework in learning meaningful user preferences



(a) User Representations



(b) Bundle Representations

Figure 4: Distribution of user/bundle representations learned from the iFashion dataset.

under the denoising augmentation of conditional bundle diffusion model. Specifically, we randomly sample $2,000$ users and bundles from the iFashion dataset and map their learned representations to 2-dimensional normalized vectors on the unit hypersphere using t-SNE (Van der Maaten and Hinton 2008). We also used Kernel Density Estimation to plot the feature distributions, aiming to present the density estimation of angles for each point on the unit hypersphere more clearly. In Figure 4, it is observed that compared to MultiCBR and BGCN, **DCBR** is capable of learning more uniformly distributed user and bundle representations, thereby preserving the intrinsic features of users and bundles.

Conclusion

In this work, we present the disentangled contrastive bundle recommendation (**DCBR**) framework. The conditional bundle diffusion model we proposed plays a pivotal role in denoising the user-bundle interaction graph, ensuring that the essential information remains intact during optimization. This is complemented by our triple-view denoised graph learning module, which leverages multiple perspectives to derive more robust user/bundle representations. Furthermore, the dual-level disentangled contrastive learning paradigm allows us to capture complex relationships between multiple views and within a single view, generating high-quality contrastive signals that facilitate better learning despite the inherent challenges of sparsity and noise. The results of our extensive experiments on multiple benchmark datasets demonstrate the effectiveness of **DCBR**, demonstrating its superiority over existing state-of-the-art methods.

References

Austin, J.; Johnson, D. D.; Ho, J.; Tarlow, D.; and Van Den Berg, R. 2021. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34: 17981–17993.

Cai, X.; Huang, C.; Xia, L.; and Ren, X. 2023. LightGCL: Simple Yet Effective Graph Contrastive Learning for Recommendation. In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR)*.

Cao, D.; Nie, L.; He, X.; Wei, X.; Zhu, S.; and Chua, T.-S. 2017. Embedding factorization models for jointly recommending items and user generated lists. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*, 585–594.

Chang, J.; Gao, C.; He, X.; Jin, D.; and Li, Y. 2020. Bundle recommendation with graph convolutional networks. In *Proceedings of the 43rd international ACM SIGIR conference on Research and development in Information Retrieval*, 1673–1676.

Chang, J.; Gao, C.; He, X.; Jin, D.; and Li, Y. 2021. Bundle recommendation and generation with graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 35(3): 2326–2340.

Chen, L.; Liu, Y.; He, X.; Gao, L.; and Zheng, Z. 2019a. Matching user with item set: Collaborative bundle recommendation with deep attention network. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2095–2101.

Chen, W.; Huang, P.; Xu, J.; Guo, X.; Guo, C.; Sun, F.; Li, C.; Pfadler, A.; Zhao, H.; and Zhao, B. 2019b. POG: personalized outfit generation for fashion recommendation at Alibaba iFashion. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2662–2670.

Cremonesi, P.; Koren, Y.; and Turrin, R. 2010. Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*, 39–46.

Deng, Q.; Wang, K.; Zhao, M.; Zou, Z.; Wu, R.; Tao, J.; Fan, C.; and Chen, L. 2020. Personalized bundle recommendation in online games. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2381–2388.

Du, X.; Qian, K.; Ma, Y.; and Xiang, X. 2023. Enhancing Item-level Bundle Representation for Bundle Recommendation. *ACM Transactions on Recommender Systems*.

Epstein, D.; Jabri, A.; Poole, B.; Efros, A.; and Holynski, A. 2023. Diffusion self-guidance for controllable image generation. *Advances in Neural Information Processing Systems*, 36: 16222–16239.

Glorot, X.; and Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 249–256. JMLR Workshop and Conference Proceedings.

He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; and Wang, M. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 639–648.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. In *Advances in neural information processing systems*, 6840–6851.

Jeon, H.; Lee, J.-e.; Yun, J.; and Kang, U. 2024. Cold-start Bundle Recommendation via Popularity-based Coalescence and Curriculum Heating. In *Proceedings of the ACM on Web Conference 2024*, 3277–3286.

Jiang, Y.; Yang, Y.; Xia, L.; and Huang, C. 2024. Diffkg: Knowledge graph diffusion model for recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 313–321.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.

Kipf, T. N.; and Welling, M. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*.

Li, J.; and Wang, H. 2024. Graph Diffusive Self-Supervised Learning for Social Recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2442–2446.

Li, M.; Li, L.; Tao, X.; and Huang, J. X. 2024. MealRec+: A Meal Recommendation Dataset with Meal-Course Affiliation for Personalization and Healthiness. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 564–574.

Lin, Z.; Tian, C.; Hou, Y.; and Zhao, W. X. 2022. Improving graph collaborative filtering with neighborhood-enriched contrastive learning. In *Proceedings of the ACM web conference 2022*, 2320–2329.

Lugmayr, A.; Danelljan, M.; Romero, A.; Yu, F.; Timofte, R.; and Van Gool, L. 2022. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11461–11471.

Ma, Y.; He, Y.; Wang, X.; Wei, Y.; Du, X.; Fu, Y.; and Chua, T.-S. 2024. MultiCBR: Multi-view Contrastive Learning for Bundle Recommendation. *ACM Transactions on Information Systems*, 42(4): 1–23.

Ma, Y.; He, Y.; Zhang, A.; Wang, X.; and Chua, T.-S. 2022. Crosscbr: Cross-view contrastive learning for bundle recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1233–1241.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Proceedings of the Neural Information Processing Systems Conference (NeurIPS)*, 32.

Rendle, S.; Freudenthaler, C.; Gantner, Z.; and Schmidt-Thieme, L. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, 452–461.

Rendle, S.; Freudenthaler, C.; and Schmidt-Thieme, L. 2010. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th international conference on World wide web*, 811–820.

Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, 2256–2265.

Sun, M.; Li, L.; Li, M.; Tao, X.; Zhang, D.; Wang, P.; and Huang, J. X. 2024. A Survey on Bundle Recommendation: Methods, Applications, and Challenges. *arXiv preprint arXiv:2411.00341*.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Wang, W.; Xu, Y.; Feng, F.; Lin, X.; He, X.; and Chua, T.-S. 2023. Diffusion recommender model. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 832–841.

Wang, X.; He, X.; Wang, M.; Feng, F.; and Chua, T.-S. 2019. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, 165–174.

Wei, Y.; Liu, X.; Ma, Y.; Wang, X.; Nie, L.; and Chua, T.-S. 2023. Strategy-aware bundle recommender system. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1198–1207.

Wu, C.; Yuan, H.; Zhao, P.; Qu, J.; Sheng, V. S.; and Liu, G. 2023. Dual-Supervised Contrastive Learning for Bundle Recommendation. *IEEE Transactions on Computational Social Systems*.

Wu, J.; Wang, X.; Feng, F.; He, X.; Chen, L.; Lian, J.; and Xie, X. 2021. Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 726–735.

Yu, J.; Xia, X.; Chen, T.; Cui, L.; Hung, N. Q. V.; and Yin, H. 2023. XSimGCL: Towards extremely simple graph contrastive learning for recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 36(2): 913–926.

Yu, J.; Yin, H.; Xia, X.; Chen, T.; Cui, L.; and Nguyen, Q. V. H. 2022a. Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, 1294–1303.

Yu, Z.; Li, J.; Chen, L.; and Zheng, Z. 2022b. Unifying multi-associations through hypergraph for bundle recommendation. *Knowledge-Based Systems*, 255: 109755.

Zhang, Y.; Sang, L.; and Zhang, Y. 2024. Exploring the individuality and collectivity of intents behind interactions for graph collaborative filtering. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1253–1262.

Zhao, J.; Wenjie, W.; Xu, Y.; Sun, T.; Feng, F.; and Chua, T.-S. 2024. Denoising diffusion recommender model. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1370–1379.

Zhao, S.; Wei, W.; Zou, D.; and Mao, X. 2022. Multi-view intent disentangle graph networks for bundle recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 4379–4387.