Hypergraph-Based Session Modeling: A Multi-Collaborative Self-Supervised Approach for Enhanced Recommender Systems

Xiangping Zheng[1,3,5], Bo Wu[2*], Alex X. Zhang[4], Wei Li[1,3,5]
1.College of Computer Science and Technology, Harbin Engineering University
2.R&D and External Relations Department, Xiangjiang Laboratory
3.Modeling and Emulation in E-Government National Engineering Laboratory
4.Harvest Fund Management Co.,Ltd.
5.Qingdao Innovation and Development Base of Harbin Engineering University
{xpzheng,wei.li}@hrbeu.edu.cn, bowuchn@163.com, zhangxuan01@jsfund.cn

Abstract

Session-based recommendation (SBR) is a challenging task that involves predicting a user's next item click based on their recent session history. Presently, many state-of-the-art methodologies employ graph neural networks to model item transitions. Notwithstanding their impressive performance, graph-based models encounter significant challenges when confronted with intricate session dependencies and data sparsity in real-world scenarios, ultimately constraining their capacity to enhance recommendation accuracy. In recognition of these challenges, we introduce an innovative methodology known as 'Mssen,' which stands for Multi-collaborative self-supervised learning in hypergraph neural networks. Mssen is meticulously crafted to adeptly discern user intent. Our approach initiates by representing session-based data as a hypergraph, adeptly capturing intricate, high-order relationships. Subsequently, we employ self-supervised learning on item-session hypergraphs to mitigate the challenges of data sparsity, all without necessitating manual fine-tuning, extensive search, or domain-specific expertise in augmentation selection. Comprehensive experimental analyses conducted across multiple datasets consistently underscore the superior performance of our approach when compared to existing methodologies.

Keywords: Graph neural networks, High-order Relations, Hypergraph

1. Introduction

In recent times, consumers have increasingly favored online product selection over in-person retail experiences. However, this shift brings with it the challenge of information overload (Lyu et al., 2021). Many contemporary transactional websites use anonymous user identities and involve brief purchase sessions, underscoring the importance of modeling user behavior within single sessions for effective recommendations (Krishnan et al., 2022; Zheng et al., 2023c). Traditional recommendation methods, like collaborative filtering (Su et al., 2021), heavily reliant on extensive user data and long-term interactions, often struggle to provide accurate recommendations in such scenarios (Wu et al., 2019; Zheng et al., 2023a).

Session-based recommendation (SBR) methods (Wang et al., 2021; Zheng et al., 2023b), renowned for their high practical relevance, are specifically engineered to delineate user intent by analyzing the behavioral sequences of users within sessions. More recently, the advent of graph neural networks (GNNs) (Zheng et al., 2022) has generated considerable interest, attributed to their remarkable effectiveness in diverse domains, including SBR. GNN-based approaches,(Wu et al., 2019; Qiu et al., 2019), map each session to a subgraph respectively, and then use the subgraph as the input of GNN to further capture the dependencies of nodes in the subgraph to provide suggestions for the next project. They model item transitions as pairwise relations, offering a more adaptable approach to accommodating temporal dependencies among items. Although these GNN-based models have exhibited promising performance in SBR, two pivotal issues warrant further in-depth investigation:

First challenge i) Modeling High-Order Item Relations: In GNN-based approaches, item data is commonly interconnected via pairwise relationships. Nonetheless, real-world transactions frequently feature complex item structures characterized by high-order interconnections (Xia et al., 2021). While GNN-based models can propagate long-term relational dependencies, considered to be of higher order, across k-hop neighbors using layers corresponding to each hop, they fall short in formulating and capturing intricate high-order user relationship patterns extending beyond mere pairwise links (Yu et al., 2021). For instance, a simplistic pairwise relationship between 'strawberry' and 'apple' does not suffice to infer a user's intention to purchase an assortment of fruits. Consequently, the development of a more generalized architecture that proficiently learns representations of items in higher-order relationships becomes vital.

Second challenge ii) Mitigating Item Data

*Corresponding author

LREC-COLING 2024, pages 8493–8504
20-25 May, 2024.

Sparsity: Data sparsity in item interactions is a widespread problem in practical recommendation scenarios, primarily due to limited user interactions with a large number of items (Lin et al.). Many GNN-based models struggle with sparse data, especially in short user sessions. While some self-supervised learning (SSL) methods (HaoChen et al., 2021) have been used to tackle data sparsity, they often involve augmenting the data by adding or removing edges and nodes. Unfortunately, such methods, though effective in some contexts, are not well-suited for session-based recommendation (SBR) as they can disrupt the relationships between items, making the dataset even sparser (Sun et al., 2019; Wu et al., 2021). Additionally, these augmentation-based SBR methods heavily depend on the choice of augmentation strategies and hyperparameters, limiting their effectiveness. Therefore, it is crucial to explore innovative SSL approaches to improve item representations and address data sparsity in session-based recommendation.

Motivated by the aforementioned discussions, we present a novel approach named Multi-collaborative self-supervised learning in hypergraph neural networks, or **Mssen**, designed to explore user intent. To tackle the first challenge, which revolves around capturing high-order relations among sessions, we introduce an innovative hypergraph modeling technique. Specifically, our methodology employs hypergraph representation, enabling more effective inference of user intent during their current session. In this hypergraph model, each session is depicted as a hyperedge, interconnecting all items within that session. The emphasis is on capturing the inherent coherence among these items, without being constrained by strict ordering requirements. Addressing the second challenge related to data sparsity, we delve into the realm of self-supervised learning (SSL) within the context of item-session hypergraphs. We propose two types of noise augmentation strategies: multiplicative noise and additive noise. These strategies involve introducing noise directly into the representation, which proves to be effective for generating diverse views of the data. By jointly optimizing these tasks, we observe substantial improvements in recommendation performance.

In summary, our contributions can be succinctly described as follows: **1)** We introduce Multi-collaborative self-supervised learning in hypergraph neural networks, denoted as **Mssen**, for session-based recommendation. Mssen excels in capturing intricate, high-order data correlations within its structure, thereby enhancing the inference of user intent in current sessions. **2)** We propose two types of noise augmentation strategies: multiplicative noise and additive noise. These strategies involve introducing noise directly into the representation, which proves to be effective for generating diverse views of the data. **3)** Our approach is validated through extensive experiments on multiple real-world datasets. The results demonstrate not only competitive performance but also the absence of a requirement for manual data augmentation.

2. Related Work

Session-based Recommendation. Session-based recommendation (SBR) systems, which aim to predict a user's next action based on their previous activity sequence, are garnering escalating scholarly interest (Li et al., 2022). Initial SBR methodologies were inspired by the observation that users with similar behaviors are likely to exhibit similar purchasing patterns, leading to the adoption of nearest neighbor-based algorithms (Dias and Fonseca, 2013; Sarwar et al., 2001). A canonical example of this approach is the item-based neighborhood recommendation method (Sarwar et al., 2001) , which gauges item similarities through co-occurrence within sessions. Advancements in deep learning have catalyzed significant enhancements in SBR (Li et al., 2017; Hidasi et al., 2016; Liu et al., 2018). Pioneering the utilization of Recurrent Neural Networks (RNN) in SBR, Hidasi et al. (Hidasi et al., 2016)presented GRU4Rec, which leverages Gated Recurrent Units (GRUs) alongside session-parallel mini-batches and pairwise ranking loss, yielding promising improvements over traditional models. Despite RNNs achieving notable success, their ability to encapsulate collective item dependencies remains limited. Conversely, Graph Neural Networks (GNNs) utilize graph structures to elegantly represent item transition dynamics. Wang et al. (Wang et al., 2020)contributed the GCE-GNN model, which intricately captures user preferences by leveraging transition information across sessional and global graph levels, thus facilitating a refined inference of current user interests . While these advancements have propelled the field forward, they exhibit restricted capacity in grasping the intricate higher-order relationships between items. Our model sets itself apart by strategically employing hypergraphs to intricately map the higher-order dependencies present in sessions and items.

Hypergraph Learning. Although the graph neural network approaches have achieved successful results in capturing high-order relations in various tasks (You et al., 2020; Zhu et al., 2021), these approaches are only appropriate in pairwise connections, which has limitation in expressing complex structures of data (Zhang et al., 2021). Recently, constructing hypergraphs to learn the data structure become a popular approach. A hypergraph is a generalization of a simple graph in which a hyperedge can connect more than two nodes.

Hayashi et al. (Hayashi et al., 2020) propose a flexible framework for clustering hypergraph-structured data based on recently proposed random walks utilizing edge-dependent vertex weights. Xue et al. (Xue et al., 2021) develop an unsupervised Dual-HGCN that transforms the multiplex bipartite network into two sets of homogeneous hypergraphs, along with intra- and inter-message passing strategies to promote information exchange within and across domains. To revisit user mobility and social relationships based on a large-scale LBSN dataset collected over a long-term period, Dingqi (Yang et al., 2019) proposes LBSN2Vec, a hypergraph embedding approach designed specifically for LBSN data for automatic feature learning. In contrast to these prior works, our approach focuses on enhancing supervised signals by facilitating contrastive learning specifically between hyperedges and their corresponding global graph representation. This targeted approach significantly improves the efficiency of supervised signal processing.

Self-supervised Learning. Self-supervised learning's success is figuring out a way to leverage the tremendous amounts of unlabeled data that becomes available to dig out the representation of general data. Existing graph contrastive learning (Fang et al., 2021; Zhu et al., 2021) is a class of self-supervised approaches, which train an encoder to measure the divergence in latent space by contrasting samples from a distribution that contains depict statistical dependencies of interest and those that do not (Hassani and Ahmadi, 2020). The main idea of graph contrastive learning is to treat each sample as a distinct category and learn how to distinguish them (Fang et al., 2021). For instance, Yuning (You et al., 2021) proposes a unified bilevel optimization framework to automatically, adaptively and dynamically select data augmentations when performing GraphCL on specific graph data. The general framework, dubbed Joint Augmentation Optimization (JOAO), is instantiated as min-max optimization. Sheng Wan (Wan et al., 2021) proposes a novel GCN-based SSL algorithm is presented in this paper to enrich the supervision signals by utilizing both data similarities and graph structure. Their main goal is to improve graph representations through diverse graph augmentation strategies. In contrast, our work stands out by focusing on learning node representations without the need for graph augmentations.

3. Methodology

As shown in **Figure 1**, our Mssen consists of two critical tasks: one is the main task for the recommendation, and the other is SSL acted as the auxiliary task to boost the former.

3.1. Hypergraph Network for SBR Task

Hypergraph Construction. In our exploration of Session-based Recommendation (SBR) systems, we embrace an advanced hypergraph structure, symbolized as $G_h = (V, E, \mathbf{W})$, where we map sessions to expansive hyperedges. Each hyperedge in this structure has the capability to interlink an extensive array of vertices. Delving into the technical definition, we consider every individual item as $i_{s,m} \in V$ and construct the session's hyperedge as $\epsilon = [i_{s,1}, i_{s,2}, \cdots, i_{s,k}, \cdots, i_{s,m}] \in E$. The use of hypergraphs transcends the conventional graph model by providing a highly adaptable framework. To illustrate, traditional graph models impose a condition wherein a sequential link between items $i_{s,k}$ and $i_{s,k+1}$ exists exclusively if the user engages with $i_{s,k}$ followed by $i_{s,k+1}$. On the contrary, our hypergraph approach enables a broader connection where any pair of items within the same session are interwoven into the hyperedges. Moreover, while graphs typically encounter difficulties in representing the multifaceted semantic relationships of items across diverse sessions, our hypergraph paradigm adeptly captures these variable semantic linkages.

Hypergraph Convolutional Network. After hypergraph construction, we further develop a hypergraph neural network (HGNN) to capture the item-level high-order relations. We concatenate the hyperedge groups to generate the hypergraph incidence matrix **H**. Referring to the spectral hypergraph convolution proposed in (Feng et al., 2019), we can build an HGNN in the following formulation:

$$\mathbf{X}^{(l+1)} = \mathbf{Q}\mathbf{X}^{(l)}\Theta^{(l)}, \mathbf{Q} = \hat{\mathbf{D}}^{-1}\mathbf{H}\mathbf{W}\mathbf{B}^{-1}\mathbf{H}^{T} \quad (1)$$

where $\mathbf{X}_h^{(l)}$ represents the l-th layer's item embeddings. $\Theta^{(l)}$ is the learnable filter matrix. Denote that $\hat{\mathbf{D}}^{-1}$ and $\mathbf{B}^{-1}$ play a role of normalization. Here, hypergraph convolution can be conceptualized as a two-stage process, involving a "nodes-hyperedges-nodes" feature transformation, which effectively refines features based on the hypergraph structure. Specifically, item features are initially aggregated according to the hyperedges, resulting in hyperedge features obtained by multiplying the transpose of the matrix $\mathbf{H}^T$ (**Stage 1: nodes to hyperedges**). Subsequently, the final node features are derived by aggregating their respective related hyperedge features, achieved through the multiplication of the matrix **H** (**Stage 2: hyperedges to nodes**). To enhance the expressive power of Hypergraph Convolutional Neural Networks (HGCN) and mitigate potential over-smoothing, we introduce flexibility in the fixed coefficients, a crucial strategy to prevent over-smoothing. We employ an initial residual technique, proven effective in mitigating over-smoothing in encoders (Chen et al., 2020). Consequently, the l-th layer of HGCN can be formally denoted as:

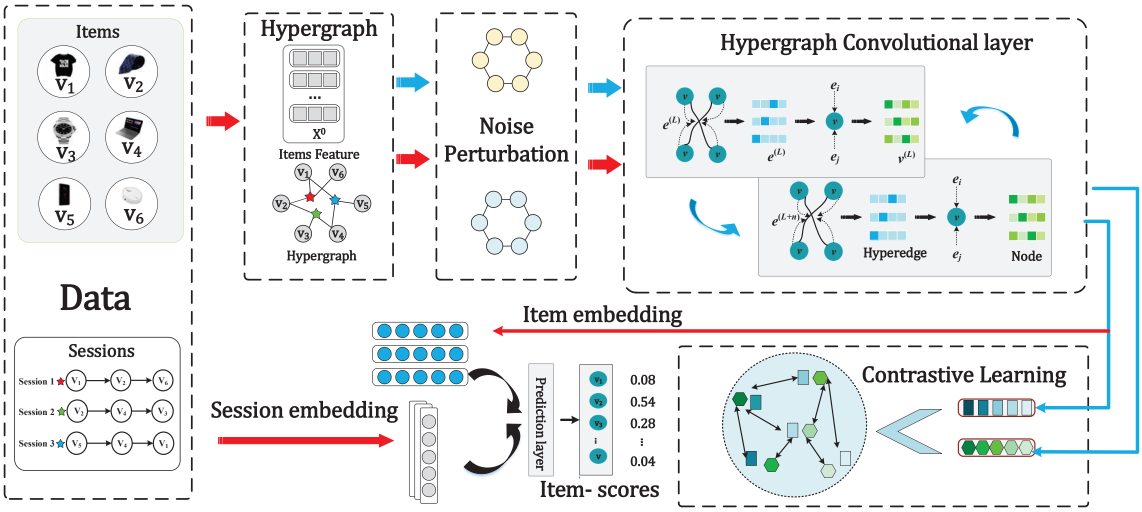


Figure 1: Overview of Mssen. All red solid arrows refer to the recommendation task, while the blue solid arrows denote the auxiliary task (self-supervised learning). By jointly optimizing the two tasks, the performance of the task achieves decent gains.

$$\mathbf{X}^{(l+1)} = ((1-\gamma_l)\mathbf{Q}\mathbf{X}^{(l)} + \gamma_l \mathbf{X}^{(0)})\Theta^{(l)} \qquad (2)$$

where γ_l is a hyper-parameter. $\mathbf{X}^{(0)} \in \mathbb{R}^{|\mathcal{V}|\times d}$ is the initial feature. We add the initial feature to each layer to compensate for the hyper nodes' heterogeneity. Hyper-parameter γ_l indicates how much each layer's initial feature information can carry. Although we stack many layers, it can receive at least γ_l proportion message from the input layer, which ensures that the performance of at least one layer of the model.

Recommendation Generation.

Embracing insights derived from the methodology in SR-GNN (Wu et al., 2019), we enhance the embedding process for a given session s. Acknowledging the variable significance of the embedded information, we integrate a soft-attention mechanism designed to more accurately encapsulate the representational quality of items within a session. Upon refining the embeddings for each session, we then turn our attention to scoring potential items. For each candidate item within the item set , we calculate a score $\hat{\mathbf{z}}_i$ through the execution of an inner product operation with the item's embedding vector $\mathbf{X}$. Subsequently, we leverage the softmax function to ascertain the likelihood of each item's candidacy as the succeeding session item $\hat{\mathbf{y}} = softmax(\hat{\mathbf{z}})$: . We employ a cross-entropy-based loss function for every session graph to quantify the disparity between our model's predictions and the actual sequence outcomes.

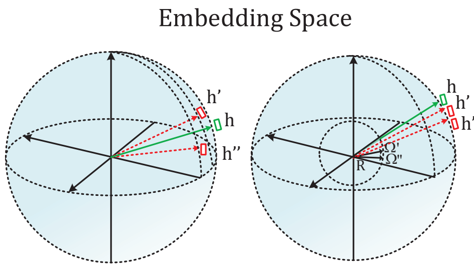
$$\mathcal{L}_t = -\sum_{i=1}^{N} \mathbf{y}_i log(\hat{\mathbf{y}}_i) + (1-\mathbf{y}_i) log(1-\hat{\mathbf{y}}_i) \qquad (3)$$

where $\mathbf{y}$ is the one-hot encoding vector of the ground truth.

3.2. Enhancing SBR with Self-Supervised Task

In comparison to other recommendation paradigms, session-based recommendation is particularly vulnerable to the challenge of data sparsity, primarily due to the limited short-term interactions. Additionally, while hypergraph modeling has demonstrated substantial improvements in performance, we hypothesize that the inherent issue of data sparsity might impede the full potential of hypergraph modeling, ultimately leading to suboptimal recommendation results. Drawing inspiration from the successful applications of self-supervised learning in graph-related tasks, which have shown promise in addressing data sparsity concerns, we introduce an innovative integration of self-supervised learning into the network to enhance the performance of session-based recommendation.

Noise Perturbation. Data augmentation has proven to be highly effective in the domain of image data, involving techniques like random cropping and rotation invariance. However, it's important to acknowledge that augmentations are not universally applicable across all scenarios, especially when dealing with graphs, as the structural information and semantics of graphs can vary significantly from one domain to another (Xia et al., 2022). For instance, randomly dropping edges in social graphs can lead to substantial semantic changes, particularly when these edges are associated with central nodes (Lee et al., 2021). Additionally, designing appropriate augmentation strategies for graphs can be a challenging and time-consuming task. Given these considerations, we've opted to forgo graph augmentations and instead focus on the feature embedding space. Taking inspiration from adversarial attacks (Goodfellow et al., 2015), which leverage noise distortion to enable the network to learn rich representations from various contexts, we propose two types of noise augmentation strategies: multiplicative noise and additive noise. These strategies involve introducing noise directly into the representation, which proves to be effective for generating diverse views of the data.



(a) Multiplicative Noise (b) Additive Noise

Figure 2: An illustration of two random noise-based data augmentation strategies: one is the Multiplicative Noise and the other is the Additive Noise.

- **Multiplicative Noise.** Formally, given a node i and its representation h in the d-dimensional embedding space, we can implement Multiplicative Noise representation-level augmentation, which as exhibited in **Figure 2 (a)**. For the attribute-level noise distortion, we first sample a random noise matrix $N \in \mathbb{R}^{N\times D}$ from a Gaussian distribution $\mathcal{N}(1, 0.1)$. Then the resulting corrupted noise augmentations $\{\mathbf{H}', \mathbf{H}''\} \in \mathbb{R}^{N\times D}$can be as:

$$\mathbf{H}' = \mathbf{H} \odot \mathbf{N}, \mathbf{H}'' = \mathbf{H} \odot \mathbf{N} \tag{4}$$

  where $\odot$ denotes the Hadamard product (Kim et al., 2017). Note that, for each node representation, the multiplied random noises are different.

- **Additive Noise.** Likewise, as shown in **Figure 2 (b)**, we add noise vectors $\mathbf{\Omega}'$ and $\mathbf{\Omega}''$ that are subject to $\|\mathbf{\Omega}\|_2 = R$, $\mathbf{\Omega} \in \mathbb{R}^{N\times d}$. We can implement the following representation-level additive noise augmentation:

$$\mathbf{H}' = \mathbf{H} + \mathbf{\Omega}' \odot \mathbf{sign}(\mathbf{H}), \mathbf{H}'' = \mathbf{H} + \mathbf{\Omega}'' \odot \mathbf{sign}(\mathbf{H}) \tag{5}$$

  There are two key constraint controls in our approach. The first control is related to the magnitude of the matrix $\mathbf{\Omega}$. This constraint essentially limits the values within $\mathbf{\Omega}$ to a hypersphere with a relatively small radius denoted as R. The second constraint ensures that both $\mathbf{H}$ and $\mathbf{\Omega}$ fall within the same hyperoctant. This restriction is in place to prevent the addition of noises from causing significant deviations in $\mathbf{H}$, which might lead to less valid positive samples. As we're dealing with relatively small levels of noise, the augmented representation manages to retain the majority of the information from the original representation while introducing some variance, which can be beneficial for our approach (Yu et al., 2022).

Then, a non-linear transformation $g(\cdot)$ named projection head maps the representations to another latent space and can enhance the performance. we obtain $\mathbf{z}'$ and $\mathbf{z}''$:

$$\mathbf{z}' = g(\mathbf{H}'), \mathbf{z}'' = g(\mathbf{H}'') \tag{6}$$

To enforce maximizing the consistency between positive pairs $\{\mathbf{z}', \mathbf{z}''\}$ compared with negative pairs, we adopt the noise-contrastive estimation loss (van den Oord et al., 2018). For each mini-batch including n sessions in training, if two-session embeddings both denote the same session in two views, we label this pair as the ground truth $\{\mathbf{z}', \mathbf{z}''\}$. Otherwise, we label it as the negative samples $\mathbf{z}_k$. Then we employ a similarity metric function $sim(\cdot, \cdot)$ to calculate the similarity of positive pair $\{\mathbf{z}', \mathbf{z}''\}$ and the negative pair $\{\mathbf{z}', \mathbf{z}_k\}$. Based on this, the loss function is as follows:

$$\mathcal{L}_{NCE} = - log\frac{exp(sim(\mathbf{z}', \mathbf{z}'')/\tau)}{exp(sim(\mathbf{z}', \mathbf{z}'')/\tau) + \sum_{i=1}^{k} exp(sim(\mathbf{z}', \mathbf{z}_k)/\tau)} \tag{7}$$

where τ denotes the temperature parameter. To simplify the calculation, we use dot product as the similarity metric function $sim(\cdot, \cdot)$. By minimizing $\mathcal{L}_{NCE}$ with Adam (Kim, 2014), we can get high-quality session-based recommendation.

Creating another self-supervised signal. In the pursuit of reinforcing the supervised learning signals, we use a dual hypergraph Infomax (DHI) following the mechanism Hyperedge-to-Node (H2N) (Zheng et al., 2023c), and drawing inspiration from the principles outlined in deep graph infomax (DGI) (Velickovic et al., 2019). To elaborate, within the designated hypergraph layer, denoted as L, we denote the features of hyperedges with e^L. It bears mention that hypergraph convolution unfolds across two pivotal phases involving a *nodes–hyperedges–nodes* transformation of features. Initially, we procure the hyperedge-specific features, followed by garnering those pertaining to hypernodes. The inherent mechanics of hypergraph encoding serve to amalgamate insights from neighbors exhibiting structural semblance, thereby offering an efficient methodology to pinpoint neighboring sessions or items. To discretely address the representations of hyperedges (or hypernodes) vis-à-vis their entire hypergraph counterparts, we deduce the terminal hyperedge feature representation, s_e . Pursuant the methodology delineated in DGI, we start by calculating the mean of all hyperedge feature representations. Post-mean calculation, we invoke a sigmoid function to transform the amalgamated representation, thereby rendering the final form as follows:

$$\mathbf{s_e} = \sigma\left(\frac{1}{n_e}\sum_{j=1}^{n_e}\mathbf{e}_j^{(L+1)}\right) \tag{8}$$

The loss of hyperedge-level can be defined:

$$\mathcal{L}_{he} = -\frac{1}{2n}\sum_{i=1}^{n}\left(\mathbb{E}_{\mathcal{G}} log\mathcal{D}\left(\mathbf{e}_i^{(L)}, \mathbf{s_e}\right) + \mathbb{E}_{\tilde{\mathcal{G}}} log\left(1-\mathcal{D}\left(\tilde{\mathbf{e}}_i^{(L)}, \mathbf{s_e}\right)\right)\right) \tag{9}$$

The core of our framework includes a discrimination component, $\mathcal{D}$, tasked with calculating the compatibility scores between paired representations, specifically hyperedge to hypergraph associations. We stimulate the hypergraph $\tilde{\mathcal{G}}$ into existence by performing a row-centric permutation on the initial feature matrix e_i^L. This creative manipulation yields a novel node representation, symbolized as $\tilde{\mathrm{e}}_i^{(L)}$, which, when juxtaposed with the overarching hypergraph representation s, serves as a negative exemplar. Similar principles apply to deducing hypernode-level discrepancies, denoted $\mathcal{L}_{hn}$. onsequently, we get multi-scale self-supervised signals:

$$\mathcal{L}_{SSL} = \mathcal{L}_{NCE} + \mathcal{L}_{DHI} = \mathcal{L}_{NCE} + \mathcal{L}_{he} + \mathcal{L}_{hn} \tag{10}$$

3.3. Model Optimization.

Finally, we unify the recommendation task and this self-supervised task into a primary&auxiliary learning framework, where the former is the primary task and the latter is the auxiliary task. Formally, the joint learning objective:

$$\mathcal{L}_{LOSS} = \mathcal{L}_t + \alpha\mathcal{L}_{SSL} \tag{11}$$

where α is a learnable weight to control the magnitude of the self-supervised task. It should be noted that, we jointly optimize the two throughout the training.

4. Experiments

4.1. Experimental Settings

Datasets. For the verification of our methodology, we employed five authentic benchmark datasets (please refer to **Table 1** for more details). These datasets include *Tmall*[1], *Nowplaying*[2], *Diginetica*[3],

Statistics	Tmall	Nowplaying	Diginetica
# Sessions (Training)	351,268	825,304	719,470
# Sessions (Testing)	25,898	89,824	60,858
# Items	40,728	60,417	43,097
Avg. Length of Sessions	6.69	7.42	5.12

Table 1: Statistics of the datasets used in our experiments.

all of which are frequently exploited for testing session-based recommendation approaches. A brief overview of each dataset: *Tmall* consists of obscured shopping details from users on the Tmall online shopping platform and was formed for the IJCAI-15 competition. In contrast, the *Nowplaying* dataset, derived from (Zangerle et al., 2014), presents an insight into the music preferences of users. Lastly, the *Diginetica* dataset harbors online retail transactions and was originally released in line with the 2016 CIKM Cup.

Baseline Methods. The following models, including the state-of-art and closely related works, are used as representative baselines to evaluate the performance of the proposed model.

- S^2**-DHCN** (Xia et al., 2021): It proposes a hypergraph convolutional network and devises another line graph of the hypergraph to improve Session-based recommendation.
- **GCE-GNN** (Wang et al., 2020): It proposes a global-level item representation learning layer, which employs a session-aware attention mechanism to recursively incorporate the neighbors' embeddings of each node on the global graph.
- **FGNN** (Qiu et al., 2019): It proposes a weighted attention graph layer and a Readout function to learn embeddings of items and sessions for the next item recommendation.
- **STAMP** (Liu et al., 2018): It proposes a novel short-term attention/memory priority model as a remedy, which is capable of capturing users' general interests from the long-term memory of a session context.
- **NARM** (Li et al., 2017): It explores a hybrid encoder with an attention mechanism to model the user's sequential behavior and capture the user's main purpose in the current session.
- **GRU4Rec** (Hidasi et al., 2016): It is RNN-based model for SBR, which considers practical aspects of the task and introduces several modifications to classic RNNs.
- **FPMC** (Rendle et al.): It is based on personalized transition graphs over underlying Markov chains.

[1]https://tianchi.aliyun.com/dataset/dataDetail?dataId=42
[2]http://dbis-nowplaying.uibk.ac.at/#nowplaying
[3]https://competitions.codalab.org/competitions/11161

Methods		Tmall				Nowplaying				Diginetica			
		P@10	MRR@10	P@20	MRR@20	P@10	MRR@10	P@20	MRR@20	P@10	MRR@10	P@20	MRR@20
Traditional	Item-KNN	6.65	3.11	9.15	3.31	10.96	4.55	15.94	4.91	25.07	10.77	35.75	11.57
	FPMC	13.10	7.12	16.06	7.32	5.28	2.68	7.36	2.82	15.43	6.20	26.53	6.95
RNNs	GRU4Rec	9.47	5.78	10.93	5.89	6.74	4.40	7.92	4.48	17.93	7.33	29.45	8.33
	NARM	19.17	10.42	23.30	10.70	13.60	6.62	18.59	6.93	35.44	15.13	49.70	16.17
	STAMP	22.63	13.12	26.47	13.36	13.22	6.57	17.66	6.88	33.98	14.26	45.64	14.32
GNNs	FGNN	20.67	10.07	25.24	10.39	13.89	6.80	18.78	7.15	37.72	15.95	50.58	16.84
	GCE-GNN	28.01	15.08	33.42	15.42	16.94	8.03	22.37	8.40	41.16	18.15	54.22	19.04
	S^2-DHCN	26.22	14.60	31.42	15.05	17.35	7.87	23.50	8.18	40.21	17.59	53.66	18.51
	Ours	**33.53**	**18.98**	**38.51**	**19.60**	**18.22**	**9.35**	**24.11**	**9.81**	**42.33**	**19.88**	**55.17**	**19.64**

Table 2: Comparison of Different Models, their results are obtained from the corresponding original papers. All the results are in percentage (%). The best performing method in each column is **boldfaced**.

That means for each user an own transition matrix is learned.

- **Item-KNN** (Sarwar et al., 2001): It defines similarity as the co-occurrence number of two items in sessions divided by the square root of the product of the number of sessions in which either item occurs, to recommend items.

Evaluation Metrics. As recommender systems can only recommend a few items at once, the actual item a user might pick should be amongst the first few items of the list (Hidasi et al., 2016). We adopt two widely used ranking based metrics: **P@K** and **MRR@K** by following previous works (Wang et al., 2020; Xia et al., 2021).Specifically, we mainly choose to use top-10 and top-20 items to evaluate a recommender system.

Hyper-parameters Setup. Following previous methods (Wang et al., 2020; Xia et al., 2021), we standardized the embedding dimension at 100, alongside consistently setting the batch size at 100 across all experimental models. We applied L_2 regularization at a rate of 10^{-5} and controlled for the hyper-parameters among the models to ensure equitable benchmarking. Within our approach, the weighting matrices originate from a normal distribution $N\left(0, 0.05^2\right)$, with biases initialized to zero. Concurrent with these parameters, we commence with item embeddings derived from a Gaussian Distribution $N\left(0, 0.1\right)$, progressing to a synchronous optimization with the remaining parameters. The hyperparameter γ_l is designated at the value of 0.2. Optimization is conducted through the Adam algorithm, commencing with an initial learning rate of 0.001 and implemented with a decay factor of 0.1 subsequent to every sequence of 5 epochs. In addition, the number of hypergraph convolutional layers is different in different datasets. For *Diginetica*, a two-layer setting is the best, while for others, a three-layer setting achieves the best performance. We train each model for 30 epochs or until the loss no longer decreases after 5 epochs.

4.2. Overall Comparison

To evaluate the performance of our proposed model, we conducted a comprehensive comparison with state-of-the-art item recommendation approaches, as detailed in **Table 2**. Here are the key observations from our experiments:

- **Consistently Strong Performance:** Across all datasets and metrics (with K=10 and 20), our proposed model consistently outperformed the existing baselines. This robust performance underscores the effectiveness of our approach. Notably, even on the Tmall dataset, where existing baselines had already achieved high performance, our method managed to push the performance boundary further.

- **Remarkable Performance Compared to Traditional and RNN Methods:** Our model exhibited remarkable performance when compared to traditional and RNN-based approaches. This suggests that our approach, which converts sequential item transitions into graph-structured data to capture the inherent order of item-transition patterns, delivers superior results.

- **Competitive Results Against Graph-Based Baselines:** Our method also achieved competitive results when compared to graph-based baselines. It's worth noting that both our approach and S^2-DHCN feature a hypergraph architecture. However, the improvements in our method primarily stem from our innovative contrastive learning strategies. By constructing hyperedge-level and node-level contrastive objectives to focus on fine-grained supervised signals, we enhance the learning process. This is in contrast to S^2-DHCN, which employs two types of hypergraphs for sessions embedding in contrastive learning, potentially resulting in weaker signals. Our approach also outperformed other strong baselines, such as GCE-GNN, and FGNN, further confirming the effectiveness of hypergraphs.

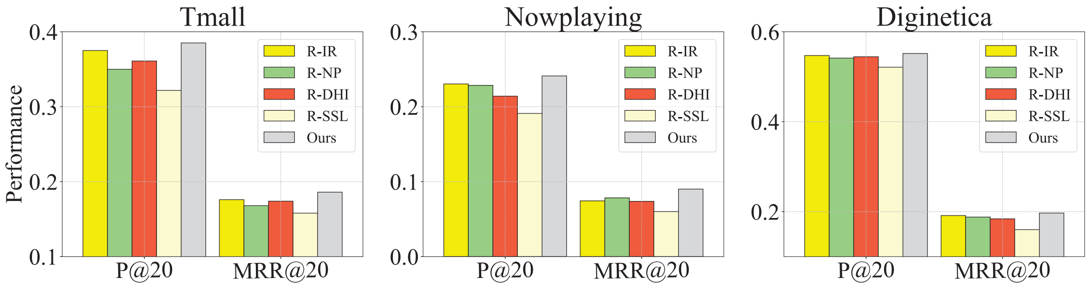


Figure 3: Ablation Study.

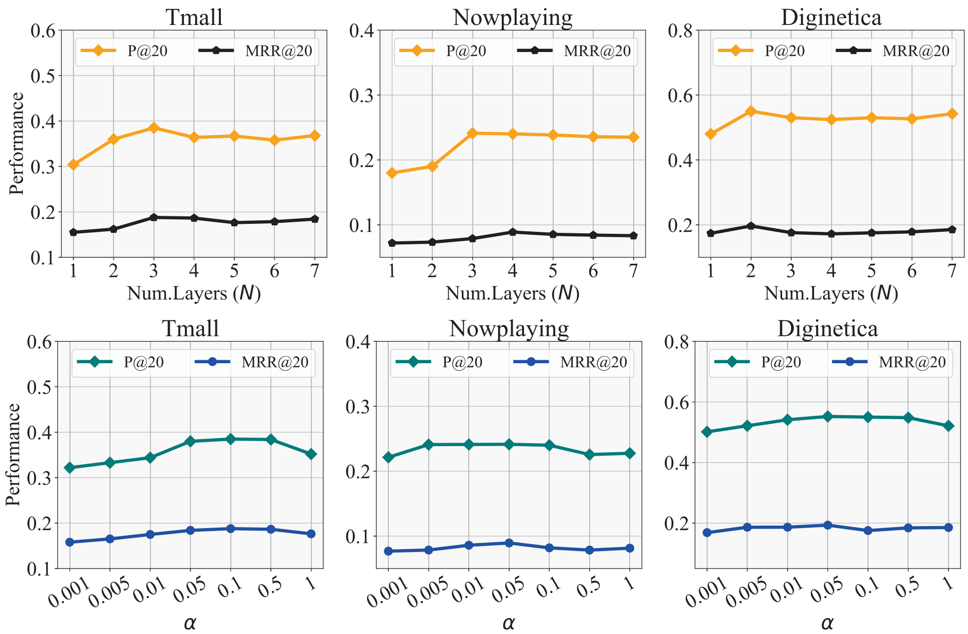


Figure 4: Two hyperparameters N and α.

4.3. Ablation Study

In this section, we conduct experiments to investigate the contribution of each component in our model. Specially, we design four variant versions: **R-IR**: We remove the initial residual for each hypergraph layer when high-order relations propagation. **R-NP**: We remove the noise perturbation (NP) for constructing self-supervised signals. **R-DHI**: We remove the dual hypergraph Infomax (DHI) for constructing self-supervised signals. **R-SSL**: We remove all SSL signals. The analysis of the components and their contributions in **Figure 3** reveals important insights into the performance of our model:

- **Self-Supervised Contrastive Learning:** Self-supervised contrastive learning plays a pivotal role in driving the performance improvement of the base model. Removing this component results in a significant drop in performance on both metrics, underscoring its importance.

- **Effectiveness of Two Contrastive Objectives:** The utilization of two contrastive objectives proves to be effective in achieving better performance compared to a single contrastive objective across the three datasets. This indicates that these two contrastive objectives can work in tandem and mutually complement each other.

- **Initial Residual for Hypergraph Layer:** The incorporation of the initial residual technique in the hypergraph layer serves to prevent oversmoothing during propagation. While removing this component leads to a performance drop on both metrics, it's worth noting that a careful balance is needed in the number of hypergraph layers, as excessive layers may not be beneficial.

- **Overall Model Effectiveness:** Across all variants, our model consistently outperforms them for K=20, showcasing the robustness and effectiveness of our model's design for session-based recommendation.

- **In summary:** Each component contributes significantly to the model's overall performance, with self-supervised contrastive learning playing a

Methods	Tmall		Nowplaying		Diginetica	
	Short Session	Long Session	Short Session	Long Session	Short Session	Long Session
FGNN	31.84	35.65	27.62	19.33	50.99	51.28
GCE-GNN	42.28	34.22	30.51	21.97	54.40	52.16
DHCN	36.47	31.73	30.56	23.68	53.29	52.43
Ours	**44.93**	**35.22**	**32.54**	**24.51**	**56.37**	**54.11**

Table 3: The performance of methods with different session lengths with P@20 (%).

central role in driving the improvements. The use of two contrastive objectives, the initial residual technique, and the overall model design contribute to the superior performance of our approach in session-based recommendation.

4.4. Hyperparameter Analysis

There are two critical hyperparameters employed by MHS-SSL as follows. (1) **The impact of the number of hypergraph layers N.** We summarize the results in **Figure 4** by ranging N within $\{1, 2, 3, 4, 5, 6, 7\}$. We can see that for both Tmall and Nowplaying, a three-layer setting achieves the best performance. For the Diginetica, stacking more than three layers will worsen the performance since the sessions in this dataset are generally short. Additionally, we notice that performance will not drop significantly when the number becomes larger. The main reason is that we adopt the technique of initial residual to prevent over-smoothing when propagation. (2) **The impact of the hyperparameters α.** We report the performance with a set of representative α values in $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1\}$ to control the magnitude of the SSL tasks. According to the results in **Figure 4**, the recommendation task achieves decent gains when jointly optimized with the self-supervised task. With the rise of α, the performance increases first and then declines. We think it is due to the gradient conflicts between the two tasks. Besides, when $\alpha = 0.05$, we get the best performance.

4.5. Analysis on Session Lengths

The analysis of different models' performance on sessions of varying lengths, as presented in **Table 3**, yields several important findings: **1) Superiority in Short Sessions:** In sessions with a length less than or equal to 5 items (Short), our model consistently achieves the best performance among the graph-based methods, as indicated by the highest Precision at 20 (P@20) scores. This underscores the model's adaptability and effectiveness in improving recommendations for sessions with fewer items, which are common in real-world scenarios. **2) Effectiveness Across Session Lengths:** While the primary focus is on enhancing recommendations for short sessions, our model also demonstrates commendable performance on long sessions (more than 5 items) compared to other graph embedding methods. This versatility is a notable advantage of our approach. **3) Short Sessions vs. Long Sessions:** The performance of our model in short sessions is notably better than that in long sessions. This finding is consistent with the primary objective of our approach to boost recommendations for sessions with fewer items, which are prevalent in real-world session-based recommendation scenarios. **In summary**, the results emphasize the effectiveness of our model in improving session-based recommendations, particularly in the context of short sessions. While the focus is on short sessions, our model maintains competitive performance in long sessions, highlighting its adaptability and suitability for various session lengths encountered in practical scenarios.

5. Conclusion

This paper introduces a novel approach, Mssen, for session-based recommendation, addressing data sparsity by employing SSL on item-session hypergraphs. Extensive empirical evaluations consistently show its superiority over existing methods. It's worth noting that hypergraph modeling in session-based recommendation is an emerging field with broader applications in graph-related research, offering ample room for further exploration and development.

6. Acknowledgements

This research was sponsored by the Stable Supporting Fund of National Key Laboratory of Underwater Acoustic Technology, JCKYS 2023604SSJS013, and Natural Science Foundation of Heilongjiang Province, LH2023F020, National Natural Science Foundation of China, 62272126.

7. References

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. 2020. Simple and deep graph convolutional networks. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 1725–1735. PMLR.

Ricardo Dias and Manuel J. Fonseca. 2013. Improving music recommendation in session-based collaborative filtering by using temporal context. In *ICTAI*, pages 783–788. IEEE Computer Society.

Gongfan Fang, Jie Song, Xinchao Wang, Chengchao Shen, Xingen Wang, and Mingli Song. 2021. Contrastive model inversion for data-free knowledge distillation. *CoRR*, abs/2105.08584.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. 2019. Hypergraph neural networks. In *AAAI*, pages 3558–3565. AAAI Press.

Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. 2015. Explaining and harnessing adversarial examples. In *ICLR*.

Jeff Z. HaoChen, Colin Wei, Adrien Gaidon, and Tengyu Ma. 2021. Provable guarantees for self-supervised deep learning with spectral contrastive loss. In *NeurIPS*, pages 5000–5011.

Kaveh Hassani and Amir Hosein Khas Ahmadi. 2020. Contrastive multi-view representation learning on graphs. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 4116–4126. PMLR.

Koby Hayashi, Sinan G. Aksoy, Cheong Hee Park, and Haesun Park. 2020. Hypergraph random walks, laplacians, and clustering. In *CIKM*, pages 495–504. ACM.

Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based recommendations with recurrent neural networks. In *ICLR*.

Pyungkang Kim, Kyung-Soo Kim, and Soohyun Kim. 2017. Modified nonnegative matrix factorization using the hadamard product to estimate real-time continuous finger-motion intentions. *IEEE Trans. Hum. Mach. Syst.*, 47(6):1089–1099.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *EMNLP*, pages 1746–1751. ACL.

Ashwin Krishnan, Manoj Nambiar, Nupur Sumeet, and Sana Iqbal. 2022. Performance model and profile guided design of a high-performance session based recommendation engine. In *ICPE*, pages 133–144. ACM.

Namkyeong Lee, Junseok Lee, and Chanyoung Park. 2021. Augmentation-free self-supervised learning on graphs. *CoRR*, abs/2112.02472.

Ansong Li, Zhiyong Cheng, Fan Liu, Zan Gao, Weili Guan, and Yuxin Peng. 2022. Disentangled graph neural networks for session-based recommendation. *CoRR*, abs/2201.03482.

Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural attentive session-based recommendation. In *CIKM*, pages 1419–1428. ACM.

Zihan Lin, Changxin Tian, Yupeng Hou, and Wayne Xin Zhao. Improving graph collaborative filtering with neighborhood-enriched contrastive learning. In *WWW*, pages 2320–2329. ACM.

Qiao Liu, Yifu Zeng, Refuoe Mokhosi, and Haibin Zhang. 2018. STAMP: short-term attention/memory priority model for session-based recommendation. In *SIGKDD*, pages 1831–1839. ACM.

Yanzhang Lyu, Hongzhi Yin, Jun Liu, Mengyue Liu, Huan Liu, and Shizhuo Deng. 2021. Reliable recommendation with review-level explanations. In *ICDE*, pages 1548–1558. IEEE.

Ruihong Qiu, Jingjing Li, Zi Huang, and Hongzhi Yin. 2019. Rethinking the item order in session-based recommendation with graph neural networks. In *CIKM*, pages 579–588. ACM.

Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. In *WWW*, pages 811–820. ACM.

Badrul Munir Sarwar, George Karypis, Joseph A. Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *WWW*, pages 285–295. ACM.

Yixin Su, Rui Zhang, Sarah M. Erfani, and Junhao Gan. 2021. Neural graph matching based collaborative filtering. In *SIGIR*, pages 849–858. ACM.

Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In *CIKM*, pages 1441–1450. ACM.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *CoRR*, abs/1807.03748.

Petar Velickovic, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. 2019. Deep graph infomax. In *ICLR*. OpenReview.net.

Sheng Wan, Shirui Pan, Jian Yang, and Chen Gong. 2021. Contrastive and generative graph convolutional networks for graph-based semi-supervised learning. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 10049–10057. AAAI Press.

Jianling Wang, Kaize Ding, Ziwei Zhu, and James Caverlee. 2021. Session-based recommendation with hypergraph attention networks. In *SDM*, pages 82–90. SIAM.

Ziyang Wang, Wei Wei, Gao Cong, Xiao-Li Li, Xianling Mao, and Minghui Qiu. 2020. Global context enhanced graph neural networks for session-based recommendation. In *SIGIR*, pages 169–178. ACM.

Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. 2021. Self-supervised graph learning for recommendation. In *SIGIR*, pages 726–735. ACM.

Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based recommendation with graph neural networks. In *AAAI*, pages 346–353. AAAI Press.

Jun Xia, Lirong Wu, and Jintao Chen. 2022. Simgrace: A simple framework for graph contrastive learning without data augmentation. In *WWW*, pages 1070–1079. ACM.

Xin Xia, Hongzhi Yin, Junliang Yu, Qinyong Wang, Lizhen Cui, and Xiangliang Zhang. 2021. Self-supervised hypergraph convolutional networks for session-based recommendation. In *AAAI*, pages 4503–4511. AAAI Press.

Hansheng Xue, Luwei Yang, Vaibhav Rajan, Wen Jiang, Yi Wei, and Yu Lin. 2021. Multiplex bipartite network embedding using dual hypergraph convolutional networks. In *WWW*, pages 1649–1660. ACM / IW3C2.

Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudré-Mauroux. 2019. Revisiting user mobility and social relationships in lbsns: A hypergraph embedding approach. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pages 2147–2157. ACM.

Yuning You, Tianlong Chen, Yang Shen, and Zhangyang Wang. 2021. Graph contrastive learning automated. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 12121–12132. PMLR.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. 2020. Graph contrastive learning with augmentations. In *Neural*.

Junliang Yu, Hongzhi Yin, Jundong Li, Qinyong Wang, Nguyen Quoc Viet Hung, and Xiangliang Zhang. 2021. Self-supervised multi-channel hypergraph convolutional network for social recommendation. In *WWW*, pages 413–424. ACM / IW3C2.

Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2022. Are graph augmentations necessary?: Simple graph contrastive learning for recommendation. In *SIGIR*, pages 1294–1303. ACM.

Eva Zangerle, Martin Pichl, Wolfgang Gassler, and Günther Specht. 2014. #nowplaying music dataset: Extracting listening behavior from twitter. In *WISMM*, pages 21–26. ACM.

Junwei Zhang, Min Gao, Junliang Yu, Lei Guo, and Jundong Li. 2021. Double-scale self-supervised hypergraph learning for group recommendation. In *CIKM*, pages 2557–2567. ACM.

Xiangping Zheng, Xun Liang, and Bo Wu. 2023a. Select the best: Enhancing graph representation with adaptive negative sample selection. In *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2023, Rhodes Island, Greece, June 4-10, 2023*, pages 1–5. IEEE.

Xiangping Zheng, Xun Liang, Bo Wu, Junlan Feng, Yuhui Guo, and Sensen Zhang. 2023b. Intent does matter! propagating high-order relations for exploring interest preferences. In *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2023, Rhodes Island, Greece, June 4-10, 2023*, pages 1–5. IEEE.

Xiangping Zheng, Xun Liang, Bo Wu, Yuhui Guo, and Xuan Zhang. 2022. Graph capsule network with a dual adaptive mechanism. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, pages 1859–1864. ACM.

Xiangping Zheng, Xun Liang, Bo Wu, Jun Wang, Yuhui Guo, Sensen Zhang, and Yuefeng Ma.

2023c. Modeling high-order relation to explore user intent with parallel collaboration views. In *Database Systems for Advanced Applications - 28th International Conference, DASFAA 2023, Tianjin, China, April 17-20, 2023, Proceedings, Part II*, volume 13944 of *Lecture Notes in Computer Science*, pages 489–504. Springer.

Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. 2021. Graph contrastive learning with adaptive augmentation. In *WWW*, pages 2069–2080. ACM / IW3C2.