TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214
DOI: 10.26599/TST.2025.9010058

Research Article

Structure-Aware Representation Enhancement for Incomplete Bundle Recommendation

Yanchao Tan, Zhijun Chen, Yonghui Yang, Xinting Liao, Weiming Liu, and Guofang Ma*

Abstract: Bundle recommendation systems are able to enhance user experience and increase sales profits by offering curated item collections. However, the complexity inherent in bundling items poses significant challenges in accurately collecting user interactions with bundles. This limitation often results in incomplete user-bundle interaction data, making it hard to accurately capture user preferences within the context of bundles. In this work, we propose a framework named Structure-aware representation enhancement for Incomplete Bundle Recommendation (SIBR). Specifically, we first conduct pretraining with user-item interactions and bundle-item affiliations to learn representations of users, bundles, and items, which are consistent with the underlying structure. Furthermore, we introduce a graph-signal enhancement strategy focused on user-bundle interactions. This strategy aims to reveal the distinct relations between user-user and bundle-bundle, enabling the refinement of user and bundle representations. The ultimate goal is to leverage these enhanced representations to improve the accuracy of bundle recommendations. Extensive experiments on three real-world benchmark datasets show that our proposed framwork achieves significant performance gains, which consistently achieves an average of 20.1% improvement over state-of-the-art competitors in both Recall and NDCG metrics.

Key words: bundle recommendation; incomplete representation learning; random walk

1 Introduction

Recommender Systems (RSs) are instrumental in mitigating the issue of information overload[1–6]. Traditional RSs primarily focus on suggesting individual items to users[7]. However, in numerous scenarios of electronic commerce, particularly within the fashion shopping domain, the strategy of item bundling has become prevalent[8]. Item bundling, as a marketing tactic, aims to incorporate less frequently purchased items into a bundle, thereby increasing the overall sales revenue for platforms. This approach not only diversifies the user's selection but also enhances their overall experience.

Traditional bundle recommendations regard item and bundle recommendation as distinct tasks and associate

• Yanchao Tan and Zhijun Chen are with College of Computer and Data Science, Fuzhou University, Fuzhou 350116, China, and also with the Fujian Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou University, Fuzhou 350116, China. E-mail: yctan@fzu.edu.cn; wu563655@gmail.com.
• Yonghui Yang is with School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China. E-mail: yyh.hfut@gmail.com.
• Xinting Liao and Weiming Liu are with College of Computer Science, Zhejiang University, Hangzhou 310027, China. E-mail: 12021080@zju.edu.cn; 21831010@zju.edu.cn.
• Guofang Ma is with School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou 310018, China. E-mail: maguofang@zjgsu.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2024-03-18; revised: 2024-09-26; accepted: 2025-03-27

© The author(s) 2026. The articles published in this open access journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

them by sharing model parameters[9–11]. However, these works often struggle to capture the intricate relations among users, bundles, and items. In contrast, recent graph-based methods leverage the expressive power of Graph Neural Networks (GNNs) to learn fine-grained representations of users and bundles[12, 13]. For instance, BundleNet[14] combines insights from both item and bundle domains into a unified tripartite graph, employing GNNs to merge adjacent information into comprehensive user and bundle profiles. AGF[15] designs a bundle-specific GNN-based mechanism to achieve refined aggregation for representations. Similarly, DGMAE[16] introduces a transformer-enhanced GNN encoder for in-depth global relation analysis, featuring an innovative edge masking strategy and reconstruction target focused on key interactions.

Despite their effectiveness, the aforementioned approaches may suffer from incomplete bundle limitations, given the difficulty in consistently securing high-quality user-bundle interaction data. As shown in Fig. 1, David has interacted with items v_1 and v_3 but does not interact with bundle b_1, which contains v_1 and v_3. This missing user-bundle interaction obscures David's preference for specific categories (i.e., fast food) in the context of bundles. Moreover, this gap in interaction data hinders the ability to discern shared preferences among users, such as between individuals with similar interests but incomplete interaction records. For example, David and Emma share fast food and photography favors in the context of items, but the missing u_1-b_1 interaction suggests they are not similar in the context of bundles. The above analysis shows the negative effect of missing user-bundle interaction on bundle recommendation.

Recognizing the Incomplete Bundle Recommendations (IBRs) issue, it becomes imperative

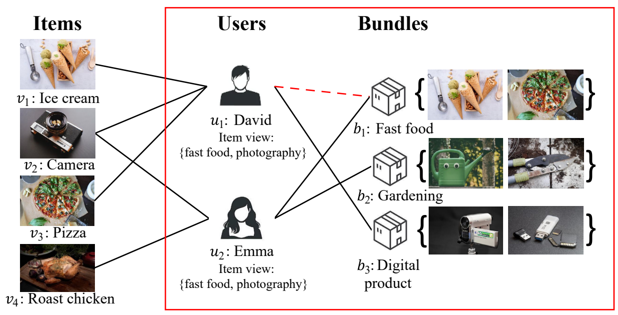


Fig. 1 A toy example of incomplete bundle recommendations. The red box is the context of bundles, and the red dash line represents the missing user-bundle interaction.

to explore methodologies that can navigate through the complexities of incomplete data. However, this novel task poses several unique challenges for us, as follows:

Challenge 1: Collaboration information utilization. Current methods[12, 13] fail to fully exploit the collaboration information embedded within user-item interactions and bundle-item affiliations. Although GNNs are employed to facilitate effective message passing within individual graphs, a direct linkage between the user-item and bundle-item graphs is missing. This absence of direct connection leads to inconsistent collaboration between the two graphs, resulting in representations that may not fully reflect the complex user preferences and bundle affinities.

Challenge 2: Distinct relation modeling. The inherent distinct relations among user and bundle nodes are often simplified or ignored by existing bundle recommendation approaches[17, 18]. Traditional methods predominantly utilize binary adjacency matrices to model interactions, assigning uniform weight to all connections. This oversimplification overlooks the differences in relation strengths and types, leading to less accurate user preferences.

In this work, we introduce Structure-aware representation enhancement for Incomplete Bundle Recommendation (SIBR), which consists of two primary modules: (1) Structure-aware Node Representation Learning (SNRL); (2) Structure-aware Graph-Signal Enhancement (SGSE). Specifically, in SNRL, we deploy a parameterized random walk mechanism based on the joint user-bundle-item graph, fully capturing collaborative information and shared structural similarities among user, bundle, and item nodes, empowering them with structure information-embedded representations in a unified manner. In SGSE, we propose a graph-signal enhancement strategy to learn distinct relations among users and bundles. This strategy leverages underlying structures inherent in user-user and bundle-bundle relations to augment the binary adjacency matrix. By enhancing the graph signal, we can refine the convolution process and facilitate the development of more precise representations of users and bundles. We evaluate our proposed SIBR with extensive experiments on three real-world benchmark datasets for incomplete bundle recommendation. The results show that our SIBR has great potential compared other state-of-the-art baselines and it achieves a improvement up to 39.3% in Recall@40.

Our overall contributions in this work are summarized as follows:

- **Analysis:** We analyze the emergence of the IBRs issue and how it influences the bundle recommendation. This analysis motivates us to explore implicit information within user-item interactions and bundle-item affiliations as incomplete data supplements.
- **Model:** We propose the SIBR model to uncover the structural information of collaboration and distinct node relations latent in the available data, which alleviates the negative impacts of incomplete data and thereby optimizes the recommendation performance.
- **Experiments:** We conduct extensive experiments on three benchmark datasets with various sizes, proving that SIBR outperforms the state-of-the-art methods by a large margin with incompleteness.

2 Related Work

2.1 Bundle recommendation

Bundle recommendation focuses on suggesting collections of items to users rather than individual items. Initial efforts did not distinguish between bundles and individual items, treating bundles simply as single entities[19]. The later works such as LIRE[10] and EFM[9] recognized the uniqueness of bundles, taking both item lists and individual items into consideration under the BPR[20] framework. Chen et al.[21] jointly modeled user-bundle and user-item interactions in a multi-task framework. However, these studies did not fully exploit the crucial bundle-item affiliation relations and failed to capture the higher-order information in the user-bundle-item graph.

Recently, Graph Convolutional Networks (GCNs) have achieved remarkable performance on graph data by their perception of graph structures[22, 23], which inspires researchers to explore their application in bundle recommendation. BGCN[12] was the first framework to utilize GCN in bundle recommendation, leveraging bundle-item affiliation relations to learn user preferences with different views. Ma et al.[18] followed BGCN and modeled the cooperative association between bundle and item views through cross-view contrastive learning. Zhao et al.[13] disentangled users' intents from dual view implicit feedback to improve bundle recommendation performance. Ke et al.[24] encoded entity relations in hyperbolic space with a GCN-based model. Ren et al.[16] proposed a knowledge distillation framework, utilizing GCNs to perform representation extraction and knowledge integration. These GCN-based techniques have set new benchmarks by significantly enhancing performance. However, they overlook the issue of incomplete user-bundle interactions, which could lead to suboptimal outcomes.

2.2 Incomplete representation learning

Incomplete representation learning focuses on extracting meaningful representations from datasets with missing information for downstream applications. Given the prevalence of incomplete data in real-world scenarios, developing effective incomplete representation learning is attracting a surge of interest in academia and industry circles[25]. Previous studies concentrated on two main approaches: Non-imputation and imputation methods, applying them to multi-view learning. Non-imputation methods handle incomplete data without filling in missing values[26]. For example, Li et al.[27] presented the first Non-negative Matrix Factorization (NMF)-based solution to learn both common and private latent representations for complete and incomplete instances, respectively. To accommodate datasets with more than two views, weighted NMF-based approaches emerged. Hu and Chen[28] proposed a weighted semi-NMF algorithm for aligning different view basis matrices, minimizing the impact of missing data. Xia et al.[29] conducted auto-weighted fusion in partition space to achieve an overall optimal incomplete multi-view clustering result. Wang et al.[30] proposed a multi-modal strategy to distinguish shared and unique features in data inputs.

Compared to non-imputation methods, imputation methods, despite potential bias introduction, are compatible with most existing general frameworks (i.e., complete data). Well-executed imputation methods can alleviate the effects of incomplete data and improve model performance[31]. Xu et al.[32] proposed an adversarial framework to seek the common latent representations of multi-view data for reconstructing raw data and inferring missing data for clustering. Zhang et al.[33] designed a modality adversarial and contrastive framework to address the modality-missing problem in multimodel knowledge graph completion.

The community of recommendation systems also adapts incomplete representation learning into various tasks. Mafrur et al.[34] proposed a selective imputation technique to determine which missing values should be imputed in incomplete data, optimizing visualization recommendation performance within a budget. Zhang et al.[35] explored a greedy approach to perform effective recommender system attacks with limited data collection. Zhang et al.[36] incorporated the large dialogue corpus to enhance the incomplete knowledge graphs in conversational recommender systems, enriching semantic information for better recommendations. Despite these advancements, to the best of our knowledge, the integration of incomplete representation learning into bundle recommendation, which notably suffers from incomplete data challenges, remains unexplored.

3 Methodology

In this section, we present our proposed SIBR framework (shown in Fig. 2). In the following subsections, firstly, we illustrate the incomplete bundle recommendation statement (Subsection 3.1); secondly, we introduce the critical SNRL module (Subsection 3.2), focusing on the detailed random walk process on the unified graph; then, we detail the SGSE module (Subsection 3.3), expounding upon representation refinement with a graph-signal enhancement strategy; finally, we elaborate on the model optimization process (Subsection 3.4).

3.1 Problem statement

We define the set of users as $\boldsymbol{U}=\{u_1, u_2, \ldots, u_M\}$, bundles as $\boldsymbol{B}=\{b_1, b_2, \ldots, b_N\}$, and items as $\boldsymbol{V}=\{v_1, v_2, \ldots, v_O\}$, where M, N, and O denote the numbers of users, bundles, and items, respectively. The user-item implicit feedback matrix can be defined as $\boldsymbol{X}=\{x_{uv}|u\in\boldsymbol{U}, v\in\boldsymbol{V}\}$, where $x_{uv}=1$ represents the interaction between user u and item v; the bundle-item affiliation matrix can be defined as $\boldsymbol{Y}=\{y_{bv}|b\in\boldsymbol{B}, v\in\boldsymbol{V}\}$, where $y_{bv}=1$ denotes that bundle b contains item v; and the user-bundle implicit feedback matrix can be defined as $\boldsymbol{Z}=\{z_{ub}|u\in\boldsymbol{U}, b\in\boldsymbol{B}\}$, where $z_{ub}=1$ represents the interaction between user u and bundle b. The overall description of notations is shown in Table 1.

The proposed SIBR aims to perform accurate bundle recommendations under incomplete interaction data. The input and output are defined as follows:

- **Input:** User-item interaction matrix $\boldsymbol{X}$, bundle-item affiliation matrix $\boldsymbol{Y}$, and incomplete user-bundle interaction matrix $\boldsymbol{Z}$.
- **Output:** The predicted user-bundle probability matrix $\hat{\boldsymbol{Z}}$ for recommendation.

3.2 Structure-aware node representation learning module

In the domain of bundle recommendation, the available user-item interactions and bundle-item affiliations can play a pivotal role. Collaboration between these two graphs helps capture accurate user preferences, thereby substantially augmenting the recommendation of incomplete bundles through the filling of missing user-

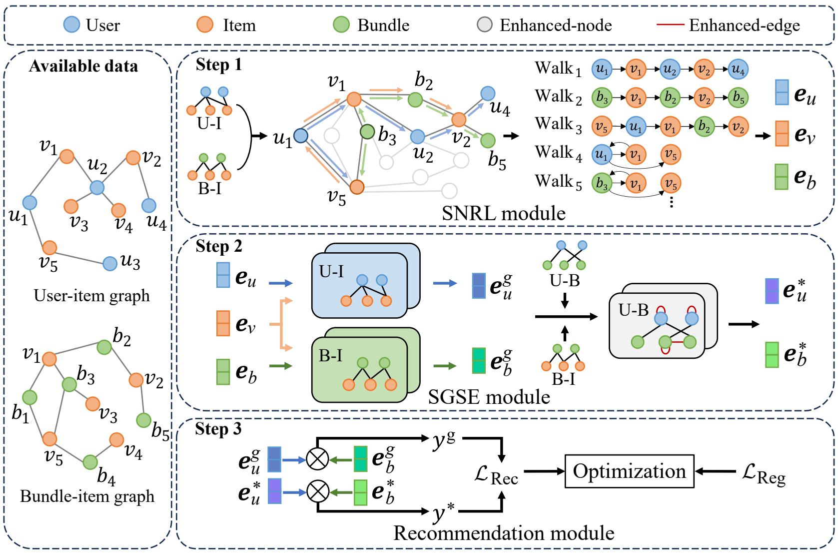


Fig. 2 Overall framework of our proposed SIBR. U-I and B-I denote the user–item interaction graph and the Bundle–Item affiliation graph, respectively. Nodes with thickened borders indicate enhanced nodes, and red straight or curved lines denote enhanced edges.

Table 1 Notations and their explanations.

Notation	Explanation
M, N, O	The number of users, bundles, and items
$\boldsymbol{U}, \boldsymbol{B}, \boldsymbol{V}$	The set of users, bundles, and items
$\boldsymbol{X} \in \mathbb{R}^{M\times O}$	The user-item adjacency matrix
$\boldsymbol{Y} \in \mathbb{R}^{N\times O}$	The bundle-item adjacency matrix
$\boldsymbol{Z} \in \mathbb{R}^{M\times N}$	The user-bundle adjacency matrix
$\hat{\boldsymbol{Z}} \in \mathbb{R}^{M\times N}$	The predicted user-bundle matrix

bundle interactions. Drawing inspiration from the popular graph algorithm Node2Vec[37], we integrate the user-item and bundle-item graphs into a unified input for the algorithm. This allows us to investigate the complex structural relations among nodes (i.e., users, bundles, and items), which helps to enhance the node representations without additional supervision.

As shown in Fig. 2, nodes in blue denote users, in green represent bundles, and in orange signify items. Within the SNRL module, contemplating the bolded nodes u_1, b_3, and v_5, we utilize a parameterized random walk strategy with u_1, b_3, and v_5 as source nodes. The walk obtained from one sampling of u_1 (Walk_1) contains $\{u_1, u_2, u_4, v_1, v_2\}$, the walk from one sampling of b_3 (Walk_2) includes $\{b_2, b_3, b_5, v_1, v_2\}$, and the walk from one sampling of v_5 (Walk_3) includes $\{v_1, v_2, v_5, u_1, b_2\}$. It can be observed that all three walks contain subsequences in the form $\{u/b - v_1 - u/b - v_2\}$, capturing structural equivalence shared in the region where u_1, b_3, and v_5 are located. As for Walk_4 and Walk_5, u_1 and b_3 are directly connected to v_1 and v_5 in the corresponding walk, indicating that they belong to the same network community, which further explores the consistent underlying structure for u_1 and b_3. By feeding the walks into Skip-Gram[38], the collaborative information among nodes is embedded into the representations. Ultimately, we effectively refine the representations of similar users, bundles, and items.

Specifically, we ensure the differentiation among users, bundles, and items by assigning them distinct IDs during random walk sampling. Following Node2Vec, we implement a parameterized node transition strategy for each node. The definition of this strategy is listed as follows:

$$P(s_i = x | s_{i-1} = v) = \begin{cases} \dfrac{\pi_{vx}}{Z_p}, & \text{if } (v,\ x) \in E_{\mathrm{uni}}; \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where s_i denotes the i-th node in the walk, Z_p is the normalizing constant, E_{uni} is the edge set of the unified graph, v and x are nodes in the graph, and π_{vx} is the unnormalized transition probability between v and x. π_{vx} can be formulated as follows:

$$\pi_{vx} = \alpha_{\mathrm{pq}}(t,\ x) \cdot w_{vx} \tag{2}$$

where $w_{vx} = 1$ means node v has interacted with x, otherwise $w_{vx} = 0$. The $\alpha_{pq}(t,\ x)$ is utilized to control the direction of the random walk, where t represents the previous node of v in the walk. The formulation of $\alpha_{pq}(t,\ x)$ is given as follows:

$$\alpha_{\mathrm{pq}}(t,\ x) = \begin{cases} \dfrac{1}{p}, & \text{if } d_{\mathrm{tx}} = 0; \\ 1, & \text{if } d_{\mathrm{tx}} = 1; \\ \dfrac{1}{q}, & \text{if } d_{\mathrm{tx}} = 2 \end{cases} \tag{3}$$

where d_{tx} denotes the shortest path distance between node t and x in the graph. Specifically, when d_{tx} equals 0, nodes t and x are the same node. The parameter p controls the likelihood of returning to the former node, facilitating the exploration of the node's vicinity. The parameter q encourages sampling to visit nodes further away from the former node when q comes to a low value ($q < 1$), achieving outward exploration. Following the above sampling strategy, the Alias[39] algorithm is applied to enhance the efficiency of the sampling process,

$$\mathrm{Walks} = \mathrm{Alias}(P,\ T,\ E_{\mathrm{uni}}) \tag{4}$$

where T denotes the length of one sample walk and Walks denotes the overall sample walks which can be represented as

$$\mathrm{Walks} = [\mathrm{Walk}_1,\ \mathrm{Walk}_2,\ \ldots,\ \mathrm{Walk}_K] \tag{5}$$

where K denotes the total count of users, bundles, and items. Walk_K denotes the sampled walk set of the K-th node. The sampled walks are fed into the Skip-Gram, generating pre-trained node embeddings $\boldsymbol{e}_u$, $\boldsymbol{e}_v$, and $\boldsymbol{e}_b$, which is formulated as

$$\max \prod_{k=1}^{K} \prod_{w=1}^{W} \prod_{t=1}^{T} F(v_{kwt+c} | v_{kwt}) \tag{6}$$

where v_{kwt} denotes the t-th node of the w-th walk in the k-th walk set, and c denotes the context size. $F(v_j|v_i)$ represents the transition probability of node v_i to node v_j, which can be formulated as

$$F(v_j|v_i) = \frac{\mathrm{e}^{z_{ij}}}{\sum_{k=1}^{K} \mathrm{e}^{z_{ik}}} \tag{7}$$

Algorithm 1 SIBR algorithm

Input: The user-item matrix $\boldsymbol{X}$, bundle-item matrix $\boldsymbol{Y}$, and incomplete user-bundle matrix $\boldsymbol{Z}$

Output: The predicted user-bundle matrix $\hat{\boldsymbol{Z}}$

1: Generate the tripartite graph G to abatin the edge set E_{uni} based on the
interaction matrices $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$;

2: Determine the return p, out q, and the length of one sample walk T;

3: P = TransProbabilityComputation(p, q, G); // Eq. (1) over the tripartite graph G

4: Walks = AliasSample(P, T, G); // Eq. (4)

5: $\boldsymbol{e}_u$, $\boldsymbol{e}_v$, $\boldsymbol{e}_b$ = Skip-Gram(Walks); // Pre-trained node embeddings

6: $\tilde{\boldsymbol{A}}$ = GraphEnhancement($\boldsymbol{Z}$, $\boldsymbol{Y}$); // Structure-aware graph-signal enhancement

7: **while** not converged **do**

8: $\quad\boldsymbol{e}_u^g$, $\boldsymbol{e}_b^g$ = GraphConvolution($\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{e}_u$, $\boldsymbol{e}_b$, $\boldsymbol{e}_v$);

9: $\quad\boldsymbol{e}_u^*$, $\boldsymbol{e}_b^*$ = GraphEnhancedConvolution($\tilde{\boldsymbol{A}}$, $\boldsymbol{e}_u^g$, $\boldsymbol{e}_b^g$);

10: $\quad$Calculate the overall loss function L of Eq. (19);

11: $\quad$Update the node representations;

12: **end while**

13: Calculate the predicted user-bundle matrix $\hat{\boldsymbol{Z}}$

where z_{ij} denotes the inner product of representations v_i and v_j. Upon optimization convergence, we obtain pre-trained representations $\boldsymbol{e}_u$, $\boldsymbol{e}_v$, and $\boldsymbol{e}_b$ for the user u, item v, and bundle b.

3.3 Structure-aware graph-signal enhancement module

Based on the pre-trained user and item representations $\boldsymbol{e}_u$ and $\boldsymbol{e}_v$ obtained in the SNRL module (Subsection 3.2), we employ the excellent GNN-based recommendation framework LightGCN[40] to learn high-order information in the user-item graph. Specifically, we perform information propagation over the graph, and the l-th layer's propagation can be formulated as

$$\begin{cases} \boldsymbol{e}_u^{(l)} = \sum\limits_{v\in N_u} \dfrac{1}{\sqrt{|N_u|}\sqrt{|N_v|}}\boldsymbol{e}_v^{(l-1)}, \\ \boldsymbol{e}_v^{(l)} = \sum\limits_{u\in N_v} \dfrac{1}{\sqrt{|N_v|}\sqrt{|N_u|}}\boldsymbol{e}_u^{(l-1)} \end{cases} \tag{8}$$

where $\boldsymbol{e}_u^{(l)}$ and $\boldsymbol{e}_v^{(l)}$ denote the refined representations of user u and item v after the l-th layers propagation, and N_u and N_v denote the neighbors of user u and item v, respectively. $\boldsymbol{e}_u^{(0)}$ and $\boldsymbol{e}_v^{(0)}$ are initialized by $\boldsymbol{e}_u$ and $\boldsymbol{e}_v$. We then utilize all layers' representations to perform different depth information integration,

$$\boldsymbol{e}_u^g = \frac{1}{L}\sum_{l=0}^{L-1}\boldsymbol{e}_u^{(l)} \tag{9}$$

where L denotes the total layer number. We simultaneously conduct the identical propagation process on the bundle-item affiliation graph and obtain refined representation $\boldsymbol{e}_b^g$ for the bundle b.

Then, we utilize the user-bundle graph for the final information propagation. He et al.[40] proposed a matrix form for convolution-based recommendation graph propagation, which is formulated as

$$\boldsymbol{E}^{(k+1)} = \left(\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\right)\boldsymbol{E}^{(k)} \tag{10}$$

where k denotes the k-th layer of propagation, $\boldsymbol{D}$ denotes the degree matrix of the adjacency matrix $\boldsymbol{A}$, and $\boldsymbol{E}$ denotes the concatenation of user and item embeddings in the general scenario. $\boldsymbol{A}$ can be formulated as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \tag{11}$$

where $\boldsymbol{X}$ denotes the user-item implicit feedback matrix, and $\boldsymbol{0}$ denotes an all-zero matrix of appropriate size. The matrix $\boldsymbol{A}$ shows the limitation in that it assigns equal weight to each interaction which is biased based on the analysis of the introduction. Since item bundling is manually and thoughtfully defined, in the bundle recommendation scenario, we propose to leverage the bundle-item affiliation relations to learn the distinct relations between bundle nodes for an enhanced adjacency matrix $\boldsymbol{A}$ (i.e., complement for the lower-right corner of $\boldsymbol{A}$).

Specifically, we start graph normalization on the original bundle-item affiliation graph,

$$\boldsymbol{W}_n = \mathrm{Norm}(\boldsymbol{Y}) \tag{12}$$

The normalization alleviates the influence of high-degree nodes which is beneficial to the stability and generalization of the following graph propagation. Then, we perform the matrix multiplication between the processed matrix and its transpose, yielding the bundle-bundle adjacency matrix,

$$\boldsymbol{W}_{\mathrm{BB}} = \mathrm{Multiply}(\boldsymbol{W}_n,\ \boldsymbol{W}_n^{\mathrm{T}}) \tag{13}$$

$\boldsymbol{W}_{\mathrm{BB}}$ captures distinct relations among bundles and so helps to learn more accurate user preferences. For the upper-left corner of $\boldsymbol{A}$ which represents user nodes' relations, given the considerable noise in user-item interactions[41], we choose to omit them in modeling to avoid performance degradation. Instead, we make the replacement with an identity diagonal matrix to add self-loops, keeping the users' information and minimizing computation in the propagation process,

$$\boldsymbol{W}_{\mathrm{UU}} = \boldsymbol{I}_{M\times M} \tag{14}$$

where $\boldsymbol{I}_{M\times M}$ denotes the $M\times M$ identity matrix. The updated matrix $\boldsymbol{A}$ for user-bundle propagation can be formulated as follows:

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{W}_{\mathrm{UU}} & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{T}} & \boldsymbol{W}_{\mathrm{BB}} \end{bmatrix} \tag{15}$$

Upon acquiring the enhanced $\tilde{\boldsymbol{A}}$, we use the refined representations $\boldsymbol{e}_u^g$ and $\boldsymbol{e}_b^g$ for user-bundle graph propagation. The matrix form of the final propagation can be formulated as follows:

$$\boldsymbol{E}^* = \left(\boldsymbol{D}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\boldsymbol{D}^{-\frac{1}{2}}\right)\boldsymbol{E}^g \tag{16}$$

where $\boldsymbol{D}$ denotes the degree matrix of the enhanced matrix $\tilde{\boldsymbol{A}}$, and $\boldsymbol{E}^g$ denotes the concatenation of refined user and bundle representations. We split $\boldsymbol{E}^*$ to obtain the graph-signal enhanced representations $\boldsymbol{e}_u^*$ for user u and $\boldsymbol{e}_b^*$ for bundle b.

3.4 Recommendation module

Considering the comprehensive user preferences, we calculate individual recommendation scores $y_{u,b}$ with two sets of representations obtained in the SGSE module (Subsection 3.3), which from item preference view (i.e., user-item and bundle-item graphs) and bundle preference view (i.e., user-bundle graph), respectively, and then merge them for the final recommendation,

$$\begin{aligned} y_{u,b} &= y^g + y^*, \\ y^g &= \boldsymbol{e}_u^g \boldsymbol{e}_b^{g\mathrm{T}}, \\ y^* &= \boldsymbol{e}_u^* \boldsymbol{e}_b^{*\mathrm{T}} \end{aligned} \tag{17}$$

where y^g denotes the recommendation score from the item preference view, and y^* denotes the recommendation score from the bundle preference view.

We apply Bayesian personalized ranking[20] loss as the recommendation loss for optimization,

$$\mathcal{L}_{\mathrm{Rec}} = \sum_{(u,\, b,\, b_n)\in Q} -\ln\sigma(y_{u,b} - y_{u,b_n}) \tag{18}$$

where $Q = \{(u, b, b_n) | u\in \boldsymbol{U}, b, b_n \in \boldsymbol{B}, z_{ub} = 1, z_{ub_n} = 0\}$, y_{u,b_n} denotes the predicted score between user u and an unobserved bundle b_n, and $\sigma(\cdot)$ denotes the sigmoid function. We adopt L_2 regularization ($||\cdot||^2$) on the model parameters, ultimately achieve the final loss $\mathcal{L}$ for our framework,

$$\mathcal{L} = \mathcal{L}_{\mathrm{Rec}} + \lambda||\Theta||^2 \tag{19}$$

where λ is the weight of L_2 regularization, Θ is the parameter set of the model. To provide a clearer description of the overall procedure of the proposed method, the workflow is summarized in Algorithm 1.

4 Experiment

In this section, we evaluate SIBR framework by focusing on the following four research questions:

- **RQ1:** How does SIBR perform in comparison to state-of-the-art bundle recommendation methods?
- **RQ2:** What are the effects of the model components on performance?
- **RQ3:** How do the hyperparameters affect the bundle recommendation results?
- **RQ4:** How do the various degrees of incompleteness affect the bundle recommendation?

4.1 Experimental settings

4.1.1 Datasets

We use three real-world bundle recommendation datasets to verify the effectiveness of compared methods, i.e., Youshu[21], NetEase[9], and iFashion[42]. Youshu is a book bundle dataset that is from a Chinese book review site. NetEase is a music bundle dataset, and iFashion is a fashion outfit dataset from Alibaba. All three datasets Youshu, NetEase, and iFashion contain the user's preference for individual items and various bundles. The statistics of the original datasets are summarized in Table 2.

4.1.2 Evaluation protocols

For prediction evaluation, we use the Recall and Normalized Discounted Cumulative Gain (NDCG)

Table 2 Statistics of the three bundle datasets. Average I/B denotes the average number of items contained in a bundle, and average B/U denotes the average number of bundles a user interacted with.

Dataset	Number of user	Number of bundle	Number of item	Number of U-I pairs	Number of U-B pairs	Average I/B	Average B/U
Youshu	8039	4771	32 770	138 515	49 351	37.03	6.14
NetEase	18 528	22 864	123 628	1 128 065	302 276	77.80	16.31
iFashion	53 897	27 694	42 563	2 290 645	1 679 709	3.86	31.16

metrics. Intuitively, the Recall metric considers whether the ground truth is ranked among the top-K_c items, where K_c denotes the cutoff, i.e., the number of recommendations generated for each user. The formulation is defined as follows:

$$\text{Recall@}K_c = \frac{\text{TP@}K_c}{\text{TP@}K_c + \text{FN@}K_c} \tag{20}$$

where TP@K_c refers to the count of positive instances correctly identified within the top-K_c recommendations, while FN@K_c denotes the number of positive instances missed within the top-K_c recommendations.

$$\text{DCG@}K_c = \sum_{i=1}^{K_c} \frac{2^{F(v_i,\,T(u))} - 1}{\log_2(i+1)} \tag{21}$$

$$\text{IDCG@}K_c = \sum_{i=1}^{K_c} \frac{1}{\log_2(i+1)} \tag{22}$$

$$\text{NDCG@}K_c = \frac{\text{DCG@}K_c}{\text{IDCG@}K_c} \tag{23}$$

where NDCG@K_c is a position-aware ranking metric. The function $F(v_i, T(u))$ serves as an indicator function to determine the presence of item v in the set $T(u)$. A value of 1 is assigned if the item is present. Otherwise, the value is 0. DCG denotes Discounted Cumulative Gain, while IDCG represents Ideal Discounted Cumulative Gain. The training/validation/testing sets are randomly split with the ratio of 70%, 10%, and 20%.

4.1.3 Methods for comparison

We compare our SIBR against a diverse set of competitive baselines, including two matrix factorization methods and six graph-based methods.

- **MFBPR[20]:** Matrix Factorization Bayesian Personalized Ranking employs the matrix factorization method for implicit feedback.
- **DAM[21]:** DeepAttentive Multi-task uses a factorized attention network to model user-bundle and user-item interactions in a multi-task manner.
- **GCN[43]:** GCN leverages node relations to capture the graph structural information for representation extraction.
- **NGCF[44]:** Neural Graph Collaborative Filtering utilizes a GCN-based method to capture high-order collaborative signals in user-item interactions (user-bundle interactions in our experiments).
- **BasConv[17]:** Basket recommendationwith graph Convolutional neural network devises aggregators to learn node representations in heterogeneous graphs.
- **BGCN[12]:** Bundle Recommendation with Graph Convolutional Network proposes a GCN-based framework to model complex relations between user, item, and bundle entities.
- **MIDGN[13]:** Multi-view Intent Disentangle Graph Network aims to disentangle users' intents from dual view implicit feedbacks for improvement in bundle recommendation performance.
- **CrossCBR[18]:** Cross-view Contrastive learning for Bundle Recommendation captures the cooperative association by cross-view contrastive learning and mutually enhances the view-aware node representations for bundle recommendation.

4.1.4 Implementation details

We implement the proposed SIBR with PyTorch. Implementations of the general recommendation methods are from open-source projects. We optimize the compared baselines with a standard Adam optimizer. We tune all hyperparameters through grid search. In particular, learning rate in $\{1\times10^{-5}, 5\times10^{-5}, 1\times10^{-4}, 1\times10^{-3}, 1\times10^{-2}\}$, and the weight decay for L_2 regularization in $\{1\times10^{-7}, 5\times10^{-7}, 1\times10^{-6}, 5\times10^{-6}, 1\times10^{-5}\}$. We set the embedding dimension D to 64. The batch size is set to 16 384. We carefully tuned the hyperparameters of all baselines on the test set as suggested in the original papers to achieve their best performance.

4.2 Overall performance comparison (RQ1)

In this part, we compare the recommendation results of our SIBR framework to those of the baseline models. Table 3 shows the Recall@K and NDCG@K values for the three datasets, respectively.

In general, SIBR consistently outperforms eight baseline methods across all evaluation metrics on all datasets. This answers RQ1, showing that our proposed framework is capable of performing effective incomplete bundle recommendation. In terms of the Youshu dataset, compared with the second-best model (i.e., MIDGN), the performance gains of our model are consistently significant, with improvements ranging from 17.2% (achieved with Recall@40) to 19.6% (achieved with NDCG@20) across various metrics. As for the NetEase dataset, compared with the second-best model (i.e., CrossCBR), the performance gains of SIBR on it range from reasonably large (23.3% achieved with NDCG@20) to significantly large (39.3% achieved with Recall@40), and for the

Table 3 Overall performance for the three datasets in Recall and NDCG metrics. Recall and NDCG are abbreviated as *R* and *N*. The best performance is in boldface and the second runners are underlined. Improvement denotes the relative performance gain of our method over the best competing baseline. Note that the training set is randomly missing at a rate of 70% to highlight the IBRs.

Method	Youshu				NetEase				iFashion			
	R@20	N@20	R@40	N@40	R@20	N@20	R@40	N@40	R@20	N@20	R@40	N@40
MFBPR	0.1666	0.0738	0.2613	0.1052	0.0283	0.0123	0.0412	0.0147	0.0416	0.0201	0.0781	0.0343
DAM	0.1716	0.0843	0.2562	0.1105	0.0288	0.0134	0.0435	0.0156	0.0378	0.0176	0.0755	0.0321
GCN	0.1705	0.0784	0.2602	0.1001	0.0297	0.0125	0.0476	0.0149	0.0437	0.0199	0.0767	0.0352
NGCF	0.1678	0.0756	0.2597	0.1043	0.0311	0.0139	0.0506	0.0163	0.0399	0.0208	0.0697	0.0321
BasConv	0.1693	0.0675	0.2655	0.0991	0.0282	0.0122	0.0497	0.0155	0.0445	0.0194	0.0742	0.0315
BGCN	0.1976	0.0881	0.2823	0.1074	0.0319	0.0134	0.0523	0.0179	0.0463	0.0245	0.0757	0.0325
MIDGN	<u>0.2157</u>	<u>0.1146</u>	<u>0.3061</u>	<u>0.1388</u>	0.0288	0.0148	0.0483	0.0198	0.0373	0.0262	0.0614	0.0346
CrossCBR	0.1678	0.0915	0.2391	0.1107	<u>0.0409</u>	<u>0.0223</u>	<u>0.0631</u>	<u>0.0279</u>	<u>0.0738</u>	<u>0.0552</u>	<u>0.1082</u>	<u>0.0673</u>
Ours	**0.2562**	**0.1371**	**0.3588**	**0.1643**	**0.0538**	**0.0275**	**0.0879**	**0.0364**	**0.0819**	**0.0582**	**0.1261**	**0.0738**
Improvement (%)	18.8	19.6	17.2	18.4	31.5	23.3	39.3	30.5	10.9	5.4	16.5	9.6

iFashion dataset, compared with the second-best model (i.e., CrossCBR), the performance gains of SIBR on it range from reasonably large (5.4% achieved with NDCG@20) to significantly large (16.5% achieved with Recall@40).

In particular, the graph-based methods (e.g., BGCN, MIDGN, and CrossCBR) perform generally better than matrix factorization methods (e.g., DAM). We attribute this to the capability of graph-based methods in capturing high-order connectivity.

Among the three datasets, the proposed SIBR performs best on the NetEase dataset which possesses the largest average number of bundle-item affiliations (i.e., 77.80). The metric performance SIBR gains in the NetEase dataset is improved by at least 23.3% compared to the second-best baseline. We attribute this to the fact that the larger the size of bundle-item affiliations, the clearer the preferences embodied in bundles. As a user misses interactions with such bundles, it will be difficult to capture accurate user interests in the user-bundle context, leading to suboptimal bundle recommendations. The proposed SIBR learns collaborative information on the joint user-bundle-item graph, effectively utilizing the cross-graph knowledge as missing interaction supplements. The result shows the excellent effectiveness of SIBR within incomplete bundle recommendations.

4.3 Model ablation (RQ2)

To better understand our proposed framework, i.e., SNRL and SGSE, we study SIBR as follows:

- **Backbone** applies the basic backbone network structure only, without the SNRL module and graph enhancement in the SGSE module;
- **Backbone+SNRL** utilizes node-enhanced embeddings instead of randomly initialized ones;
- **Backbone+SGSE** employs a graph-signal enhancement strategy to perform distinct relations for more accurate user and bundle representations;
- **Backbone+SNRL+SGSE (SIBR)** integrates the proposed two enhancement components: Structure-aware node-enhanced embeddings and structure-aware signal-enhanced graph, so as to alleviate the missing data problem for better incomplete bundle recommendations.

From Table 4, we have the following observations:

Compared with Backbone, on Youshu, Backbone+SNRL gains the performance ranges from 3.64% (achieved in Recall@40) to 10.24% (achieved in Recall@20), and Backbone+SGSE gains the performance ranges from 1.21% (achieved in Recall@40) to 4.98% (achieved in Recall@20). These results show that both SNRL and SGSE are significantly effective in incomplete bundle recommendations. The SNRL component contributes to a more direct and successful cross-graph

Table 4 Experimental results of ablation studies on Youshu, where Recall and NDCG are abbreviated as *R* and *N*.

Method	R@20	N@20	R@40	N@40
Backbone	0.2207	0.1209	0.3214	0.1474
Backbone+SNRL	0.2433	0.1293	0.3331	0.1536
Backbone+SGSE	0.2317	0.1284	0.3253	0.1535
Backbone+SNRL+SGSE (ours)	0.2562	0.1371	0.3588	0.1643

collaborative information utilization, learning more accurate user preferences in the pretraining phase. The SGSE component explores the graph signal enhancement for the user-bundle bipartite graph in incomplete bundle recommendations, extracting distinct user-user and bundle-bundle relations. Both of them improve the accuracy of bundle recommendation performance.

We made comparisons between the Backbone+SNRL and the Backbone+SGSE in the following. Considering the Recall metric on Youshu, the performance gains of Backbone+SNRL over Backbone+SGSE range from 2.39% (achieved in Recall@40) to 5.01% (achieved in Recall@20). In terms of the NDCG metric, both methods exhibit nearly consistent performance. The result indicates that when considering the order of the recall list, the performance of both is comparable, and the Backbone+SNRL outperforms without order restriction.

The proposed method integrates the SNRL and SGSE module, leading to performance gains ranging from 11.46% (achieved in NDCG@40) to 16.08% (achieved in Recall@20) compared with Backbone, since the joint modeling of available user-item interactions and bundle-item affiliations helps to learn node structural patterns and distinct relations among users and bundles. The experimental result proves the excellent effectiveness of the integration of the two proposed components.

4.4 Effect of hyperparameters (RQ3)

Our proposed SIBR framework mainly introduces three hyperparameters, i.e., T, p, and q. We set their initial values to 20, 1, and 1, respectively. When analyzing the effect of a single hyperparameter, the other two are held constant at their initial settings.

From Fig. 3, along with the statistical data in Table 2, we have the following observations: (1) T determines the number of nodes visited during one random walk, where we found the optimal T is about 20. For datasets of varying scales, sufficient nodes in a random walk are effective in capturing the structural information, thereby improving performance. Conversely, when data is sparse (e.g., iFashion with a low average item number in one bundle), a short walk can lead to

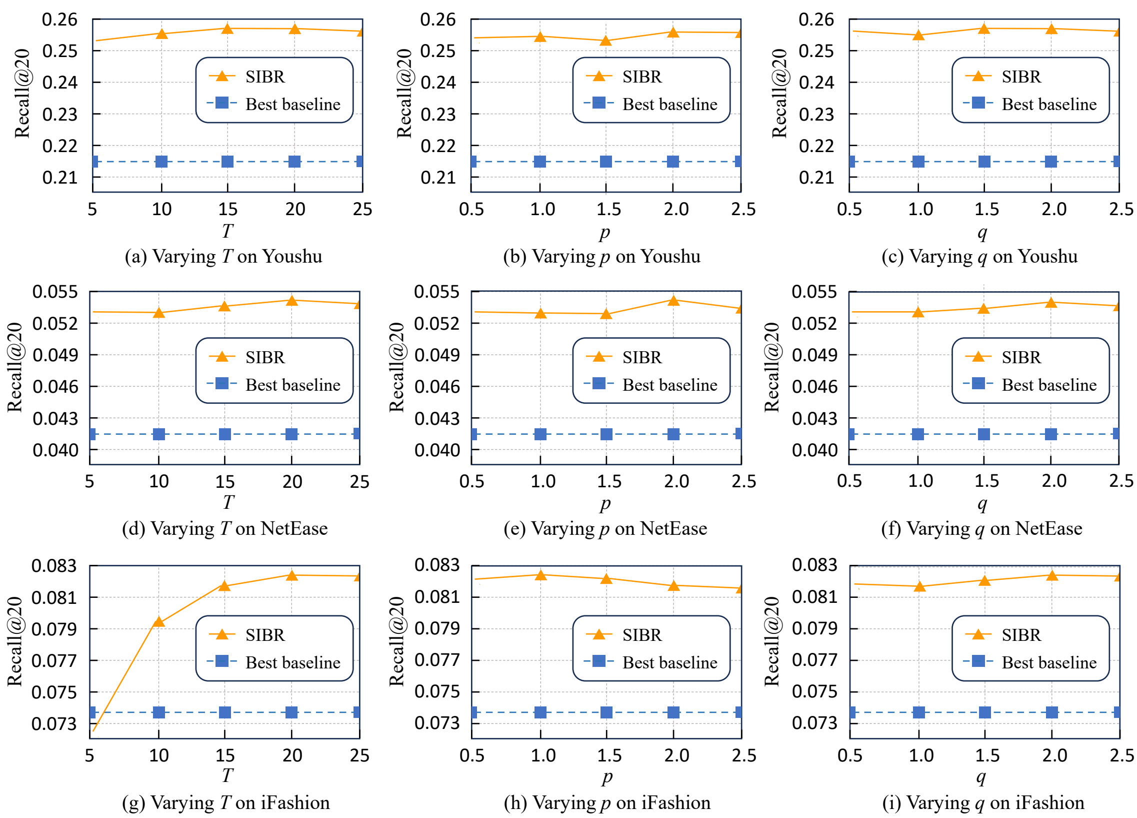


Fig. 3 Performance regarding Recall@20 of the best baseline and proposed SIBR with varying *T*, *p*, and *q* on Youshu, NetEase, and iFashion.

substantial performance degradation, as illustrated in Fig. 3g with T set to 5; (2) p specifies the probability of returning to the former node during the sequence sampling, and q determines the probability of moving away from the former node during the sampling. The optimal p and q are contingent upon the characteristics of the data. A higher average item number in one bundle indicates the potentially richer local and global structure information (e.g., Youshu and NetEase). It is essential to balance the opposite walk directions controlled by p and q in this case, for comprehensive structure information capture. The experiments demonstrate that setting p and q both to 2 yields the best performance. For datasets with converse characteristics (e.g., iFashion), encouraging random walks to move away from the origin may introduce significant noise. In this case, a lower p and a higher q can facilitate improved model performance. The optimal values of p and q for iFashion are about 1 and 2, respectively. The rules for selecting T, p, and q can be a rule of thumb in practice across the datasets of various scales.

4.5 Effect of incompleteness (RQ4)

To validate the adaptability of SIBR to user-bundle interaction data with various degrees of incompleteness, we run different proportions of random missing value for the Youshu and NetEase training sets to perform comparative experiments with the overall best baseline. From Fig. 4, we have the following observations:

With a low degree of missing interactions (20%), the proposed SIBR slightly surpasses the best baseline by no substantial margin. Nevertheless, as the proportion of missing interactions escalates, SIBR consistently exhibits robust performance. In stark contrast, the performance of the best baseline significantly declined, nearly halving when the missing rate reached 70% compared to its performance at a 20% missing rate. The result shows our proposed SIBR's superior capability to mitigate the impact of missing interactions. This effectiveness is attributed to its exploitation of collaborative information, node relations in user-item interactions, and bundle-item affiliations, compensating effectively for the incompleteness of interactions.

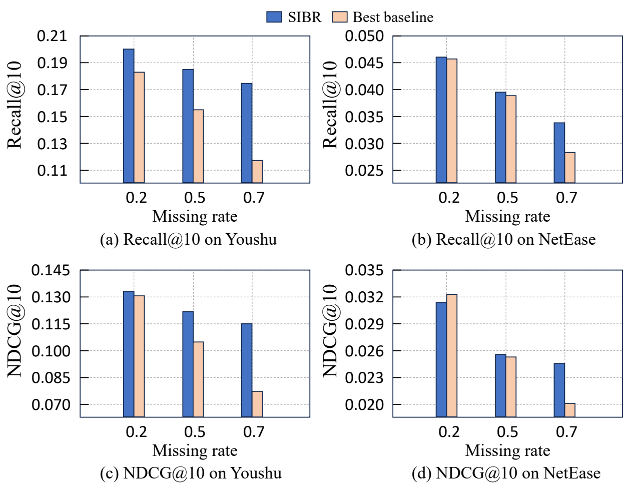


Fig. 4 Performance regarding Recall@10 and NDCG@10 of the best baseline and SIBR at different missing rates.

5 Conclusion

In this study, we introduce the SIBR framework, which addresses the challenges posed by incomplete user-bundle interaction data through the integration of user-item interactions and bundle-item affiliations. Specifically, we employ a parameterized random walk strategy to learn the consistent information in the joint graph, enabling the precise acquisition of user preferences for bundles. Furthermore, we capture distinct relations among user and bundle nodes, thereby refining the convolution process for more accurate representation. Extensive experiments have demonstrated clear improvements of SIBR. In the future, it would be interesting to consider the incorporation of node attributes when they are available, improving the adaptability of our work in diverse data scenarios.

Acknowledgment

This work was supported by the Fujian Provincial Artificial Intelligence Industry Development Technology Project (No. 2025H0042), the Fujian Provincial Natural Science Foundation of China (No. 2025J01540), the National Natural Science Foundation of China (No. 62302098), the Zhejiang Provincial Natural Science Foundation of China (No. LQ23F020007), and the Zhejiang Provincial Department of Agriculture and Rural Affairs Project (No. 2024SNJF044).

References

[1] W. Wei, J. Liu, X. Mao, G. Guo, F. Zhu, P. Zhou, and Y. Hu, Emotion-aware chat machine: Automatic emotional response generation for human-like emotional interaction, in *Proc. 28th ACM Int. Conf. Information and Knowledge Management*, Beijing, China, 2019, pp. 1401–1410.

[2] D. Li and J. A. Esquivel, Trust-aware hybrid collaborative recommendation with locality-sensitive hashing, *Tsinghua Science and Technology*, vol. 30, no. 4, pp. 1421–1434, 2025.

[3] M. Zhang, X. Zhang, W. Pedrycz, S. Wang, and G. Wu, Learning fine-grained user preference for personalized recommendation, *Tsinghua Science and Technology*, vol. 30, no. 6, pp. 2544–2556, 2025.

[4] F. Wang, H. Zhu, G. Srivastava, S. Li, M. R. Khosravi, and L. Qi, Robust collaborative filtering recommendation with user-item-trust records, *IEEE Trans. Computat. Soc. Syst.*, vol. 9, no. 4, pp. 986–996, 2022.

[5] Y. Liu, H. Wu, K. Rezaee, M. R. Khosravi, O. I. Khalaf, A. A. Khan, D. Ramesh, and L. Qi, Interaction-enhanced and time-aware graph convolutional network for successive point-of-interest recommendation in traveling enterprises, *IEEE Trans. Ind. Inf.*, vol. 19, no. 1, pp. 635–643, 2023.

[6] X. Zhang, B. Xu, F. Ma, C. Li, L. Yang, and H. Lin, Beyond co-occurrence: Multi-modal session-based recommendation, *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 4, pp. 1450–1462, 2024.

[7] L. Hu, G. Song, Z. Xie, and K. Zhao, Personalized recommendation algorithm based on preference features, *Tsinghua Science and Technology*, vol. 19, no. 3, pp. 293–299, 2014.

[8] T. Zhu, P. Harrington, J. Li, and L. Tang, Bundle recommendation in eCommerce, in *Proc. 37th Int. ACM SIGIR Conf. Research & Development in Information Retrieval*, Gold Coast, Australia, 2014, pp. 657–666.

[9] D. Cao, L. Nie, X. He, X. Wei, S. Zhu, and T. S. Chua, Embedding factorization models for jointly recommending items and user generated lists, in *Proc. 40th Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Tokyo, Japan, 2017, pp. 585–594.

[10] Y. Liu, M. Xie, and L. V. S. Lakshmanan, Recommending user generated item lists, in *Proc. 8th ACM Conf. Recommender Systems*, Silicon Valley, CA, USA, 2014, pp. 185–192.

[11] A. Pathak, K. Gupta, and J. McAuley, Generating and personalizing bundle recommendations on *steam*, in *Proc. 40th Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Tokyo, Japan, 2017, pp. 1073–1076.

[12] J. Chang, C. Gao, X. He, D. Jin, and Y. Li, Bundle recommendation with graph convolutional networks, in *Proc. 43rd Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Virtual Event, 2020, pp. 1673–1676.

[13] S. Zhao, W. Wei, D. Zou, and X. Mao, Multi-view intent disentangle graph networks for bundle recommendation, in *Proc. 36th AAAI Conf. Artificial Intelligence*, Virtual Event, 2022, pp. 4379–4387.

[14] Q. Deng, K.Wang, M. Zhao, Z. Zou, R.Wu, J. Tao, C. Fan, and L. Chen, Personalized bundle recommendation in online games, in *Proc. 29th ACM Int. Conf. Information & Knowledge Management*, Virtual Event, 2020, pp. 2381–2388.

[15] X. L. Li, W. D. Xi, X. X. Xing, and C. D. Wang, Auto graph filtering for bundle recommendation, in *Proc. 2023 IEEE Int. Conf. Data Mining* (*ICDM*), Shanghai, China, 2023, pp. 299–308.

[16] Y. Ren, H. Zhang, L. Fu, X. Wang, and C. Zhou, Distillation-enhanced graph masked autoencoders for bundle recommendation, in *Proc. 46th Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Taipei, China, 2023, pp. 1660–1669.

[17] Z. Liu, M. Wan, S. Guo, K. Achan, and P. S. Yu, BasConv: Aggregating heterogeneous interactions for basket recommendation with graph convolutional neural network, in *Proc. 2020 SIAM Int. Conf. Data Mining, SDM 2020*, Cincinnati, OH, USA, 2020, pp. 64–72.

[18] Y. Ma, Y. He, A. Zhang, X. Wang, and T. S. Chua, CrossCBR: Cross-view contrastive learning for bundle recommendation, in *Proc. 28th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Washington, DC, USA, 2022, pp. 1233–1241.

[19] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, Factorizing personalized Markov chains for next-basket recommendation, in *Proc. 19th Int. Conf. World Wide Web*, Raleigh, NC, USA, 2010, pp. 811–820.

[20] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, BPR: Bayesian personalized ranking from implicit feedback, in *Proc. 25th Conf. Uncertainty in Artificial Intelligence*, Montreal, Canada, 2009, pp. 452–461.

[21] L. Chen, Y. Liu, X. He, L. Gao, and Z. Zheng, Matching user with item set: Collaborative bundle recommendation with deep attention network, in *Proc. 28th Int. Joint Conf. Artificial Intelligence*, Macao, China, 2019, pp. 2095–2101.

[22] Y. Liu, X. Zhou, H. Kou, Y. Zhao, X. Xu, X. Zhang, and L. Qi, Privacy-preserving point-of-interest recommendation based on simplified graph convolutional network for geological traveling, *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 4, p. 76, 2024.

[23] X. Liu, S. Meng, Q. Li, L. Qi, X. Xu, W. Dou, and X. Zhang, SMEF: Social-aware multi-dimensional edge features-based graph representation learning for recommendation, in *Proc. 32nd ACM Int. Conf. Information and Knowledge Management*, Birmingham, UK, 2023, pp. 1566–1575.

[24] H. Ke, L. Li, P. Wang, J. Yuan, and X. Tao, Hyperbolic mutual learning for bundle recommendation, in *Proc. 28th Int. Conf. Database Systems for Advanced Applications*, Tianjin, China, 2023, pp. 417–433.

[25] L. Kong, G. Li, W. Rafique, S. Shen, Q. He, M. R. Khosravi, R. Wang, and L. Qi, Time-aware missing healthcare data prediction based on ARIMA model, *IEEE/ACM Trans. Computat. Biol. Bioinf.*, vol. 21, no. 4, pp. 1042–1050, 2024.

[26] C. Liu, J. Wen, X. Luo, and Y. Xu, Incomplete multi-view multi-label learning via label-guided masked view-and category-aware transformers, in *Proc. 37th AAAI Conf. Artificial Intelligence*, Washington, DC, USA, 2023, pp. 8816–8824.

[27] S. Y. Li, Y. Jiang, and Z. H. Zhou, Partial multi-view clustering, in *Proc. 28th AAAI Conf. Artificial Intelligence*,

Québec City, Canada, 2014, pp. 1968−1974.

[28] M. Hu and S. Chen, Doubly aligned incomplete multi-view clustering, in *Proc. 27^{th} Int. Joint Conf. Artificial Intelligence*, Stockholm, Sweden, 2018, pp. 2262−2268.

[29] D. Xia, Y. Yang, and S. Yang, Incomplete multi-view clustering via auto-weighted fusion in partition space, *Tsinghua Science and Technology*, vol. 28, no. 3, pp. 595−611, 2023.

[30] H. Wang, Y. Chen, C. Ma, J. Avery, L. Hull, and G. Carneiro, Multi-modal learning with missing modality via shared-specific feature modelling, in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Vancouver, Canada, 2023, pp. 15878−15887.

[31] H. Cai, W. Huang, S. Yang, S. Ding, Y. Zhang, B. Hu, F. Zhang, and Y. M. Cheung, Realize generative yet complete latent representation for incomplete multi-view learning, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 5, pp. 3637−3652, 2024.

[32] C. Xu, H. Liu, Z. Guan, X. Wu, J. Tan, and B. Ling, Adversarial incomplete multiview subspace clustering networks, *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10490−10503, 2022.

[33] Y. Zhang, Z. Chen, and W. Zhang, MACO: A modality adversarial and contrastive framework for modality-missing multi-modal knowledge graph completion, in *Proc. 12^{th} National CCF Conf. Natural Language Processing and Chinese Computing*, Foshan, China, 2023, pp. 123−134.

[34] R. Mafrur, M. A. Sharaf, and G. Zuccon, VizPut: Insight-aware imputation of incomplete data for visualization recommendation, arXiv preprint arXiv: 2311.07926, 2023.

[35] H. Zhang, C. Tian, Y. Li, L. Su, N. Yang, W. X. Zhao, and J. Gao, Data poisoning attack against recommender system using incomplete and perturbed data, in *Proc. 27^{th} ACM SIGKDD Conf. Knowledge Discovery & Data Mining*, Virtual Event, 2021, pp. 2154−2164.

[36] X. Zhang, X. Xin, D. Li, W. Liu, P. Ren, Z. Chen, J. Ma, and Z. Ren, Variational reasoning over incomplete knowledge graphs for conversational recommendation, in *Proc. 16^{th} ACM Int. Conf. Web Search and Data Mining*, Singapore, 2023, pp. 231−239.

[37] A. Grover and J. Leskovec, node2vec: Scalable feature learning for networks, in *Proc. 22^{nd} ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 855−864.

[38] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space, in *Proc. 1^{st} Int. Conf. Learning Representations*, Scottsdale, AZ, USA, 2013.

[39] A. J. Walker, An efficient method for generating discrete random variables with general distributions, *ACM Trans. Math. Softw.*, vol. 3, no. 3, pp. 253−256, 1977.

[40] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, LightGCN: Simplifying and powering graph convolution network for recommendation, in *Proc. 43^{rd} Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Virtual Event, 2020, pp. 639−648.

[41] Y. Gao, Y. Du, Y. Hu, L. Chen, X. Zhu, Z. Fang, and B. Zheng, Self-guided learning to denoise for robust recommendation, in *Proc. 45^{th} Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Madrid, Spain, 2022, pp. 1412−1422.

[42] W. Chen, P. Huang, J. Xu, X. Guo, C. Guo, F. Sun, C. Li, A. Pfadler, H. Zhao, and B. Zhao, POG: Personalized outfit generation for fashion recommendation at alibaba ifashion, in *Proc. 25^{th} ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*, Anchorage, AK, USA, 2019, pp. 2662−2670.

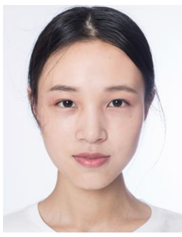
[43] R. van den Berg, T. N. Kipf, and M. Welling, Graph convolutional matrix completion, arXiv preprint arXiv: 1706.02263, 2017.

[44] X. Wang, X. He, M. Wang, F. Feng, and T. S. Chua, Neural graph collaborative filtering, in *Proc. 42^{nd} Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Paris, France, 2019, pp. 165−174.

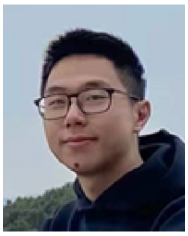
Yanchao Tan received the PhD degree from Zhejiang University, China, in 2022. She is currently an associate professor at College of Computer and Data Science and the Fujian Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou University, China. Her research interests include data mining, recommender system, and healthcare.

Yonghui Yang received the master degree from Hefei University of Technology, China, in 2021. He is currently pursuing the PhD degree at the same university. He has published several papers in top journals and conferences, such as *IEEE TKDE*, SIGIR, and IJCAI. His research interests include graph learning and recommender systems.

Zhijun Chen received the bachelor degree from Fuzhou University, China, in 2023. He is currently pursuing the master degree at the same university. His research interests include data mining and recommender system.



Xinting Liao received the bachelor degree in computer science from Sichuan University, China, in 2020. She is currently pursuing the PhD degree at College of Computer Science, Zhejiang University, China. Her research interests include cross domain recommendation and privacy-preserving machine learning.



Weiming Liu received the bachelor degree in electronic science and technology from Zhejiang University, China, in 2021. He is currently pursuing the PhD degree at College of Computer Science, Zhejiang University, China. His research interests mainly focus on transfer learning with its applications on recommendation system.

Guofang Ma received the PhD degree from Zhejiang University, China, in 2021. She is currently a lecturer at School of Computer Science and Technology, Zhejiang Gongshang University, China. Her research interests include cross-domain recommender system and healthcare.