Improving Knowledge-aware Recommendation with Multi-level Interactive Contrastive Learning

Anonymous Author(s)

ABSTRACT

Incorporating Knowledge Graphs (KG) into recommeder system as side information has attracted considerable attention. Recently, the technical trend of Knowledge-aware Recommendation (KGR) is to develop end-to-end models based on graph neural networks (GNNs). However, the extremely sparse user-item interactions significantly degrade the performance of the GNN-based models, from the following aspects: 1) the sparse interaction, itself, means inadequate supervision signals and limits the supervised GNN-based models; 2) the combination of sparse interactions (CF part) and redundant KG facts (KG part) further results in an unbalanced information utilization. Besides, the GNN paradigm aggregates local neighbors for node representation learning, while ignoring the non-local KG facts and making the knowledge extraction insufficient. Inspired by the recent success of contrastive learning in mining supervised signals from data itself, in this paper, we focus on exploring contrastive learning in KGR and propose a novel multi-level interactive contrastive learning mechanism, to alleviate the aforementioned challenges. Different from traditional contrastive learning methods which contrast nodes of two generated graph views, interactive contrastive mechanism conducts layer-wise self-supervised learning by contrasting layers of different parts within graphs, which is also an "interaction" action. Specifically, we first construct local and non-local graphs for user/item in KG, exploring more KG facts for KGR. Then an intra-graph level interactive contrastive learning is performed within each local/non-local graph, which contrasts layers of the CF and KG parts, for more consistent information leveraging. Besides, an inter-graph level interactive contrastive learning is performed between the local and non-local graphs, for sufficiently and coherently extracting non-local KG signals. Extensive experiments conducted on three benchmark datasets show the superior performance of our proposed method over the state-of-the-arts.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

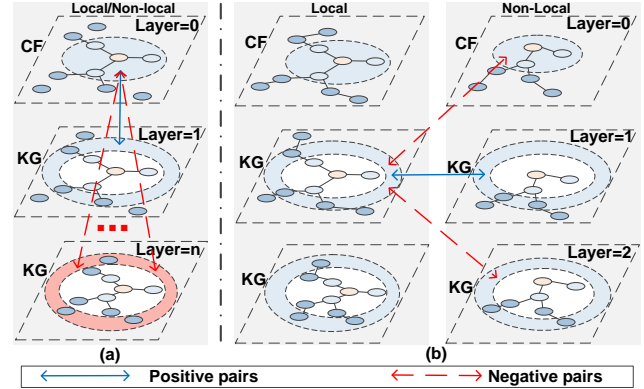Knowledge Graph, Recommendation, Contrastive Learning

Figure 1: A toy example of multi-level contrastive mechanism. (a) Our proposed intra-graph level interactive contrastive mechanism; (b) Our proposed inter-graph level interactive contrastive mechanism (Only taking Layer=1 as the anchor here for clarity).

1 INTRODUCTION

Recommender systems (RS) aim to alleviate the information explosion, proposing to recommend a small set of items to meet users' personalized interests. As an effective solution, collaborative filtering [12, 13, 19] presumes that behaviorally similar users have a similar preference on items and achieves great success. However, they severely suffer from the cold-start problem, due to treating each interaction as an independent instance while neglecting their relations, such as NFM [13], xDeepFM [19]. A commonly-adopted solution is to integrate side information with collaborative filtering models, such as a knowledge graph (KG), which contains rich facts and connections about items, to learn high-quality user and item representations for recommendation (*aka.* knowledge-aware recommendation, KGR).

Indeed, there already exists much research effort [31, 47, 48] devoted to KGR, the theme of which is how to **sufficiently and coherently** leverage the graph information of CF (*i.e.,* user-item interactions) and KG (*i.e.,* item-entity affiliations). Earlier studies [15, 31, 47] on KGR explore CF and KG in a relatively independent manner, which exploits KG information to enhance item representation in CF. They employ different *knowledge graph embedding* (KGE) models (*e.g.,* TransE [2], TransH [41]) to pre-train entity embeddings and treat them as prior or content information. However, these methods fall short in distilling sufficient knowledge signals from KG, since they treat each item-entity relation independently. Hence, some follow-on studies [14, 26, 39] focus on extracting more

sufficient KG signals to enrich CF (user-item interactions), via capturing the long-range KG connectivity, such as selecting prominent paths over KG [27] *or* representing the interactions with multi-hop paths from users to items [14, 39]. Nevertheless, most of these methods heavily rely on manually designed meta-paths, thus are hard to optimize in reality. More recent works [25, 34, 36, 37] further unify CF and KG as a heterogeneous graph (*i.e.,* user-item-entity graph), with an informative aggregation paradigm (*i.e.,* Graph Neural Networks, GNNs) to perform graph representation learning. Due to the powerful capability of GNNs in effectively generating local permutation-invariant aggregation on the neighbors of a node, these methods achieve promising performance and become the mainstream tendency in KGR.

Despite effectiveness, existing GNN-based methods still fall short in achieving the above core goal in the following three aspects:

- **Sparse Supervised Signals.** Since established in a supervised manner, existing GNN-based methods rely on the observed user-item interactions as supervision signals to perform graph representation learning on the unified heterogeneous graph. However, the user-item interactions are actually extremely sparse in real scenarios [1, 42], which makes it insufficient to achieve satisfactory performance and even results in terrible side effects, *e.g.,* degeneration problem [7] (*i.e.,* degenerating node embeddings distribution into a narrow cone, even leading to the indiscrimination of generated node representations).
- **Unbalanced Information Utilization.** When the sparse user-item interactions meet redundant knowledge facts, there occurs an unbalanced heterogeneous structure in KGR, which results in an unbalanced information utilization problem. A commonly approved fact is that, it's the CF signals that determine the user's preferences, since they are composed of user's historical interactions. But with the unbalanced information utilization, the noisy KG information, however, gets emphasized more in final user/item modeling, which makes the crucial explicit CF signals less stressed and further results in suboptimal representation learning.
- **Insufficient Knowledge Extraction.** Although there are redundant KG facts, extracting informative and helpful knowledge is actually far from sufficient in previous GNN-based models, due to the local aggregation feature of GNN. The GNN-based methods usually learn item representations by aggregating neighboring entities on their local KG structures (*i.e.,* neighboring areas of the item itself), which ignores the non-local KG facts (*i.e.,* neighboring areas of similar items) and thus results in inadequate knowledge extraction. Nevertheless, simply aggregating the non-local KG facts may introduce more irrelevant noise, further resulting in an unexpected performance decrease.

Inspired by recent success in contrastive learning, one of the classical Self-Supervised Learning (SSL) methods, we naturally propose to leverage the superiority of SSL to alleviate the first problem of Sparse Supervised Signals. A straightforward idea is that we could augment (or corrupt) the input user-item-entity graph as a graph view, and contrast the nodes in it with the original one, following the traditional contrastive paradigm analogous to [5, 11, 18]. However, such a paradigm performs contrastive learning in a relatively independent manner, which only contrasts the same part (CF or KG

part) of different graph views, hence ignores the inner interactions of different parts within a graph. That is, the improvements of CF and KG representation learning are mutually isolated, and the impact of CF part in final user/item modeling is still limited. Hence with the traditional contrastive mechanism, the above second and third problems are still far from resolution or alleviation. As a result, it is a necessity to endow the contrastive learning paradigm with the capability of effective *information interaction* between CF and KG parts, so as to coherently exploit information of each part without reliance on extra explicit labels. This motivates us to design an interactive graph contrastive mechanism tailored for KGR task, which is required to conclude the following aspects so as to address the above limitations: 1) contrasting the CF and KG parts to balance their impact on representation learning, as shown in Figure 1 (a); 2) contrasting the local and non-local graphs in KG to extract informative non-local KG facts, as shown in Figure 1 (b).

In this paper, we develop a novel model, U̲nowledge-aware Recommender System with Multi-level I̲nteractive C̲ontrastive Learning (KGIC), to solve the aforementioned limitations and challenges. Conceptually, KGIC focuses on exploring a proper in-graph contrastive learning mechanism for knowledge-aware recommendation, aiming to unify the two crucial but relatively independent parts (*i.e.,* CF and KG) in a self-supervised manner. In particular, we first perform the interactive graph contrastive learning in **intra-graph level**, which contrasts the CF signals with the KG information in each local/non-local graph of user/item, aiming to increase the information consistency between CF and KG parts. Then an **inter-graph level** interactive graph contrastive learning is conducted, which contrasts the local KG facts with non-local ones, for integrating more KG signals and denoising the non-local KG information. Empirically, KGIC outperforms the state-of-the-art models on three benchmark datasets.

Our contributions of this work can be summarized as follows:

- **General Aspects:** We emphasize the importance of incorporating self-supervised learning to unify CF and KG information for knowledge-aware recommendation, which takes layer self-discrimination as a self-supervised task to offer auxiliary signals for coherent graph representation learning.
- **Novel Methodologies:** We propose a novel model KGIC, which designs a multi-level interactive contrastive mechanism for knowledge-aware recommendation. KGIC combines multi-order CF with KG to construct local and non-local graphs for fully exploring external knowledge. KGIC then performs intra-graph level and inter-graph level interactive contrastive learning among local and non-local graphs, for a sufficient and coherent information utilization in CF and KG.
- **Multifaceted Experiments:** We conduct extensive experiments on three benchmark datasets. The results demonstrate the advantages of our KGIC in better representation learning, which shows the effectiveness of our multi-level interactive contrastive learning for KGR.

2 RELATED WORK

2.1 Knowledge-aware Recommendation

2.1.1 Embedding-based methods. Embedding-based methods [4, 15, 29, 31, 33, 47, 48] pre-train the KG entity embeddings with

knowledge graph embeddings methods (KGE) [2, 20, 41], for enriching item representations. CKE [47] combines CF module with structural, textual, and visual knowledge embeddings of items in a unified Bayesian framework. KTUP [4] jointly model user preference and perform KG completion with the TransH [41] method on user-item interactions and KG triplets. RippleNet [30] propagates users' historical clicked items along links in KG, which makes more KG facts explored to reveal the users' potential interests. Embedding-based methods show high flexibility in utilizing KG, but the utilized KGE algorithms focus more on modeling rigorous semantic relatedness, which is more suitable for link prediction rather than recommendation.

2.1.2 Path-based methods. Path-based methods [14, 26, 39, 45, 46, 49] explore various patterns of connections among items in KG to provide additional guidance for the recommendation. PER [46] and meta-graph based recommendation [14] extract the meta-path/meta-graph latent features and exploit the connectivity between users and items along different types of relation paths/graphs. KPRN [39] further automatically extracts paths between users and items, modeling these paths with RNNs. Path-based methods mostly involve the design of meta-paths for generating meaningful connectivity patterns, which requires specific domain knowledge and labor-intensive human efforts for accurate path construction.

2.1.3 GNN-based methods. GNN-based methods [14, 39, 45, 46, 49] are founded on the information aggregation mechanism of graph neural networks (GNNs) [9, 17, 43]. Typically it integrates multi-hop neighbors into node representations to capture node feature and graph structure, which hence could model long-range connectivity. KGCN [34] and KGNN-LS [32] firstly utilize graph convolutional network (GCN) on KG to obtain item embeddings by aggregating items' neighborhood information iteratively. Later, KGAT [36] combines the user-item graph with a knowledge graph as a unified heterogeneous graph, then utilizes GCN to recursively perform aggregation on it. But CKAN [40] separately propagates collaborative signals and knowledge signals on the user-item graph and KG to stress the importance of CF signals. More recently, KGIN [37] models user-item interactions at an intent level, which reveals user intents behind the KG interactions and performs GNN on the user-intent-item-entity graph. And CG-KGR [6] exploits the pre-trained collaborative signals to guide the aggregation on KG, for sufficient knowledge extraction. However, all these approaches follow the supervised learning paradigm, hence suffering from the original sparse interactions. Moreover, most of them have an unbalanced utilization of user-item-entity graph, overstressing the importance of KG and ignoring the crucial effect of CF. Besides, insufficient knowledge extraction also exists for the local aggregation feature of GNN. In contrast, our work explores self-supervised learning in KGR, to overcome the limitations brought by the spare interactions and achieve sufficient knowledge extraction.

2.2 Contrastive Learning

Contrastive learning methods [28, 38, 42] learn node representations by contrasting positive pairs against negative pairs. DGI [28] first adopts Infomax [22] in graph representation learning, and contrasts the local node embeddings with global graph embeddings.

Then GMI [23] proposes to contrast the center node with its local neighbors on both node features and topological structure. Similarly, MVGRL [10] learns node-level and graph-level node representations from first-order neighbors and a graph diffusion, and contrasts encoded embeddings between two graph views. More recently, HeCo [38] proposes to learn node representations from network schema view and meta-path view, and performs contrastive learning between them. As for contrastive learning in the traditional CF-based recommendation domain, SGL [42] generates two graph views by corrupting the user-item interactions, and performs contrastive learning between them. And NCL [21] captures potential node (structural and semantic) relatedness into contrastive learning for a neighbor-enhanced contrasting. However, little effort has been made to investigate the great potential of contrastive learning on knowledge-aware recommendation.

3 PROBLEM FORMULATION

In this section, we first formulate structural data of CF and KG parts, *i.e.,* user-item interactions and knowledge graph, then present the problem statement of knowledge-aware recommendation.

Interaction Data. In a typical recommendation scenario, let $\mathcal{U} = \{u_1, u_2, \ldots, u_M\}$ be a set of M users and $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ a set of N items. Let $\mathbf{Y} \in \mathbf{R}^{M \times N}$ be the user-item interaction matrix, where $y_{uv} = 1$ indicates that user u engaged with item v, such as behaviors like clicking or purchasing; otherwise $y_{uv} = 0$.

Knowledge Graph. A KG stores luxuriant real-world facts associated with items, *e.g.,* item attributes, or external commonsense knowledge, in the form of a heterogeneous graph [26]. Let $\mathcal{G} = \{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}\}$ be the knowledge graph, where h, r, t are on behalf of head, relation, tail of a knowledge triple correspondingly; $\mathcal{E}$ and $\mathcal{R}$ refer to the sets of entities and relations in $\mathcal{G}$. For example, the triple (Batman Begins, film.film.star, Christian Bale) means that Christian Bale is an actor of the movie Batman Begins. In many recommendation scenarios, an item $v \in \mathcal{V}$ corresponds to one entity $e \in \mathcal{E}$. So we establish a set of item-entity alignments $\mathcal{A} = \{(v, e) | v \in \mathcal{V}, e \in \mathcal{E}\}$, where (v, e) indicates that item v can be aligned with an entity e in the KG. With the alignments between items and KG entities, KG is able to profile items and offer complementary information to the interaction data.

Problem Statement. Given the user-item interaction matrix $\mathbf{Y}$ and the knowledge graph $\mathcal{G}$, knowledge-aware recommendation task aims to learn a function that can predict how likely a user would adopt an item.

4 METHODOLOGY

We now present the proposed <u>K</u>nowledge-aware Recommender System with Multi-level <u>I</u>nteractive <u>C</u>ontrastive Learning (KGIC). KGIC aims at unifying the CF and KG parts with an interactive contrastive mechanism for coherent information utilization and hence improving the user/item representation learning. Figure 2 displays the working flow of KGIC, which mainly consists of three key components: 1) Graph Constructing and Encoding. It first constructs local and non-local graphs for user/item by combining different CF signals with KG facts, then encodes each layer with a simple GNN encoder in all the graphs. 2) Intra-graph Interactive Contrastive Learning. It performs interactive contrastive Learning within local
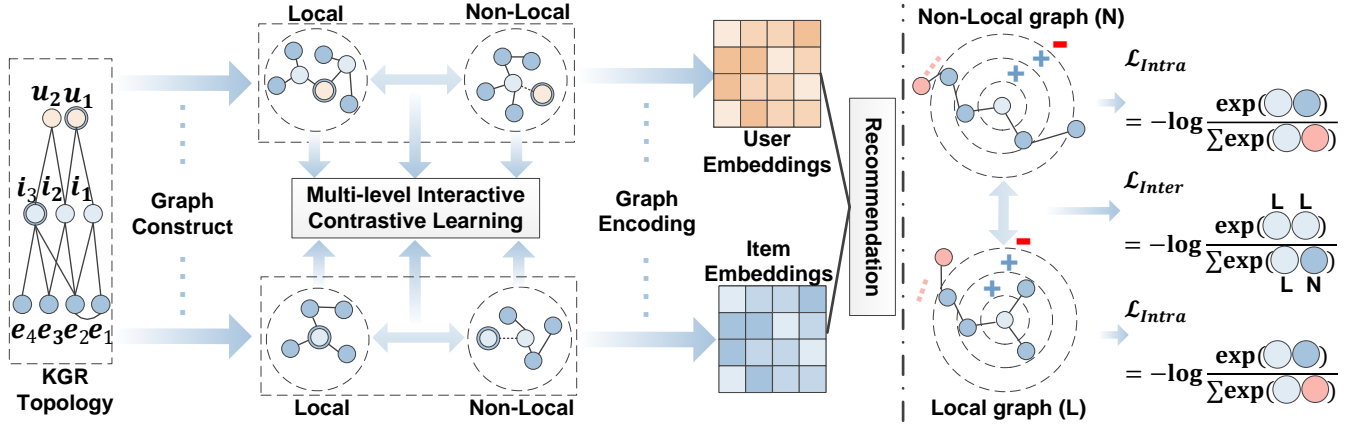
Figure 2: Illustration of the proposed KGIC model. The left subfigure shows model framework of KGIC; and the right subfigure presents the details of Multi-level Interactive Contrastive Learning. Best viewed in color.

and non-local graphs, for unifying the CF and KG information, and further having a consistent and sufficient representation learning. 3) Inter-graph Interactive Contrastive Learning. It performs interactive contrastive Learning between local and non-local graphs, aiming to extract more informative KG signals. We next present the three components in detail.

4.1 Graph Constructing and Encoding

Different from previous KG-aware recommendation methods that only integrate KG entities of local neighboring areas into user/item representation learning, we propose to incorporate the non-local KG information which could be acquired from neighboring KG entities of similar items (*i.e.,* co-occurrence items in CF), aiming to fully explore the external facts in KG. We first construct local and non-local graphs (*i.e.,* the local and non-local sub-KGs) for user/item by combining corresponding CF signals with KG. Then an attentive embedding mechanism is adopted for layer encoding in each graph.

4.1.1 Local Graph Construction. Local graph consists of first-order CF (*i.e.,* the user's interacted items or the item itself) and related KG facts for user/item. Firstly, first-order CF signals for user/item are extracted from the user-item interactions Y. Then through the item-entity alignments $\mathcal{A} = \{(v, e)|v \in \mathcal{V}, e \in \mathcal{E}\}$, the items of first-order CF signals are aligned with KG and the initial entities in KG are acquired as follows:

$$\begin{aligned}\mathcal{E}_{u,L}^0 &= \{e \mid (v, e) \in \mathcal{A}, \text{ and } v \in \{v \mid y_{uv} = 1\}\}, \\ \mathcal{E}_{v,L}^0 &= \{e \mid (v, e) \in \mathcal{A}\},\end{aligned} \quad (1)$$

where $\mathcal{E}_{u,L}^0$ and $\mathcal{E}_{v,L}^0$ represents the initial entity sets of KG in local graphs, for the user and the item respectively. After that, more layers' related KG facts are acquired through a natural propagation in KG (*i.e.,* propagating along links in KG). And by doing so, the local graphs for user/item are constructed. The triples in local graphs are obtained as follows:

$$\mathcal{S}_{o,L}^l = \left\{(h, r, t) \mid (h, r, t) \in \mathcal{G} \text{ and } h \in \mathcal{E}_{o,L}^{l-1}\right\}, l = 1, \dots, L, \quad (2)$$

where the symbol o is a uniform placeholder which means u or v, $\mathcal{S}_{u,L}^l$ and $\mathcal{S}_{v,L}^l$ represents the triple set in the local graph's l−th layer for user and item, which is composed of $(l-1)$−th layer head entity, relation and l−th layer tail entity. In this way, we obtain a L−layer local graph for user/item, which contains user-item-entity and item-entity heterogeneous structure for user and item respectively.

4.1.2 Non-Local Graph Construction. Non-local graph contains high-order CF signals (*i.e.,* item-user-item co-occurrence items in CF) and more external KG facts for user/item. Similarly, the non-local graph is constructed by aligning high-order items of CF with KG and KG propagating. Firstly, the high-order items for user/item are acquired through propagating in user-item interactions, as follows:

$$\begin{aligned}\mathcal{V}_p &= \left\{v_p \mid u \in \mathcal{U}_{\text{sim}}, \text{ and } y_{uv_p} = 1\right\}, \\ \mathcal{U}_{\text{sim}} &= \left\{u_{\text{sim}} \mid v \in \{v \mid y_{uv} = 1\} \text{ and } y_{u_{\text{sim}} v} = 1\right\}, \\ \mathcal{V}_u &= \left\{v_u \mid u \in \{u \mid y_{uv} = 1\} \text{ and } y_{uv_u} = 1\right\},\end{aligned} \quad (3)$$

where $\mathcal{V}_p$ and $\mathcal{U}_{\text{sim}}$ represent high-order items and similar users of the user, and $\mathcal{V}_u$ is the high-order items of the item. Further the aligned initial entities $\mathcal{E}_{u,N}^0$ and $\mathcal{E}_{v,N}^0$ in KG are acquired as follows:

$$\begin{aligned}\mathcal{E}_{u,N}^0 &= \left\{e \mid (v_p, e) \in \mathcal{A}, \text{ and } v_p \in \mathcal{V}_p\right\}, \\ \mathcal{E}_{v,N}^0 &= \left\{e \mid (v_u, e) \in \mathcal{A}, \text{ and } v_u \in \mathcal{V}_u\right\}.\end{aligned} \quad (4)$$

Then by propagating these initial entities in KG, the non-local graph for user/item is constructed, whose triples are formed as follows:

$$\mathcal{S}_{o,N}^l = \left\{(h, r, t) \mid (h, r, t) \in \mathcal{G} \text{ and } h \in \mathcal{E}_{o,N}^{l-1}\right\}, l = 1, \dots, L. \quad (5)$$

4.1.3 Graph Encoding. Obtaining local and non-local graphs of user/item, it's indispensable to encode each layer information into an embedding for learning a comprehensive representation. Inspired by previous KG-aware recommendation works, an attentive embedding mechanism is adopted here, which reveals the different meanings of the tail entities with different contexts and generates different attentive weights. Considering (h, r, t) the i-th triple of

the l-th layer triple set, we can get the representation of the l-th layer as follows:

$$\mathbf{E}_{o,D}^{l} = \sum_{i=1}^{m} \pi\left(e_i^h, r_i\right) e_i^t, \tag{6}$$

where the symbol D is a uniform placeholder which means L or N, $\mathbf{E}_{o,D}^{l}$ represents l-th layer's embedding of u or v in local/non-local graph. And m is the number of triples in the l-th layer, e_i^h, r_i and e_i^t are the embeddings of head entity, relation and the tail entity of the i-th triple. The weights $\pi(e_i^h, r_i)$ are acquired by the attentive mechanism which can be described as follows:

$$\pi\left(e_i^h, r_i\right) = \sigma\left(W_1\left[\sigma\left(W_0\left(e_i^h || r_i\right) + b_0\right)\right] + b_1\right),$$
$$\pi\left(e_i^h, r_i\right) = \frac{\exp\left(\pi\left(e_i^h, r_i\right)\right)}{\sum_{(h',r',t')\in\mathcal{S}_{o,D}^l} \exp\left(\pi\left(e_i^{h'}, r_i'\right)\right)}, \tag{7}$$

where $||$ is the concatenation operation, $W_* \in \mathbf{R}^{2d\times d}$ and $b_* \in \mathbf{R}^d$ are the trainable weight matrices and biases. Hence we successfully encode each layer of local and non-local graphs into embeddings.

4.2 Intra-graph Interactive Contrastive Learning

Based on the constructed local and non-local graphs of user/item, we move on to encourage a balanced information utilization in these graphs. Since the local and non-local graphs have a relatively unbalanced heterogeneous structure, which is composed of sparse user-item interactions and redundant KG connections, thus the crucial CF signals tend to have less impact on representation learning. The intra-graph interactive contrastive learning is hence proposed to make coherent use of CF and KG, by performing interactions between CF and KG information with contrastive learning.

Specifically, the intra-graph interactive contrastive learning regards the CF part (*i.e.,* the initial entities in Equation (1)) as the anchor, which is located in center layer of the local/non-local graph. The knowledge information involved in aggregation for learning user/item representation forms the positive pairs, and the other KG entities (*i.e.,* the unemployed higher layers in graphs) form the negative pairs. With the defined positive and negative pairs, we have the following contrastive loss for the user:

$$\mathcal{L}_{Intra}^{U} = \sum_{u\in\mathcal{U}} -\log \frac{\sum_{k\in L} e^{\left(\left(\mathbf{E}_{u,L}^{(0)}\cdot\mathbf{E}_{u,L}^{(k)}/\tau\right)\right)}}{\underbrace{\sum_{k\in L} e^{\left(\left(\mathbf{E}_{u,L}^{(0)}\cdot\mathbf{E}_{u,L}^{(k)}/\tau\right)\right)}}_{\text{positive pair}} + \underbrace{\sum_{k'>L} e^{\left(\left(\mathbf{E}_{u,L}^{(0)}\cdot\mathbf{E}_{u,L}^{(k')}/\tau\right)\right)}}_{\text{intra-graph negative pair}}}$$
$$+ \sum_{u\in\mathcal{U}} -\log \frac{\sum_{k\in L} e^{\left(\left(\mathbf{E}_{u,N}^{(0)}\cdot\mathbf{E}_{u,N}^{(k)}/\tau\right)\right)}}{\underbrace{\sum_{k\in L} e^{\left(\left(\mathbf{E}_{u,N}^{(0)}\cdot\mathbf{E}_{u,N}^{(k)}/\tau\right)\right)}}_{\text{positive pair}} + \underbrace{\sum_{k'>L} e^{\left(\left(\mathbf{E}_{u,N}^{(0)}\cdot\mathbf{E}_{u,N}^{(k')}/\tau\right)\right)}}_{\text{intra-graph negative pair}}}, \tag{8}$$

where τ denotes a temperature parameter in softmax. In a similar way, the intra-graph contrastive loss of the item can be obtained as

$\mathcal{L}_{Intra}^{I}$. And the complete Intra-graph interactive contrastive Loss is the sum of the above two losses:

$$\mathcal{L}_{Intra} = \mathcal{L}_{Intra}^{U} + \mathcal{L}_{Intra}^{I}. \tag{9}$$

In this way, the CF and KG signals successfully supervise each other to have a more coherent and sufficient representation learning.

4.3 Inter-graph Interactive Contrastive Learning

Although intra-graph interactive contrastive learning has achieved coherent information utilization in each single graph, it's still a challenge to integrate the local and non-local information together for that the non-local one is noisier. Since the non-local graph is composed of high-order CF signals and its corresponding KG facts, more external useful external facts along with noisy information are included. Hence the inter-graph interactive contrastive learning is proposed, to extract informative non-local information by contrasting the non-local graph with the cleaner local graph and collaboratively supervising each other.

To be more specific, the inter-graph interactive contrastive learning treats any layers of the local graph as the anchor, and the same layer in the non-local graph forms the positive pair, and the other layers in the non-local graph are naturally regarded as the negative pairs. Similar to intra-graph interactive contrastive mechanism, we have the following contrastive loss for the user:

$$\mathcal{L}_{Inter}^{U} = \sum_{u\in\mathcal{U}} \sum_{k\in L} -\log \frac{e^{\left(\left(\mathbf{E}_{u,L}^{(k)}\cdot\mathbf{E}_{u,N}^{(k)}/\tau\right)\right)}}{\underbrace{e^{\left(\left(\mathbf{E}_{u,L}^{(k)}\cdot\mathbf{E}_{u,N}^{(k)}/\tau\right)\right)}}_{\text{positive pair}} + \underbrace{\sum_{k'\neq k} e^{\left(\left(\mathbf{E}_{u,L}^{(k)}\cdot\mathbf{E}_{u,N}^{(k')}/\tau\right)\right)}}_{\text{inter-graph negative pair}}}. \tag{10}$$

Similarly, we could obtain the inter-graph contrastive loss of the item $\mathcal{L}_{Inter}^{I}$. Then the complete inter-graph contrastive loss is the sum of the user and item contrastive loss:

$$\mathcal{L}_{Inter} = \mathcal{L}_{Inter}^{U} + \mathcal{L}_{Inter}^{I}. \tag{11}$$

4.4 Model Prediction

Having updated all the layers' embeddings, we obtain every layer's embedding for user's local and non-local graph, item's local and non-local graph. Every layer's embedding expresses a part of user/item representation and stresses the influence of different layers and components. By concatenating these representation vectors of each layer, final user/item embedding is concluded for predicting their matching score through inner product, as follows:

$$\mathbf{e}_u = \mathbf{E}_{u,L}^{0} || \dots || \mathbf{E}_{u,L}^{L} || \mathbf{E}_{u,N}^{0} || \dots || \mathbf{E}_{u,N}^{L},$$
$$\mathbf{e}_i = \mathbf{E}_{i,L}^{0} || \dots || \mathbf{E}_{i,L}^{L} || \mathbf{E}_{i,N}^{0} || \dots || \mathbf{E}_{i,N}^{L}, \tag{12}$$
$$\hat{\mathbf{y}}(u,i) = \mathbf{e}_u^{\top} \mathbf{e}_i.$$

4.5 Multi-task Training

To combine the recommendation task with the self-supervised task, we treat the proposed two contrastive learning losses as supplementary. Hence a multi-task learning strategy is leveraged to jointly

		Book-Crossing	MovieLens-1M	Last.FM
User-item Interaction	# users	17,860	6,036	1,872
	# items	14,967	2,445	3,846
	# interactions	139,746	753,772	42,346
Knowledge Graph	# entities	77,903	182,011	9,366
	# relations	25	12	60
	# triplets	151,500	1,241,996	15,518
Hyper-parameter Settings	# η	4×10^{-3}	4×10^{-3}	4×10^{-3}
	# $\lambda 1$	1×10^{-6}	1×10^{-7}	1×10^{-6}
	# $\lambda 2$	1×10^{-4}	1×10^{-5}	1×10^{-4}

Table 1: Statistics and hyper-parameter settings for the three datasets.(η: learning rate, $\lambda 1$: constrastive loss weight, $\lambda 2$: L2 regularizer weight.)

train the KG-aware recommendation loss and the proposed contrastive loss. For the KG-aware recommendation task, a pairwise BPR loss [24] is adopted to reconstruct the historical data, which encourages the prediction scores of a user's historical items to be higher than the unobserved items.

$$\mathcal{L}_{\text{BPR}} = \sum_{(u,i,j) \in O} -\ln \sigma \left(\hat{y}_{ui} - \hat{y}_{uj} \right), \quad (13)$$

where $O = \left\{ (u, i, j) \mid (u, i) \in O^+, (u, j) \in O^- \right\}$ is the training dataset consisting of the observed interactions O^+ and unobserved counterparts O^-; σ is the sigmoid function. By combining the intra- and inter-graph interactive contrastive loss with BPR loss, we minimize the following objective function to learn the model parameter:

$$\mathcal{L}_{KGIC} = \mathcal{L}_{\text{BPR}} + \lambda 1(\alpha \mathcal{L}_{Intra} + \mathcal{L}_{Inter}) + \lambda 2 \|\Theta\|_2^2, \quad (14)$$

where Θ is the model parameter set, α is the hyperparameter to balance the weight of the intra- and inter-graph contrastive losses, $\lambda 1$ and $\lambda 2$ are two hyperparameters to control the contrastive loss and L_2 regularization term, respectively.

5 EXPERIMENT

Extensive experiments have been done on three public datasets, for answering the following research questions:

- **RQ1**: How does KGIC perform, compared with the state-of-the-art models?
- **RQ2**: How do the main components (*e.g.,* intra- and inter-graph interactive contrastive learning, non-local graph) affect KGIC performance?
- **RQ3**: How do different hyper-parameter settings (*e.g.,* model depth, coefficient α, temperature τ) affect KGIC?
- **RQ4**: Is the self-supervised task really improving the representation learning quality?

5.1 Experiment Settings

5.1.1 Dataset Description. Three publicly available datasets are used to evaluate the effectiveness of KGIC: Book-Crossing, MovieLens-1M, and Last.FM, which vary in size and sparsity, making our experiments more convincing. The basic statistics of the three datasets are presented in Table 1.

- Book-Crossing[1]: It consists of trenchant ratings (ranging from 0 to 10) about various books, which is collected from the book-crossing community.
- MovieLens-1M[2]: It contains approximately 1 million explicit ratings (ranging from 1 to 5) for movie recommendations.
- Last.FM[3]: It is collected from Last.FM online music systems, containing listening history with around 2 thousand users.

Note that we follow RippleNet [30] to transform the explicit feedback in three datasets into the implicit one where 1 indicates the positive samples (the positive threshold of the rating is 4 for MovieLens-1M, but no threshold is set for Last.FM and Book-Crossing due to their sparsity). For negative samples, we randomly sample unobserved items with the same size as positive ones for each user.

As for the sub-KG construction, we use Microsoft Satori[4] to construct it for the three datasets, closely following RippleNet [30] and KGCN[34]. Each sub knowledge graph follows the triple format and is a subset of the whole KG with a confidence level over 0.9. Given the sub-KG, we gather Satori IDs of all valid movies/books/musicians through matching their names with the tail of triples. Then we match the item IDs with the head of all triples and select all well-matched triples from the sub-KG.

5.1.2 Evaluation Metrics. To comprehensively evaluate our method, we conduct the evaluation in two experimental scenarios: (1) In click-through rate (CTR) prediction, two widely used metrics [30, 34] AUC and $F1$ are adopted here. (2) In top-K recommendation, we choose Recall@K to evaluate the recommended sets, where K is set to 5, 10, 20, 50, and 100 for consistency.

5.1.3 Baselines. To demonstrate the effectiveness of our proposed KGIC, we compare KGIC with the state-of-the-art recommender system methods, covering: CF-based methods (BPRMF), embedding-based method (CKE, RippleNet), path-based method (PER), and GNN-based methods(KGCN, KGNN-LS, KGAT, CKAN, KGIN, CG-KGR) as follows:

- BPRMF [24]: It's a typical CF-based method that uses pairwise matrix factorization for implicit feedback optimized by the BPR loss.
- CKE [47]: It's an embedding-based method that combines structural, textual, and visual knowledge in one framework.
- RippleNet [30]: It's a classical embedding-based method which propagates users' preferences on the KG.
- PER [46]: It's a typical path-based method which extracts meta-path-based features to represent the connectivity between users and items.
- KGCN [34]: It's a GNN-based method which iteratively integrates neighboring information to enrich item embeddings.
- KGNN-LS [32]: It is a GNN-based model which enriches item embeddings with GNN and label smoothness regularization.
- KGAT [36]: It's a GNN-based method which iteratively integrates neighbors on user-item-entity graph with an attention mechanism to get user/item representations.

---

[1]http://www2.informatik.uni-freiburg.de/~cziegler/BX/
[2]https://grouplens.org/datasets/movielens/1m/
[3]https://grouplens.org/datasets/hetrec-2011/
[4]https://searchengineland.com/library/bing/bing-satori

- CKAN [40]: It's a GNN-based method which independently propagates collaborative and knowledge information on CF and KG parts respectively.
- KGIN [37]: It's a state-of-the-art GNN-based method, which disentangles user-item interactions at the granularity of user intents, and performs GNN on the proposed user-intent-item-entity graph.
- CG-KGR [6]: It's a state-of-the-art GNN-based method which fuses the encoded collaborative signals into knowledge aggregation with GNN.

5.1.4 Parameter Settings. We implement our KGIC and all baselines in Pytorch and carefully tune the key parameters. We fix the embedding size to 64 in all models for a fair comparison. The default Xavier method [8] is adopted to initialize the model parameters. Besides, we utilize Adam [16] optimizer to optimize our method and set the batch size to 2048. We employ a grid search to confirm the optimal settings: the learning rate η is tuned among $\{0.0004, 0.0008, 0.004, 0.008\}$, λ of $L2$ regularization term among $\{10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$. The local and non-local triple set size are limited to 40 and 128 respectively, for a balance of computational effectiveness and enough information. Other hyper-parameter settings are provided in Table 1. The best settings for hyper-parameters in all comparison methods are researched by either empirical study or following the original papers.

5.2 Performance Comparison (RQ1)

The empirical results of all methods in CTR prediction and top-K recommendation are reported in Table 2 and Figure 3, respectively. The improvements and statistical significance test are performed between KGIC and the strongest baselines (highlighted with underline). By analyzing the performance comparison, we have the following observations:

- **Our proposed KGIC achieves the best results.** KGIC yields the best performance across three datasets in terms of all measures. More specifically, KGIC improves *w.r.t. AUC* by 2.51%, 0.62%, and 1.06% over the strongest baselines in Book-Crossing, MovieLens-1M, and Last.FM datasets respectively. In top-K recommendation scenario, KGIC also performs best *w.r.t. Recall@K* (k = 5, 10, 20, 50, 100) in all the cases. We attribute such improvements to the following aspects: (1) By contrasting the CF with KG signals in local/non-local graphs, the intra-graph level interactive contrastive learning performs interactions between two parts and supervises each other to improve the representation learning. (2) Through contrasting the local and non-local graphs of user/item, inter-graph level interactive contrastive learning sufficiently incorporates the non-local KG facts and learns discriminative representations from the two kinds of graphs.
- **Most KG-aware models achieve better performance.** We can observe that models with KG information mostly perform better than conventional CF methods. Comparing CKE with BPRMF, simply incorporating KG embeddings into MF boosts the model performance. This confirms the importance of bringing in side information like KG.
- **Simply integrating KG is not a guarantee of performance improvement.** Traditional CF-based methods BPRMF

works slightly better than embedding-based method CKE and path-based method PER in some scenarios. This phenomenon reveals that the overly unbalanced utilization of KG would unexpectedly degrade the model performance, which stresses the importance of making sufficient and coherent use of KG.

- **Extracting more informative KG facts boosts the model performance** GNN-based methods have a better performance than embedding-based and path-based ones in most cases, which indicates the effectiveness of extracting long-range KG facts. This fact convinces that exploring more informative KG facts highly related to user/item would facilitate the representation, which motivates us to explore more knowledge entities ignoring the limitation of local areas.

5.3 Ablation Studies (RQ2)

Towards examining the contributions of main components in our model, we compare the KGIC with the following variants: (1) $KGIC_{w/o\ intra}$: In this variant, we disable the intra-graph interactive contrastive learning module. (2) $KGIC_{w/o\ inter}$: This variant removes inter-graph interactive contrastive learning module. (3) $KGIC_{w/o\ Non-local}$ This variant removes non-local graphs, hence the intra-graph constrastive loss of non-local graph and inter-graph contrastive loss are both disabled. The experimental results are reported in Figure 4, from which we could summarize the following observations:

- Compared with KGIC, $KGIC_{w/o\ intra}$ and $KGIC_{w/o\ inter}$ consistently get a worse performance, which means removing each of the contrastive loss leads to a performance decrease. This fact demonstrates the importance of exploiting interactive contrastive learning to improve user/item representation learning. Besides, it reveals that these two contrastive losses complement each other and improve the performance in different aspects.
- Removing Non-local graph significantly degrades the model's performance, and $KGIC_{w/o\ Non-local}$ is even the least competitive model. It makes sense since $KGIC_{w/o\ Non-local}$ could only consider the intra-graph contrastive loss in local graphs, which also verifies the importance of both intra- and inter-graph contrastive learning.
- In most cases, all of the three variants surpass the baseline models across the three datasets, which further demonstrates the effectiveness of intra- and inter-graph contrastive learning.

5.4 Sensitivity Analysis (RQ3)

5.4.1 Impact of model depth. The model depth L represents the aggregation layer in the local/non-local graph, and also represents the layers of positive pairs in the interactive contrastive mechanism. To study the influence of model depth, we vary L in range of $\{1, 2, 3\}$ and demonstrate the performance comparison on book, movie, and music datasets in Table 3. MCCLK performs best when $L = 1, 2, 2$, on Book, Movie, and Music respectively. We could observe that: (1) One or two layers are the proper distance for aggregating neighboring information in the local/non-local graph, further stacking more layers only introduces more noise. (2) One or two layers' positive

Model	Book-Crossing		MovieLens-1M		Last.FM	
	AUC	*F1*	*AUC*	*F1*	*AUC*	*F1*
BPRMF	$0.6583(-11.66\%)$	$0.6117(-6.95\%)$	$0.8920(-3.32\%)$	$0.7921(-6.38\%)$	$0.7563(-10.29\%)$	$0.7010(-7.43\%)$
CKE	$0.6759(-9.90\%)$	$0.6235(-5.77\%)$	$0.9065(-1.87\%)$	$0.8024(-5.35\%)$	$0.7471(-11.21\%)$	$0.6740(-10.13\%)$
RippleNet	$0.7211(-5.38\%)$	$0.6472(-3.40\%)$	$0.9190(-0.62\%)$	$0.8422(-1.37\%)$	$0.7762(-8.30\%)$	$0.7025(-7.28\%)$
PER	$0.6048(-17.01\%)$	$0.5726(-10.86\%)$	$0.7124(-21.28\%)$	$0.6670(-18.89\%)$	$0.6414(-21.78\%)$	$0.6033(-17.20\%)$
KGCN	$0.6841(-9.08\%)$	$0.6313(-4.99\%)$	$0.9090(-1.62\%)$	$0.8366(-1.93\%)$	$0.8027(-5.65\%)$	$0.7086(-6.67\%)$
KGNN-LS	$0.6762(-9.87\%)$	$0.6314(-4.98\%)$	$0.9140(-1.12\%)$	$0.8410(-1.49\%)$	$0.8052(-5.40\%)$	$0.7224(-5.29\%)$
KGAT	$0.7314(-4.35\%)$	$0.6544(-2.68\%)$	$0.9140(-1.12\%)$	$0.8440(-1.19\%)$	$0.8293(-2.99\%)$	$0.7424(-3.29\%)$
CKAN	$0.7420(-3.29\%)$	$0.6671(-1.41\%)$	$0.9082(-1.70\%)$	$0.8410(-1.49\%)$	$0.8418(-1.74\%)$	$0.7592(-1.61\%)$
KGIN	$0.7273(-4.76\%)$	$0.6614(-1.98\%)$	$\underline{0.9190}(-0.62\%)$	$\underline{0.8441}(-1.18\%)$	$\underline{0.8486}(-1.06\%)$	$\underline{0.7602}(-1.51\%)$
CG-KGR	$\underline{0.7498}(-2.51\%)$	$\underline{0.6689}(-1.23\%)$	$0.9110(-1.42\%)$	$0.8359(-2.00\%)$	$0.8336(-2.56\%)$	$0.7433(-3.20\%)$
KGIC	$\mathbf{0.7749^*}$	$\mathbf{0.6812^*}$	$\mathbf{0.9252^*}$	$\mathbf{0.8559^*}$	$\mathbf{0.8592^*}$	$\mathbf{0.7753^*}$

Table 2: The result of *AUC* and *F1* in CTR prediction. The best results are in boldface and the second best results are underlined. * denotes statistically significant improvement by unpaired two-sample *t*-test with $p < 0.001$.
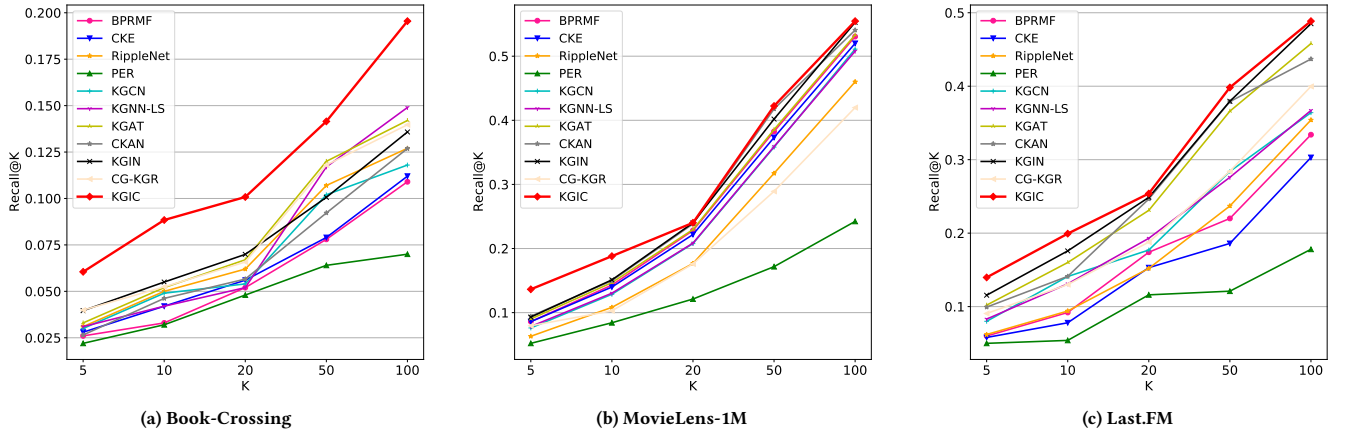

(a) Book-Crossing   (b) MovieLens-1M   (c) Last.FM

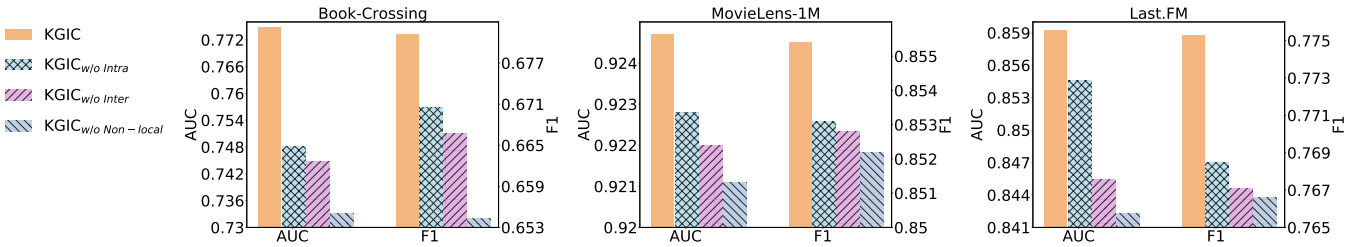**Figure 3: The result of Recall@K in top-K recommendation.**


Figure 4: Effect of ablation study.

pairs are enough for learning discriminative embeddings while performing interactions between layers of CF and KG.

5.4.2 Impact of intra-graph contrastive loss weight α. In the overall contrastive loss defined in Equation (14), the trade-off parameter α can balance the influence of intra- and inter-graph contrastive losses. To analyze the influence of coefficient α, we vary α in {0.1, 0.5, 1, 1.5, 2,} and report the results in Figure 5. From the results we could observe that: an appropriate α can effectively improve the

performance of contrastive learning. Specifically, the model performs best when $\alpha = 1$, which means there exists equal importance between the intra- and inter-graph contrastive learning. In addition, with different α, the performance of our KGIC is consistently better than other baselines, which also confirms the effectiveness of our multi-level interactive contrastive learning mechanism.

	Book		Movie		Music	
	Auc	F1	Auc	F1	Auc	F1
$L=1$	**0.7749**	**0.6812**	0.9241	0.8551	0.8482	0.7692
$L=2$	0.7689	0.6705	**0.9252**	**0.8559**	**0.8592**	**0.7753**
$L=3$	0.7513	0.6718	0.9203	0.8521	0.8511	0.7694

Table 3: Impact of model depth.

(a) Book   (b) Movie   (c) Music
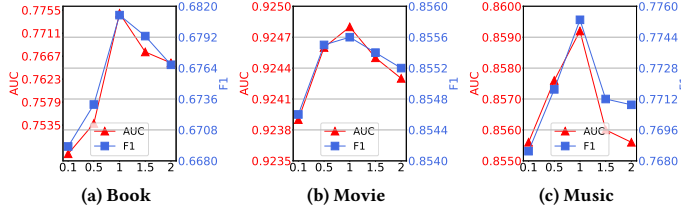
Figure 5: Impact of coefficient α.

(a) Book   (b) Music
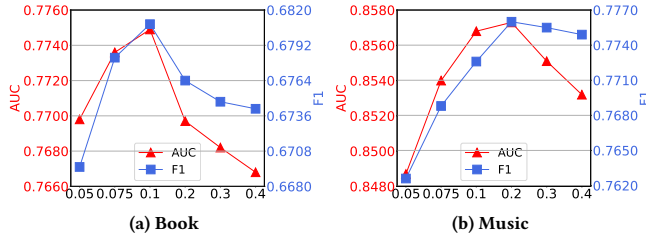
Figure 6: Impact of temperature τ

5.4.3   Impact of Temperature τ. As mentioned in previous contrastive learning work [42, 50], the temperature τ defined in Equation (8) and Equation (10) plays an important role in contrastive learning. To investigate the impact of τ, we vary it in range of {0.05, 0.075, 0.1, 0.2, 0.3, 0.4}. From the results shown in Figure 6, we can find that: a too large value of τ will cause poor performance, consistent with conclusions of previous works [42]. And generally, a temperature in the range of [0.1, 0.2] could lead to a satisfactory recommendation performance.

5.5   Visualization (RQ4)

Towards a more intuitive evaluation on how the proposed interactive contrastive mechanism affects the representation learning performance, we visualize the learned item embeddings in Figure 7. Following previous Contrastive learning work [21, 44], we plot item embeddings distributions with Gaussian kernel density estimation (KDE) [3] in two-dimensional space (where the darker the color is, the more points fall in that area), and we also plot KDE on angles (*i.e., $arctan2(y, x)$* for each point (x, y) in the above graph) towards a clearer presentation. As shown in Figure 7, we compare the visualized results of KGIC, KGIC$_{w/o\ contrast}$, CG-KGR, and CKAN on book and music, from which we can have the following observations:
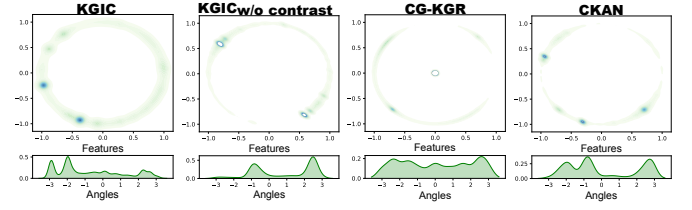
(a) Book
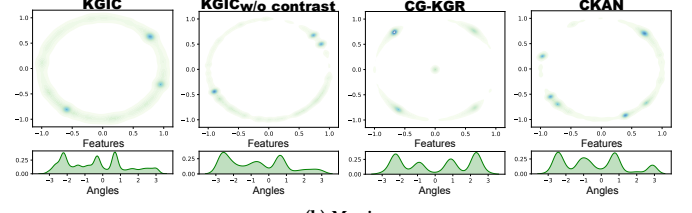
(b) Music

Figure 7: Visualization for the distribution of item embeddings.

- CG-KGR and CKAN show highly clustered features that mainly reside on some narrow arcs, while our KGIC clearly has a more uniform distribution and hence is able to present more different node feature information. As mentioned in previous work [35], in a proper range, a more uniform distribution means a better capacity to model the diversity of node features. This fact demonstrates the superiority of KGIC in better representation learning and alleviating the representation degeneration problem.
- By removing all of the contrastive loss in KGIC, the learned item embeddings are less uniform and fall into several coherent clusters. This phenomenon confirms the promotion of representation learning comes from the proposed interactive contrastive mechanism, for preserving maximal information on representations and improving the uniformity of the learned representations.

6   CONCLUSION

In this paper, we focus on incorporating contrastive learning into KG-aware recommendation, making sufficient and coherent use of CF and KG in a self-supervised manner. We propose a novel framework, KGIC, which achieves better user/item representation learning from two dimensions: (1) KGIC makes coherent utilization of CF and KG information in each local/non-local graph, by performing intra-graph interactive contrastive learning which contrasts layers of the CF and KG parts. (2) KGIC sufficiently extracts and integrates more KG facts for user/item representation learning, by constructing non-local graphs and performing inter-graph interactive contrastive learning which contrasts the local and non-local graphs. Extensive experiments on three public datasets demonstrate that KGIC significantly improves the recommendation performance over baselines on both Click-Through rate prediction and Top-K recommendation tasks.

REFERENCES

[1] Immanuel Bayer, Xiangnan He, Bhargav Kanagal, and Steffen Rendle. 2017. A generic coordinate descent framework for learning from implicit feedback. In *WWW*. 1341–1350.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Neural Information Processing Systems (NIPS)*. 1–9.

[3] Zdravko I Botev, Joseph F Grotowski, and Dirk P Kroese. 2010. Kernel density estimation via diffusion. *The annals of Statistics* (2010), 2916–2957.

[4] Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, and Tat-Seng Chua. 2019. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. In *WWW*. 151–161.

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *ICML*. 1597–1607.

[6] Yankai Chen, Yaming Yang, Yujing Wang, Jing Bai, Xiangchen Song, and Irwin King. 2022. Attentive Knowledge-aware Graph Convolutional Networks with Collaborative Guidance for Personalized Recommendation. In *ICDE*.

[7] Jun Gao, Di He, Xu Tan, Tao Qin, Liwei Wang, and Tie-Yan Liu. 2019. Representation degeneration problem in training natural language generation models. *arXiv preprint arXiv:1907.12009* (2019).

[8] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*. 249–256.

[9] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *NeurIPS*. 1025–1035.

[10] Kaveh Hassani and Amir Hosein Khasahmadi. 2020. Contrastive multi-view representation learning on graphs. In *ICML*. PMLR, 4116–4126.

[11] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *CVPR*. 9729–9738.

[12] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *SIGIR*. 639–648.

[13] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *WWW*. 173–182.

[14] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S Yu. 2018. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In *SIGKDD*. 1531–1540.

[15] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y Chang. 2018. Improving sequential recommendation with knowledge-enhanced memory networks. In *SIGIR*. 505–514.

[16] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[17] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[18] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942* (2019).

[19] Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, and Guangzhong Sun. 2018. xdeepfm: Combining explicit and implicit feature interactions for recommender systems. In *SIGKDD*. 1754–1763.

[20] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*.

[21] Zihan Lin, Changxin Tian, Yupeng Hou, and Wayne Xin Zhao. 2022. Improving Graph Collaborative Filtering with Neighborhood-enriched Contrastive Learning. In *WWW*.

[22] Ralph Linsker. 1988. Self-organization in a perceptual network. *Computer* (1988), 105–117.

[23] Zhen Peng, Wenbing Huang, Minnan Luo, Qinghua Zheng, Yu Rong, Tingyang Xu, and Junzhou Huang. 2020. Graph representation learning via graphical mutual information maximization. In *WWW*. 259–270.

[24] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618* (2012).

[25] Xiao Sha, Zhu Sun, and Jie Zhang. 2019. Attentive knowledge graph embedding for personalized recommendation. *arXiv preprint arXiv:1910.08288* (2019).

[26] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and S Yu Philip. 2018. Heterogeneous information network embedding for recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2018), 357–370.

[27] Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. 2018. Recurrent knowledge graph embedding for effective recommendation. In *RecSys*. 297–305.

[28] Petar Velickovic, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. 2019. Deep Graph Infomax. *ICLR (Poster)* (2019), 4.

[29] Hongwei Wang, Fuzheng Zhang, Min Hou, Xing Xie, Minyi Guo, and Qi Liu. 2018. Shine: Signed heterogeneous information network embedding for sentiment link prediction. In *WSDM*. 592–600.

[30] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In *CIKM*. 417–426.

[31] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep knowledge-aware network for news recommendation. In *WWW*. 1835–1844.

[32] Hongwei Wang, Fuzheng Zhang, Mengdi Zhang, Jure Leskovec, Miao Zhao, Wenjie Li, and Zhongyuan Wang. 2019. Knowledge-aware graph neural networks with label smoothness regularization for recommender systems. In *SIGKDD*. 968–977.

[33] Hongwei Wang, Fuzheng Zhang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2019. Multi-task feature learning for knowledge graph enhanced recommendation. In *WWW*. 2000–2010.

[34] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge graph convolutional networks for recommender systems. In *WWW*. 3307–3313.

[35] Tongzhou Wang and Phillip Isola. 2020. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*. PMLR, 9929–9939.

[36] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. Kgat: Knowledge graph attention network for recommendation. In *SIGKDD*. 950–958.

[37] Xiang Wang, Tinglin Huang, Dingxian Wang, Yancheng Yuan, Zhenguang Liu, Xiangnan He, and Tat-Seng Chua. 2021. Learning Intents behind Interactions with Knowledge Graph for Recommendation. In *WWW*. 878–887.

[38] Xiao Wang, Nian Liu, Hui Han, and Chuan Shi. 2021. Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning. *arXiv preprint arXiv:2105.09111* (2021).

[39] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. 2019. Explainable reasoning over knowledge graphs for recommendation. In *AAAI*, Vol. 33. 5329–5336.

[40] Ze Wang, Guangyan Lin, Huobin Tan, Qinghong Chen, and Xiyang Liu. 2020. CKAN: Collaborative Knowledge-aware Attentive Network for Recommender Systems. In *SIGIR*. 219–228.

[41] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *AAAI*.

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

[43] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In *SIGKDD*. 974–983.

[44] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2021. Are Graph Augmentations Necessary? Simple Graph Contrastive Learning for Recommendation.

[45] Xiao Yu, Xiang Ren, Quanquan Gu, Yizhou Sun, and Jiawei Han. 2013. Collaborative filtering with entity similarity regularization in heterogeneous information networks. *IJCAI HINA* (2013).

[46] Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu, Bradley Sturt, Urvashi Khandelwal, Brandon Norick, and Jiawei Han. 2014. Personalized entity recommendation: A heterogeneous information network approach. In *WSDM*. 283–292.

[47] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *SIGKDD*. 353–362.

[48] Yongfeng Zhang, Qingyao Ai, Xu Chen, and Pengfei Wang. 2018. Learning over knowledge-base embeddings for recommendation. *arXiv preprint arXiv:1803.06540* (2018).

[49] Huan Zhao, Quanming Yao, Jianda Li, Yangqiu Song, and Dik Lun Lee. 2017. Meta-graph based recommendation fusion over heterogeneous information networks. In *SIGKDD*. 635–644.

[50] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. 2020. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131* (2020).