H^2-GCN: Towards Complex Higher-order Interactions for Session-based Recommendation

1st Qi Zhang†
School of Big Data and Software Engineering
Chongqing University
Chongqing, China
qizhang@stu.cqu.edu.cn

1st Kunqi Wang†
Faculty of Engineering
The University of Sydney
Sydney, Australia
kwan0569@uni.sydney.edu.au

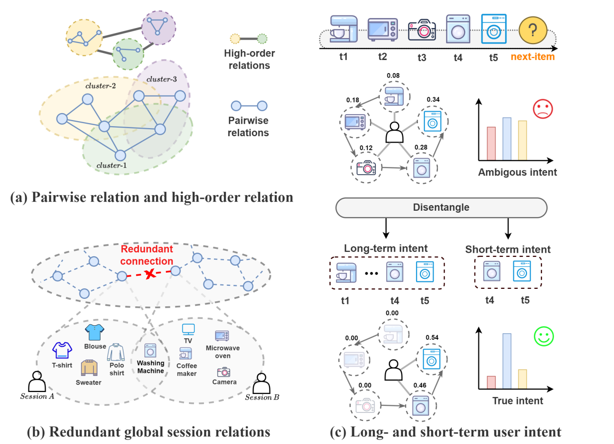
2nd Huayi Shen, and 3rd Wei Zhou*
School of Big Data and Software Engineering
Chongqing University
Chongqing, China
zhouwei@cqu.edu.cn

***Abstract*—Session-based recommendation (SBR) aims to predict the next item based on anonymous user sessions. However, existing studies still face challenges below: (1) Inadequate capture of both pairwise and high-order relationships in graph relations. (2) Spurious edges in the global graph lead to redundant global relations extraction. (3) Complex coupling of different user intents caused by noise and dynamics within sessions. To address these issues, we propose a novel Higher-order Hybrid Graph Convolutional Network (H^2-GCN) for SBR, which effectively captures multi-level relationships and refines diverse user intents, enabling effective recommendations. Specifically, we first present a higher-order graph based on simplicial complexes (SCs) to learn the multilevel complex dependencies of items and sessions separately. Moreover, two intent expert modules are designed to independently learn user intents within sessions: the short-term intent expert module is proposed to block long-term dependencies and activate true short-term intents from user noise behaviors, while the long-term intent expert module employs a dynamic-static graph collaborative propagation to filter redundant global relations and derive true long-term intents. Eventually, we guide the intents obtained from the two expert modules to form the final user intents. Extensive experiments show that H^2-GCN outperforms state-of-the-art methods in three real-world datasets.**

***Index Terms*—Session-based Recommendation, Graph Neural Networks, Expert System**

I. Introduction

Recommendation systems have demonstrated the ability to filter information from massive datasets to provide personalized recommendations. Session-based recommendation (SBR) is highly effective at capturing short-term user preferences and providing timely, accurate recommendations. SBR primarily relies on anonymous sessions that record interactions between the user and the online platform over a short period to predict the next item. Some early studies based on recurrent neural networks (RNNs) [1]–[3] predict the next item by capturing the sequential order of transitions between items. However, since a session often includes multiple user intents, this diversity and complexity make it difficult for these time-series models to accurately infer users' true intents. Later, graph neural networks (GNNs) [4]–[7] have demonstrated the superiority in modeling pairwise relations between adjacent items and high-order connectivity among clusters of items. However, these approaches still suffer from data sparsity due to short sessions and the over-mining challenge in global relations, which hinders further improvement in recommendation performance. Thus, more studies have focused on exploring intra-session and inter-session relations, using GNNs as encoders for item and session representations, meanwhile modeling long- and short-term user intent to achieve more accurate recommendations [8]–[11].



(a) Pairwise relation and high-order relation

(b) Redundant global session relations

(c) Long- and short-term user intent

Fig. 1: Illustrations of (a) there are pairwise and high-order relations among items, (b) global redundant relations between session A and session B, and (c) necessity of disentangling long- and short-term user intents.

Existing methods still suffer from the following three limitations. (1) *Inadequate relation capturing*. **Firstly**, in between items, there are often pairwise and higher-order relations, both of which play an important role in modeling item representations from the topological structures, as shown in Figure 1 (a). Previous works [5], [12] either focus on capturing the pairwise relations between items or the higher-order relations among the item clusters, ignoring the fine-grained relations within the clusters, failing to consider fusing the two relations to fully learn item representations. (2) *Redundant global dependencies*. **Secondly**, previous models [11], [12] may introduce redun-

* Wei Zhou is the corresponding author; † These authors contributed equally to this work.

dant connections when constructing global relation graphs[1], leading to redundant relation mining. Taking Figure 1 (b) as an example, when searching for clothing items, the user of *Session-A* accidentally clicks on a washing machine item, making *Session-A* related to clothing and *Session-B* related to home appliances form a redundant connection. Such a redundant connection links two sessions with different user intents, destabilizing global relations and leading to irrelevant recommendations. (3) *Coupling of long- and short-term intents*. **Thirdly**, most existing studies [5], [10] model short-term user intent from the global contextual relations within the session, as shown in Figure 1 (c). These approaches easily cause long- and short-term intents to interfere with each other, making captured user intent unclear and affecting recommendation performance. To more accurately understand user intent, it is necessary to decouple the learning processes for short- and long-term user behaviors, separately mining user intent at different time scales to reduce interference between them.

To address the above limitations, we propose a novel Higher-order Hybrid Graph Convolutional Network (H^2-GCN) for SBR, which effectively captures multi-level relationships and refines diverse user intents, thereby achieving effective recommendations. Specifically, **to address the limitation of inadequate relation capturing**, we capture the relations with a higher-order graph based on simplicial complexes (SCs) in both item- and session-level, which not only considers the higher-order but also the pairwise relations within it, thereby refining the item representations. **To address the limitation of redundant global dependencies**, we design a long-term intent expert module to perform a dynamic-static graph collaborative propagation, which enables adaptive adjustment of global dependencies to avoid irrelevant connections between sessions, achieving accurate modeling of long-term user intents. **To address the limitation caused by the coupling of long- and short-term intents**, we design a short-term intent expert module, in which we introduce an intent disentanglement learning module with a hierarchical attention mechanism to block long-term dependencies in the session and eliminate the interference of user noise behaviors to efficiently learn short-term user intents. Extensive experiments show that H^2-GCN outperforms state-of-the-art SBR models across three real-world datasets, demonstrating its effectiveness in modeling long- and short-term user intents separately.

II. Related Work

A. Session-based Recommendation

Early methods mainly rely on recurrent neural networks (RNNs) to model item transition patterns in anonymous interaction sequences [13]. Attention mechanisms were subsequently introduced into SBR. Representative models include NARM [2], which combines a GRU with soft attention to model user intent, and its extensions, such as CSRM [3]. Other attention-based approaches, including STAMP [14], SASRec [15], and CoSAN [16], further enhance intent modeling by capturing short- and long-term dependencies within or across sessions. Several studies also leverage sparse or gated attention to mitigate session noise, such as DSAN [17] and DIDN [18]. GNNs have been widely adopted for SBR due to their ability to model complex item relations. SR-GNN [7] constructs session graphs to capture item transitions, while subsequent works extend this idea by incorporating global item relations [5], virtual nodes [19], or high-order structures such as hypergraphs [10], [12]. Recent methods further address inter-session dependencies and session noise using attention or contrastive learning [11], [20]. Intent disentanglement learning reduces interference among multiple user intents and has recently shown strong effectiveness in SBR. Representative methods include Disen-GNN [21], SDHID [22], and HIDE [23], which validate the benefit of decoupled intent modeling. Despite their effectiveness, existing methods generally fail to jointly model the coupling of short- and long-term user intents and still suffer from insufficient relational modeling.

B. High-order GNNs for Recommendation

Higher-order GNNs extend vanilla GNNs to model interactions beyond pairwise relations. Recently, hypergraph learning has been introduced into recommender systems to capture high-order relations via hyperedges and improve recommendation performance [12], [24], [25]. However, hypergraph-based methods overlook fine-grained relations within hyperedges and incur high construction costs. Simplicial GNNs provide a more expressive alternative by explicitly modeling both pairwise and higher-order interactions. Existing studies, such as SCNN [26] and HiGCN [27], demonstrate the effectiveness of simplicial structures in capturing high-order connectivities. Only limited work has explored simplicial GNNs in recommender systems, with HEM-GNN [28] being the most relevant, applying them to multi-behavior recommendation. To the best of our knowledge, no prior work has explored session-based recommendation with simplicial GNNs, and our work fills this gap by modeling complex higher-order relations among items and sessions.

III. Preliminary

Problem Statement. Let $S = \{s_1, s_2, \ldots, s_m\}$ denote the set of all sessions, and $V = \{v_1, v_2, \ldots, v_w\}$ denote the set of unique items. Each session $s_t = \{v_{t,1}, v_{t,2}, \ldots, v_{t,n}\}$ records a sequence of items clicked by a user, where $v_{t,k} \in V$ is the k-th clicked item and n is the session length. Given a session s_t, the goal of session-based recommendation is to predict the next item $v_{t,n+1}$ based on the observed interaction sequence.
Definition 1. Simplicial Complexes. A simplicial complex $\mathcal{K}$ is a finite collection of node subsets that is closed under taking nonempty subsets. A p-simplex $\sigma = [v_0, v_1, \ldots, v_p] \in \mathcal{K}_p$ is a set of $p+1$ nodes, where $\mathcal{K}_p$ denotes the set of all p-simplices. In SBR, individual items are treated as 0-simplices, pairwise item connections as 1-simplices, and closed triangular

[1] In the Diginetica, Tmall, and Nowplaying datasets, connections created based on items that co-occur once in a session account for 78.54%, 87.09%, and 81.00% of the total connections, respectively.

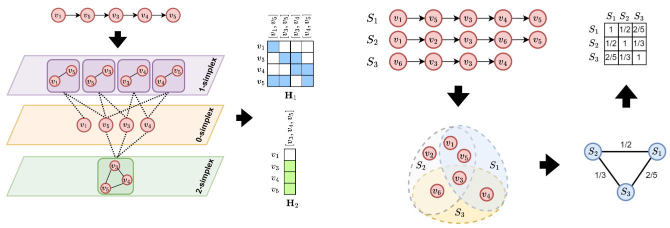


Fig. 2: Illustration of simplicial graph and global relation graph construction.

connections as 2-simplices. This formulation enables modeling both pairwise relations and higher-order interactions among items.

Definition 2. Simplicial Graph. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote a simplicial graph based on FP theory [27], where $\mathcal{V}$ is a set containing different orders of simplices and $\mathcal{E}$ is a set containing all edges that exist between 0-simplices and p-simplices ($p \geq 1$). The item-level simplicial graph can be represented as an incidence matrix $\mathbf{H}_p \in \mathbb{R}^{w \times |\mathcal{K}_p|}$ as shown in Figure 2, where $\mathbf{H}_p(v, \sigma) = 1$ indicates that node $v \in V$ is contained in the simplex $\sigma \in \mathcal{K}_p$, otherwise 0. Similarly, we also construct a session-level simplicial graph by considering sessions as nodes to better extract the session-level higher-order relations, which are represented as an incidence matrix $\mathbf{H}_p^R \in \mathbb{R}^{m \times |\mathcal{K}_p^S|}$. When at least one co-occurring item appears in both sessions, a relation exists between them.

Definition 3. Global Relation Graph. Let $\mathcal{G}_R = (\mathcal{V}_s, \mathcal{E}_s)$ denote a global relation graph, where $\mathcal{V}_s = \{v_i^s \in S\}$ and $\mathcal{E}_s = \{(v_i^s, v_j^s) | v_i^s, v_j^s \in S, |s_i \cap s_j| \geq 1\}$. An incidence matrix for $\mathcal{G}_R$ is defined as $\mathbf{A}_R \in \mathbb{R}^{m \times m}$. To distinguish the significance of different relations, we assign weights W_{ij} for all edges based on co-occurring items between sessions, where $W_{ij} = |s_i \bigcap s_j| / |s_i \bigcup s_j|$.

IV. Methodology

A. Short-term Intent Expert Module

The short-term intent expert module aims to learn item representations from item-level higher-order interactions and to disentangle short-term user intent learning from user noise behaviors and interference from long-term dependencies, thereby improving short-term intent accuracy.

1) Higher-order Item Representation Learning: We first capture item representations through an item-level simplicial graph $\mathbf{H}_p$ with a two-step random walk message propagation, following the approach of HiGCN [27]. Notably, to enhance its suitability for the SBR task, we additionally perform a gated fusion operation on updating node embeddings. The improved HiGCN can be defined as,

$$\widetilde{\mathbf{A}}_p = \mathbf{D}_{p,v}^{-1/2} \mathbf{H}_p \mathbf{D}_{p,h}^{-1} \mathbf{H}_p \mathbf{D}_{p,v}^{-1/2}, \tag{1}$$

$$\widehat{\mathbf{e}}_{v_i}^h = \text{leakyReLU}\left(\Big\Vert_{p=1}^{P} \left(\sum_{k=0}^{K} \gamma_{p,k} \widetilde{\mathbf{A}}_p^k \mathbf{e}_{v_i}^o \mathbf{W}_1 \right) \mathbf{W}_2 \right), \tag{2}$$

where $\widetilde{\mathbf{A}}_p \in \mathbb{R}^{w \times |\mathcal{K}_p|}$ is the normalized higher-order adjacency matrix. $\mathbf{D}_{p,v}$ and $\mathbf{D}_{p,\sigma}$ are diagonal matrices, where $D_{p,v_i} = \sum_{j=1}^{|\mathcal{K}_p|} \mathbf{H}_p(v_i, \sigma_j)$ and $D_{p,\sigma_j} = \sum_{i=1}^{w} \mathbf{H}_p(v_i, \sigma_j)$. $\mathbf{W}_1 \in \mathbb{R}^{d \times d_h}$ and $\mathbf{W}_2 \in \mathbb{R}^{Pd_h \times d}$ are learnable weights, γ denotes the learnable contribution of different hops, K is the largest hops under consideration, $\mathbf{e}_{\mathbf{v_i}}^{\mathbf{0}}$ represents the initial embedding of each item. Next, we perform a gated fusion to better combine the captured higher-order item representations with the original ones,

$$\begin{aligned} \rho_i &= \text{Sigmoid}\left(\mathbf{e}_{v_i}^h \mathbf{W}_3 + \mathbf{b}_1 \right), \\ \mathbf{e}_{v_i}^h &= \rho_i \widehat{\mathbf{e}}_{v_i}^h + (1 - \rho_i)\, \mathbf{e}_{v_i}^o, \end{aligned} \tag{3}$$

where $\mathbf{e}_{v_i}^h \in \mathbb{R}^d$ is the enhanced item representation incorporating both pairwise and higher-order relations. $\mathbf{W}_3 \in \mathbb{R}^{d \times 1}$ and $\mathbf{b}_1 \in \mathbb{R}$ are learnable weights. To simplify the presentation, we use $\theta(\cdot)$ later to represent the gated fusion operation. In summary, we define the whole process of improved HiGCN as,

$$\mathbf{e}_{v_i}^h = \text{HiGCN}_{\text{imp}}\left(\mathbf{e}_{v_i}^o, \mathbf{D}_{p,v}^{-1/2} \mathbf{H}_p \mathbf{D}_{p,h}^{-1} \mathbf{H}_p \mathbf{D}_{p,v}^{-1/2} \right). \tag{4}$$

2) Intent Disentanglement Learning: Considering the user noise behaviors and the long-term item dependencies within sessions, we propose a hierarchical attention mechanism, which ensures stable and efficient disentangled learning for short-term user intent. Generally, the items clicked later in the session are more important for representing current user intent. Hence, we utilize a positional embedding for each item in each session. Specifically, we define a learnable positional embedding $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n]$, where $\mathbf{p}_i \in \mathbb{R}^d$ is a positional embedding for i-th item in a session. Then, the item embeddings $\mathbf{E}_s^p = \left[\mathbf{e}_{v_1}^p, \mathbf{e}_{v_2}^p, \ldots, \mathbf{e}_{v_n}^p \right] \in \mathbb{R}^{n \times d}$ of a session, which integrate positional information, are defined as $\mathbf{e}_{v_i}^p = \mathbf{e}_{v_i}^h + \mathbf{p}_i$. Next, we adopt a self-attention mechanism based on scaled dot-product attention to learn the item dependencies within sessions. Different from the vanilla self-attention mechanism, we only utilize linear projection to obtain query vector $\mathbf{Q}$ and key vector $\mathbf{K}$, keeping the value vector $\mathbf{V}$ consistent with item embeddings $\mathbf{e}_{v_i}^p$,

$$\begin{bmatrix} \mathbf{Q} \\ \mathbf{K} \\ \mathbf{V} \end{bmatrix} = \mathbf{E}_s^p \begin{bmatrix} \mathbf{W}_Q \\ \mathbf{W}_K \\ \mathbf{I} \end{bmatrix}, \tag{5}$$

where $\mathbf{W}_Q, \mathbf{W}_K \in \mathbb{R}^{d \times d}$ are learnable weights, $\mathbf{I}$ is an identity matrix. We further introduce an auxiliary vector $\mathbf{A}$ for performing hierarchical double-attention, which is extracted from the query vector $\mathbf{Q}$ to reduce the interference of user noise behaviors and long-term dependencies,

$$\mathbf{A} = \text{leakyReLU}\left(\text{Conv}\left(\mathbf{Q} \right) \right), \tag{6}$$

$$\begin{aligned} \text{HAtt}(\mathbf{Q}, \mathbf{A}, \mathbf{K}, \mathbf{V}) &= \text{Att}\left(\mathbf{Q}, \mathbf{A}, \text{Att}\left(\mathbf{A}, \mathbf{K}, \mathbf{V} \right) \right) \\ &= \phi\left(\omega\left(\frac{\mathbf{Q}\mathbf{A}^\top}{\sqrt{d}} \right) \right) \phi\left(\omega\left(\frac{\mathbf{A}\mathbf{K}^\top}{\sqrt{d}} \right) \right) \mathbf{V}, \end{aligned} \tag{7}$$

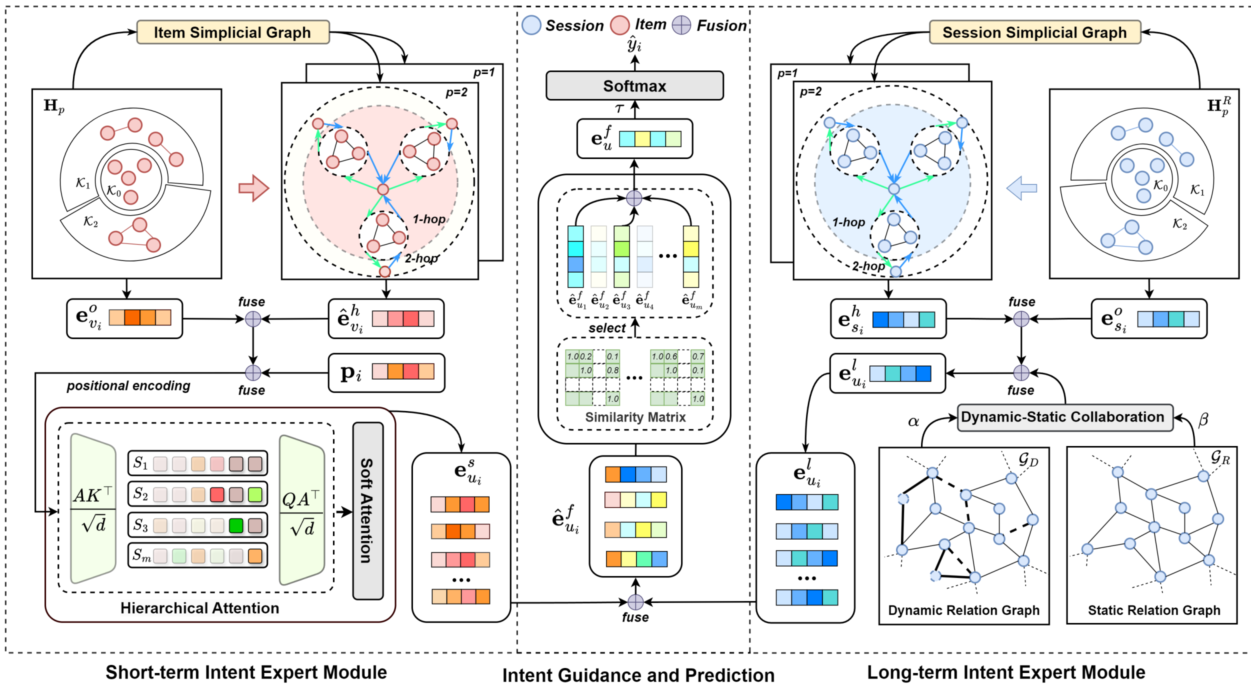


Fig. 3: Overview of H^2-GCN. Our model mainly consists of three components: (1) Short-term Intent Expert Module (**Left**), item embeddings $\mathbf{e}^o_{v_i}$ are used as the input of the simplicial graph to derive $\mathbf{e}^h_{v_i}$. In disentangling short-term intent, the light-coloured rounded rectangles represent masked items in attention calculation. (2) Long-term Intent Expert Module (**Right**), session embeddings $\mathbf{e}^o_{s_i}$ are used as the input of the session-level simplicial graph to obtain $\mathbf{e}^h_{s_i}$. In the dynamic graph $\mathcal{G}_D$, the dotted edges denote potential redundant connections, and the solid edges and dotted nodes represent new potential relations that do not occur in the static graph $\mathcal{G}_R$. (3) Intent Guidance and Prediction (**Middle**), we use similarity δ_i between intents to guide and enhance the intent representations.

where $\mathrm{Conv}(\cdot)$ denotes one-dimensional convolution with kernel size 3×1, $\phi(\cdot)$ denotes the softmax function, and $\omega(\cdot)$ denotes a tail-fixed window function that directs the allocation of attention from a whole session to the tail of a session for blocking the long-term dependencies. Here, we remove the position-wise feed-forward network from the vanilla self-attention mechanism and instead use a residual connection to minimize the random noise introduced by the linear transformation. Thus, we define the learned item embeddings as $\mathbf{E}^a_s = \mathrm{HAtt}\left(\mathbf{Q}, \mathbf{A}, \mathbf{K}, \mathbf{V}\right) + \mathbf{V}$, where $\mathbf{E}^a_s = [\mathbf{e}^a_{v_1}, \mathbf{e}^a_{v_2}, \ldots, \mathbf{e}^a_{v_n}] \in \mathbb{R}^{n \times d}$. Next, we fuse the item embeddings learned from $\mathrm{HAtt}(\cdot)$ with a soft attention mechanism,

$$\begin{aligned} \mu_i &= \left(\mathbf{e}^a_{v_n}\mathbf{W}_4 + \mathbf{e}^a_{v_i}\mathbf{W}_5 + \mathbf{b}_2\right)\mathbf{W}_6{}^\top, \\ \mathbf{e}^s_u &= \sum_{i=1}^{n} \mu_i \cdot \mathbf{e}^a_{v_i}, \end{aligned} \tag{8}$$

where $\mathbf{e}^s_u \in \mathbb{R}^d$ is the learned short-term user intent, $\mathbf{W}_4, \mathbf{W}_5 \in \mathbb{R}^{d \times d}$ and $\mathbf{W}_6, \mathbf{b}_2 \in \mathbb{R}^d$ are learnable weights.

B. Long-term Intent Expert Module

The long-term intent expert module aims to capture essential relations between sessions from multi-scale perspectives, while avoiding redundant relations, to accurately model long-term user intent.

1) Higher-order Session Representation Learning: Only considering item dependencies within sessions is insufficient for modeling user intent, as it ignores the complex yet important interactions between sessions. Therefore, we utilize improved HiGCN to capture higher-order interaction relations on the session-level simplicial graph $\mathbf{H}^R_p$,

$$\mathbf{e}^h_{s_i} = \mathrm{HiGCN}_{\mathrm{imp}}\left(\mathbf{e}^o_{s_i}, \mathbf{D}^{-1/2}_{p,v}\mathbf{H}^R_p\mathbf{D}^{-1}_{p,h}\mathbf{H}^R_p\mathbf{D}^{-1/2}_{p,v}\right), \tag{9}$$

where $\mathbf{e}^h_{s_i} \in \mathbb{R}^d$ is the learned higher-order session representation, and $\mathbf{e}^o_{s_i}$ denotes the initial embedding of each session, which is obtained as $\mathbf{e}^o_{s_i} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}^a_{v_i}$.

2) Dynamic-Static Graph Collaborative Propagation: The construction process of the static global relation graph $\mathcal{G}_R$ may introduce redundant edges, thereby hindering the effectiveness of global relation mining. In addition to $\mathcal{G}_R$, we construct a dynamic global graph $\mathcal{G}_D$ according to the learned short-term user intents to adjust $\mathcal{G}_R$. Its incidence matrix is defined as $\mathbf{A}_D = \psi\left(\mathbf{e}^s_u\right) \in \mathbb{R}^{m \times m}$, where $\psi(\cdot)$ denotes a function that measures the similarity of different short-term intents. Here, we choose cosine similarity as a measure. Based on this, we design a dynamic-static graph collaborative propagation to adjust global dependency learning from $\mathcal{G}_D$ and $\mathcal{G}_R$,

$$\begin{aligned} \mathbf{A}_s &= \alpha \cdot \mathbf{A}_D + \beta \cdot \mathbf{A}_R, \\ \mathbf{e}^{r,(l+1)}_{s_i} &= \mathrm{leakyReLU}\left(\mathbf{D}^{-1/2}\mathbf{A}_s\mathbf{D}^{-1/2}\mathbf{e}^o_{s_i}\mathbf{W}^l\right), \end{aligned} \tag{10}$$

TABLE I: Statistics of datasets.

Dataset	# Train	# Test	# Items	Avg.len
Diginetica	719,470	60,858	43,097	5.12
Tmall	351,268	25,898	40,728	6.69
Nowplaying	825,304	89,824	60,417	7.42

where $\mathbf{e}_{s_i}^{r,(l+1)} \in \mathbb{R}^d$ is the session representation that avoids redundant relations. α, β are learnable coefficients, $\mathbf{A}_R$ is the incidence matrix of G_R, $\mathbf{D}$ is the diagonal matrix and $\mathbf{W}^l \in \mathbb{R}^{d\times d}$ is learnable weights in propagation layer l. Next, we fuse the multi-scale session representations to derive the long-term user intent,

$$\mathbf{e}_u^l = \theta\left(\mathbf{e}_{s_i}^h, \mathbf{e}_{s_i}^r\right), \tag{11}$$

where $\mathbf{e}_u^l \in \mathbb{R}^d$ is the long-term user intent.

C. Intent Guidance and Prediction

Intuitively, users with similar interaction distributions tend to have homogeneous intents [29]. Therefore, we design an intent-guidance mechanism to adaptively activate potential collaboration among different user intents. Specifically, we first obtain the fused user intent as $\widehat{\mathbf{e}}_u^f = \mathbf{e}_u^s + \mathbf{e}_u^l$. For each current intent $\widehat{\mathbf{e}}_{u_i}^f$, we compute its cosine similarity with other fused intents $\widehat{\mathbf{e}}_{u_j}^f$ and use the resulting score δ_j as the guidance weight. We then aggregate the top-N best matches of the current user intent from the user intent set to obtain the final user intent $\mathbf{e}_{u_i}^f$,

$$\mathbf{e}_{u_i}^f = \sum_{j=1}^{N} \delta_j \cdot \widehat{\mathbf{e}}_{u_j}^f. \tag{12}$$

Finally, we compute the predicted probability $\hat{y}_i$ for each candidate item $v_i \in V$ by dot product,

$$\hat{y}_i = \mathrm{Softmax}\left(\left(\mathbf{e}_u^f/\tau\right)^\top \mathbf{e}_{v_i}\right), \tag{13}$$

where $\hat{y}_i \in \mathbb{R}^w$ denotes the probability of the next item that may be clicked, and τ is a scale coefficient that controls the distribution of recommendations. The cross-entropy function is employed to optimize model parameters,

$$\mathcal{L}\left(\hat{y}\right) = -\sum_{i=1}^{w} y_i \log\left(\hat{y}_i\right) + \left(1-y_i\right)\log\left(1-\hat{y}_i\right) + \lambda_1 \left\|\Theta\right\|_2^2, \tag{14}$$

where w denotes all items that appeared in sessions, y_i denotes the label of each session, λ_1 controls the weight decay, and Θ denotes all learnable parameters in the model.

V. Experiment

A. Experimental Settings

Datasets. We conduct comprehensive experiments on three real-world datasets: *Diginetica*[2], *Tmall*[3], *Nowplaying*[4]. Table I shows the statistics for all datasets. Diginetica comes from the CIKM Cup 2016, which contains interaction records from e-commerce platforms. Tmall comes from the IJCAI-15 competition and contains anonymized user interaction data from the Tmall platform. Nowplaying is derived from an online music platform that records users' listening behavior. We preprocess the datasets following previous work [7]. Specifically, sessions with fewer than two interactions and items occurring fewer than five times are removed. We use the data from the last week as the test set, and the remaining data for training.

Baselines. We compare our proposed H^2-GCN with the following three groups of SBR baselines, including: (1) General methods: **Item-KNN** [30] and **FPMC** [31]. (2) RNN-based methods: **GRU4Rec** [13], **NARM** [2], and **CSRM** [3]. (3) GNN-based methods: **SR-GNN** [7], **FGNN** [4], **GCE-GNN** [5], S^2**-DHCN** [12], **MTD** [10], **HIDE** [23], **MSGAT** [11], and **ECCL** [20].

Parameter Setting. Following [7], we set embedding size $d = 100$ and batch size to 128. Optimization is performed using Adam with a learning rate of 0.001 (decaying by 0.1 every 3 epochs) and weight decay $\lambda_1 = 10^{-5}$. Key parameters are tuned via grid search: $N \in \{1, \ldots, 7\}$, $K \in \{2, 4, \ldots, 10\}$, and window size $L \in \{1, \ldots, 5\}$. Based on [27], we fix simplex order $P = 2$. Other parameters are initialized as $\alpha = 0.5, \beta = 1.0$, with a single-layer graph collaborative propagation. The code is available at https://github.com/lucky6768/H2-GCN.

Evaluation Metric. Following the evaluation metrics used in previous works [7], we evaluate the model using $\mathrm{P@}K$ (Precision) and $\mathrm{MRR@}K$ (Mean Reciprocal Rank).

B. Overall Performance Comparison

The performance of all methods is reported in Table II, where the best results are marked in bold and the second best are underlined. From the results, we can get the following key observations:

GNN-based methods effectively capture contextual dependencies among items through graph structures, enabling more comprehensive contextual modeling and thus better recommendation performance. In contrast, general methods (e.g., Item-KNN and FPMC) lack sufficient sequential modeling capabilities, whereas RNN-based models such as GRU4Rec struggle with long-term dependency modeling due to vanishing gradients. By jointly modeling intra- and cross-session dependencies, GCE-GNN, S^2-DHCN, and MTD learn richer item representations and consistently outperform SR-GNN, which captures only local item relations, across all datasets. MSGAT identifies target items to distinguish truly relevant items from noisy interactions within sessions, while ECCL introduces an automatic search module to reduce the impact of noisy information. Their strong performance demonstrates the need to mitigate user noise behaviors for effective intent modeling.

H^2-GCN consistently outperforms all SBR baselines across all datasets, demonstrating the effectiveness of jointly modeling multi-level relations and disentangled user intents. Specifically, H^2-GCN first captures both pairwise and high-order

[2]http://cikm2016.cs.iupui.edu/cikm-cup/

[3]https://tianchi.aliyun.com/dataset/dataDetail?dataId=42

[4]http://dbis-nowplaying.uibk.ac.at/#nowplaying

TABLE II: Overall performance comparison on three datasets. The best results are marked in bold and the second best are underlined. **P** denotes precision and **M** denotes mean reciprocal rank.

Dataset	Diginetica				Tmall				Nowplaying			
Metric	**P@10**	**P@20**	**M@10**	**M@20**	**P@10**	**P@20**	**M@10**	**M@20**	**P@10**	**P@20**	**M@10**	**M@20**
Item-KNN	25.06	35.52	10.78	11.64	6.65	9.15	3.11	3.31	10.96	15.94	4.55	4.91
FPMC	15.42	22.11	6.24	6.68	13.10	16.06	7.12	7.32	5.28	7.36	2.68	2.82
GRU4Rec	17.85	30.76	7.78	8.32	9.47	10.93	5.78	5.89	6.74	7.92	4.40	4.48
NARM	35.54	48.21	15.21	16.02	19.17	23.30	10.42	10.70	13.60	18.59	6.62	6.93
CSRM	36.59	50.54	15.42	16.33	24.54	29.64	13.62	13.72	13.20	18.14	6.08	6.42
SR-GNN	38.41	50.73	16.88	17.59	23.41	27.57	13.45	13.72	14.17	18.87	7.15	7.47
FGNN	37.70	50.68	15.94	16.82	20.67	25.24	10.07	10.39	13.89	18.78	6.80	7.15
GCE-GNN	41.16	54.22	18.15	19.04	28.01	33.42	15.08	15.42	16.94	22.37	8.03	8.40
S^2**-DHCN**	40.21	53.66	17.59	18.51	26.22	31.42	14.60	15.05	17.35	23.03	7.87	8.18
MTD	40.22	53.92	17.58	18.65	25.98	30.33	14.71	15.13	16.54	20.55	7.61	7.93
HIDE	41.02	53.68	17.78	18.36	31.10	37.12	16.77	17.19	17.32	23.53	7.22	7.51
MSGAT	57.09	66.97	26.30	26.91	37.49	42.74	21.04	21.35	28.39	40.62	17.61	18.10
ECCL	53.33	64.32	24.31	25.07	38.82	44.12	22.51	22.87	27.89	35.36	11.89	12.41
H^2**-GCN**	**57.45**	**67.66**	**28.68**	**29.40**	**43.64**	**47.31**	**27.80**	**28.06**	**39.19**	**46.65**	**17.70**	**18.22**
Improv.	0.63%	1.03%	6.53%	6.29%	12.42%	7.23%	23.50%	22.69%	38.04%	14.84%	0.51%	0.66%

TABLE III: Ablation of components on H^2-GCN.

Dataset	Diginetica		Tmall		Nowplaying	
Metric	**P@20**	**M@20**	**P@20**	**M@20**	**P@20**	**M@20**
w/o HG-P	64.51	27.37	32.73	21.19	17.58	7.94
w/o HG-H	62.97	25.79	44.03	25.82	38.38	15.23
w/o DIL	65.12	26.05	43.98	23.84	40.47	13.84
w/o DSC	66.35	29.35	45.90	27.05	46.45	17.54
H^2**-GCN**	**67.66**	**29.40**	**47.31**	**28.06**	**46.65**	**18.22**

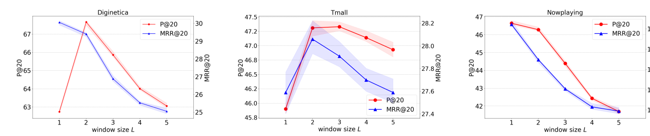


Fig. 4: Impact of size L in tail-fixed window function.

relations at the item and session levels via simplicial graph convolution, leading to more refined item and session representations. It then introduces a short-term intent expert with hierarchical attention and intent disentanglement to suppress long-term dependencies within a session and mitigate user noise, enabling accurate modeling of current user intent. To complement this, a long-term intent expert is designed with a dynamic–static graph collaborative propagation to adaptively model global dependencies while avoiding irrelevant session connections. Finally, intent-guided collaboration adaptively activates interactions among multiple intents, further enhancing recommendation performance.

C. Ablation Study

To analyze the effects of different components in our methods, we conduct ablation studies on the Diginetica, Tmall, and Nowplaying datasets. Specifically, we design four variants of H^2-GCN as follows:

w/o HG-P represents a variant that only employs the GGNN [7] to capture pairwise relations without the simplicial graph.

w/o HG-H removes the simplicial graph and only utilizes the hypergraph to capture high-order relations.

w/o DIL is a variant without an intent disentanglement learning module that is replaced with a vanilla self-attention mechanism.

w/o DSC removes the dynamic-static collaborative propagation and uses only a static global relation graph without adjustment.

The ablation results are shown in Table III. Specifically, (1) **w/o HG-P** and **w/o HG-H** both suffer performance degradation compared to H^2-GCN, indicating that modeling only a single type of relation is suboptimal. Notably, **w/o HG-H** exhibits a larger performance drop than **w/o HG-P**, highlighting the importance of capturing high-order relationships among items or sessions. Moreover, despite employing a hypergraph to model high-order relations, **w/o HG-H** still underperforms H^2-GCN, suggesting that conventional hypergraph-based approaches are limited in that they overlook fine-grained pairwise relations within hyperedges. (2) **w/o DIL** shows a large performance loss compared to H^2-GCN, which further illustrates the superiority of the approach we used to decouple intent learning by completely blocking the interference of user long-term intents within sessions and using user collaborative intents learned from the graph topologies across sessions that are compatible with short-term intents as long-term intents. (3) **w/o DSC** shows varying degrees of performance degradation across datasets, with the largest drop on Tmall, followed by Nowplaying, and only a slight decrease on Diginetica. This trend aligns with the proportion of redundant global connections formed by single co-occurrence items in each dataset, underscoring the need to remove noisy relations from the global graph.

D. Hyper-parameter Analysis

1) Impact of L in tail-fixed window $\omega(\cdot)$*:* The window size L determines the focus range of the hierarchical attention mechanism, playing a vital role in disentangling long- and short-term intent. From the results shown in Figure 4, we

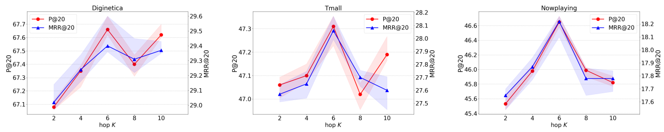


Fig. 5: Impact of hop K in simplicial GNNs.

TABLE IV: Comparison of time costs for training and inference per epoch across different methods.

Dataset	**Diginetica**		**Tmall**		**Nowplaying**	
Methods	Train	Infer	Train	Infer	Train	Infer
SR-GNN	10m22s	27s	4m29s	15s	10m44s	42s
GCE-GNN	9m22s	15s	3m11s	8s	7m46s	22s
MSGAT	20m58s	1m17s	10m22s	44s	23m23s	2m14s
H^2-GCN	5m32s	28s	2m27s	39s	5m54s	58s

find that L is generally shorter, the better, indicating the effectiveness of blocking long-term dependencies within the session. However, a shorter L yields better recommendation ranks for datasets with simpler interactions (e.g., Diginetica) but reduces accuracy due to insufficient short-term patterns within the session (e.g., Nowplaying).

2) Impact of hop K: Different hops in simplicial GNNs yield distinct message-passing effects. The larger the hop, the more information the node aggregates; conversely, the smaller the hop, the less information the node aggregates. The experimental results are shown in Figure 5, where it can be observed that the model performs worst when the hop $K = 2$, in which the model only considers nodes that can be reached within two steps from the current node. With hop $K = 6$, the model achieves the best performance. The recommendation performance decreases as the number of hops increases, indicating that larger random-walk hops may cause feature over-smoothing, thereby affecting performance.

E. Complexity Analysis

To comprehensively evaluate the efficiency of our proposed approach, we analyze the computational efficiency of H^2-GCN and report its training and inference time across different datasets in Table IV. Overall, H^2-GCN exhibits computational efficiency comparable to other GNN-based methods in both training and inference phases, demonstrating that the proposed high-order modeling does not introduce significant additional overhead.

F. Item Embedding Distributions

Figure 6 illustrates the item embedding distributions via Kernel Density Estimation (KDE) on $\mathbb{R}^2$ and the hypersphere $\mathcal{S}^1$. MSGAT produces nearly uniform embedding distributions, whereas SR-GNN concentrates embeddings within limited regions. In contrast, H^2-GCN yields a balanced distribution that smooths extreme peaks while preserving discriminative structure. These observations indicate that uniformly distributed embeddings do not necessarily benefit session-based recommendation, while overly clustered embeddings may amplify the long-tail bias.

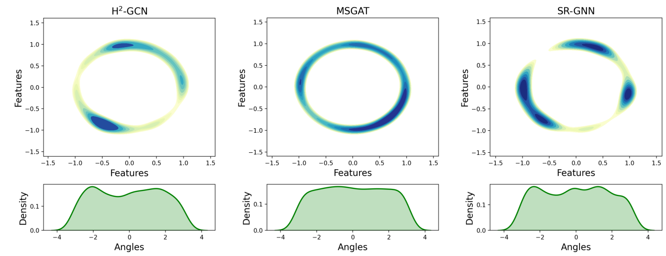
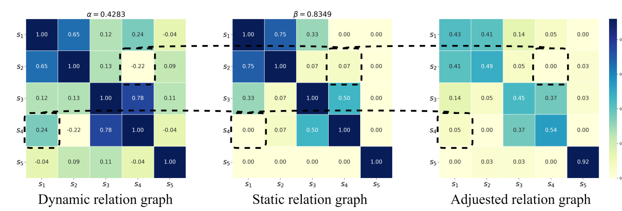
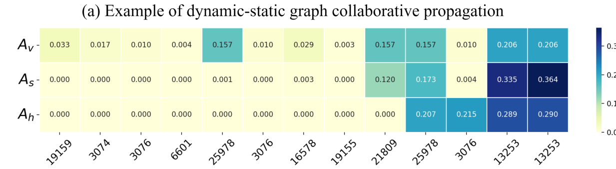


Fig. 6: Distribution of item embeddings learned from the Nowplaying dataset.



(a) Example of dynamic-static graph collaborative propagation



(b) Three different attention mechanism

Fig. 7: Case study of H^2-GCN.

G. Case Study

To examine the effectiveness of the DSC module, we randomly select several sessions $(s_1, \ldots, s_5)$ from the Tmall dataset. As shown in Figure 7(a), the dynamic and static graphs exhibit different session relations, where some weak or irrelevant connections in the static graph are removed (e.g., (s_2, s_4)), while meaningful short-term relations absent in the static graph are introduced (e.g., (s_1, s_4)). After dynamic–static collaborative propagation, noisy relations are suppressed and informative relations are strengthened, leading to more accurate global session dependencies. To demonstrate the effectiveness of the DIL module, we further visualize the attention scores of the same session under three attention mechanisms, as shown in Figure 7(b). The vanilla self-attention (A_v) is heavily affected by long-term dependencies and noisy behaviors, while sparse self-attention (A_s) partially alleviates this issue but fails to fully eliminate interference. In contrast, the proposed hierarchical self-attention (A_h) effectively suppresses irrelevant dependencies, enabling more accurate short-term intent learning.

VI. Conclusion

This paper proposes a novel H^2-GCN for SBR, which achieves effective recommendations by capturing multi-level relationships and refining the diverse user intents. Specifically, we first capture higher-order and fine-grained pairwise relations within simplicial structures at both the item- and session-levels through simplicial GNNs. Moreover, the short-term intent expert module is designed to block long-term dependencies and mitigate noisy user behaviors, thereby enabling accurate modeling of short-term intents. The long-term intent expert module is proposed to filter redundant global relations and derive true long-term intents. Extensive experiments on three real-world datasets validate the effectiveness of H^2-GCN, which outperforms state-of-the-art methods.

References

[1] D. Jannach and M. Ludewig, "When recurrent neural networks meet the neighborhood for session-based recommendation," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, ser. RecSys '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 306–310.

[2] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, "Neural attentive session-based recommendation," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ser. CIKM '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1419–1428.

[3] M. Wang, P. Ren, L. Mei, Z. Chen, J. Ma, and M. de Rijke, "A collaborative session-based recommendation approach with parallel memory modules," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 345–354.

[4] R. Qiu, Z. Huang, J. Li, and H. Yin, "Exploiting cross-session information for session-based recommendation with graph neural networks," *ACM Trans. Inf. Syst.*, vol. 38, no. 3, 2020.

[5] Z. Wang, W. Wei, G. Cong, X.-L. Li, X.-L. Mao, and M. Qiu, "Global context enhanced graph neural networks for session-based recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 169–178.

[6] Z. Wang, W. Wei, D. Zou, Y. Liu, X.-L. Li, X.-L. Mao, and M. Qiu, "Exploring global information for session-based recommendation," *Pattern Recognition*, vol. 145, p. 109911, 2024.

[7] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 346–353.

[8] J. Chen, H. Li, X. Zhang, F. Zhang, S. Wang, K. Wei, and J. Ji, "Sr-hetgnn: session-based recommendation with heterogeneous graph neural network," *Knowledge and Information Systems*, vol. 66, no. 2, pp. 1111–1134, 2024.

[9] Q. Chen, Z. Guo, J. Li, and G. Li, "Knowledge-enhanced multi-view graph neural networks for session-based recommendation," in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, p. 352–361.

[10] C. Huang, J. Chen, L. Xia, Y. Xu, P. Dai, Y. Chen, L. Bo, J. Zhao, and J. X. Huang, "Graph-enhanced multi-task learning of multi-level transition dynamics for session-based recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, 2021, pp. 4123–4130.

[11] S. Qiao, W. Zhou, J. Wen, H. Zhang, and M. Gao, "Bi-channel multiple sparse graph attention networks for session-based recommendation," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, ser. CIKM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 2075–2084.

[12] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, and X. Zhang, "Self-supervised hypergraph convolutional networks for session-based recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, 2021, pp. 4503–4511.

[13] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," *arXiv preprint arXiv:1511.06939*, 2016.

[14] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, "Stamp: Short-term attention/memory priority model for session-based recommendation," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1831–1839.

[15] W.-C. Kang and J. McAuley, "Self-attentive sequential recommendation," in *2018 IEEE International Conference on Data Mining (ICDM)*, 2018, pp. 197–206.

[16] A. Luo, P. Zhao, Y. Liu, F. Zhuang, D. Wang, J. Xu, J. Fang, and V. S. Sheng, "Collaborative self-attention network for session-based recommendation." in *IJCAI*, 2020, pp. 2591–2597.

[17] J. Yuan, Z. Song, M. Sun, X. Wang, and W. X. Zhao, "Dual sparse attention network for session-based recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, 2021, pp. 4635–4643.

[18] X. Zhang, H. Lin, B. Xu, C. Li, Y. Lin, H. Liu, and F. Ma, "Dynamic intent-aware iterative denoising network for session-based recommendation," *Information Processing & Management*, vol. 59, no. 3, p. 102936, 2022.

[19] Z. Pan, F. Cai, W. Chen, H. Chen, and M. de Rijke, "Star graph neural networks for session-based recommendation," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, ser. CIKM '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1195–1204.

[20] G. An, J. Sun, Y. Yang, and F. Sun, "Enhancing collaborative information with contrastive learning for session-based recommendation," *Information Processing & Management*, vol. 61, no. 4, p. 103738, 2024.

[21] A. Li, Z. Cheng, F. Liu, Z. Gao, W. Guan, and Y. Peng, "Disentangled graph neural networks for session-based recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 7870–7882, 2023.

[22] R. Gao, Y. Tao, Y. Yu, J. Wu, X. Shao, J. Li, and Z. Ye, "Self-supervised dual hypergraph learning with intent disentanglement for session-based recommendation," *Knowledge-Based Systems*, vol. 270, p. 110528, 2023.

[23] Y. Li, C. Gao, H. Luo, D. Jin, and Y. Li, "Enhancing hypergraph neural networks with intent disentanglement for session-based recommendation," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 1997–2002.

[24] J. Yu, H. Yin, J. Li, Q. Wang, N. Q. V. Hung, and X. Zhang, "Self-supervised multi-channel hypergraph convolutional network for social recommendation," in *Proceedings of the Web Conference 2021*, ser. WWW '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 413–424.

[25] X. Zhang, B. Xu, L. Yang, C. Li, F. Ma, H. Liu, and H. Lin, "Price does matter! modeling price and interest preferences in session-based recommendation," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 1684–1693.

[26] M. Yang, E. Isufi, and G. Leus, "Simplicial convolutional neural networks," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 8847–8851.

[27] Y. Huang, Y. Zeng, Q. Wu, and L. Lü, "Higher-order graph convolutional network with flower-petals laplacians on simplicial complexes," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 11, 2024, pp. 12 653–12 661.

[28] Q. Hao, C. Wang, Y. Xiao, and H. Lin, "Simplices-based higher-order enhancement graph neural network for multi-behavior recommendation," *Information Processing & Management*, vol. 61, no. 5, p. 103790, 2024.

[29] J. Sun, Y. Zhang, C. Ma, M. Coates, H. Guo, R. Tang, and X. He, "Multi-graph convolution collaborative filtering," in *2019 IEEE International Conference on Data Mining (ICDM)*, 2019, pp. 1306–1311.

[30] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International Conference on World Wide Web*, ser. WWW '01. New York, NY, USA: Association for Computing Machinery, 2001, p. 285–295.

[31] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, "Factorizing personalized markov chains for next-basket recommendation," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 811–820.