Quad-Tier Entity Fusion Contrastive Representation Learning for Knowledge Aware Recommendation System

Rongqing Kenneth Ong
Nanyang Technological University
Singapore
rongqing001@e.ntu.edu.sg

Wei Qiu
Nanyang Technological University
Singapore
qiuwei@ntu.edu.sg

Andy W. H. Khong
Nanyang Technological University
Singapore
andykhong@ntu.edu.sg

ABSTRACT

Knowledge graph (KG) has recently emerged as a powerful source of auxiliary information in the realm of knowledge-aware recommendation (KGR) systems. However, due to the lack of supervision signals caused by the sparse nature of user-item interactions, existing supervised graph neural network (GNN) models suffer from performance degradation. Moreover, the over-smoothing issue further limits the number of GNN layers or hops required to propagate messages—these models ignore the non-local information concealed deep within the knowledge graph. We propose the **Q**uad-**T**ier **E**ntity **F**usion **C**ontrastive **R**epresentation **L**earning (**QTEF-CRL**) knowledge-aware framework to achieve learning of deep user preferences from four perspectives: the collaborative, semantic, preference, and structural view. Unlike existing methods, the proposed tri-local and single-global quad-tier architecture exploits the knowledge graph holistically to achieve effective self-supervised representation learning. The newly-introduced preference view constructed from the collaborative knowledge graph (CKG) comprises a preference graph and preference-guided GNN that are specifically designed to capture non-local information explicitly. Experiments conducted on three datasets highlight the efficacy of our proposed model.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

Recommender System, Knowledge Graph, Users Preferences, Contrastive Learning

ACM Reference Format:
Rongqing Kenneth Ong, Wei Qiu, and Andy W. H. Khong. 2023. Quad-Tier Entity Fusion Contrastive Representation Learning for Knowledge Aware Recommendation System. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3583780.3615020

1 INTRODUCTION

Recommendation systems empower users with the ability to explore and discover items aligned with their interests. These systems play a critical role in online services (e.g., e-commerce and social media) by generating a concise and personalized set of items that maximizes user experience. Conventional techniques that adopt collaborative filtering (CF) [25, 33, 35, 42, 46] exploit the similarity of user preferences who are associated with similar items. Despite its effectiveness, CF-based models such as neural collaborative filtering (NCF) [14] and xDeepFM [19] rely heavily on historical user interactions—they do not consider user and item features that influence the user's choice. Notwithstanding the above, the highly sparse user-item interaction makes it challenging for CF-based models to perform accurate recommendations [18, 28]. To alleviate such issues, knowledge graphs that provide semantic information by establishing relationships between items and real-world entities have been incorporated.

Undoubtedly, the rich auxiliary information endowed by knowledge graphs has attracted increasing attention. Earlier works that utilize knowledge graphs for recommendation focus on generating meaningful embeddings derived from knowledge graph triplets [1, 20, 38, 49]. These embeddings are then used to complement the item representation learning. Following that, pre-defined meta-paths [3, 26, 36, 45] have been proposed to enhance user representations. However, defining these meta-paths require domain expertise and resource-intensive feature engineering. Such an approach also limits the model's robustness and generalizability to different domains. More recent works exploit the hierarchical structure of the interaction data and the development of an end-to-end model via graph neural networks (GNNs) [29, 31, 33, 34, 37], with its success being dependent on the design of the aggregation function. Such an approach is effective in capturing the high-order connectivity within the heterogeneous graph structure, where the high-order collaborative signal is collected and propagated collectively.

While existing architectures leverage information derived from relations within the knowledge graph as a proxy to infer user interests [27, 29, 33, 34], they do not sufficiently capture deep user preferences—existing systems generate items based on relational preferences that may not correctly reflect user interests. Although it is important to model user preferences explicitly to achieve a holistic and accurate depiction of user interest, capturing entity-level preferences is not straightforward. This is because the entity information resides deeply within the user-item-relation-entity heterogeneous graph. Moreover, to capture non-local deep collaborative signal information, the number of neighboring hops has to be increased significantly (e.g., user-item-entity-item-entity). Such an increase will lead to the over-smoothing phenomenon [4, 7, 21, 41], where node representations become highly similar to their neighboring nodes—the recommender system can no longer differentiate

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10.
https://doi.org/10.1145/3583780.3615020

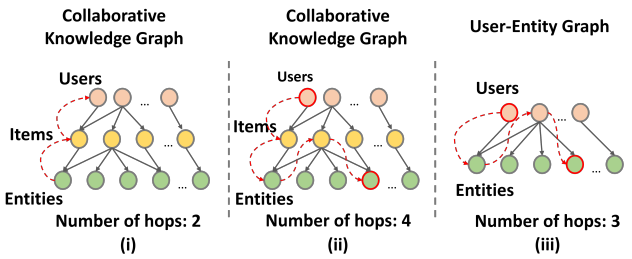

Figure 1: An illustrative example illustrating the capture of non-local knowledge graph information. (i) A two-hop network that captures the closest neighboring entities, (ii) a four-hop knowledge graph capturing non-local information (i.e., deep entities), and (iii) a preference module in QTEF-CRL that constructs a graph that bridges deep entities closer to user nodes.

between users. As depicted in Fig. 1(i), a standard two-hop aggregation will capture entities within an item's immediate neighbor. However, capturing non-local information as illustrated in Fig. 1(ii) will require deeper hops. As such, a preference module must be designed meticulously to strike a good balance between capturing user preferences and mitigating the detrimental effects of over-smoothing. Moreover, due to the highly-sparse labels, an effective learning module must be incorporated to ensure that user preferences are well captured.

To this end, we propose a **Q**uad-**T**ier **E**ntity **F**usion **C**ontrastive **R**epresentation **L**earning (**QTEF-CRL**) knowledge-aware framework to achieve deep learning of user preferences. QTEF-CRL adopts a unified quad-tier architecture that exploits the knowledge graph holistically by taking various intra-graph relations into account: 1) user-item (collaborative view), 2) item-entity (semantic view), 3) user-entity (preference view), and 4) user-item-entity collaborative knowledge graph (structural view). The first three relations are characterized by local views while the last by a global view. Each of the tri-local and single-global architecture reveals unique information about the users and items that can enhance the overall learned representations. To learn the user and item representations effectively across the different views for the extraction of vital and significant signals, we then incorporate the cross-view contrastive learning paradigm. More importantly, we capture deep user preferences at the same granularity of the entities by formulating a new augmented view. This view is termed as the Preference View and comprises the user-entity graph constructed from the collaborative knowledge graph (CKG), where non-local information resides. With this new preference view, a reduced number of hops is required to capture the entity preferences. However, due to the noisy nature of knowledge graphs, the direct application of GNN to capture high-order signals may adversely affect the recommender's performance. To alleviate this problem, QTEF-CRL reconstructs the user-entity graph using a low-rank singular value decomposition (SVD) approximation. A preference-guided GNN then captures the non-local information effectively. With such information, the deeply encapsulated user preferences enable the recommendation model to distill intricate deep collaborative signals, unlike existing models. Experiments conducted show that QTEF-CRL outperforms the state-of-the-art models on three real-world datasets.

We summarise the contributions of this work as follows:

- **Unified architecture:** We highlight the importance of utilizing the knowledge graph for the construction of different views and propose a quad-tier architecture that incorporates a cross-view contrastive representation learning paradigm;
- **Novel methodologies:** We develop and propose a new perspective of modeling deep user preferences that are encompassed in non-local information.

2 RELATED WORKS

2.1 Knowledge Graph Recommendation

Integrating knowledge graph schemas into recommender systems improves the accuracy and interpretability of recommendations. Embedding-based methods such as CKE [47] exploits the structured knowledge of knowledge graphs to intricately weave diverse side information—item content, users' social connections, and knowledge graph —into a collaborative filtering architecture. Path-based approaches such as MEIRec [6] utilize manually designed meta-path to guide representation learning and to infer the user's intentions. Predominantly, GNN-based methods such as KGCN [31] and KGAT [33] leverage the structured information embedded in knowledge graphs to establish higher-order statistical relations between user and item. For instance, KGCN incorporates user preferences by implementing convolutional operations to uncover intricate connections between items. In contrast, KGAT incorporates an attention mechanism that aggregates higher-order neighborhood information of items, thereby modeling the multiple relationships that connect items within the knowledge graphs. In addition, KGIN [34] infers the user's latent intention by exploiting KG relations and infusing it into the aggregation layer which enhances the model's performance and provides attention-level explainability.

The exploitation of entity-level preferences in recommendation systems has also gained prominence since users are more likely to choose an item because of a certain attribute entity. For instance, EPKG [5] attempts to model the user's entity preferences by computing the number of entity connections and assigning it as weights for user preference learning. Models such as the LKGCN [32] and ECRN [15] employ entity-level information to improve the accuracy and interpretability of recommendations. In particular, LKGCN leverages item and entity-level information directly from the tripartite graph to propagate relational messages, while ECRN exploits pre-defined meta-paths to propagate entity affiliation signals for user representation learning, thereby refining the prediction of user-item interactions.

2.2 Contrastive Learning for Recommender System

Contrastive learning has recently been proposed in recommender systems [43, 44] to learn highly robust embeddings by maximizing the correlation between similar instances while distancing dissimilar instances. SGL [40] generates contrastive views by performing random perturbation such as node dropout, edge dropout and random walk, before employing the InfoNCE loss [9, 23] to encourage

feature alignment across different views. KGIC [53] implements a two-stage contrastive strategy that improves recommendation performance by contrasting layers of different parts within graphs at both inter- and intra-graph levels. Moreover, MCCLK [52] integrates contrastive learning across multiple levels in knowledge-aware recommender systems. It introduces a contrastive objective that considers both intra- and inter-view consistency across different data views, such as user-item interactions and KG-anchored relationships. However, in light of these existing works, we argue that i) less effort has been focused on modeling the deep user preferences at the granularity of entities and ii) that existing models fail to fully capitalize on the knowledge graphs for the generation of semantically rich graph contrastive views.

3 TASK FORMULATION

In a standard recommendation setting, users' interaction with the items may exist in either explicit or implicit form. In this work, we focus on the implicit interaction data (e.g., purchase, click, like). We let $\mathcal{U} = (u_1, u_2, ..., u_M)$ be the set representing the users, where M is the total number of users. Similarly, we define the set of items as $\mathcal{I} = (i_1, i_2, ..., i_N)$, with N being the number of items. We can then obtain an interaction matrix denoted by $\mathbf{Y} \in \mathbb{R}^{MxN}$, where element $Y_{ui} = 1$ implies an observed interaction between the user and item, while 0 signifies otherwise.

A knowledge graph stores rich auxiliary information pertaining to real-world facts, which encompasses attributes such as entity relationships, scientific, and extrinsic commonsense knowledge. It provides information associated with the relationship(s) between items and can be represented by two sets—a set containing real-world entities $\mathcal{E}$ and the relations $\mathcal{R}$ between the entities. The knowledge graph can then be defined as

$$\mathcal{G}_k = \big\{(h, r, t)|h, t \in \mathcal{E}, r \in \mathcal{R}\big\}, \tag{1}$$

where h and t correspond to the entities' head and tail, respectively, while r represents the relationship between the two entities. An additional item-entity alignment set can further be defined as

$$\mathcal{A} = \big\{(i, e)|i \in \mathcal{I}, e \in \mathcal{E}\big\}, \tag{2}$$

which implies that an item i is associated with an entity e from the knowledge graph.

The CKG [33] seeks to unify the interaction data within the knowledge graph. By modeling the user-interaction data as an *interact* relation, it can be combined into the knowledge graph as a form of a heterogeneous graph. As CKG is used specifically in our structural view of QTEF-CRL, we denote the unified graph as

$$\mathcal{G}_s = \big\{(h, r, t)|h, t \in \mathcal{E}', r \in \mathcal{R}'\big\}, \tag{3}$$

where $\mathcal{E}' = \mathcal{E} \cup \mathcal{U}$ and $\mathcal{R}' = \mathcal{R} \cup \{Interaction\}$. Given interaction data Y, the knowledge graph $\mathcal{G}_k$ and the CKG $\mathcal{G}_s$, our objective is to learn a function to predict the probability $\widehat{y}_{ui}$ that a user u will interact with an item i.

4 METHODOLOGY

In contrast to existing methods that solely focus on extracting collaborative signals from the global CKG [33, 34], the proposed QTEF-CRL is a unified model that exploits the knowledge graph holistically from different aspects for the construction of local and global graph that is subsequently used for contrastive representation learning. More importantly, QTEF-CRL incorporates the preference view—a rarely explored yet crucial view constructed from the CKG. This results in QTEF-CRL being a quad-tier architecture that jointly considers and optimizes user preferences at different granularity levels. Our tri-local and single-global tiers each possess distinct and highly meaningful properties that will enable the model to distill different user preference levels.

As shown in Fig. 2, QTEF-CRL is partitioned into four tiers comprising the following components: the *collaborative view*, *semantic view*, *preference view*, and *structural view*. Each view seeks to capture distinct latent factors of the users and items. Firstly, the collaborative view reflects the user-item interactions, where collaborative signals can be extracted through an adequately designed encoder. Secondly, the semantic view reveals the rich semantic information associated with entities that enhance the items. The third preference view, which has received lower attention in existing works, reveals deep user preferences that are modeled through non-local KG information. This view is critical since it captures user interests through entity affiliations. Lastly, the global structural view comprises a heterogeneous CKG that enriches the user-interaction graph through the relations and external entities. By modeling the quad views explicitly, QTEF-CRL is able to distill salient information for effective representation learning. Thereafter, the respective generated user and item representations are contrasted with each other, using a cross-view contrastive learning mechanism that contrasts against the inter and intra representations.

4.1 Tier 1: Collaborative View (Local Level)

The collaborative view focuses on extracting the collaborative signals from the interaction data. This can be achieved by adopting an encoder that propagates information across the user's and the item's neighbourhood. Following the recent success of collaborative-based models [13, 22, 39], we employ the LightGCN as the collaborative view encoder due to its inherently simplistic design, which has been empirically proven to be effective. LightGCN incorporates a straightforward mechanism for message passing and aggregation without the need for feature transformation and non-linear activation. The propagation rule may be expressed as

$$e_u^{(k)} = \sum_{i \in \mathcal{N}_u} \frac{1}{\sqrt{|\mathcal{N}_u||\mathcal{N}_i|}} e_i^{(k-1)}, \qquad e_i^{(k)} = \sum_{u \in \mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_u||\mathcal{N}_i|}} e_u^{(k-1)}, \tag{4}$$

where $e_u^{(k)}$ and $e_i^{(k)}$ are the kth layer representations of the users and items, respectively, $\frac{1}{\sqrt{|\mathcal{N}_u||\mathcal{N}_i|}}$ is the symmetric normalization to prevent overscaling of the representations. Thereafter, these representations are summed across K layers to form the overall local collaborative representations, i.e.,

$$z_u^{local,c} = e_u^{(0)} + ... + e_u^{(K)}, \qquad z_i^{local,c} = e_i^{(0)} + ... + e_i^{(K)}. \tag{5}$$

4.2 Tier 2: Semantic View (Local Level)

The basis of the semantic view lies in the effectiveness of leveraging the knowledge graph to generate useful item embeddings. Most of the existing works [34, 52] employ $e_r \odot e_v$ for item representation learning, where the element-wise product aims to "memorize" the relational signals which are propagated from the knowledge graph. The attention mechanism has also been incorporated to emphasize the importance of different relations and entities. However, simply

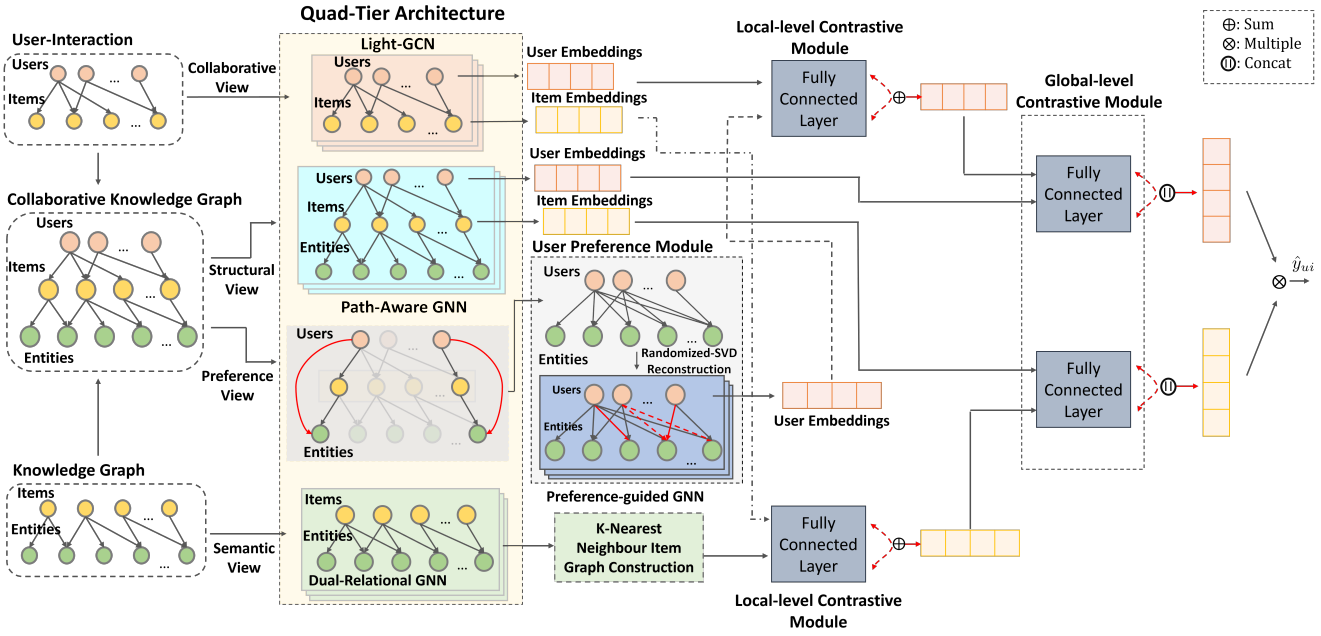

Figure 2: Illustration of the proposed QTEF-CRL architecture comprising of a quad-tier level structure, with collaborative, semantic, preference, and structural views.

weighing the relations between the items and entities is insufficient since, in most cases, relationships between the entities play an important role. For instance, a user watches a movie (item) because two actors (entities) v_1, v_2 are being starred. Removing either v_1 or v_2 will reduce the interest from the user. These intra-relations are often unaccounted for and we propose a dual-relational GNN to weigh both the inter and intra-relationship. This is achieved by incorporating a dual-attention mechanism that collaboratively emphasizes different viewpoints such that

$$\begin{aligned}\mathbf{e}_i^{(k)} &= \frac{1}{|\mathcal{N}_i|}\sum_{r,v\in\mathcal{N}_i}\mathbf{e}_r\odot\mathbf{e}_v^{(k-1)},\\ \mathbf{e}_v^{(k)} &= \frac{1}{|\mathcal{N}_v|}\Bigg(\sum_{(r,v)\in\mathcal{N}_v}\alpha(\mathbf{v},\mathbf{r},\mathbf{i})\mathbf{e}_r\odot\mathbf{e}_i^{(k-1)}\\ &\qquad+\sum_{(r,i)\in\mathcal{N}_v}\beta(\mathbf{v},\mathbf{r},\mathbf{v}')\mathbf{e}_r\odot\mathbf{e}_v^{(k-1)}\Bigg),\end{aligned} \tag{6}$$

where $e_v^{(k)}$ denotes the kth layer entity representation. The variables

$$\begin{aligned}\alpha(v,r,i) &= \frac{exp\left((e_v\odot e_r)^\top\cdot(e_i\odot e_r)\right)}{\sum_{(i',r)\in\widehat{\mathcal{N}}_{(v)}}exp\left((e_v\odot e_r)^\top\cdot(e_i\odot e_r)\right)},\\ \beta(v,r,v') &= \frac{exp\left((e_v\odot e_r)^\top\cdot(e_{v'}\odot e_r)\right)}{\sum_{(v',r)\in\widehat{\mathcal{N}}_{(v)}}exp\left((e_v\odot e_r)^\top\cdot(e_{v'}\odot e_r)\right)}\end{aligned} \tag{7}$$

denote the inter and intra-relations between the neighboring entity-item and entity-entity, respectively. Here, $\widehat{\mathcal{N}}_{(v)}$ is the set of entities within the neighborhood (including itself), $e_v\odot e_r$ and $e_i\odot e_r$ are the relational signals stored in the entity.

Inspired by [48, 52], we then construct an item-item cosine similarity-based graph such that:

$$S_{ij} = (\mathbf{e}_i^{(K')})^\top\mathbf{e}_j^{(K')}\Big/\left\|\mathbf{e}_i^{(K')}\right\|\left\|\mathbf{e}_j^{(K')}\right\|. \tag{8}$$

With the constructed graph, the k-nearest-neighbour sparsification on the densely connected graph is performed to eliminate noise that exists in the semantic graph [52]. In particular, we select the top-k nearest neighbor such that the sparse item-item matrix is given by

$$\widehat{S}_{ij} = \begin{cases} S_{ij}, & S_{ij}\in \text{top}-k(S_i);\\ 0, & \text{otherwise,}\end{cases} \tag{9}$$

Defining $D\in\mathbb{R}^{N\times N}$ as the degree matrix of $\widehat{S}_{ij}$, symmetric normalisation is subsequently performed on the adjacency matrix $\widetilde{S} = (D)^{-\frac{1}{2}}\widehat{S}(D)^{-\frac{1}{2}}$ to mitigate the gradient vanishing problem before LightGCN is being employed as the semantic encoder to obtain the lth layer item representations

$$\mathbf{e}_i^{(l)} = \sum_{i'\in\mathcal{N}(i)}\widetilde{S}\mathbf{e}_{i'}^{(l-1)}. \tag{10}$$

Here, $e_{i'}^{(l-1)}$ is defined as the item representation at the previous layer and $\widetilde{S}$ is the normalised item-item adjacency matrix. The local level semantic item representation is obtained and summed across all layers, i.e.,

$$z_i^{local,s} = e_i^{(0)} + e_i^{(1)} + \ldots + e_i^{(L)}. \tag{11}$$

4.3 Tier 3: Preference View (Local Level)

The third tier of QTEF-CRL models the deep user interest. The generated preference view reveals the deep local collaborative signals that reflect user interest consistent with the granularity of entities. One of the most crucial aspects, which are often unaccounted for

in existing works, is user-collaborative signals that are concealed deep within the knowledge graph. The proposed preference view emphasizes the importance of capturing such information that will deeply influence user decisions and behavior consistent with their interaction with the items. There are two key components under the preference view in the preference module: preference graph construction and user preference-aware representation learning.

The preference graph aims to bridge the gap between the users and entities, allowing easier capturing of non-local information through high-order propagation. Assigning item i as the intermediary node, a meta-path may be defined directly if there exists a connection between a user u and an entity v. We can therefore construct a user-entity preference matrix representing the user-entity graph directly from CKG. In particular, we denote the user-entity preference matrix as $\mathbf{X}$, where $x_{ue} = n$ for n number of connections that links u to v via i given that $i, v \in \mathcal{A}, u \in \mathcal{U}$, and $i \in \mathcal{I}$. With this approach, entries in the user-entity matrix indicate the degree of the user's interest in the entities, thereby reflecting the intensity of their deep interests.

In terms of user preference-aware learning and with the user-entity preference graph being constructed, we employ GNN to extract the deep collaborative signals. However, since the knowledge graph is constructed externally, significant amount of noise is present, rendering the process of recursive propagation a challenge. While singular value decomposition (SVD) may be adopted to reconstruct the user-entity preference graph and eliminate potential noises deemed by the algorithm, direct application of the SVD for dimensionality reduction may still require prohibitively high computational resources. Inspired by previous works [2, 24], we adopt the randomized SVD [10] by performing an initial approximation of the user-entity adjacency matrix $\mathbf{X}$ using a low-rank orthonormal matrix before applying SVD on this reduced matrix, i.e.,

$$\widehat{\mathcal{X}}_{SVD} = \widehat{\mathbf{U}}_t \widehat{\Sigma}_t \widehat{\mathbf{V}}_t^\top. \tag{12}$$

Here, t denotes for the low-rank approximation, $\widehat{\mathbf{U}}_t \in \mathbb{R}^{M \times t}$ and $\widehat{\mathbf{V}}_t \in \mathbb{R}^{S \times t}$ are, respectively, the left and right singular matrices containing the eigenvectors of X. The term $\widehat{\Sigma}_t \in \mathbb{R}^{t \times t}$ comprises the t highest (truncated) singular values and $\widehat{\mathcal{X}}_{SVD}$ denotes the approximated reconstructed matrix. Employing such a reconstruction paradigm can effectively extract and preserve the crucial deep collaborative signals by taking into account each user-entity's degree of interest. With the user-entity preference graph $\widehat{\mathcal{X}}_{SVD}$, we propose to model the long-range connectivity by employing a preference-guided GNN that performs aggregation P number of times recursively via

$$e_u^{(p)} = W_p\left(\widehat{\mathcal{X}}_{SVD}^\top e_v^{(p-1)}\right) = W_p\left(\widehat{\mathbf{U}}_t \widehat{\Sigma}_t \widehat{\mathbf{V}}_t^\top e_v^{(p-1)}\right), \tag{13}$$

where $W_p \in \mathbb{R}^{M \times d}$ is a trainable linear transformation matrix, d is the dimension size, and $e_u^{(p)}$ is the pth layer user-preference representation.

In contrast to conventional graph convolution aggregation paradigm [11, 17], the non-linear activation function is omitted after the linear transformation process since it does not yield any benefits [12, 13]. Furthermore, to optimize the computational process, we employ the low-rank approximation matrices $\widehat{\mathbf{U}}_t \widehat{\Sigma}_t \widehat{\mathbf{V}}_t$ to propagate high-order signal collectively. Doing so does not require storing the dense matrix $\widehat{\mathcal{X}}_{SVD}$, which is significantly huge due to the fully-connected nature of SVD. It is also worth highlighting that reconstruction of the user-entity graph is performed during the pre-processing stage and that only the low rank matrices $\widehat{\mathbf{U}}_t \widehat{\Sigma}_t$ and $\widehat{\mathbf{V}}_t \widehat{\Sigma}_t$ are computed and propagated prior to the model training. This increases the model's overall efficiency compared to storing the dense matrix. The final local user-preference representation is obtained via

$$z_u^{local,p} = e_u^{(0)} + e_u^{(1)} + \ldots + e_u^{(P)}. \tag{14}$$

4.4 Tier 4: Structural View (Global Level)

The fourth and final tier of QTEF-CRL comprises a global-level architecture that generates a structural view. In this view, the model aims to encode the semantically rich heterogeneous graph to extract long-range connectivity relational information. Such a heterogeneous graph is constructed by combining the user-item interaction graph with the knowledge graph, resulting in a unified relational CKG. Similar to [34, 52], we adopt the path-aware GNN that seeks to propagate the local neighborhood information while retaining the relational information. This is achieved via the relational operator $e_r \odot e_v$ such that the aggregation rule is given by

$$e_u^{(l)} = \frac{1}{|\mathcal{N}_u|} \sum_{i \in \mathcal{N}_u} e_i^{(l-1)}, \; e_i^{(l)} = \frac{1}{|\mathcal{N}_i|} \sum_{(r,v) \in \mathcal{N}_i} \alpha(i,r,v) e_r \odot e_v^{(l-1)}, \tag{15}$$

where $e_u^{(l)}, e_i^{(l)}$ denote, respectively, the lth layer user and item representations, while $\mathcal{N}_u$ and $\mathcal{N}_i$ denote the neighborhood of user u and item i, respectively. With $\widehat{\mathcal{N}}_{(v)}$ defined after (7), the attention mechanism

$$\alpha(i,r,v) = \frac{exp\left((e_v \odot e_r)^\top \cdot (e_i \odot e_r)\right)}{\sum_{(v',r) \in \widehat{\mathcal{N}}_{(v)}} exp\left((e_v \odot e_r)^\top \cdot (e'_v \odot e_r)\right)} \tag{16}$$

assigns individual weights to each relation and entity. As opposed to the path-aware GNN [52], we adopted $e_i \odot e_r$ to capture the item-entity affiliation more effectively. Consequently, the overall global structural user and item representations are obtained by summing up the representation across each layer, i.e.,

$$\begin{aligned} z_u^{global,s} &= e_u^{(0)} + e_u^{(1)} + \ldots + e_u^{(L)}, \\ z_i^{global,s} &= e_i^{(0)} + e_i^{(1)} + \ldots + e_i^{(L)}. \end{aligned} \tag{17}$$

4.5 Local Level Contrastive Learning

Having learned the set of local user and item representations $\left\{z_u^{local,c}, z_i^{local,c}, z_u^{local,p}, z_i^{local,s}\right\}$ from the tri-local tiers, we contrast (i) the user representation across the collaborative and preference views and (ii) the item representation across the collaborative and semantic views to learn discriminative features across the different tiers for each user and item. To this end, we adopt a cross view contrastive learning mechanism [50–52] to jointly supervise the different views. For the users and items representation learning, the embeddings are intially passed through a multi-layer perceptron (MLP) such that

$$\begin{aligned} \widehat{\mathbf{z}}_u^{local,c} &= W_2\sigma\left(W_1 \mathbf{z}_u^{local,c} + b_1\right) + b_2, \\ \widehat{\mathbf{z}}_u^{local,p} &= W_2\sigma\left(W_1 \mathbf{z}_u^{local,p} + b_1\right) + b_2, \\ \widehat{\mathbf{z}}_i^{local,c} &= W_2\sigma\left(W_1 \mathbf{z}_i^{local,c} + b_1\right) + b_2, \\ \widehat{\mathbf{z}}_i^{local,s} &= W_2\sigma\left(W_1 \mathbf{z}_i^{local,s} + b_1\right) + b_2, \end{aligned} \tag{22}$$

$$\mathcal{L}_u^{local} = -log \frac{e^{s(\widehat{z}_u^{local,p}, \widehat{z}_u^{local,c})/\tau}}{\underbrace{e^{s(\widehat{z}_u^{local,p}, \widehat{z}_u^{local,c})/\tau}}_{\text{positive pair}} + \underbrace{\sum_{k \neq i} e^{s(\widehat{z}_i^{local,p}, \widehat{z}_k^{local,p})/\tau}}_{\text{intra-view negative pairs}} + \underbrace{\sum_{k \neq i} e^{s(\widehat{z}_i^{local,p}, \widehat{z}_k^{local,c})/\tau}}_{\text{inter-view negative pairs}}} \tag{18}$$

$$\mathcal{L}_i^{local} = -log \frac{e^{s(\widehat{z}_i^{local,s}, \widehat{z}_i^{local,c})/\tau}}{\underbrace{e^{s(\widehat{z}_i^{local,s}, \widehat{z}_i^{local,c})/\tau}}_{\text{positive pair}} + \underbrace{\sum_{k \neq i} e^{s(\widehat{z}_i^{local,s}, \widehat{z}_k^{local,s})/\tau}}_{\text{intra-view negative pairs}} + \underbrace{\sum_{k \neq i} e^{s(\widehat{z}_i^{local,s}, \widehat{z}_k^{local,c})/\tau}}_{\text{inter-view negative pairs}}} \tag{19}$$

$$\mathcal{L}_u^{global*} = -log \frac{e^{s(\widehat{z}_u^{global}, \widehat{z}_u^{local})/\tau}}{\underbrace{e^{s(\widehat{z}_i^{global}, \widehat{z}_k^{global})/\tau}}_{\text{positive pair}} + \underbrace{\sum_{k \neq i} e^{s(\widehat{z}_i^{global}, \widehat{z}_k^{global})/\tau}}_{\text{intra-view negative pairs}} + \underbrace{\sum_{k \neq i} e^{s(\widehat{z}_i^{global}, \widehat{z}_k^{local})/\tau}}_{\text{inter-view negative pairs}}} \tag{20}$$

$$\mathcal{L}_u^{local*} = -log \frac{e^{s(\widehat{z}_u^{local}, \widehat{z}_u^{global})/\tau}}{\underbrace{e^{s(\widehat{z}_u^{local}, \widehat{z}_u^{global})/\tau}}_{\text{positive pair}} + \underbrace{\sum_{k \neq i} e^{s(\widehat{z}_i^{local}, \widehat{z}_k^{local})/\tau}}_{\text{intra-view negative pairs}} + \underbrace{\sum_{k \neq i} e^{s(\widehat{z}_i^{local}, \widehat{z}_k^{global})/\tau}}_{\text{inter-view negative pairs}}} \tag{21}$$

where $W_{(.)} \in \mathbb{R}^{d \times d}$ and $b_{(.)} \in \mathbb{R}^{d \times 1}$ denote the trainable parameters, and σ is the ReLU activation function. The variables $\widehat{\mathbf{z}}_u^{local,c}$ and $\widehat{\mathbf{z}}_u^{local,p}$, therefore, correspond to the transformed user collaborative and preference representations, while $\widehat{\mathbf{z}}_i^{local,c}$ and $\widehat{\mathbf{z}}_i^{local,s}$ correspond to the transformed item collaborative and semantic representations, respectively.

Prior to the computation of contrastive loss, we establish the positive and negative samples as follows:

- For any given node in a particular view, the corresponding node embedding acquired from the alternate view is considered a positive sample;
- Conversely, in the two-view scenario, the node embeddings other than the targeted node are treated as negative samples.

By leveraging these predefined positive and negative samples, the contrastive loss functions for the users and items are formulated as shown in (18) and (19), respectively. The variable τ denotes the temperature hyperparameter that controls the smoothness of the distribution, $s(\cdot)$ is the cosine similarity function to quantify the affinity between the embeddings. With the local-level loss for the users and items being computed, the tri-local loss is then given by

$$\mathcal{L}^{tri-local} = \frac{1}{M} \sum_{u=1}^{M} \mathcal{L}_u^{local} + \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i^{local}. \tag{23}$$

4.6 Global-Level Contrastive Learning & Prediction

The global-level contrastive learning paradigm aims to learn discriminative features from the user and item representation in the structural view propagated from $\mathcal{G}_s$ with respect to the aggregated representations learned from the tri-local views. The local-level aggregated embeddings are first summed up for users and items such that

$$z_u^{local} = z_u^{local,c} + z_u^{local,p}, \qquad z_i^{local} = z_i^{local,c} + z_i^{local,s}, \tag{24}$$

where z_u^{local} and z_i^{local} denote the overall aggregated users and items local representations, respectively. Similar to (22), the global and aggregated user and item embeddings serve as inputs to the MLP to obtain the respective $\widehat{z}_u^{global,s}$, $\widehat{z}_u^{local}$, $\widehat{z}_i^{global,s}$, and $\widehat{z}_i^{local}$ representations. As with the local-level contrastive learning module, the contrastive loss for the users in both global and local levels are given by (20) and (21), where the variables $\mathcal{L}_i^{global*}$ and $\mathcal{L}_i^{local*}$ are similarly calculated by replacing the users with the items and the "$*$" denote for the global loss. Finally, the globally- and locally-computed losses are added to obtain the resultant global loss

$$\mathcal{L}^{global} = \frac{1}{2M} \sum_{u=1}^{M} (\mathcal{L}_u^{global*} + \mathcal{L}_u^{local*}) + \frac{1}{2N} \sum_{i=1}^{N} (\mathcal{L}_i^{global*} + \mathcal{L}_i^{local*}). \tag{25}$$

After obtaining representations from the different tiers of QTEF-CRL, the respective user representations within their locality are summed up to encapsulate the different preferences learned. Embeddings are then concatenated globally as

$$\begin{aligned} z_u^* &= z_u^{global,s} || \big(z_u^{local,c} + z_u^{local,p}\big), \\ z_i^* &= z_i^{global,s} || \big(z_i^{local,c} + z_i^{local,s}\big), \\ \widehat{y}(u,i) &= z_u^{*\top} z_i^*. \end{aligned} \tag{26}$$

For model optimization, we adopt the multi-training paradigm with contrastive representation learning as the pretext task and the knowledge-aware recommendation as the downstream task. For the KGR task, we employ the pairwise BPR loss [25] to reconstruct the historical data, where observed items will be allocated higher scores in contrast to unobserved items, i.e.,

$$\mathcal{L}_{BPR} = \sum_{(u,i,j) \in O} -\ln \sigma(\widehat{y}_{ui} - \widehat{y}_{uj}), \tag{27}$$

where $O = \{(u,i,j) | (u,i) \in O^+, (u,j) \in O^-\}$ comprises the observed O^+ and unobserved O^- sets of interactions and σ is the sigmoid function. Thereafter, combining the self-supervised and recommendation tasks, the overall objective function is defined by

$$\mathcal{L}_{QTEF-CRL} = \mathcal{L}_{BPR} + \beta \Big(\alpha \mathcal{L}^{tri-local} + (1-\alpha) \mathcal{L}^{global} \Big) + \lambda \|\Theta\|_2^2, \tag{28}$$

		Book -Crossing	Last.FM	MovieLens -1M
User-Item Interaction	#users	17,860	1,872	6,036
	#items	14,967	3,846	2,445
	#interactions	139,746	42,346	753,772
Knowledge Graph	#entities	77,903	9,366	182,011
	#relations	25	60	12
	#triplets	151,500	15,518	1,241,996
Hyper-parameter Settings	# α	0.2	0.2	0.2
	# β	0.001	0.001	0.001
	# K	2	3	2
	# K'	3	2	1
	# L	3	3	1
	# L'	3	2	1
	# P	3	4	2

Table 1: Statistics and optimal QTEF-CRL settings for tri-local contrastive loss α, global contrastive loss β, collaborative layers K, semantic aggregation K', semantic layers L, structural aggregation L', and preference layer P.

where α and β are the hyperparameters used to control the local and global level loss, respectively, λ controls the L_2 regularization term, and Θ is the model hyperparameter.

5 EXPERIMENT

We present empirical results to validate the efficacy of QTEF-CRL. The experiments are designed to answer the following:

- **RQ1:** How well does QTEF-CRL perform compared to existing state-of-the-art KGR models?
- **RQ2:** How does each collaborative, semantic, preference, and structural view in QTEF-CRL affect the overall performance?
- **RQ3:** How does the aggregation depth in the individual tiers and the hyperparameters influence QTEF-CRL?

5.1 Datasets, Baselines, and Parameters

Experiments were conducted on three datasets: the Book-Crossing dataset, the MovieLens-1M dataset, and the Last-FM dataset. Each of these datasets possesses unique attributes, providing a diverse range of experimental environments. The Book-Crossing dataset[1] includes explicit and implicit user ratings of books. Explicit ratings range from 1 to 10, while implicit user interactions are marked with 0. For our experiments, we segregate this data into training and testing sets with an 80/20 split. The MovieLens-1M benchmark dataset[2] contains 1 million ratings (ranging from 1 to 5) from 6,000 users on over 2,000 movies. This dataset is segmented into three categories—ratings, users, and movies. For consistency, we adopt the same division as the Book-Crossing dataset by splitting the data into an 80/20 training/testing set. The Last-FM music listening dataset[3] is collected from Last.FM online music systems. The dataset contains about 15,000 ratings from 2,000 users to 4,000 items.

We evaluate our method in click-through rate (CTR) prediction by applying the trained model to predict each interaction in the test set. AUC and F1 score are applied to evaluate the performance of the baselines and the proposed model. To validate the effectiveness of our proposed model, we compared it with the following baselines

- **BPRMF** [25], a CF-based method that uses pairwise matrix factorization for implicit feedback.
- **CKE** [47] collaboratively learns the embeddings of users, items, and entities to generate recommendations.
- **RippleNet** [29] propagates users' direct preferences throughout the knowledge graph to infer their potential preferences, generating personalized item rankings.
- **PER** [45] is designed for entity-level recommendations by generating granular and personalized recommendations.
- **KGCN** [31] extends the GNN model by leveraging the structure and attribute information in knowledge graphs. KGCN captures complex connections between entities in the graph via convolutional operations.
- **KGNN-LS** [30] is a GNN-based model, integrating label semantics into the recommendation process.
- **KGAT** [33] is based on a knowledge graph with an attention mechanism to aggregate high-order neighborhood entity information.
- **KGIN** [34] combines knowledge graph with user-item interaction information to enhance recommendation accuracy.
- **KGIC** [53] utilizes a contrastive learning framework to improve the performance of knowledge graph-based recommender systems.
- **MCCLK** [52] employs a contrastive objective function that considers both intra- and inter-view consistency across user-item interactions and KG-based relationships.

To ensure fairness and consistency, we set the embedding size $d = 64$ and used the same seed as per baseline open-sourced implementation. We initialize the trainable parameters in our model using the Xavier technique [8] and adopted the Adam optimizer [16]. The batch size is consistently set to 2048. We performed grid search to obtain the best fine-tuned settings. Specifically, we varied the learning rate within the range of 3×10^{-4} to 3×10^{-3} and tabulated the hyperparameters in Table 1.

5.2 Performance Comparison (RQ1)

We compare the effectiveness of QTEF-CRL with other baselines in Table 2. These results highlighted the following:

- *The proposed QTEF-CRL model consistently outperforms all baseline models.* In particular, QTEF-CRL achieves an AUC improvement over KGIC by 4.38% for the Book-Crossing dataset, MCCLK by 1.96% and 0.50% for Last-FM and MovieLens-1M datasets, respectively. This improvement is attributed to the modeling of user preferences consistent with the granularity of entities—the proposed preference view allows QTEF-CRL to capture deep user interest from the non-local knowledge graph information. On the contrary, existing approaches do not explicitly model user preferences deeply, while KGIN only models user preferences at the relational level. The tri-local and single-global model of QTEF-CRL also facilitates the distillation of distinct preferences at various granularities associated with user interest.

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

Table 2: Overall performance comparison on three datasets in terms of click-through rate prediction

	Book-Crossing		Last-FM		Movielens-1M	
	AUC	F1	AUC	F1	AUC	F1
BPRMF	0.6583 (-16.04%)	0.6117 (-12.57%)	0.7563 (-13.96%)	0.7010 (-11.47%)	0.8920 (-4.81%)	0.7921 (-7.59%)
CKE	0.6759 (-14.28%)	0.6235 (-11.39%)	0.7471 (-14.88%)	0.6740 (-14.17%)	0.9065 (-3.36%)	0.8024 (-6.56%)
RippleNet	0.7211(-9.76%)	0.6472(-9.02%)	0.7762(-11.97%)	0.7025(-11.32%)	0.9190(-2.11%)	0.8422(-2.58%)
PER	0.6048 (-13.46%)	0.5726 (-16.48%)	0.6414 (-25.45%)	0.6033(-21.24%)	0.7124 (-22.77%)	0.6670 (-20.10%)
KGCN	0.6841 (-13.46%)	0.6313 (-10.61%)	0.8027 (-9.32%)	0.7086 (-10.71%)	0.9090 (-3.11%)	0.8366(-3.14%)
KGNN-LS	0.6762 (-14.25%)	0.6314 (-10.60%)	0.8052 (-9.07%)	0.7224 (-9.33%)	0.9140 (-2.61%)	0.8410(-2.70%)
KGAT	0.7314 (-8.73%)	0.6544 (-8.3%)	0.8293 (-6.66%)	0.7424 (-7.33%)	0.9140 (-2.61%)	0.8440(-2.40%)
KGIN	0.7273 (-9.14%)	0.6614 (-7.60%)	0.8486 (-4.73%)	0.7602 (-5.55%)	0.9190 (-2.11%)	0.8441(-2.39%)
KGIC	0.7749(-4.38%)	0.6812(-5.62%)	0.8592 (-3.67%)	0.7753 (-4.04%)	0.9252 (-1.49%)	0.8559(-1.21%)
MCCLK	0.7625 (-5.62%)	0.6777 (-5.97%)	0.8763(-1.96%)	0.8008 (-1.49%)	0.9351(-0.50%)	0.8631(-0.49%)
QTEF-CRL	**0.8187***	**0.7374***	**0.8959***	**0.8157***	**0.9401***	**0.8680***

- *Models that leverage on knowledge graph achieve higher performance.* Both KGCN and KGNN-LS achieve higher performance than BPRMF due to the effectiveness of knowledge graph in capturing item-entity affiliations. In addition, KGIN, that leverages relational information to infer user intention, outperforms most baselines. Similarly, MCCLK, which explores non-local information from the perspective of enriching item representation, outperforms other baselines.
- *Models that adopt self-supervised learning as the pretext task achieve higher performance.* We observe that MCCLK and KGIC outperform other baselines which are trained in a supervised manner. This stems from the inherent sparsity issue that exists within the dataset and that incorporating self-supervised learning tasks can alleviate such an issue.

5.3 Ablation Tests (RQ2)

We perform ablation studies by segmenting QTEF-CRL into five variants: QTEF-CRL, QTEF-CRL without preference tier, QTEF-CRL without semantic tier, QTEF-CRL without local tier (removing preference and semantic local level) and, QTEF-CRL without global tier. Results plotted in Fig. 3 highlighted that:

- *Removing any tiers in QTEF-CRL reduces its performance across all datasets.* Each tier in QTEF-CRL serves to provide unique properties that reflect user preferences and enhance an item's semantic information through the connection of the knowledge graph.
- *Preference View plays a key role in extracting deep user preferences.* From the three datasets being bench-marked against, we observe a significant decrease in performance when the preference tier is removed from QTEF-CRL. The removal of the preference tier will lead to its inability to capture profound user preference resulting in reduced performance.
- *Removing preference and semantic views results in worse performance than removing the global level tier.* We note from Table 1 that Book-Crossing and Movielens-1M possess significant number of relations and entities. Since preference and semantic tiers aim to capture deep user preference and item-entity affiliations, respectively, removing these tiers in $QTEF\text{-}CRL_{w/o\ pref\ and\ semantic}$ results in worse performance compared to other variants.

Table 3: Variation of QTEF-CRL performance with semantic depth L

	Book-Crossing		Last-FM		Movielens-1M	
	AUC	F1	AUC	F1	AUC	F1
L=1	0.7897	0.7075	0.8806	0.7936	0.9382	0.8664
L=2	0.8042	0.7056	0.8860	0.8058	**0.9383**	**0.8666**
L=3	**0.8051**	**0.7095**	**0.8926**	**0.8117**	0.9353	0.8665
L=4	0.7995	0.6954	0.8917	0.8025	0.9329	0.8654

Table 4: Variation of QTEF-CRL performance with structural aggregation layer L'

	Book-Crossing		Last-FM		Movielens-1M	
	AUC	F1	AUC	F1	AUC	F1
L'=1	0.7882	0.6968	0.8851	0.8079	**0.9392**	**0.8683**
L'=2	0.8043	0.7184	**0.8916**	**0.8101**	0.9383	0.8659
L'=3	**0.8121**	**0.7266**	0.8880	0.7952	0.9369	0.8636
L'=4	0.7689	0.6588	0.8870	0.7999	0.9356	0.8611

Table 5: Varation of QTEF-CRL performance with preference depth P

	Book-Crossing		Last-FM		Movielens-1M	
	AUC	F1	AUC	F1	AUC	F1
P=1	0.7717	0.6875	0.8812	0.8019	0.9357	0.8613
P=2	0.8021	0.7192	0.8861	0.8062	**0.9385**	**0.8668**
P=3	**0.8103**	**0.7308**	0.8907	0.7964	0.9382	0.8654
P=4	0.8098	0.7296	**0.8917**	**0.8024**	0.9384	0.8665

Table 6: Variation of QTEF-CRL performance with α

	Book-Crossing		Last-FM		Movielens-1M	
	AUC	F1	AUC	F1	AUC	F1
0.1	0.8011	0.7162	0.8939	0.8064	0.9392	0.8671
0.2	**0.8121**	**0.7266**	**0.8947**	**0.8142**	**0.9398**	**0.8680**
0.4	0.8090	0.7253	0.8878	0.8064	0.9392	0.8679
0.6	0.8086	0.7230	0.8924	0.8094	0.9393	0.8673
0.8	0.8006	0.7112	0.8921	0.7980	0.9392	0.8675
1	0.8005	0.7106	0.8814	0.7969	0.9390	0.8641

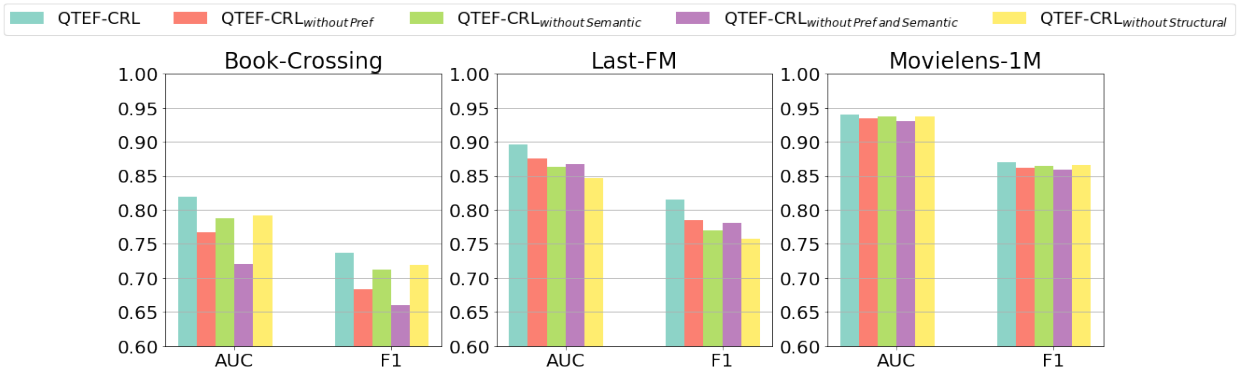

Figure 3: Ablation studies on the proposed QTEF-CRL model with different variants.

Table 7: Variation of QTEF-CRL performance with β

	Book-Crossing		Last-FM		Movielens-1M	
	AUC	F1	AUC	F1	AUC	F1
0.1	0.8101	0.7212	0.8954	0.8034	0.9381	0.8661
0.01	0.8117	0.7261	0.8906	0.7983	0.9396	0.8678
0.001	**0.8187**	**0.7374**	**0.8959**	**0.8157**	**0.9401**	**0.8680**
0.0001	0.8186	0.7373	0.8924	0.8070	0.9395	0.8679
0.00001	0.8068	0.7223	0.8933	0.8002	0.9393	0.8673

5.4 Selection of Hyperparameters (RQ3)

We first evaluate the impact of the number of semantic layers on QTEF-CRL by varying $1 \leq L \leq 4$ as shown in Table 3. We observed that, for all datasets, the performance reduces beyond $L = 3$ due to unwanted noise being introduced during recursive propagation. For the Book-crossing and Last-FM datasets that have more relations, applying two to three layers exhibits good performance. For Movielens-1M that has significant number of entities triplets within the knowledge graph, applying two layers can sufficiently capture the item-entity affiliation.

We next assess the influence of the structural layer on our model for $1 \leq L' \leq 4$. With reference to Table 4, we note that, similar to the number of semantic layers, the performance of QTEF-CRL reduces with large L' due to the noise introduced by $\mathcal{G}_s$ in CKG. Similar trend is observed where Movielens-1M achieves good performance with a single layer, whereas for Book-Crossing and Last-FM datasets, two or three layers are sufficient to capture significant associations to enhance the user and item representation learning.

We next vary the number of preference layers $1 \leq P \leq 4$ as shown in Table 5. We observe that, as opposed to the higher amount of noise with increasing number of layers, stacking GNN layers generally lead to over-smoothing and performance degradation. However, a further increase with $P = 4$ results in a higher AUC and F1 score. This is due to ability of randomized-SVD in reconstructing the user-entity graph—noise is reduced and that each user is now connected with entities of significant importance resulting in high-order propagation being effective across all datasets.

The hyperparameter α defined in (28) governs the influence of the tri-local contrastive loss within the overall loss. We examine the effect of α on the performance of QTEF-CRL by varying $0.1 \leq \alpha \leq 1$ as shown in Table 6. We observe that QTEF-CRL achieves the best performance for $\alpha = 0.2$ beyond which its performance starts to reduce. This reduction in performance is due to the removal of the global loss defined in (28). Since $\alpha = 1$ will result in $\mathcal{L}^{global}$ being removed from the optimization, this highlights the importance of structural information in the global view.

The significance of global contrastive loss is weighted by β and variation in QTEF-CRL performance for $1 \times 10^{-1} \leq \beta \leq 1 \times 10^{-5}$ is tabulated in Table 7. We note that QTEF-CRL achieves its optimal performance when $\beta = 1 \times 10^{-3}$ and suffers from worst performance when $\beta = 1 \times 10^{-5}$. These results imply that a low value of β reduces the impact of the global loss—β must be tuned with the recommendation loss defined in the objective function to achieve good performance.

6 CONCLUSION

We propose the QTEF-CRL for knowledge-aware recommendation that is trained via a pretext task for representation learning and a downstream task for recommendation. The tri-local, single-global architecture aims to fully utilize the knowledge graph holistically to construct augmented views with distinct properties. Through the proposed encoders, the model is able to distill intricate information about user preferences in both relational and deep interests level, allowing QTEF-CRL to outperform the state-of-the-art baselines. We also propose a new view that is designed to extract the user's deep interests at the level of non-local information concealed deep within the knowledge graph that is commonly unaccounted for by existing works. It is also worthwhile highlighting that QTEF-CRL involves a high number of hyperparameters, and that balancing them can be particularly complex. Thus, the challenge of simplifying such a sophisticated design while preserving its essential properties remains an open research question.

REFERENCES

[1] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems* 26 (2013).

[2] Xuheng Cai, Chao Huang, Lianghao Xia, and Xubin Ren. 2023. LightGCL: Simple Yet Effective Graph Contrastive Learning for Recommendation. In *The Eleventh International Conference on Learning Representations*.

[3] Rose Catherine and William Cohen. 2016. Personalized recommendations using knowledge graphs: A probabilistic logic programming approach. In *Proceedings of the 10th ACM conference on recommender systems*. 325–332.

[4] Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. 2020. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 3438–3445.

[5] Pengfei Chen, Qi Wang, and Yuan Tian. 2022. Exploring Entity-level User Preference on the Knowledge Graph for Recommender System. In *Proceedings of the 2022 5th International Conference on Algorithms, Computing and Artificial Intelligence*. 1–7.

[6] Shaohua Fan, Junxiong Zhu, Xiaotian Han, Chuan Shi, Linmei Hu, Biyu Ma, and Yongliang Li. 2019. Metapath-guided heterogeneous graph neural network for intent recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2478–2486.

[7] Chen Gao, Xiang Wang, Xiangnan He, and Yong Li. 2022. Graph neural networks for recommender system. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 1623–1625.

[8] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 249–256.

[9] M. U. Gutmann and A Hyvärinen. 2012. Noise-Contrastive Estimation of Unnormalized Statistical Models, with Applications to Natural Image Statistics. *JMLR.org* (2012).

[10] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. 2011. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review* 53, 2 (2011), 217–288.

[11] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

[12] Li He, Xianzhi Wang, Dingxian Wang, Haoyuan Zou, Hongzhi Yin, and Guandong Xu. 2023. Simplifying Graph-based Collaborative Filtering for Recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 60–68.

[13] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. 639–648.

[14] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*. 173–182.

[15] Ruoran Huang, Chuanqi Han, and Li Cui. 2021. Entity-aware collaborative relation network with knowledge graph for recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 3098–3102.

[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. [n. d.]. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*.

[18] Xuan Nhat Lam, Thuc Vu, Trong Duc Le, and Anh Duc Duong. 2008. Addressing cold-start problem in recommendation systems. In *Proceedings of the 2nd international conference on Ubiquitous information management and communication*. 208–211.

[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 *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 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 *Proceedings of the AAAI conference on artificial intelligence*, Vol. 29.

[21] Meng Liu, Hongyang Gao, and Shuiwang Ji. 2020. Towards deeper graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 338–348.

[22] Kelong Mao, Jieming Zhu, Xi Xiao, Biao Lu, Zhaowei Wang, and Xiuqiang He. 2021. UltraGCN: ultra simplification of graph convolutional networks for recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 1253–1262.

[23] A. Mnih and K. Kavukcuoglu. 2013. Learning word embeddings efficiently with noise-contrastive estimation. In *Proceedings Advances in Neural Information Processing Systems*. 2265–2273.

[24] Ajit Rajwade, Anand Rangarajan, and Arunava Banerjee. 2012. Image denoising using the higher order singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 4 (2012), 849–862.

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

[26] Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu. 2018. Recurrent knowledge graph embedding for effective recommendation. In *Proceedings of the 12th ACM conference on recommender systems*. 297–305.

[27] Chang-You Tai, Meng-Ru Wu, Yun-Wei Chu, Shao-Yu Chu, and Lun-Wei Ku. 2020. Mvin: Learning multiview items for recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 99–108.

[28] Riku Togashi, Mayu Otani, and Shin'ichi Satoh. 2021. Alleviating cold-start problems in recommendation through pseudo-labelling over knowledge graph. In *Proceedings of the 14th ACM international conference on web search and data mining*. 931–939.

[29] 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 *Proceedings of the 27th ACM international conference on information and knowledge management*. 417–426.

[30] 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 *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 968–977.

[31] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge graph convolutional networks for recommender systems. In *The world wide web conference*. 3307–3313.

[32] Qinqin Wang, Elias Tragos, Neil Hurley, Barry Smyth, Aonghus Lawlor, and Ruihai Dong. 2022. Entity-Enhanced Graph Convolutional Network for Accurate and Explainable Recommendation. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 79–88.

[33] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 950–958.

[34] 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 *Proceedings of the Web Conference 2021*. 878–887.

[35] Xiang Wang, Hongye Jin, An Zhang, Xiangnan He, Tong Xu, and Tat-Seng Chua. 2020. Disentangled graph collaborative filtering. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 1001–1010.

[36] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. 2019. Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5329–5336.

[37] Ze Wang, Guangyan Lin, Huobin Tan, Qinghong Chen, and Xiyang Liu. 2020. CKAN: collaborative knowledge-aware attentive network for recommender systems. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. 219–228.

[38] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 28.

[39] Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. 2019. Simplifying graph convolutional networks. In *International conference on machine learning*. PMLR, 6861–6871.

[40] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. 2021. Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 726–735.

[41] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems* 32, 1 (2020), 4–24.

[42] Xin Xin, Xiangnan He, Yongfeng Zhang, Yongdong Zhang, and Joemon Jose. 2019. Relational collaborative filtering: Modeling multiple item relations for recommendation. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*. 125–134.

[43] Yuhao Yang, Chao Huang, Lianghao Xia, and Chenliang Li. 2022. Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1434–1443.

[44] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2022. Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*. 1294–1303.

[45] 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 *Proceedings of the 7th ACM international conference on Web search and data mining*. 283–292.

[46] Fajie Yuan, Xiangnan He, Alexandros Karatzoglou, and Liguang Zhang. 2020. Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. In *Proceedings of the 43rd International ACM SIGIR conference*

on research and development in Information Retrieval. 1469–1478.

[47] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 353–362.

[48] Jinghao Zhang, Yanqiao Zhu, Qiang Liu, Shu Wu, Shuhui Wang, and Liang Wang. 2021. Mining latent structures for multimedia recommendation. In *Proceedings of the 29th ACM International Conference on Multimedia*. 3872–3880.

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

[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).

[51] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. 2021. Graph contrastive learning with adaptive augmentation. In *Proceedings of the Web Conference 2021*. 2069–2080.

[52] Ding Zou, Wei Wei, Xian-Ling Mao, Ziyang Wang, Minghui Qiu, Feida Zhu, and Xin Cao. 2022. Multi-level cross-view contrastive learning for knowledge-aware recommender system. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1358–1368.

[53] 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 & Knowledge Management*. 2817–2826.