ABSTRACT
Incorporating Knowledge Graphs (KG) into recommender 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. The implementations are available at: https://github.com/CCIIPLab/KGIC.

CCS CONCEPTS
• Information systems → Recommender systems.

KEYWORDS
Knowledge Graph, Recommendation, Contrastive Learning

ACM Reference Format:
Ding Zou, Wei Wei, Ziyang Wang, Xian-Ling Mao, Feida Zhu, Rui Fang, and Dangyang Chen. 2022. Improving Knowledge-aware Recommendation with Multi-level Interactive Contrastive Learning. In Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM ’22), October 17–21, 2022, Atlanta, GA, USA. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3511808.3557358

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, 18, 24, 43, 53] presumes that behaviorally similar users have
Despite effectiveness, existing GNN-based methods still fall short in achieving the above core goal in the following three aspects: (a) 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.

In this paper, we develop a novel model, Knowledge-aware Recommender System with Multi-level Interactive Contrastive Learning (KGIC), to solve the aforementioned limitations and challenges. A key insight is that the user-item interactions are actually extremely sparse in real scenarios [1, 45], 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. With the unbalanced information utilization, the noisy KG information, however, gets emphasized more in final user/item modeling, which makes the crucial CF signals less stressed and further results in suboptimal representation learning.

• 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, 45], 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).

Figure 1: A toy example of multi-level contrastive mechanism. (a) Intra-graph level interactive contrastive mechanism; (b) Inter-graph level interactive contrastive mechanism (Only taking Layer=1 as the anchor here for clarity).
Conceptually, KGIC focuses on exploring a proper in-graph contrastive learning mechanism for KGR, 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 the 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, 14, 31, 33, 35, 50, 51] pre-train the KG entity embeddings with knowledge graph embeddings methods (KGE) [2, 19, 44], for enriching item representations. CKE [50] 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 [44] method on user-item interactions and KG triplets. RippleNet [32] propagates users’ historical clicked items along links in KG, exploring more KG facts to reveal the users’ potential interests. Embedding-based methods show high flexibility in utilizing KG, but the utilized KGE 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 [13, 28, 41, 48, 49, 52] explore various patterns of connections among items in KG to provide additional guidance for the recommendation. PER [49] and meta-graph based recommendation [13] extract the meta-path/meta-graph latent features and exploit the connectivity between users and items along different types of relation paths/graphs. KPRN [41] 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 [13, 41, 48, 49, 52] are founded on the information aggregation mechanism of graph neural networks (GNNs) [9, 16, 46]. Typically it integrates multi-hop neighbors into node representations to capture node feature and graph structure, modeling long-range connectivity. KGNN [36] and KGNN-LS [34] firstly utilize GNN on KG to obtain item embeddings by aggregating items’ neighborhood information iteratively. Later, KGAT [38] combines the user-item graph with the KG as a unified heterogeneous graph, then utilizes GNN to recursively perform aggregation on it. But CKAN [42] separately propagates collaborative signals and knowledge signals on the user-item graph and KG to stress the importance of CF signals. More recently, KGIn [39] 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 [30, 40, 45] learn node representations by contrasting positive pairs against negative pairs. DGI [30] first adopts Infomax [21] in graph representation learning, and contrasts the local node embeddings with global graph embeddings. Then GM [25] proposes to contrast the center node with its local neighbors on both node features and topological structure. Similarly, MVRG [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 [40] 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 [45] generates two graph views by corrupting the user-item interactions, and performs contrastive learning between them. And NCL [20] 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 \( Y \in \mathbb{R}^{M \times N} \) 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 [28]. Let \( G = \{(h, r, t) \mid h, t \in \mathbb{E}, r \in \mathbb{R}\} \) be the knowledge graph, where \( h, r, t \) are on behalf of head, relation, tail of a knowledge triple correspondingly; \( E \) and \( R \) refer to the sets of entities and relations in \( G \). In many recommendation scenarios, an entity \( e \in \mathbb{V} \) corresponds to one entity \( e \in \mathbb{E} \). So we establish a set of item-entity alignments \( A = \{(u, e) \mid u \in \mathbb{V}, e \in \mathbb{E}\} \), where \( (u, e) \) indicates that item \( e \) can be aligned with an entity \( u \) 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 \( Y \) and the knowledge graph \( G \), 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 Knowledge-aware Recommender System with Multi-level Interactive Contrastive 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 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-occurance 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 \( A = \{(u, e) \mid u \in \mathbb{V}, e \in \mathbb{E}\} \), the items of first-order CF signals are aligned with KG and the initial entities in KG are acquired as follows:

\[
E_{u,L}^0 = \{ (u, e) \mid e \in A, u \in \{v \mid y_{uv} = 1\}\},
E_{v,L}^0 = \{ (v, e) \mid e \in A\}.
\]

where \( E_{u,L}^0 \) and \( E_{v,L}^0 \) 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:

\[
S_{o,L}^l = \{ (h, r, t) \mid (h, r, t) \in G \text{ and } h \in E_{o,L}^{l-1} \}, l = 1, \ldots , L.
\]

where the symbol \( o \) is a uniform placeholder which means \( u \) or \( v \), \( S_{u,L}^l \) and \( S_{o,L}^l \) 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:

\[
V_p = \{ v_p \mid u \in \mathbb{U}_{dim}, y_{uvp} = 1 \},
\]

\[
\mathbb{U}_{dim} = \{ u \in \{ v \mid y_{uv} = 1 \}, y_{u,v} = 1 \}=
\]

\[
V_e = \{ v_e \mid u \in \{ v \mid y_{uv} = 1 \}, y_{u,v} = 1 \}.
\]

where \( V_p \) and \( \mathbb{U}_{dim} \) represent high-order items and similar users of the user, and \( V_e \) is the high-order items of the item. Further the aligned initial entities \( E_{u,N}^0 \) and \( E_{v,N}^0 \) in KG are acquired as follows:

\[
E_{u,N}^0 = \{ e \mid (v_p, e) \in A, v_p \in V_p \},
E_{v,N}^0 = \{ e \mid (v_e, e) \in A, v_e \in V_e \}.
\]

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

\[
S_{o,N}^l = \{ (h, r, t) \mid (h, r, t) \in G \text{ and } h \in E_{o,N}^{l-1} \}, l = 1, \ldots , L.
\]

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

\[
E_{v,D}^l = \sum_{i=1}^{m} \pi(D_i) e_{i}^h, r_i e_{i},
\]

where the symbol \( D \) is a uniform placeholder which means \( L \) or \( N \), \( E_{v,D}^l \) 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
We move on to encourage a balanced information utilization in where the intra-graph interactive contrastive learning is hence proposed, to extract informative non-local information by making coherent use of CF and KG, by performing interactions between CF and KG information with contrastive learning.

### 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:

\[
L^U_{\text{Intra}} = \sum_{u \in U} \log \left( \frac{e^{\frac{1}{\tau} \sum_{k \in L} e^{(i_{u,L}^0, i_{u,L}^k) // (i_{u,L}^0, i_{u,L}^k)}}}{\sum_{k \in L} e^{(i_{u,L}^0, i_{u,L}^k) // (i_{u,L}^0, i_{u,L}^k)}} + \sum_{k' > L} e^{(i_{u,L}^0, i_{u,L}^k) // (i_{u,L}^0, i_{u,L}^k)}} \right)
\]

\[
\quad + \sum_{u \in U} \log \left( \frac{e^{\frac{1}{\tau} \sum_{k \in L} e^{(i_{u,N}^0, i_{u,N}^k) // (i_{u,N}^0, i_{u,N}^k)}}}{\sum_{k \in L} e^{(i_{u,N}^0, i_{u,N}^k) // (i_{u,N}^0, i_{u,N}^k)}} + \sum_{k' > L} e^{(i_{u,N}^0, i_{u,N}^k) // (i_{u,N}^0, i_{u,N}^k)}} \right)
\]

where \( \tau \) denotes a temperature parameter in softmax. In a similar way, the intra-graph contrastive loss of the item can be obtained as \( L^I_{\text{Intra}} \). And the complete Intra-graph interactive contrastive Loss is the sum of the above two losses:

\[
L_{\text{Intra}} = L^U_{\text{Intra}} + L^I_{\text{Intra}}.
\]

In this way, 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: \[ L_{\text{Inter}} = \sum_{u \in U} \sum_{k \in L} -\log \frac{e^{u,k} e^{u,N}}{\sum_{k \in L} e^{u,k} e^{u,N}} \]

Similarly, we could obtain the inter-graph contrastive loss of the item \( L_{\text{Inter}}^I \). Then the complete inter-graph contrastive loss is the sum of the user and item contrastive loss:

\[ L_{\text{Inter}} = L_{\text{Inter}}^U + L_{\text{Inter}}^I. \]

### 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:

\[
\begin{align*}
e_u &= E_u^0 || E_u^1 || \cdots || E_u^L || E_u^N, \\
e_i &= E_i^0 || E_i^1 || \cdots || E_i^L || E_i^N, \\
y(u, i) &= e_u^T e_i.
\end{align*}
\]

### 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 train the KG-aware recommendation loss and the proposed contrastive loss. For the KG-aware recommendation task, a pairwise BPR loss [26] 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.

\[ L_{\text{BPR}} = \sum_{(u, i) \in O} -\log \sigma(y_{ui} - y_{ui}), \]

where \( O = \{(u, i) \mid (u, i) \in O^+, (u, i) \in O^-\} \) is the training dataset consisting of the observed interactions \( O^+ \) and unobserved counterparts \( O^- \); \( \sigma \) 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:

\[ L_{\text{KGIC}} = L_{\text{BPR}} + \lambda_1 (\alpha L_{\text{Intra}} + L_{\text{Inter}}) + \lambda_2 \| \Theta \|_2^2, \]

where \( \Theta \) is the model parameter set, \( \alpha \) 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) affect KGIC performance?
- **RQ3**: How do different hyper-parameter settings (e.g., model depth, coefficient \( \alpha \), temperature \( \tau \)) 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**
- **MovieLens-1M**
- **Last.FM**

### 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 [32, 36] 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 methods, covering: CF-based methods (BPRMF), embedding-based methods (CKE, RippleNet), path-based methods (PER), and GNN-based methods (KGCN, KGNN-LS, KGAT, CKAN, KGIN, CG-KGR) as follows:

- BPRMF [26]: It’s a typical CF-based method that uses pairwise matrix factorization for implicit feedback optimized by BPR loss.
- CKE [50]: It’s an embedding-based method that combines structural, textual, and visual knowledge in one framework.
- RippleNet [32]: It’s a classical embedding-based method which propagates users’ preferences on the KG.
- PER [49]: It’s a path-based method which extracts meta-path features to represent the connectivity between users and items.
- KGNN-LS [34]: It’s a GNN-based method which iteratively integrates neighboring information to enrich item embeddings.
- KGAT [38]: It’s a GNN-based method which iteratively integrates neighbors on user-item-entity graph with an attention mechanism to get user/item representations.
- CKAN [42]: It’s a GNN-based method which independently propagates collaborative and knowledge signals on CF and KG parts.
- KGIN [39]: 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 user-intent-item-entity graph.
- CG-KGR [6]: It’s the latest GNN-based method which fuses the 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 [15] optimizer and set the batch size to 2048. The local and non-local triple set size are limited to 40 and 128 respectively. Other hyper-parameter settings are provided in Table 1. The best settings for hyperparameters in all methods are researched by either empirical study or the original papers.

5.2 Performance Comparison (RQ1)

The empirical results of all methods in two scenarios 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 LastFM 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 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 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 contrastive 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 performance, and KGIC_{w/o Non-local} is the least competitive model. It makes sense since KGIC_{w/o Non-local} only considers 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. KGIC 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 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 $\alpha$. In the overall contrastive loss defined in Equation (14), the trade-off parameter $\alpha$ can balance the influence of intra- and inter-graph contrastive losses. To analyze the influence of coefficient $\alpha$, we vary $\alpha$ in $\{0.1, 0.5, 1.5, 2\}$ and report the results in Figure 5. From the results we could observe that: an appropriate $\alpha$ can effectively improve the
5.4.3 Impact of Temperature $\tau$. As mentioned in previous contrastive learning work [45, 54], the temperature $\tau$ defined in Equation (8) and Equation (10) plays an important role in contrastive learning. To investigate the impact of $\tau$, 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 $\tau$ will cause poor performance, consistent with conclusions of previous works [45]. 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 [20, 47], 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 have the following observations:

- 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 [37], 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 the 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.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61602197, Grant No.L1924068, Grant No.61772076, in part by CCF-AFSG Research Fund under Grant No.RF20210005, and in part by the fund of Joint Laboratory of HUST and Pingan Property & Casualty Research (HPL).
