Hyper Meta-Path Contrastive Learning for Multi-Behavior Recommendation

Haoran Yang*¶, Hongxu Chen*†¶, Lin Li‡, Philip S. Yu§ and Guandong Xu†¶

¶*School of Computer Science, University of Technology Sydney, Sydney, Australia*
‡*School of Computer and Artificial Intelligence, Wuhan University of Technology, Wuhan, China*
§*Department of Computer Science, University of Illinois at Chicago, Chicago, U.S.A*
haoran.yang-2@student.uts.edu.au, {hongxu.chen, guandong.xu}@uts.edu.au
cathylilin@whut.edu.cn, psyu@cs.uic.edu

Abstract—**User purchasing prediction with multi-behavior information remains a challenging problem for current recommendation systems. Various methods have been proposed to address it via leveraging the advantages of graph neural networks (GNNs) or multi-task learning. However, most existing works do not take the complex dependencies among different behaviors of users into consideration. They utilize simple and fixed schemes, like neighborhood information aggregation or mathematical calculation of vectors, to fuse the embeddings of different user behaviors to obtain a unified embedding to represent a user's behavioral patterns which will be used in downstream recommendation tasks. To tackle the challenge, in this paper, we first propose the concept of hyper meta-path to construct hyper meta-paths or hyper meta-graphs to explicitly illustrate the dependencies among different behaviors of a user. How to obtain a unified embedding for a user from hyper meta-paths and avoid the previously mentioned limitations simultaneously is critical. Thanks to the recent success of graph contrastive learning, we leverage it to learn embeddings of user behavior patterns adaptively instead of assigning a fixed scheme to understand the dependencies among different behaviors. A new graph contrastive learning based framework is proposed by coupling with hyper meta-paths, namely HMG-CR, which consistently and significantly outperforms all baselines in extensive comparison experiments.**

Keywords-**Hyper Meta-Path; Graph Contrastive Learning; Recommendation Systems;**

I. INTRODUCTION

Online shopping is becoming more and more essential nowadays, which generates a large volume of user behavioral data depicting users' purchasing motivations, interests, behavioral patterns, etc. However, many traditional recommendation systems [1], [2] pay significant attention to purchasing alone, leaving other associated behavioral data unexploited. Though recent works [3]–[7] reveal the gap and try to leverage multi-behavior information to improve recommendation quality, there are still limitations. For instance, some path-based works [3], [8] leverage meta-paths [9], [10] to extract recommendation context to better characterise users' multiple behaviors. However, there exist many meta-path schemes observed in heterogeneous graphs, resulting in the

difficulty of finding out the best one from multiple meta-path schemes via exhaustive search or learning a specific rule or scheme from the heterogeneous graphs to construct meta-paths [8]. Selecting effective and meaningful meta-path scheme in this case is time-consuming, and the reasons of the selection are usually unknown.

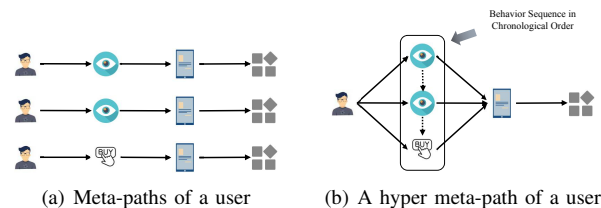(a) Meta-paths of a user   (b) A hyper meta-path of a user

Figure 1. (a) Meta-paths of a user in the recommendation system denotes the user's different behaviors. (b) A hyper meta-path constructed by previous meta-paths denotes the user's behavior pattern when he is purchasing a smartphone.

To overcome the above limitation of existing path-based approaches, we propose a new concept of hyper meta-path that consolidates multiple paths in a well-organised and holistic way. Similar to hyperedge in hypergraph [11]–[14] where an edge can connect more than two nodes, a hyper meta-path is a composition of multiple meta-paths between specified two end nodes in a heterogeneous network. As shown in Figure 1, let us assume that before purchasing a phone, the user had viewed the item twice. If meta-path based approaches are adopted to model the different shopping-related behaviors, three independent meta-path instances (Fig. 1.(a)) can be discovered to characterise this purchasing context. Though from these three meta-paths we can learn the user has viewed and purchased the item, it dose not explicitly reflect the exact behavioral pattern of the user. That is, it is unclear whether the user purchased the item directly or viewed the item carefully before purchasing.

In contrast, the proposed hyper meta-path is capable of achieving such a goal. As shown in part (b) of Fig. 1, a hyper meta-path between the user and the phone consolidates all three behaviors in a sequential order, which explicitly shows that the user carefully viewed the phone twice before the final purchasing action. Note that the way of consolidating related meta-paths into a hyper meta-path can be

*: Equal contribution
†: Corresponding author

flexible and generalized to any reasonable rule depending on particular application scenario. Besides, the concept of hyper meta-path is also useful for differentiating different behavior patterns between different users or when facing various categories of items presented to a user. For instance, normally technical people may research on different substitutable electronic products and take longer time to compare them, while non-technical people are not keen on investigating them and would probably directly buy one based on someone's recommendation. Even for the same user no matter their age and gender, they usually exhibit quite diverged buying patterns when facing different categories of products. For example, a user may have totally different buying patterns when purchasing large items (e.g., white goods like fridge or TV) and small fast-moving consumers goods (e.g., periodically buying tissues from online market without viewing again and again).

Nevertheless, it is not straightforward to incorporate the modelling of hyper meta-path into existing learning frameworks. Currently, graph-based unsupervised learning approaches are mainly used for path-based recommendation. For example, GNNs based approaches [4] are a popular means for multi-behavior recommendation via aggregating information passed from different types of edges or nodes in heterogeneous information networks [15]. Despite its popularity, these methods usually fuse the learned features of different behaviors independently, which is too naive to reflect hyper meta-path context for recommendation. Moreover, multi-task learning based models [5], [6] are also possible ways that introduce additional supervision signals from the observed multiple behavior data to improve recommendation quality. However, extra efforts on well-elaborated tasks are tricky, and researchers have to carefully work out the effective dependencies among related tasks. For example, taking purchasing prediction as a primary task while modelling the *add-to-cart* behavior prediction as an auxiliary task might not be always right as some users may buy some items directly without putting them into the cart.

Thus, to further reveal and capture the differences between buying patterns, together with hyper meta-paths, we innovatively leverage graph contrastive learning [16], [17] paradigm for the multi-behavior recommendation problem. The main idea of graph contrastive learning is to distinguish the differences among graphs to obtain the useful structure information of each graph, raising a recent surge of interest [16], [17]. The rationale of incorporating contrasitive learning with our proposed hyper meta-paths is that, a user may have multiple hyper meta-paths explicitly illustrating his/her behavioral patterns when facing different products. Since hyper meta-path explicitly describes users' behaviors towards purchasing different items, graph contrastive learning becomes a best fit for comparing and extracting the key structures in the graph consisting of hyper meta-paths.

More specifically, we combine multiple hyper meta-paths

of a user to construct several hyper meta-graphs. Each hyper meta-graph contains different number of types of behaviors, For example, the first hyper meta-graph contains *buy* and the second hyper meta-graph contains *buy* and *page view*. In this case, different hyper meta-graphs reflect different behavioral patterns regarding different products of the user. Then, we conduct graph contrastive learning among the constructed hyper meta-graphs to adaptively obtain the complex dependencies among different behaviors and the embeddings representing different behavioral patterns. For instance, in HMG-CR, we first build the target contrastive graph that only contains *buy* interactions between users and items as it is the target behavior for recommendation systems, and the other contrastive hyper meta-graphs are added for comparison by incrementally introducing auxiliary behaviors to the precedent hyper meta-graph. After that, we conduct graph contrastive learning between the constructed contrastive hyper meta-graphs to successively obtain progressive and comprehensive representations for each types of behaviors. Finally, the recommendation will be performed based on those discovered behavior patterns and features.

The contributions of this paper can be summarized into three aspects:

- We propose the concept of hyper meta-path explicitly illustrating the logical relations among a collection of meta-paths, which tackles the limitation of meta-path that is insufficient to model the interactions among meta-paths. Hyper meta-path can be regarded as an approach to enrich graph structures.
- We innovatively utilize graph contrastive learning to capture the complex behavior patterns of users adaptively, alleviating existing methods' limitation.
- We propose a novel and flexible recommendation framework coupling graph contrastive learning with hyper meta-path, which achieves the superior performances in the comprehensive comparison experiments.

II. PRELIMINARIES

This section will introduce some preliminaries regarding the proposed concept of hyper meta-path.

A. Meta-Path

Heterogeneous networks have been intensively studied by a lot of researchers due to its ability of utilizing multi-model multi-typed graph data. To illustrate the power of hetergeneous networks, Sun et al. [9] proposed the concept of meta-path, which is widely used by many existing works [10], [18] in the research area of heterogeneous networks modeling. Each meta-path captures the features among the nodes on the meta-path from a particular semantic perspective. Due to the diversity of meta-paths in a heterogeneous graph, for the target (e.g., a node or an edge), there exist multiple meta-paths. Thus, the informative meta-paths give heterogeneous

network models the chance to obtain the multi-model multi-typed features of nodes and their relations. This kind of data structure indeed shows the advantage in many real-world graph data mining applications [19], [20]. However, there are limitations existing in meta-paths mentioned in Section I, failing to capture the interaction information among multiple meta-paths.

B. Hyper Meta-Path

Though we can build extra meta-paths based on the interactions among existing meta-paths, we cannot take an exhaustive method to compute on every meta-paths since the computation complexity is unaffordable. Inspired by the concepts of hyperedge and hypergraph, we find a way to integrate interaction information among meta-paths into the target. According to the limitation of conventional meta-path mentioned above and advantages of hyperedge and hypergraph, we propose the concept of hyper meta-path to capture meta-path features and interaction information among them simultaneously.

Definition 1: **Hyper meta-path**. A *hyper meta-path* is a logical composition of multiple meta-path schemas connecting two end nodes in a heterogenous information network. Hyper meta-path has the following properties:

- It describes the logical relations (e.g., chronological order, spatial order and topological order) among a sort of meta-paths with the same end nodes.
- Multiple hyper meta-paths, which have the same start node, compose a *hyper meta-graph*.

III. METHODOLOGY

In this section, we will introduce details of our proposed framework as shown in Figure 2, **H**yper **M**eta-**G**raph **C**ontrastive learning for **R**ecommendations, namely HMG-CR.

A. Hyper Meta-Graph Generation

GNNs based recommendation methods [21] have recently achieved tremendous success due to the power of GNNs. Data is critical to neural network models' performances. One way to leverage GNN based recommendation models is to construct proper graphs for them. The most common way to construct graphs in recommendation systems is building bipartite graphs via user-item interaction history. Since user-item interaction graphs are bipartite graphs, they are lack of semantic information because of their simple structure. To tackle this limitation, researchers have taken measures to further enrich semantic information carried by graphs. For example, adding side information into the graph [22], utilizing meta-paths existing in the graph [8], and constructing more complex graph structures (e.g., hypergraph) [23], [24].

To improve recommendation results, in our work, we utilize the proposed concept of hyper meta-path to construct hyper meta-graphs carrying rich semantic information. Next

we will introduce details to construct hyper meta-graphs for our proposed recommendation framework.

Given a set of interaction records in a recommendation system, $\{(u_j, r_k, i_q)|u_j \in \mathbf{U}, i_q \in \mathbf{I}, r_k \in \mathbf{R}\}$, where $\mathbf{U} = \{u_0, u_1, \cdots, u_n\}$ denotes the set of all users, $\mathbf{I} = \{i_0, i_1, \cdots, i_m\}$ denotes the set of all items, and $\mathbf{R} = \{r_0, r_1, \cdots, r_l\}$ denotes the set of all different kinds of user behaviors. According to the number of types of different user behaviors, we construct $|\mathbf{R}| = l + 1$ hyper meta-graphs for each user. For the t-th hyper meta-graph of user u_j, it is defined as $\mathbf{HG}_t^j = \{(u_j, (r_a, r_b, \cdots, r_c), i_q)|i_q \in \mathbf{I}, \forall r \in \{r_0, r_1, \cdots, r_{t-2}, r_l\}\}$, where behavior sequence $(r_a, r_b, \cdots, r_c)$ is sorted in chronological order, and each behavior r in the sequence solely bridges user u_j and item i_q. Hence, we will have a set of hyper meta-graphs $\mathcal{HG}^j = \{\mathbf{HG}_0^j, \mathbf{HG}_1^j, \cdots, \mathbf{HG}_l^j\}$ to illustrate user-item interactions of user u_j in the recommendation system. Note that the order among different behaviors in $\mathbf{R}$ is based on the distance between behaviors and behavior *buy* in the semantic space. For example, four types of common user behaviors, *page view*, *favorite*, *add to cart*, *buy*. Behavior *page view* is farthest from behavior *buy*, we can define a sorted set of behaviors here, $\{r_{pv}, r_{fav}, r_{cart}, r_{buy}\}$. So, the first hyper meta-graph of a user solely contains the behavior of *buy*, the second one contains *page view* and *buy*, the third one contains all the behaviors except *add to cart*, and the last hyper meta-graph of the user contains all of four types of behaviors.

B. Graph Encoders

Graph encoder is the essential part of the whole framework since it determines whether the framework can learn representative embeddings for users' behavior patterns from hyper meta-graphs. GNN models are widely used graph encoders, e.g., GCN [25] and GAT [26]. Technically, any GNN models can be used in our framework with sufficient information (e.g., edge types and node features). Note that we do not need exquisite GNN models since we have built hyper meta-graphs which carry rich semantic information (e.g., geometric information, topological structures). In the practice of our framework, we could apply geometric or topoloy based GNNs, like GIN [27] and TAG [28], as the graph encoder because of their simplicity and effectiveness. Further more, they can leverage the structure information from proposed hyper meta-path and advantages of structure-level contrastive learning [16].

As mentioned in the previous section, each user in the recommendation system have $|\mathbf{R}|$ hyper meta-graphs. We assign $|\mathbf{R}|$ independent graph encoders to process these hyper meta-graphs accordingly, note that these graph encoders are shared among different users. Given the t-th hyper meta-graph of user u_j and a graph encoder $g_t(\cdot)$, where $\cdot$ denotes a hyper meta-graph, we will have the embedding of the t-th
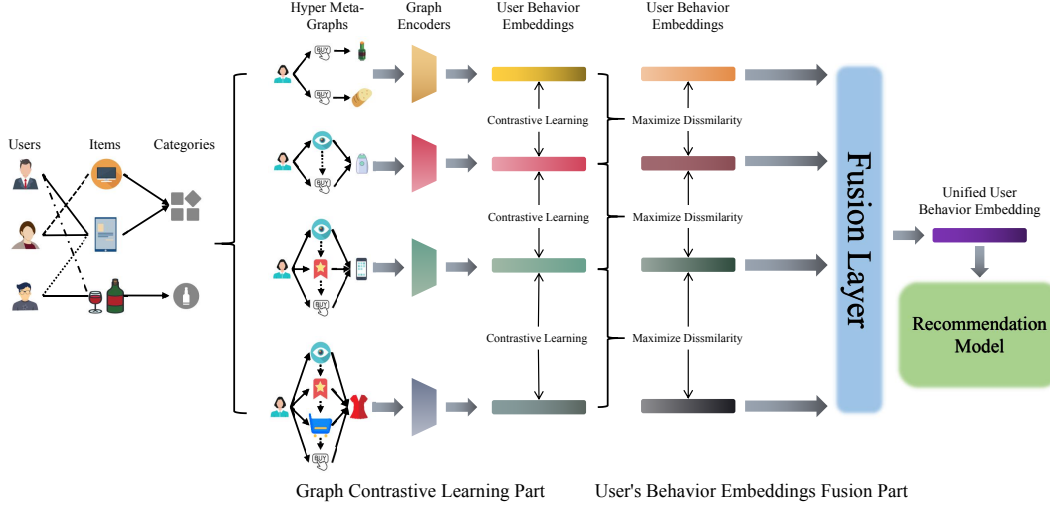
Figure 2. The overview of hyper meta-graph contrastive learning framework for recommendantions.

hyper meta-graph of user u_j:

$$\mathbf{h}_t^j = g_t(\mathbf{HG}_t^j), \qquad (1)$$

where $\mathbf{h}_t^j \in \mathbf{R}^h$ and h denotes the hidden dimension of the user behavior pattern embeddings and item embeddings.

C. Hyper Meta-Graph Contrastive Learning

For each user, we build several hyper meta-graphs. The graphs carry the interaction records of the user. We can capture this information via graph encoders learning on the hyper meta-graphs separately. However, the behavior patterns of a user would be complicated. According to the example mentioned in the previous section, we have four different hyper meta-graphs for each user. The complexity of the hyper meta-graph is increasing following the number of behavior types it contains. For example, the first hyper meta-graph solely includes *buy*, and the second hyper meta-graph includes *page view* and *buy*. The second hyper meta-graph contains at least two purchasing patterns: buying the item directly, which is also contained in the first hyper meta-graph, and buying the item after viewing. Suppose we adopt graph encoders to learn on each hyper meta-graph separately. In that case, different behavior patterns in the same hyper meta-graph will be fused. This result may neglect the performances when using the learned behavior patterns for the recommendation. It is critical to extract different behavior patterns from a sequence of hyper meta-graphs whose complexities are cascadingly increasing. A potential solution is to contrast the hyper meta-graph with its previous one to obtain the differences (e.g., different behavior patterns) between these two adjacent hyper meta-graphs.

Thanks to the recent success of contrastive learning in graph learning, we propose to utilize *hyper meta-graphs*

discrimination as the solution to obtain different behavior patterns and InfoNCE as the contrastive learning objective.

We give an example here. For the user u_j, we give out two adjacent hyper meta-graphs of u_j, which are HG_{t-1}^j and HG_t^j. We assign $g_{t-1}(\cdot)$ and $g_t(\cdot)$ as their graph encoders, respectively. Hence, we will have embeddings of two hyper meta-graphs:

$$\mathbf{h}_{t-1}^j = g_{t-1}(\mathbf{HG}_{t-1}^j), \qquad (2)$$

$$\mathbf{h}_t^j = g_t(\mathbf{HG}_t^j). \qquad (3)$$

In this example, $\mathbf{h}_{t-1}^j$ and $\mathbf{h}_t^j$ compose the negative pair. To satisfy the setting of InfoNCE, we must construct the positive pair to fulfill the contrastive learning process. Following the graph contrastive learning settings in GCC [16], we use $g_{t-1}(\cdot)$ to encode HG_t^j to obtain $\hat{\mathbf{h}}_t^j$:

$$\hat{\mathbf{h}}_t^j = g_{t-1}(\mathbf{HG}_t^j), \qquad (4)$$

which is together with $\mathbf{h}_t^j$ to compose the positive pair. In our work, we adopt InfoNCE such that:

$$\mathcal{L}_{t-1,t}^j = -\ln \frac{\exp\left(d(\mathbf{h}_t^j, \hat{\mathbf{h}}_t^j)\right)}{\exp\left(d(\mathbf{h}_t^j, \hat{\mathbf{h}}_t^j)\right) + \exp\left(d(\mathbf{h}_t^j, \mathbf{h}_{t-1}^j)\right)}, \quad (5)$$

where $d(\cdot, \cdot)$ denotes the metrics measuring the distance between two vectors. For the recommendation system having $n+1$ users and $l+1$ different types of user behaviors, we will have a total contrastive learning objective:

$$\mathcal{L}_{contra} = \frac{1}{n+1} \sum_{j=0}^{n} \sum_{t=1}^{l} \mathcal{L}_{t-1,t}^j. \qquad (6)$$

The intuitions of adopting such a strategy are twofold:

- **Avoid generating the negative pair via graph augmentation**. Some works [17] utilize graph augmentation to generate negative pairs. But in the recommendation scenario, graph augmentation would disturb the

790

users' interaction records and affect behavior pattern generation, which may cause misleading results in the downstream recommendation tasks. Such a strategy is an alternative solution for us to generate the negative pair without disturbing original semantics.

- **Bridge two contrasting hyper meta-graphs**. It is hard for us to link the embeddings generated from different graph encoders with different graphs in semantic space. However, with such a strategy, we can build an implicit connection between contrasting hyper meta-graphs in the contrastive learning process.

We will have $|\mathbf{R}|$ user behavior embeddings for a user after contrastive learning process, which will be fed into fusion layer and downstream recommendation tasks.

D. Users' Multi-behavior Patterns Fusion

After obtaining $|\mathbf{R}|$ different embeddings which denote different behavior patterns of a user, we have to fuse them and obtain a unified embedding to conduct recommendations. There is a sort of widely used linear fusion methods, like *sum* and *mean*. And there is another type of fusion method, which is neural network-based methods (e.g., Multi-Layer Perceptron (MLP) and Personalized Non-Linear Fusion (PNLF) [29]). Given a fusion function $f(*)$, we can have a unified behavior pattern embedding for the user:

$$\mathbf{h}^j_{uni} = f(\mathbf{h}^j_0, \mathbf{h}^j_1, \cdots, \mathbf{h}^j_l) \in \mathbf{R}^h. \tag{7}$$

E. Recommendation Task

There are plenty of collaborative filtering-based recommendation frameworks, which leverage the explicit or implicit feedback of users, like [1]. To fully demonstrate the ability of the proposed model, we use vector product to make prediction instead of those complex and state-of-the-art models to avoid the improvement brought by the recommendation model.

Let $\mathbf{h}^k$ denotes the embedding of item i_k. With the unified behavior pattern embedding of user u_j, we can obtain the predicted score between the item and the user via:

$$\hat{p}_{u_j,i_k} = {\mathbf{h}^j_{uni}}^T \boldsymbol{W} \mathbf{h}^k, \tag{8}$$

where the trainable weight matrix $\boldsymbol{W} \in \mathbf{R}^h$. The matrix $\boldsymbol{W}$ is used to map the unified behavior pattern to the space where item embeddings in for score prediction.

To train model parameters, we take the negative logarithm [1] of the likelihood function:

$$
\begin{aligned}
\mathcal{L}_{rec} = - \sum_{(u_j,i_k) \in \mathcal{Y} \cup \mathcal{Y}^-} & p_{u_j,i_k} \log \hat{p}_{u_j,i_k} \\
& + (1 - p_{u_j,i_k}) \log(1 - \hat{p}_{u_j,i_k}),
\end{aligned} \tag{9}
$$

to normalize the loss value of loss function on recommendation tasks, we take

$$\mathcal{L}_{ave\_rec} = \frac{\mathcal{L}_{rec}}{|\{(u_j,i_k)|(u_j,i_k) \in \mathcal{Y} \cup \mathcal{Y}^-\}|} \tag{10}$$

Dataset	Taobao	Tmall
#users	48946	9368
#items	1500839	302722
#pv (percentage)	7723217 (85.17%)	1510303 (92.14%)
#fav (percentage)	436715 (4.82%)	102419 (6.25%)
#cart (percentage)	527221 (5.81%)	24557 (1.50%)
#buy (percentage)	380877 (4.20%)	104360 (6.37%)
#total	9068030	1639220
#ave_pv	157.79	161.22
#ave_fav	8.92	10.93
#ave_cart	10.77	2.62
#ave_buy	7.78	11.14
#ave_total	185.27	174.98

as the objective, where $\mathcal{Y}$ and $\mathcal{Y}^-$ denote postive interaction records and sampled negative interaction records, $p_{u_j,i_k} \in \{0,1\}$ represents if there is an interaction between user u_j and item i_k.

To train the model from end to end, we couple two objectives as the total loss function:

$$\mathcal{L} = (1 - \beta) \cdot \mathcal{L}_{contra} + \beta \cdot \mathcal{L}_{ave\_rec}, \tag{11}$$

where β is a hyperparameter contoling the significance of two objectives in the total training objective.

IV. EXPERIMENTS

This section evaluates HMG-CR on recommendation tasks with two real-world datasets. We will first report the comparison experiment results of HMG-CR and baselines. Then, we analyze how the graph contrastive learning works in our model. Lastly, we conduct ablation studies on graph encoder and fusion layer in the model.

A. Datasets

We evaluate the proposed framework on two real-world datasets, which have high quality and are widely used. including Taobao[1] and Tmall[2]. To ensure the quality of the datasets, we follow the customary practice [30] to discard users and items with less than five interactions of *buy*. We also filter users with too much interactions of *page view* in Tmall to discard noise. The statistics of the filtered datasets are shown in Table I.

B. Baselines

To verify the effectiveness of the proposed framework, we compare it with three categories of baselines. The first category is conventional GNNs including GCN [25] and GraphSAGE [2], which cannot distinguish different types of edges in the graph, they treat different user behaviors as the same. The second category is edge types-aware GNNs including GAT [26] and RGCN [4], which can process various types of edges in the graph explicitly or implicitly to capture the features of different user behaviors.

---

[1]https://tianchi.aliyun.com/dataset/dataDetail?dataId=649
[2]https://tianchi.aliyun.com/dataset/dataDetail?dataId=47

The last category is novel multi-behavior recommendation frameworks, NMTR [5] and EHCF [6] which achieve state-of-the-art performances on multi-behavior recommendation tasks.

C. Experiment Settings

For reproducibility, we introduce the details of the hyper-parameter settings of the proposed framework. We train our model on dataset Taobao with learning rate $lr = 0.0001$, weight decay $wd = 0.000001$, hidden dimension $h = 16$ advised by [31], and 3-layer TAG as graph encoder. As to dataset Tmall, we tune our model with the same learning rate, weight decay and hidden dimension. We take 3-layer GIN as graph encoder for dataset Tmall. To ensure the fairness in the comparison studies, we follow the widely used *leave-one-out* strategy [1] to conduct comparison. The metrics we adopt are Recall@K and NDCG@K, which show the recommendation quality of top-K recommended items. For more details, you can refer to the source code of this project via this link[3]. All experiments are conducted on NVIDIA TITAN Xp.

D. Comparison Experiment Results

Table II lists the comparison experiment results for all methods on two datasets. Overall, the proposed framework HMG-CR with different graph encoders consistently and significantly outperforms all baselines in terms of all metrics. Particularly, our proposed framework has more significant improvement on the metric NDCG, which shows that our proposed framework pays more attention to sorting recommended items. Note that HMG-CR on dataset Taobao slightly outperforms that on dataset Tmall. According to the statistics of the two datasets, as shown in Table I, we note that the average numbers of total interactions for each user are close in two datasets, but there are differences among the distribution of numbers of different user behaviors. The ratio of *add to cart* in dataset Tmall is much less than that in dataset Taobao. Each user in both datasets has four hyper meta-graphs since there are four different types of user behaviors. Due to lack of *add to cart* in dataset Tmall, the third hyper meta-graph for a user, including *page view*, *add to cart*, and *buy*, is similar to the second hyper meta-graph for the user, including *page view* and *buy*. Under such a scenario, it is hard for graph contrastive learning to maximize the dissimilarities between the second hyper meta-graph and the third hyper meta-graph. Hence, the user behavior pattern embedding generated in this part would be misleading for unified user behavior pattern embedding generation.

Graph neural network-based methods have unsatisfying performances in our experiment. The interaction graphs for each user in the recommendation systems have simple structures (e.g., bipartite graphs). Conventional GNN models,

like GCN and GraphSAGE, may be insufficient to capture user behavior pattern embeddings on such simple graph structures. Edge types-aware GNN models, like GAT and RGCN, slightly outperform GCN and GraphSAGE since they integrate fruitful side information regarding different types of user behaviors. Overall, two categories of GNN models have no significant gaps, because *page view* takes the most of place in the datasets. Message passing and aggregation are not capable to capture sophisticated relations among different types of user behaviors, since semantics of *page view* would conceal other information.

NMTR and EHCF are state-of-the-art multi-behavior recommendation frameworks. They leverage the well-designed recommendation models and multi-task learning strategy to utilize the supervision signals from all types of user behaviors. However, there is a limitation for both frameworks. Both of them have an assumption that each type of user behaviors has strong connections with precedent types of user behaviors. This assumption is not solid, because users' behavioral patterns are complex as shown in a toy example in Section II-B. The proposed HMG-CR adopts a more flexible manner to utilize graph contrastive learning to capture the dependencies among different types of user behaviors instead of assuming there are strong connections between a behavior and the precedent one. Because of it, even without multi-task supervision signals and well-designed recommendation models, the proposed HMG-CR still outperforms NMTR and EHCF with leveraging the advantages of hyper meta-graphs and graph contrastive learning.

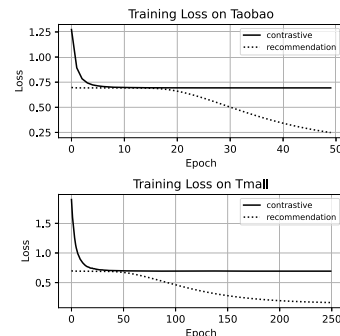## E. Analysis of Graph Contrastive Learning

Figure 3. The contrastive loss and recommendation loss of HMG-CR on both datasets in training process

In this section, we will introduce to you the detailed mechanism of graph contrastive learning in our proposed framework. First, as shown in the Figure 3, we demonstrate the training loss of the proposed framework on two datasets during the training process. The training loss is twofold, contrastive loss and recommendation loss. We observe a

Dataset	Taobao				Tmall			
Methods / Metrics	Recall@5	Recall@10	NDCG@5	NDCG@10	Recall@5	Recall@10	NDCG@5	NDCG@10
GCN	0.2577	0.3589	0.1842	0.2167	0.2544	0.3775	0.1763	0.2163
GraphSAGE	0.2751	0.3826	0.1965	0.2312	0.2588	0.3695	0.1813	0.2170
GAT	0.2782	0.3921	0.1972	0.2339	0.2561	0.3735	0.1777	0.2158
RGCN	0.2714	0.3767	0.1946	0.2285	0.2725	0.4144	0.1749	0.2215
NMTR	0.2215	0.3781	0.1513	0.2012	0.2780	0.4230	0.1798	0.2265
EHCF	0.2882	0.4166	0.1945	0.2359	0.2451	0.4115	0.1581	0.2113
HMG-CR(SG)	0.3050	0.4417	0.2162	0.2608	0.2943	0.4329	0.1863	0.2321
HMG-CR(GCN)	0.3039	0.4441	0.2154	0.2613	0.2954	0.4332	0.1869	0.2324
HMG-CR(GAT)	0.3460	0.4390	0.2443	0.2746	0.3163	0.4320	0.2224	0.2604
HMG-CR(GIN)	0.3141	0.3627	0.2029	0.2191	**0.3547**	0.4313	**0.2642**	**0.2891**
HMG-CR(TAG)	**0.3588**	**0.4464**	**0.2639**	**0.2926**	0.2964	**0.4350**	0.1902	0.2359
Improvement	24.50%	7.15%	33.82%	24.04%	27.59%	2.84%	45.73%	27.64%

clear tendency that contrastive loss drops first and remains stabilized, and then the recommendation loss starts to decrease. This phenomenon reflects that our proposed framework first maximizes the dissimilarity among hyper meta-graphs to obtain user behavior pattern embeddings and updates parameters on the recommendation task. And the contrast among hyper meta-graphs is maintained within the whole training process.
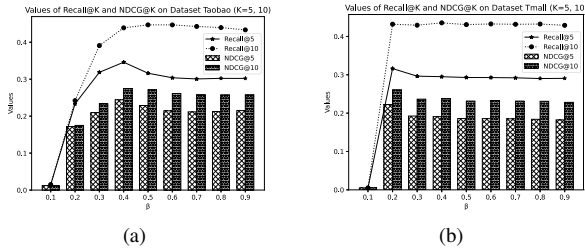

Figure 4.   (a) Values of Recall@K and NDCG@K on dataset Taobao variation trend with different β. (b) Values of Recall@K and NDCG@K on dataset Tmall variation trend with different β. (Graph encoder and fusion layer of HMG-CR here are GAT and MEAN, respectively.)

To further illustrate the impact of graph contrastive learning on our proposed framework, we conduct hyperparameter studies on β, which controls the relative significance of graph contrastive learning tasks and recommendation tasks. The experimental results of hyperparameter studies are shown in the Figure 4. Overall, our proposed framework is not that sensitive to β as long as β is not too small. However, we note that the proposed framework has worse results when β takes the boundary values (e.g., $\beta = 0.1$). When β is too low, the model pays less attention to the recommendation tasks. The model cannot acquire sufficient supervision signals from training data to update the parameters. Under this scenario, it is difficult for our model to converge quickly and precisely on recommendation tasks. With β increasing, the performances of the proposed framework increase accordingly. When β is larger than some specific values, e.g., $\beta = 0.4$ for dataset Taobao and $\beta = 0.2$ for dataset Tmall, the performances start to decrease slightly. Large β values

will neglect graph contrastive learning tasks, which would undermine the ability of the model to acquire user behavior pattern embeddings from sophisticated hyper meta-graphs. This phenomenon also verifies that graph contrastive learning is helpful to our proposed framework. In summary, graph contrastive learning tasks and recommendation tasks should have a relatively balanced significance and β should not be too large in our proposed framework to avoid decreasing model performances and cannot be too small in which case the framework may not work.

F. Ablation Studies

Graph Encoder Choosing a proper graph encoder for the framework determines whether it can achieve good performances. We select three common categories GNNs, conventional message passing based GNNs including SG [32] and GCN, attention mechanism based GNNs including GAT, and graph topological or geometric structure aware GNNs including TAG [28] and GIN [27]. The experiment results are shown in the Table II. According to the results, the proposed HMG-CR with any graph encoders outperforms all baselines. Specifically, HMG-CR with SG or GCN slightly outperforms baselines since conventional message passing based GNNs are insufficient to capture complex user behavior features from the constructed hyper meta-graphs. Despite the user-item interactions in the hyper meta-graphs, there are also chronological dependencies among different user behaviors. With such sophisticated relations in the hyper meta-graphs, GAT leverages attention mechanism to learn user behavior embeddings via adaptively distinguish different relations (edges) in the hyper-meta graphs. However, we replaced different types of edges, which represent user behaviors, with different types of nodes in the hyper meta-graphs. We explicitly add the information of interactions among users and items into the hyper meta-graph. It means that the improvement brought by attention mechanism, distinguishing different edges is limited. Note that the hyper meta-graphs have a structure which is similar to tree topology. Hence, the hyper meta-graphs have not only fruitful semantic information but also excellent structure. HMG-CR

Table III
SUPPLEMENTARY EXPERIMENT RESULTS OF GIN AND TAG

Dataset	Taobao				Tmall			
Methods \ Metrics	Recall@5	Recall@10	NDCG@5	NDCG@10	Recall@5	Recall@10	NDCG@5	NDCG@10
GIN	0.2682	0.3779	0.1892	0.2246	0.2817	0.4236	0.1878	0.2340
TAG	0.2784	0.3863	0.1994	0.2342	0.2845	0.4235	0.1869	0.2323
HMG-CR(GIN)	0.3141	0.3627	0.2029	0.2191	**0.3547**	0.4313	**0.2642**	**0.2891**
HMG-CR(TAG)	**0.3588**	**0.4464**	**0.2639**	**0.2926**	0.2964	**0.4350**	0.1902	0.2359

Table IV
PERFORMANCES OF HMG-CR(GAT) WITH DIFFERENT FUSION LAYERS ON BOTH DATASETS
($\beta = 0.4$ FOR DATASET TAOBAO AND $\beta = 0.2$ FOR DATASET TMALL)

Dataset	Taobao				Tmall			
Fusion \ Metrics	Recall@5	Recall@10	NDCG@5	NDCG@10	Recall@5	Recall@10	NDCG@5	NDCG@10
MEAN	**0.3460**	0.4390	**0.2443**	**0.2746**	**0.3163**	0.4320	**0.2224**	**0.2604**
SUM	0.3012	**0.4427**	0.2118	0.2580	0.2939	0.4345	0.1879	0.2343
MLP	0.3024	0.4344	0.2150	0.2579	0.2946	**0.4349**	0.1873	0.2336
PNLF	0.3046	0.4363	0.2157	0.2586	0.2944	0.4344	0.1865	0.2327

with graph structure aware graph encoders leverages the advantages of the hyper meta-graphs and achieve the best results in our experiments. To verify the improvement is brought by our proposed framework instead of TAG or GIN solely, we conduct supplementary experiments with TAG and GIN shown in the Table III.

Fusion Layer Fusion layer is the output layer of the proposed framework. We focus on two categories of fusion layers, linear fusion layer, *mean* and *sum*, and non-linear fusion layer, MLP and PNLF. The experimental results are shown in the Table IV. According to the results, HMG-CR taking *mean* as the fusion layer achieves the best result. Overall, HMG-CR with linear fusion layer has better performances in our experiments. We note that there are mapping layers in MLP and PNLF. In this component, mapping mechanism may disturb the user behavior pattern obtained in the space in which the graph contrastive learning was conducted. Hence, we should take linear fusion layer to output the unified user behavior pattern embeddings to avoid disturbing caused by conducting fusion in another embedding space.

V. RELATED WORKS

A. Graph Contrastive Learning

Graph contrastive learning recently attracts attention from researchers leveraging contrastive learning idea to enhance existing GNNs. The core idea of contrastive learning is to maximize representation agreement among sampled and transformed data. Some works [16], [17] introduce contrastive learning to graph representation learning and had achieved promising results. There are two main measures to generate contrasting pairs, including graph perturbation and sub-graph sampling. However, how to adpatively construct contrasting pairs instead of randomly and how to implement graph contrastive learning in real-world problems are not fully explored.

B. Multi-behavior Recommendation

Multi-behavior recommendation utilizes multiple user-item feedback for enhancing recommendation on target behaviors. There are different approaches to make use of users' multi-behavior information. In a multi-behavior interactions graph, [4], [26], [33] assign different weights to different types of edges, representing different types of behaviors, before conducting aggregation. Graph based recommendation methods [4], [26], [34] achieve good performances in recommendation tasks with leveraging advantages of GNNs. Moreover, [5], [6] adopt multi-task learning techniques to acquire more supervision signals from multi-behavior data. For embedding generation, they assume that one behavior is strongly related to the precedent behaviors and embeddings of different types of user behaviors are adjacent in the embedding space. Both aggregation in [4], [26] and assumptions in [5], [6] are insufficient to capture complex relationships among different behaviors of users.

VI. CONCLUSION

In this paper, we propose the concept of hyper meta-path and a novel framework, HMG-CR, which first utilizes graph contrastive learning techniques into recommendation systems. Leveraging the advantages of hyper meta-path and HMG-CR, we achieve the SOTA performances on the task of purchasing prediction on both datasets. We also conduct extensive analysis on HMG-CR and fully demonstrate the details of it. The concept of hyper meta-path and the framework are flexible and can be used in other heterogeneous graph mining tasks.

REFERENCES

[1] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. Chua, "Neural collaborative filtering," in *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, R. Barrett, R. Cummings, E. Agichtein, and E. Gabrilovich, Eds. ACM, 2017, pp. 173–182. [Online]. Available: https://doi.org/10.1145/3038912.3052569

[2] W. L. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds., 2017, pp. 1024–1034. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract.html

[3] H. Zhao, Q. Yao, J. Li, Y. Song, and D. L. Lee, "Meta-graph based recommendation fusion over heterogeneous information networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017*. ACM, 2017, pp. 635–644. [Online]. Available: https://doi.org/10.1145/3097983.3098063

[4] M. S. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, ser. Lecture Notes in Computer Science, A. Gangemi, R. Navigli, M. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, and M. Alam, Eds., vol. 10843. Springer, 2018, pp. 593–607. [Online]. Available: https://doi.org/10.1007/978-3-319-93417-4_38

[5] C. Gao, X. He, D. Gan, X. Chen, F. Feng, Y. Li, T.-S. Chua, L. Yao, Y. Song, and D. Jin, "Learning to recommend with multiple cascading behaviors," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2019.

[6] C. Chen, M. Zhang, Y. Zhang, W. Ma, Y. Liu, and S. Ma, "Efficient heterogeneous collaborative filtering without negative sampling for recommendation," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 19–26. [Online]. Available: https://aaai.org/ojs/index.php/AAAI/article/view/5329

[7] H. Yin, H. Chen, X. Sun, H. Wang, Y. Wang, and Q. V. H. Nguyen, "SPTF: A scalable probabilistic tensor factorization model for semantic-aware behavior prediction," in *2017 IEEE International Conference on Data Mining, ICDM 2017, New Orleans, LA, USA, November 18-21, 2017*, V. Raghavan, S. Aluru, G. Karypis, L. Miele, and X. Wu, Eds. IEEE Computer Society, 2017, pp. 585–594. [Online]. Available: https://doi.org/10.1109/ICDM.2017.68

[8] H. Chen, Y. Li, X. Sun, G. Xu, and H. Yin, "Temporal meta-path guided explainable recommendation," in *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, L. Lewin-Eytan, D. Carmel, E. Yom-Tov, E. Agichtein, and E. Gabrilovich, Eds. ACM, 2021, pp. 1056–1064. [Online]. Available: https://doi.org/10.1145/3437963.3441762

[9] Y. Sun and J. Han, "Mining heterogeneous information networks: Principles and methodologies," *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 3, no. 2, pp. 1–159, 2012.

[10] A. J. Bose, A. Jain, P. Molino, and W. L. Hamilton, "Meta-graph: Few shot link prediction via meta learning," *CoRR*, vol. abs/1912.09867, 2019. [Online]. Available: http://arxiv.org/abs/1912.09867

[11] X. Sun, H. Yin, B. Liu, H. Chen, J. Cao, Y. Shao, and N. Q. V. Hung, "Heterogeneous hypergraph embedding for graph classification," in *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, L. Lewin-Eytan, D. Carmel, E. Yom-Tov, E. Agichtein, and E. Gabrilovich, Eds. ACM, 2021, pp. 725–733. [Online]. Available: https://doi.org/10.1145/3437963.3441835

[12] X. Sun, H. Yin, B. Liu, H. Chen, Q. Meng, W. Han, and J. Cao, "Multi-level hyperedge distillation for social linking prediction on sparsely observed networks," in *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, Eds. ACM / IW3C2, 2021, pp. 2934–2945. [Online]. Available: https://doi.org/10.1145/3442381.3449912

[13] Y. Li, H. Chen, X. Sun, Z. Sun, L. Li, L. Cui, P. S. Yu, and G. Xu, "Hyperbolic hypergraphs for sequential recommendation," *CoRR*, vol. abs/2108.08134, 2021. [Online]. Available: https://arxiv.org/abs/2108.08134

[14] Y. Luo, Z. Huang, H. Chen, Y. Yang, and M. Baktashmotlagh, "Interpretable signed link prediction with signed infomax hyperbolic graph," *CoRR*, vol. abs/2011.12517, 2020. [Online]. Available: https://arxiv.org/abs/2011.12517

[15] C. Yang, Y. Xiao, Y. Zhang, Y. Sun, and J. Han, "Heterogeneous network representation learning: Survey, benchmark, evaluation, and beyond," *CoRR*, vol. abs/2004.00216, 2020. [Online]. Available: https://arxiv.org/abs/2004.00216

[16] J. Qiu, Q. Chen, Y. Dong, J. Zhang, H. Yang, M. Ding, K. Wang, and J. Tang, "Gcc: Graph contrastive coding for graph neural network pre-training," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1150–1160. [Online]. Available: https://doi.org/10.1145/3394486.3403168

[17] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020,*

795

December 6-12, 2020, virtual, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., 2020. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/3fe230348e9a12c13120749e3f9fa4cd-Abstract.html

[18] Y. Dong, N. V. Chawla, and A. Swami, "Metapath2vec: Scalable representation learning for heterogeneous networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 135–144. [Online]. Available: https://doi.org/10.1145/3097983.3098036

[19] H. Jiang, Y. Song, C. Wang, M. Zhang, and Y. Sun, "Semi-supervised learning over heterogeneous information networks by ensemble of meta-graph guided random walks," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 2017, pp. 1944–1950. [Online]. Available: https://doi.org/10.24963/ijcai.2017/270

[20] C. Shi, X. Kong, Y. Huang, P. S. Yu, and B. Wu, "Hetesim: A general framework for relevance measure in heterogeneous networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2479–2492, 2014.

[21] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "Lightgcn: Simplifying and powering graph convolution network for recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 639–648. [Online]. Available: https://doi.org/10.1145/3397271.3401063

[22] A. Pfadler, H. Zhao, J. Wang, L. Wang, P. Huang, and D. L. Lee, "Billion-scale recommendation with heterogeneous side information at taobao," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, 2020, pp. 1667–1676.

[23] J. Wang, K. Ding, L. Hong, H. Liu, and J. Caverlee, "Next-item recommendation with sequential hypergraphs," in *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, J. Huang, Y. Chang, X. Cheng, J. Kamps, V. Murdock, J. Wen, and Y. Liu, Eds. ACM, 2020, pp. 1101–1110. [Online]. Available: https://doi.org/10.1145/3397271.3401133

[24] J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, L. Zhang, and X. He, "Music recommendation by unified hypergraph: combining social media information and music content," in *Proceedings of the 18th International Conference on Multimedia 2010, Firenze, Italy, October 25-29, 2010*, A. D. Bimbo, S. Chang, and A. W. M. Smeulders, Eds. ACM, 2010, pp. 391–400. [Online]. Available: https://doi.org/10.1145/1873951.1874005

[25] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. [Online]. Available: https://openreview.net/forum?id=SJU4ayYgl

[26] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. [Online]. Available: https://openreview.net/forum?id=rJXMpikCZ

[27] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. [Online]. Available: https://openreview.net/forum?id=ryGs6iA5Km

[28] J. Du, S. Zhang, G. Wu, J. M. F. Moura, and S. Kar, "Topology adaptive graph convolutional networks," *CoRR*, vol. abs/1710.10370, 2017. [Online]. Available: http://arxiv.org/abs/1710.10370

[29] C. Shi, B. Hu, W. X. Zhao, and P. S. Yu, "Heterogeneous information network embedding for recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 2, pp. 357–370, 2019. [Online]. Available: https://doi.org/10.1109/TKDE.2018.2833443

[30] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang, "Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, W. Zhu, D. Tao, X. Cheng, P. Cui, E. A. Rundensteiner, D. Carmel, Q. He, and J. X. Yu, Eds. ACM, 2019, pp. 1441–1450. [Online]. Available: https://doi.org/10.1145/3357384.3357895

[31] S. Zhang, H. Chen, X. Ming, L. Cui, H. Yin, and G. Xu, "Where are we in embedding spaces? A comprehensive analysis on network embedding approaches for recommender systems," *CoRR*, vol. abs/2105.08908, 2021. [Online]. Available: https://arxiv.org/abs/2105.08908

[32] F. Wu, A. H. S. Jr., T. Zhang, C. Fifty, T. Yu, and K. Q. Weinberger, "Simplifying graph convolutional networks," in *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 2019, pp. 6861–6871. [Online]. Available: http://proceedings.mlr.press/v97/wu19e.html

[33] Z. Cui, H. Chen, L. Cui, S. Liu, X. Liu, G. Xu, and H. Yin, "Reinforced kgs reasoning for explainable sequential recommendation," *World Wide Web*, 06 2021.

[34] A. Li, B. Yang, H. Chen, and G. Xu, "Hyperbolic neural collaborative recommender," *CoRR*, vol. abs/2104.07414, 2021. [Online]. Available: https://arxiv.org/abs/2104.07414