



Preference-aware Graph Attention Networks for Cross-Domain Recommendations with Collaborative Knowledge Graph

YAKUN LI, LEI HOU, and JUANZI LI, Department of Computer Science and Technology,
Tsinghua University

Knowledge graphs (KGs) can provide users with semantic information and relations among numerous entities and nodes, which can greatly facilitate the performance of recommender systems. However, existing KG-based approaches still suffer from severe data sparsity and may not be effective in capturing the preference features of similar entities across domains. Therefore, in this article, we propose a **Preference-aware Graph Attention network** model with **Collaborative Knowledge Graph (PGACKG)** for cross-domain recommendations. Preference-aware entity embeddings with some collaborative signals are first obtained by exploiting the graph-embedding model, which can transform entities and items in the collaborative knowledge graph into semantic preference spaces. To better learn user preference features, we devise a preference-aware graph attention network framework that aggregates the preference features of similar entities within domains and across domains. In this framework, multi-hop reasoning is employed to assist in the generation of preference features within domains, and the node random walk based on frequency visits is proposed to gather similar preferences across domains for target entities. Then, the final preference features of entities are fused, while a novel **Cross-domain Bayesian Personalized Ranking (CBPR)** is proposed to improve cross-domain recommendation accuracy. Extensive empirical experiments on four real-world datasets demonstrate that our proposed approach consistently outperforms state-of-the-art baselines. Furthermore, our PGACKG achieves strong performance in different ablation scenarios, and the interaction sparsity experiments also demonstrate that our proposed approach can significantly alleviate the data sparsity issue.

CCS Concepts: • **Information systems** → **Recommender systems**;

Additional Key Words and Phrases: Knowledge graph, graph attention network, preference-aware embeddings

ACM Reference format:

Yakun Li, Lei Hou, and Juanzi Li. 2023. Preference-aware Graph Attention Networks for Cross-Domain Recommendations with Collaborative Knowledge Graph. *ACM Trans. Inf. Syst.* 41, 3, Article 80 (January 2023), 26 pages.

<https://doi.org/10.1145/3576921>

This work is supported by the New Generation Artificial Intelligence of China (grant no. 2020AAA0106501) and a grant from the Institute for Guo Qiang, Tsinghua University (grant no. 2019GQB0003).

Authors' address: Y. Li, L. Hou, and J. Li (corresponding author), Tsinghua University, 30 Shuangqing Rd, Haidian District, Beijing, China; emails: {liyakun, houlei}@mails.tsinghua.edu.cn, lijuanzi@tsinghua.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Association for Computing Machinery.

1046-8188/2023/02-ART80 \$15.00

<https://doi.org/10.1145/3576921>

1 INTRODUCTION

Recommender systems can assist users to quickly find the desired information from a large amount of data in many application scenarios, such as book recommendations on Amazon, paper recommendations on AMiner, and video recommendations on YouTube. However, the inherent data sparsity issue still seriously hinders performance improvement of existing recommendation approaches [1]. Therefore, cross-domain recommendations (CDRs) provide another promising solution [2, 3], which can leverage preference knowledge in the source domain to assist in improving entity recommendation performance in the target domain. Despite many research efforts devoted to developing CDRs, most existing approaches [4, 5] consider interactive rating records based only on overlapping entities, which makes it difficult to substantially improve the accuracy of recommendations.

A knowledge graph (KG) is a heterogeneous information network [6] that can capture structured semantic information and contextual preference features between nodes, where nodes correspond to entities, items, and so forth, and edges correspond to relations. Essentially, knowledge graph embedding (KGE) can expose user preference features to some extent and yield satisfactory representations for network training [7]. Therefore, collaborative knowledge graphs with rich preference information in some of the literature [8, 9] have been demonstrated to effectively improve recommendation performance.

Currently, most KG-based recommendations are mainly classified into two categories: path-based schemes and embedding-based schemes. Path-based schemes can extract higher-order connectivity information of nodes carrying rich preference signals, which often requires manual design of traversal principles. While these approaches contribute to the interpretability of recommendations, their performance relies heavily on domain knowledge, resulting in less reliable scalability. Recently, motivated by the development of deep training networks, embedding-based schemes that aggregate the feature embeddings for entities in knowledge graphs have been proposed to improve recommendation performance in an end-to-end manner. However, these schemes still have too many parameter dependencies and are limited to single-domain recommendations [10], which cannot be applied well to cross-domain scenarios.

To tackle this challenge, in this article, we attempt to expose preference signals from entity interactions in the collaborative knowledge graph to learn similar preference features within and across domains for target entities, which can enhance the recommendation performance of the target domain. Figure 1 illustrates our cross-domain recommendation between music and movie domains with the collaborative knowledge graph. Suppose that for a given overlapping user u_3 in the movie domain B, our recommender system needs to recommend movies that the user may like. Since u_3 also belongs to users in the music domain A, according to a known path $u_3 \rightarrow i_2 \rightarrow e_2 \rightarrow i_3$ in the collaborative knowledge graph, the system judges that u_3 is likely to prefer i_3 . Then, based on the preferences of overlapping entities across domains, we can generate final predictions for the target entity u_3 . Similarly, for those non-overlapping entity preferences, we can use overlapping entities as bridging nodes to connect the two domains to make preference predictions.

Therefore, in this article, we propose a novel preference-aware graph attentive model with a collaborative knowledge graph for cross-domain recommendations that is equipped with knowledge graphs to capture entity semantic information and user potential long-distance preferences across domains. Specifically, we leverage a trainable and personalized graph representation scheme to transform entities or items into preference-aware embeddings. Considering the limitations of existing solutions, a preference-aware graph attention network is then devised to aggregate preference features of similar entities within domains and across domains. Finally, the fused entity features with rich contextual preference information are obtained, and the cross-domain Bayesian personalized ranking is proposed to generate predictive results for different cross-domain

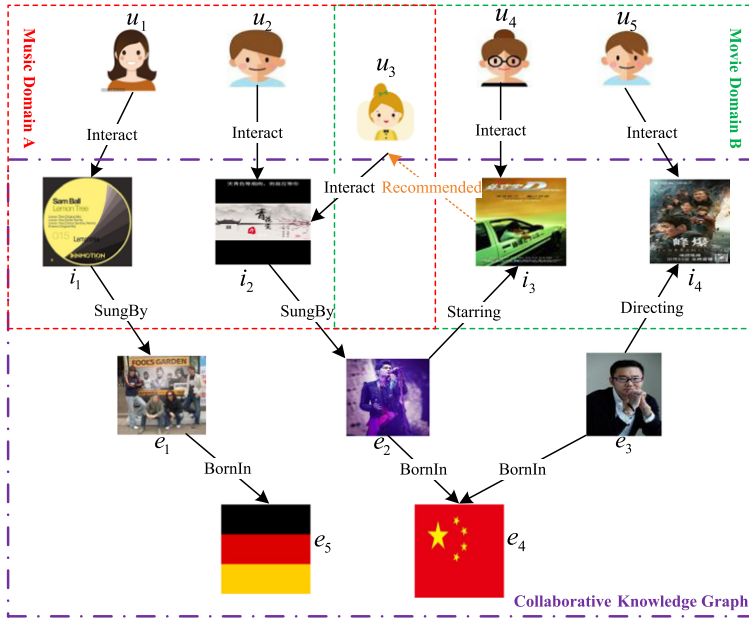


Fig. 1. An illustration of our cross-domain recommendation between music and movie domains with the collaborative knowledge graph.

recommendation scenarios. In addition, experimental results also show that our proposed Preference-aware Graph Attention network model with Collaborative Knowledge Graph (PGACKG) achieves significant gains over state-of-the-art approaches in recommendation accuracy.

Our work makes the following main contributions:

- We propose the PGACKG, a new cross-domain recommendation framework based on graph attention networks, which can leverage similar entity preferences across domains to improve prediction performance in target domains. To the best of our knowledge, this is the first attempt to apply knowledge graphs to cross-domain recommender systems to enhance their performance.
- We propose to learn preference-aware embeddings for network training by exploiting a graph representation model. Some collaborative preference signals are exposed to learn entities with different preferences in the collaborative knowledge graph.
- We propose a preference-aware graph attention network model to aggregate the preferences from similar entities within domains and across domains. Intuitively, we employ a multi-hop reasoning process to extract entity preference features within domains, and develop a node random walking scheme to model preference features across domains.
- We conduct extensive experiments on several real datasets to demonstrate the effectiveness of the PGACKG. Experimental results indicate that our proposed PGACKG consistently outperforms state-of-the-art baseline recommendation approaches. Meanwhile, the data sparsity issue is greatly alleviated to improve recommendation performance in the target domain.

The remainder of this article is organized as follows. Section 2 reviews some latest works relevant to our research topic. Section 3 introduces the formulation of the cross-domain recommendation problem studies in this article. Section 4 presents our proposed PGACKG model in detail.

Extensive experiments performed on four real-world datasets demonstrate the effectiveness of our approach in Section 5. Our conclusions and future research directions are summarized in Section 6.

2 RELATED WORK

In this section, we review some existing works related to our research, graph neural networks, recommendations with knowledge graphs, and cross-domain recommendations.

2.1 Graph Neural Networks

Based on the differences in graph representation training architecture, most graph neural network (GNN) models are divided into graph convolutional neural networks and graph generative neural networks.

Graph convolutional neural networks (GCNs) can leverage convolution operations to generalize complex types of data into graph data, which is the basis for many derived GNN models. In the recent literature, spectral-based solutions have attracted considerable attention, and they generally formalize convolution operations for graph representations from the perspective of graph signal processing. For example, the authors of [11] propose a new spectral domain convolutional architecture, which can efficiently compute spectral filters on graphs by handling different constructions of Laplacian operators. To train learnable parameters and consider graph signals of multiple dimensions, the authors of [12] devise spectral convolution layers to constrain channel filters. The authors of [13] and [14] define a convolutional filter as Chebyshev polynomials of the diagonal matrix of eigenvalues. Compared with these spectral GCNs, non-spectral GCN solutions can directly employ graph convolution operations to propagate node information and relations along those edges. The authors of [15] and [16] pioneered spatial-based frameworks, which can recursively stack and propagate representations of node neighborhoods to memorize interlayer information. To enhance the inherent transduction of nodes and naturally generalize to unseen nodes, the authors of [17] present a novel general inductive framework, GraphSAGE, which leverages node feature information to efficiently generate node embeddings by sampling and aggregating features from a node's local neighborhood. The learnable graph convolutional layer is employed to automatically select neighbor nodes to perform convolution operations on generic graphs [18].

The graph-generating neural network (GGN) is a generative learning architecture that aims at re-encoding node or graph information to generate plausible structures from data. Many approaches to GGNs exploit observed data distributions to learn plausible data representations. For instance, the authors of [19] propose a novel deep training scheme for learning generative models over graphs, which can leverage GNNs to express probabilistic dependencies for nodes and edges, thereby learning distributions over any arbitrary graph and updating the graph representation. The literature [20] applies generative adversarial networks to generate discrete output samples by learning the distribution of biased random walks over the input graph. To address meaningless training constraints, the authors of [21] and [22] adopt different graph generative models to consistently encode graph structure information for node embeddings.

2.2 Recommendations with Knowledge Graphs

In general, existing KG-based recommendations aim to leverage KGs to improve prediction accuracy. These approaches are almost all applied to single-domain recommender systems and are roughly categorized into two groups: embedding-based solutions and path-based solutions.

Embedding-based solutions leverage the KG embedding model to capture KG structure or semantic information and then learn the entity embeddings to achieve the final recommendation. For example, a novel collaborative knowledge embedding model [23] is proposed to learn item latent representations in collaborative filtering and entity representations from KGs that can boost the

recommendation performance. To further optimize the knowledge base embeddings, the authors of [24] leverage knowledge-guided interaction sequences to enhance the representation power in capturing complicated entity preferences for improving sequential recommendations. In [25], considering KGs as the source of side information, entity latent features and their interactions can be associated by cross-compress units through the KG embedding mechanism. The authors of [26] propose a new translation-based representation model that can transfer entity relation information in KGs and jointly train them with a KG completion model. However, due to the lack of explicit modeling capabilities, embedding-based schemes are not guaranteed to capture long-range node dependencies and find it difficult to interpret high-order semantic relations between entities for recommendations.

Path-based solutions extract entity paths rich in high-order node information in the KG to provide some preference signals for entity recommendations. For instance, knowledge-aware path recurrent networks [8] are exploited to recursively generate path representations to discriminate the importance of different paths by composing the semantics of both entities and relations for explainable recommendations. In [27], the meta-path-based context is constructed to learn effective representations for entities in heterogeneous information networks, whereas a co-attention mechanism is adopted to assist in improving Top-N recommendations. Similarly, the authors of [28] leverage the matrix factorization and factorization machine to generate latent features for entities in heterogeneous information networks, thereby automatically learning from the known ratings to efficiently select useful feature-based meta-graphs. Meta-path-based random walking and heterogeneous skip-gram strategies are proposed to further enable structural and semantic associations of entities in KGs [29]. Some other approaches [30–32] have also explored various patterns of connections in KGs to assist various recommendation scenarios. While these approaches facilitate and enrich the diversity of recommendations in an intuitive way, they heavily rely on domain knowledge and manually designed path-based patterns, making them difficult for large-scale industrial applications.

Recently, some promising reinforcement learning (RL)-based approaches over the KG have also been proposed to improve recommendation performance. For example, the authors of [33] devise a demonstration-based KG reasoning framework for explainable recommendation, and also propose an ADversarial Actor-Critic model for demonstration-guided path finding. To improve sampling efficiency and user experiences, the prior knowledge of item correlations learned from KGs is exploited to enrich item representations, and user preferences are propagated for interactive recommender systems [34]. Motivated by the availability of KGs, the authors of [35] propose a knowledge-guided RL model to fuse KG information into the RL architecture for sequence recommendations. The authors of [36] develop a negative sampling model in which a KG policy network is designed to explore high-quality negative samples. Naturally, deep RL-based recommendation models over the KG have shown strong advantages in solving complex tasks and dealing with complex data. However, these methods cannot be applied to cross-domain recommendation scenarios due to distribution drift and data sparsity issues.

2.3 Cross-Domain Recommendations

CDRs aim to leverage rich data from auxiliary domains to assist in solving the long-standing data sparsity issue in the target domain to improve recommendation accuracy. According to different knowledge transfer strategies [37], most existing CDR approaches are mainly divided into two categories: feature-based solutions and content-based solutions.

Feature-based solutions mainly attempt to exploit collaborative filtering techniques or some deep training models to simultaneously learn entity latent features of the two domains to improve the recommendation performance of the target domain. For example, the authors of [38] propose

a deep transfer collaborative filtering architecture in which the collective matrix factorization and deep transfer learning techniques are integrated to generate efficient entity latent representations for CDRs. To alleviate data sparsity, the authors of [39] employ a multilayer perceptron to capture the nonlinear mapping relationship across domains and then learn the domain-specific features of entities. In [40], a deep sparse autoencoder recommendation framework based on adversarial learning is proposed to enhance the quality of recommendations in a CDR system. In addition, the authors of [41] devise a CDR model based on a probabilistic knowledge transfer mechanism to profile both domain-shared and domain-specific entity features, thereby improving recommendation accuracy. However, although many example solutions [42–46] have also been proposed to analyze entity features and provide some preference knowledge for sparse target domains, how to generate accurate representations for entities is still a crucial challenge.

Content-based solutions tend to leverage content information (entity attributes, side information, semantic associations, social relations, etc.) to connect the auxiliary domain and target domain to generate the preference knowledge for recommendations in the target domain. For instance, the authors of [47] introduce a novel CDR model based on tag semantic association, which can automatically capture the semantic relationship between nonidentical tags, and similar entities across domains are identified for recommendations by transferring preference knowledge from the auxiliary domain. Targeting improvement of the accuracy of CDRs, multiple types of media information are exploited to transfer user interests through a proposed Bayesian hierarchical scheme based on Latent Dirichlet Allocation [48]. Nevertheless, these recommendation effects are not always as expected in various application scenarios. Therefore, a framework of content-based metadata features is implemented with various classifiers to segment different user groups and provide recommendations for target domains in [49]. The authors of [50] employ an unsupervised domain adaptation technique to analyze cross-domain relations and preference behaviors among entities based on the historical contents browsed by users. In addition, there are some recent content-based approaches [51–54] that can leverage explicit or implicit attributes to improve the quality of CDRs at different levels.

3 PROBLEM FORMULATIONS

In those typical CDR scenarios, it is assumed that there is a dense source domain S and a sparse target domain T . Generally, each domain has a set of M users $U = \{u_1, u_2, \dots, u_m\}$, a set of N items $I = \{i_1, i_2, \dots, i_n\}$, and an interaction rating matrix $R \in R^{M \times N}$. $r_{mn} = 1$ indicates some historical preference behavior (browsing, liking, forwarding, etc.) of user u_m on item i_n ; otherwise, $r_{mn} = 0$. Additionally, the existence of some shared entities (users or items) in the source domain and the target domain is a prerequisite for CDRs. For convenience, we assume that the two domains have some overlapping users in this article.

Collaborative Knowledge Graph (CKG). Inspired by previous work [55], a collaborative KG is defined to facilitate our work, which can encode and project user behaviors and item information into a multi-relational graph. The CKG not only reflects the preference interaction between users and items in the recommendation system, but also exposes the relationship between entities in the real world. Normally, it is composed of entity-relation-entity triples (h, r, t) , where $h \in E, r \in \Omega, t \in E$ denote the head entity, relation, tail entity of a knowledge triple, and E and Ω , respectively, represent the set of entities and relationships in the CKG. For example, the triplet $(\text{Vin Diesel}, \text{ActorOf}, \text{Fast \& Furious 9})$ presents the fact that Vin Diesel is an actor of the movie “Fast & Furious 9.” Items in the recommender system are regarded as a special type of entity in the CKG, that is, i belongs to E .

Problem Description. Given a dense source domain interaction matrix X , a sparse target domain interaction matrix Y and the CKG G_c , we aim to predict whether user u in the target domain

Table 1. Main Symbols Used in This Article

Symbol	Description
U	set of users in recommender systems
I	set of items in recommender systems
S	dense source domain
T	sparse target domain
R	interaction rating matrix
E	set of entities in the KG
Ω	set of relations in the KG
h	entity in E
e_h	preference embedding for h
e_r	preference embedding for r
Γ, Γ'	golden and negative triplet sets
τ, τ'	golden and negative triplets
ϑ	margin separating golden and negative triplets
v	node in the KG
$N_u(v)$	set of similar neighbors of node v
h_v	vector feature of node v
h_k	vector feature of node k
W_{vk}	shared attention parameter for h_v and h_k
a_{vk}	preference attention coefficient for nodes v and k
h_v^c	preference attentive embedding across domains
h_v^w	preference attentive embedding within domains
h_v^f	fused preference representation
Θ	all trainable parameters
η	regularization parameter that constrains entities and relations
L_{PGACKG}	final objective function

Y has potential interest with which the user has had no interaction before. In other words, the problem to be solved in this article is to predict the probability that u in the target domain Y would adopt item i based on the known rating matrices in the source and target domains. We list the main symbols used in this article in Table 1.

4 THE PROPOSED METHOD

In this section, we present the proposed PGACKG approach, the architecture of which is shown in Figure 2. Our framework consists of four main components: (a) preference embedding transformation that transforms entities and nodes in the CKG into vector representations; (b) preference-aware graph attention layer that propagates and leverages embeddings from the source and target domains to generate user-attentive preference features within domains and across domains; (c) entity feature fusion that aggregates the preference features of similar entities from different domains; and (d) model prediction that employs the CKG to construct cross-domain prediction models and outputs the predictable matching ratings.

4.1 Preference Embedding Transformation

The graph-embedding model can provide users with semantic preference information for cross-domain recommendations while preserving the graph structure of the CKG. The different attributes of entity nodes in the graph reveal different interaction preferences of users. Therefore, we

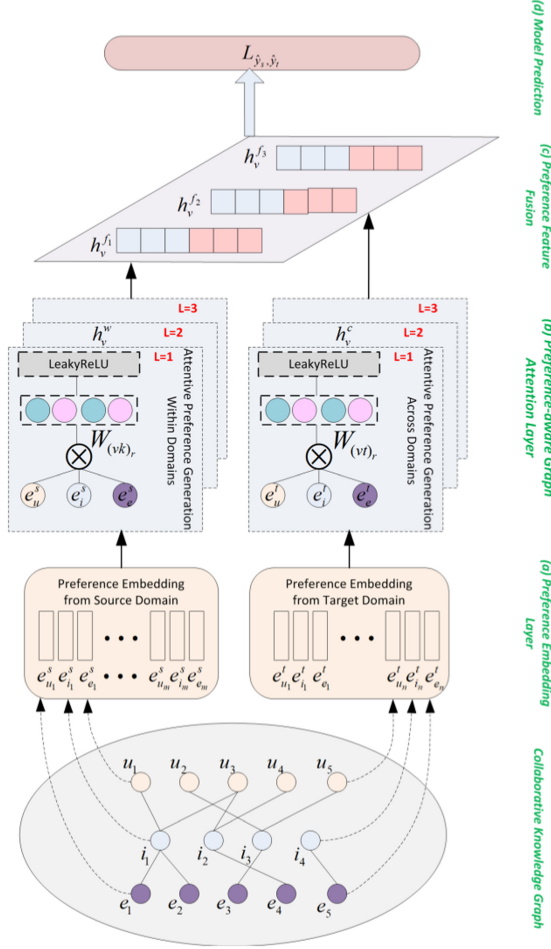


Fig. 2. The architecture of our proposed cross-domain recommendation PGACKG.

employ an advanced embedding model TransD [56] to characterize user preferences. To be more specific, based on the well-known transformation principle $e_h^r + e_r = e_t^r$, multi-type attributes of entities and relations are learned if and only if a triple (h, r, t) exists in the KG. Therefore, the score function for a given triple is defined as follows.

$$f(h, r, t) = \|W_{rh}e_h + e_r - W_{rt}e_t\|_2^2. \quad (1)$$

Herein, W_{rh} and W_{rt} are two mapping matrices that can project entities from entity spaces into the relation spaces. e_h, e_r and e_t are the preference embeddings for h, r , and t , respectively. As can be seen from Equation (1), a lower score reveals that the triple is more likely to be true, and vice versa.

To obtain real training results, the training of TransD encodes all available triples in the KG, including the golden triples and negative triples. Thus, the margin-based ranking loss is defined as the objective function to train the sample set. Γ and Γ' denote the golden and negative triplet sets, respectively:

$$L_{CKG} = \sum_{\tau \in \Gamma} \sum_{\tau' \in \Gamma'} (\vartheta + f(\tau) - f(\tau')), \quad (2)$$

where τ and τ' are the golden and negative triplets, respectively. ϑ represents the margin separating golden triplets and negative triplets. We can employ the preference embeddings as input to guide the generation of user preferences within domains and across domains, and L_{CKG} can be utilized as the regularization to constrain the training of entities and relations to accelerate convergence and avoid overfitting. The main reason that TransD is adopted as our KG representation approach is that it can capture multi-type relations of entities and use fewer training parameters. This makes it possible to apply large-scale KGs in CDR systems.

4.2 Preference-Aware Graph Attention Layer

4.2.1 Attentive Preference Generation within Domains. User preference features within domains play a fundamental role in modeling the ultimate preferences of a target user because intra-domain node relations can reveal its intrinsic intentions. Therefore, the intra-domain preference acquisition of the target user needs to aggregate the preferences from similar neighbor nodes in the same domain. In this article, a multi-hop reasoning mechanism [57] is adopted to aggregate similar neighbors of target nodes in the same domain. Motivated by the advantages of the Transformer architecture [58] in capturing node semantics, the graph attention approach is leveraged to learn the preference weights of nodes in different aspects. In addition, there may be multiple relations between two entities, and target users may be interested in items because of a certain relation. To this end, assume that Ω' is the set of relations between the two entities and $N_u(v)$ is the set of similar neighbors of node v within domains. Then, the preference attention weight of node v for each neighbor in $N_u(v)$ is calculated as follows.

$$e_{(vk)_r} = A\left(W_{(vk)_r} \cdot h_v \parallel W_{(vk)_r} \cdot h_k\right), k \in N_u(v) \text{ \& } r \in \Omega', \quad (3)$$

where $W_{(vk)_r}$ is the shared attention parameter for the vector features of nodes h_v and h_k . $\parallel$ denotes a concatenation operation of low-dimensional vectors. $A(\cdot)$ denotes the attention function, which can expose the similarity between different grouped nodes to some extent. In our approach, dot-product attention is adopted as an attention function because it can be much faster and more space efficient. We can normalize the preference attention weight as follows:

$$a_{(vk)_r} = \frac{\exp\left(\text{LeakyReLU}\left(e_{(vk)_r}\right)\right)}{\sum_{j \in N_u(v)} \exp\left(\text{LeakyReLU}\left(e_{(vj)_r}\right)\right)}. \quad (4)$$

The preference attention coefficient $a_{(vk)_r}$ shows the contribution degree of the node k within domains to the preference weight of the central node v under the constraints of the relation r . In other words, it can directly reveal the differentiated preferences between different nodes and the central node in a separate domain. Then, the low-dimensional preference representations of similar nodes within domains are weighted and summed to obtain a new preference embedding of node v in a separate domain:

$$h_v^w = \sigma\left(\sum_{k \in N_u(v)} a_{(vk)_r} W h_k\right). \quad (5)$$

Note that these attentive preference embeddings are generated based on a first-order preference-aware graph attention mechanism within domains. Thus, to capture more semantic preferences of similar neighbors in a separate domain, a k-hop-based approach [59] is utilized to explore high-order user preferences. Finally, high-order preference embeddings are obtained and injected with collaborative preference signals into representation learning and network training.

4.2.2 Attentive Preference Generation Across Domains. Intuitively, similar user preferences across domains can transfer and provide similar preference information for cold-start users in the target domain, thereby alleviating the low recommendation accuracy of the target domain. In view of the domain heterogeneity and the difference in data distribution in two different recommendation systems, a node-walking approach [60] is adopted to explore the cross-domain similar nodes in the collaborative KG. This scheme aims to aggregate the preference features of similar nodes across domains to more accurately model the profiles of target users and enhance the generalization performance of the overall model. In Equation (6), we assume that t is a cross-domain similar neighbor of the target node, $N'_u(v)$ is the set of all cross-domain similar neighbors, r is the relation between two entities, and n is the number of cross-domain similar neighbors. Based on the attention mechanism, Mean Aggregator [61] is employed to obtain the attentive preference vectors across domain of target nodes.

$$h_{across} = \frac{1}{n} \sum_{t \in N'_u(v)} \left(A \left(W_{(vt)_r} h_t \right) \right), \quad r \in \Omega'', \quad (6)$$

where $W_{(vt)_r}$ denotes the trainable attentive weight to distill preference information for propagation, $A(\cdot)$ denotes the attention function, and Ω'' is the set of relations between the two entities. The low-dimensional embeddings of the current target node v are propagated and combined to perceive user preferences, and the final attentive preference embeddings across domains are obtained through the representation propagation of the multi-full connection layer.

$$h_v^c = \text{LeakyReLU}([h_v \parallel h_{across}]), \quad (7)$$

where h_v^c is the preference attentive embedding across domains. The neighbors of the cold-start nodes may not have nodes across domains and the multi-hop reasoning model has difficulty finding cross-domain similar neighbors of such nodes. As a result, the node-walking algorithm across domains based on frequency visits (FVNW) is designed for our cross-domain scenario, as shown in Algorithm 1.

ALGORITHM 1: FVNW

Input: The target node t , the length of the global walk L , the sampling number of a node n and the set of edges in a graph E

Output: The collection of similar nodes across domains for the target node $N'_u(v)$

1: **Initialize** the collection of similar nodes across domains $N'_u(v)$

2: **For** walkingStep = 0, ..., L **do**

3: currNode = t

4: **For** $j = 1, \dots, \text{walkingLength}$ **do**

5: t_{Next} = Samples randomly from the edge set of the current graph E

6: **IF** $t_{\text{Next}} \in N_{\text{Cross}}$

7: $n++$; $t_{\text{Next}} \rightarrow N'_u(v)$

8: **Else:** $n++$

9: **End For**

10: **End For**

11: **Return** $N'_u(v)$

The following acquisition process of the target node's similar neighbors across domains is mainly as follows. Given the current target node t , the next neighbor node is randomly selected to start the deep walk. When the target node walks the specified length L , the visited nodes are sorted according to the visited frequency of nodes during the whole walk. Then, the algorithm judges whether the traversed node belongs to a cross-domain node. Finally, the nodes with higher

frequency are selected as the similar neighbors of the target node to assist in remodeling the preference attentive embeddings across domains. The advantage of our proposed FVNW approach is that it can better assist cold users in finding cross-domain similar neighbors, explore more preferred items across domains, and enhance the generalization ability of entity preferences. Therefore, Algorithm 1 is the premise and foundation for acquiring attentive embeddings across domains. A well-designed attentive preference-generation approach essentially reflects high-quality cross-domain similar neighbor sets.

4.3 Entity Preference Feature Fusion

Naturally, the preference features of similar entities from different domains have different contributions to remodeling the user's final preferences. Based on the embeddings of similar nodes within domains and across domains for the target node v , its new preference representations are finally fused by nonlinear mapping as shown next.

$$h_v^f = \text{LeakyReLU} [\xi h_v^w + (1 - \xi) h_v^c], \quad (8)$$

where ξ denotes a balance parameter of preference representations, which is used to control the degree of preference transfer across domains. In particular, when ξ is set to 1, the nodes across domains do not participate in the model learning, that is, any cross-domain entity and item information is not exploited to provide preference features for recommendations. In contrast, when ξ is controlled to 0, preference learning of the target node originates from cross-domain nodes and the feature information of similar nodes within domains will not be applied to the final preference predictions. Therefore, a new user preference-aware representation approach that merges the preferences of similar neighbors within domains and across domains (UPRWA) is shown in Algorithm 2.

The main steps of Algorithm 2 are as follows. First, some nodes are randomly selected from the entity set to form a batch. Then, according to the sampling depth, the corresponding P-order similar neighbors within domains and Q-order similar neighbors across domains are obtained based on the mapping representation function. In accordance with the aggregated similar neighbors, new low-dimensional representations of target nodes are generated to facilitate the network training, while propagating entity representations from first-order to higher-order connections can assist the model in more reasonably exploring user potential preference features in a deeper and broader way. Finally, based on the devised preference-aware graph attention layer, the fused embeddings are obtained to model user final preferences.

4.4 Cross-domain Recommendation and Model Optimization

Naturally, the preference features of similar entities from different domains have different contributions to remodeling the user's final preferences. Based on the embeddings of similar nodes within domains and across domains for the target node v , its new preference representations are finally fused by nonlinear mapping as shown here.

After performing L layers in cross-domain node propagation, multiple representations are obtained for user and item nodes in the source domain, that is, $\{e_{u_s}^{(1)}, e_{u_s}^{(2)}, \dots, e_{u_s}^{(L_1)}\}$ and $\{e_{i_s}^{(1)}, e_{i_s}^{(2)}, \dots, e_{i_s}^{(L_1)}\}$. Similar to the multiple node representations in the source domain, $\{e_{u_t}^{(1)}, e_{u_t}^{(2)}, \dots, e_{u_t}^{(L_2)}\}$ and $\{e_{i_t}^{(1)}, e_{i_t}^{(2)}, \dots, e_{i_t}^{(L_2)}\}$ are obtained for the multiple representations of user and item nodes in the target domain under the constraint $L_2 = L - L_1$. Since the embeddings of nodes on different layers emphasizes the preference messages over different connections, they have different contributions to the final fusion preferences of the target node. Therefore, the layer-aggregation strategy [62] is

ALGORITHM 2: UPRWA

Input: The Original node representation h , sampling depth within domains P , sampling depth across domains Q and the mapping representation function O

Output: The new user preference-aware representation h_v^f

```

1: Initialize intra-domain and inter-domain preference representations  $h_v^w$  and  $h_v^c$  of the target node  $u$ 
2: For  $i = 1, \dots, P$  do
3:    $N_u^i(v) \leftarrow N_u^{i-1}(v) + O_w(u)$ 
4:    $[h_u^w]^i \leftarrow [h_u^w]^{i-1}$ 
5: End For
6: For  $j = 1, \dots, Q$  do
7:    $[N_u'(v)]^j \leftarrow [N_u'(v)]^{j-1} + O_c(u)$ 
8:    $[h_u^c]^j \leftarrow [h_u^c]^{j-1}$ 
9: End For
10: Obtaining the new user preference-aware representation  $h_v^f$  using Equation (8)
11: Return  $h_v^f$ 

```

adopted to concatenate the representations within domains and across domains, respectively.

$$e_{u_s}^* = e_{u_s}^{(0)} \parallel e_{u_s}^{(1)} \dots \parallel e_{u_s}^{(L_1)}, \quad e_{i_s}^* = e_{i_s}^{(0)} \parallel e_{i_s}^{(1)} \dots \parallel e_{i_s}^{(L_1)} \quad (9)$$

$$e_{u_t}^* = e_{u_t}^{(0)} \parallel e_{u_t}^{(1)} \dots \parallel e_{u_t}^{(L_2)}, \quad e_{i_t}^* = e_{i_t}^{(0)} \parallel e_{i_t}^{(1)} \dots \parallel e_{i_t}^{(L_2)}, \quad (10)$$

where s and t denote the source and target domain labels, respectively. $\parallel$ is the concatenation operation. Based on this processing, we can not only obtain the preference-aware representations of similar nodes within domains and across domains, but also flexibly control the propagation range by adjusting L_1 and L_2 . In addition, the advantage of the concatenation mechanism is that it does not need to learn too many parameters, plus the training process is very simple and efficient. Therefore, the user and item representations of the source and target domains are respectively conducted inner products to obtain their respective matching scores:

$$\hat{y}_s(u_s, i_s) = [e_{u_s}^*]^T e_{i_s}^* \quad (11)$$

$$\hat{y}_t(u_t, i_t) = [e_{u_t}^*]^T e_{i_t}^*. \quad (12)$$

The Bayesian personalized ranking (BPR) strategy as an advanced baseline recommendation model [63] is often exploited to learn model parameters and predict entity ratings. Its basic principle is that the observed entity ratings, which can better expose users' true preferences, should be given a higher predictive weight than unobserved ones in a single-domain recommendation system. Therefore, based on the analysis presented earlier, a Cross-domain Bayesian Personalized Ranking method is proposed that can add similar preference features in the source domain as a regularization term to the BPR for predicting the unobserved interaction ratings in the target domain:

$$L_{CBPR} = \sum_{(u_t, i_t, j_t) \in \kappa} -\ln \sigma(\hat{y}_t(u_t, i_t) - \hat{y}_t(u_t, j_t)) + \lambda \sum_{(u_s, i_s, j_s) \in \kappa'} -\ln \sigma(\hat{y}_s(u_s, i_s) - \hat{y}_s(u_s, j_s)), \quad (13)$$

where κ and κ' denote the training samples in the source and target domains, respectively; $\sigma(\cdot)$ is the sigmoid function; λ is the regularization parameter controlling the fusion degree of similar preferences across domains. Thus, the final objective function is presented as follows:

$$L_{PGACKG} = L_{CBPR} + \eta L_{CKG} + \beta \|\Theta\|_2^2, \quad (14)$$

where Θ represents all trainable parameters and η is a regularization parameter that constrains entities and relations in the collaborative KG. In order to improve training efficiency, the Mini-Batch Gradient Descent strategy [64] is adopted to optimize the embedding loss and the proposed cross-domain prediction model. For a batch of randomly sampled triples, their preference representations within domains and across domains are established after multi-step propagations. Finally, the proposed model updates the parameters by the gradients of the prediction function. The optimization process of the proposed PGACKG approach is summarized in Algorithm 3.

ALGORITHM 3: PGACKG Optimization

Input: Observed source domain interactions X_s , observed target domain interactions Y_t , collaborative Knowledge Graph $\mathcal{G}$

Output: Prediction function $\mathcal{F}(u_t, i; \Theta)$

```

1: Initialize all trainable parameters
2: Construct the training sets  $\hat{X}_s$  and  $\hat{Y}_t$  based on  $X_s$  and  $Y_t$ 
3: For  $p_s = 1, 2, 3, \dots, M$  do
4:   For  $q_t = 1, 2, 3, \dots, N$  do
5:     Sample a mini-batch of training data  $\mathcal{D}_s$  from  $\hat{X}_s$ ;
6:     Sample a mini-batch of training data  $\mathcal{D}_t$  from  $\hat{Y}_t$ ;
7:     Parameter gradients computed by back-propagation according to Equations (13) and (14);
8:     Updated  $\Theta$  by Gradient Descent based on tuples in  $\mathcal{D}_s$  and  $\mathcal{D}_t$ ;
9:   End For
10: End For
11: Return  $\mathcal{F}(u_t, i; \Theta)$ 
  
```

5 EXPERIMENTS

In this section, we conduct empirical experiments on real datasets to verify the effectiveness of our proposed approach and answer the following research questions:

- **RQ1:** Does our proposed PGACKG method achieve better performances on different datasets than existing advanced baseline recommendations?
- **RQ2:** How does the PGACKG perform over different user groups with different interaction sparsity levels?
- **RQ3:** How do different parameter settings (e.g., Model Depth, Iteration Times for Layers, Dropout, Random Walks) affect the proposed PGACKG?
- **RQ4:** How do different variants or key components affect the PGACKG?
- **RQ5:** How does the PGACKG perform in visualization experiments?

5.1 Dataset Descriptions

To effectively verify the recommendation performance of our proposed PGACKG approach and some baseline methods, four public real-world datasets are adopted as our experimental datasets: three Amazon datasets¹ (AmazonMusic, AmazonMovie, AmazonBook) and the Book-Crossing dataset.² For these datasets, we keep only those items that have more than 10 interactions with the users to ensure that the model has enough dense entities for training. This article mainly focuses on how CKGs can improve the prediction accuracy of cross-domain recommendations,

¹<http://jmcauley.ucsd.edu/data/amazon/>.

²<http://www2.informatik.unifreiburg.de/~ciegler/BX/>.

Table 2. Statistics of the Four Experimental Datasets

	Book-Crossing	Amazon		
Domains	Books	Books	Movie	Music
# Users	16,327	75,872	42,763	2,876
# Items	17,915	89,640	6,029	4,901
# Interactions	143,007	10,651,427	4,760,416	59,327
# Entities	21,341	64,228	38,225	7,573
# Relations	14	31	19	48
# CKG triples	52,832	113,705	150,094	12,799

rather than how to construct knowledge maps. Thus, Microsoft Satori,³ a widely used structured tool, is adopted to directly construct a CKG for each dataset.

- ☆ **Book-Crossing** is an online book community. Its original dataset can provide over 1 million ratings (explicit/implicit) about more than 270,000 books. The constructed CKG contains 21,341 entities, 14 relations, and 52,832 CKG triples.
- ☆ **Amazon** is one of the largest online e-commerce companies in the United States, where users are free to rate products from 1 to 5. In the experiment, we use data from three domains to verify recommendation performance: books, music, and movies. The statistics for these datasets are shown in Table 2.

5.2 Experimental Settings

5.2.1 Setup and Metrics. Based on partially overlapping user or item entities, three cross-domain recommendation tasks in different scenarios are designed to more comprehensively verify the validity of our proposed approach. Their detailed tasks are described here.

- ✓ **Task 1:** Book-Crossing (BCr) versus AmazonBook (ABo); there are some overlapping item entities.
- ✓ **Task 2:** AmazonBook (ABo) versus AmazonMovie (AMo); there are some overlapping user entities.
- ✓ **Task 3:** AmazonMovie (AMo) versus AmazonMusic (AMu); there are some overlapping user entities.

For task 1, when BCr is used as auxiliary data, ABo is treated as the target data and vice versa. Our proposed PGACKG framework and those baseline approaches are implemented on TensorFlow. For fair comparison, we optimize the parameters of our proposed model and those of the baselines in their original articles. For example, the embedding size is fixed to 64 as suggested in [62], and all models are optimized by exploiting the Adam optimizer. According to the grid search, the learning rate is tuned in $\{0.00001, 0.0001, 0.001, 0.01\}$, the regularization parameter is searched in $\{0.001, 0.002, 0.004, 0.008, 0.01\}$, and the mini-batch size is set to 1024. To better model preference signals encoded in multi-order connectivity and reduce some incidental noise, the node depth of the proposed PGACKG is controlled at 2. Furthermore, for each dataset, 70% of entity interaction records are randomly selected for model training, while the two remaining 15% of entity interaction records are used for parameter fine-tuning and testing data, respectively.

To effectively observe Top-K performance and preference ranking, two commonly used ranking indicators are adopted to evaluate recommendation quality: Hit Ratio (HR@K) and Normalized Discounted Cumulative Gain (NDCG@K) [65]. The HR@K measures whether the predicted item

³<https://searchengineland.com/library/bing/bing-satori>.

is in the Top-K list, whereas the NDCG@K measures the ranking quality of the hit position. As we all know, higher values of these metrics reveal better ranking results. To obtain a fairer ranking result, the evaluation protocols adopted in our experiments are the full-ranking protocols during testing. This is because although the sampled protocols can speed up the computation of ranking metrics during testing, the model generally only receives biased results, which can hardly reflect the real tendency of recommendation performance, as mentioned in the literature [66]. Therefore, the full-ranking protocols are adopted to evaluate the recommendation quality of the proposed PGACKG and those baseline approaches during testing.

5.2.2 Baseline Methods. To demonstrate the effectiveness, we compare our proposed PGACKG with the following 7 baseline approaches in two groups: Single-Domain Recommendations (NGCF, KGAT, and CGAT) and Cross-Domain recommendations (KerKT, SemStim, GFM, and CD-GNN). All baseline models are representative or advanced methods in terms of KG or graph structure enabling recommendations. Their detailed descriptions are as follows.

NGCF [62]: This approach can leverage graph structure to project collaborative signals to an embedding propagation process based on a user-item KG, and a sufficient entity-embedding mechanism can be generated to promote a collaborative filtering recommendation.

KGAT [55]: This is a recommendation scheme that combines matrix factorization and KG techniques, which can model high-order connectivities in the KG and recursively refine the node embeddings by employing a graph attention network.

CGAT [67]: It is a state-of-the-art single-domain recommendation framework in which both local and non-local graph contexts are captured simultaneously by exploiting graph attention networks in the item KG.

KerKT [42]: This is a representative cross-domain recommendation solution. In this solution, diffusion kernel completion is used to associate the source and target domain knowledge to improve the accuracy of rating prediction based on overlapping entities.

SemStim [68]: This approach exploits semantic links generated in a KG (e.g., DBpedia) to assist the target domain in making cross-domain recommendations with an unsupervised graph algorithm.

GFM [69]: This model applies graph factorization machines on the KG structure to compute entity embeddings by propagating and aggregating multi-order interactions from the neighborhood in the source and target domains.

CD-GNN [70]: This framework adopts a graph neural network encoder to capture more representations for inactive entities and relations; the preference features of multi-order neighbors in the KG are considered for cross-domain recommendations.

5.3 Performance Comparison (for RQ1)

To answer RQ1, we observe the average experimental results and compare the performance of our proposed PGACKG model with those of other baselines. The top number of full-ranking item lists for these approaches is set to 5, 15, and 30, respectively. Note that for single-domain recommendation models (NGCF, KGAT, and CGAT), we train them and report their results on each domain. Tables 3 to 5 show the performance of all methods for the three tasks, followed by a summary of our observations.

- ☆ The CGAT outperforms other SDR models in all datasets. For example, it improves over the NGCF with regard to NDCG@5,15,30 by 7.61%, 7.23%, 9.17%, when Book-Crossing is used as the target domain in task 1. By contrast, the PGACKG outperforms other CDR models in all datasets. For example, it improves over the KerKT with regard to HR@5,15,30 by 9.90%, 10.77%, 11.26%, when AmazonBook is used as the target domain in task 1. The reason for

Table 3. Comparative Performance for Task 1

Methods	Book-Crossing						AmazonBook					
	HR			NDCG			HR			NDCG		
	5	15	30	5	15	30	5	15	30	5	15	30
NGCF	0.4012	0.4064	0.4155	0.2901	0.2958	0.3020	0.3317	0.3363	0.3426	0.2189	0.2230	0.2252
KGAT	0.4040	0.4091	0.4177	0.3100	0.3143	0.3168	0.3361	0.3419	0.3466	0.2200	0.2231	0.2284
CGAT	0.4078	0.4119	0.4194	0.3126	0.3172	0.3297	0.3406	0.3432	0.3485	0.2241	0.2269	0.2311
KerKT	0.4305	0.4343	0.4381	0.3334	0.3376	0.3428	0.3545	0.3567	0.3583	0.2428	0.2440	0.2472
SemStim	0.4287	0.4313	0.4374	0.3351	0.3408	0.3443	0.3672	0.3696	0.3727	0.2409	0.2431	0.2453
GFM	0.4336	0.4384	0.4427	0.3439	0.3475	0.3499	0.3720	0.3754	0.3786	0.2462	0.2499	0.2533
CD-GNN	<u>0.4373</u>	<u>0.4424</u>	<u>0.4465</u>	<u>0.3481</u>	<u>0.3530</u>	<u>0.3572</u>	<u>0.3748</u>	<u>0.3775</u>	<u>0.3807</u>	<u>0.2508</u>	<u>0.2536</u>	<u>0.2581</u>
PGACKG	0.4513	0.4549	0.4595	0.3604	0.3639	0.3688	0.3896	0.3951	0.3987	0.2734	0.2763	0.2810

Table 4. Comparative Performance for Task 2

Methods	AmazonBook						AmazonMovie					
	HR			NDCG			HR			NDCG		
	5	15	30	5	15	30	5	15	30	5	15	30
NGCF	0.5342	0.5375	0.5416	0.4109	0.4141	0.4180	0.5802	0.5840	0.5888	0.4557	0.4599	0.4624
KGAT	0.5383	0.5424	0.5467	0.4135	0.4174	0.4211	0.5846	0.5873	0.5931	0.4590	0.4642	0.4683
CGAT	0.5416	0.5455	0.5498	0.4163	0.4215	0.4240	0.5885	0.5937	0.5970	0.4627	0.4654	0.4691
KerKT	0.5672	0.5731	0.5764	0.4400	0.4436	0.4483	0.5988	0.6001	0.6035	0.4747	0.4769	0.4820
SemStim	0.5704	0.5750	0.5799	0.4421	0.4468	0.4516	0.6032	0.6068	0.6122	0.4721	0.4763	0.4806
GFM	0.5727	0.5765	0.5790	0.4456	0.4481	0.4539	0.6137	0.6174	0.6208	0.4933	0.4975	0.5008
CD-GNN	<u>0.5801</u>	<u>0.5843</u>	<u>0.5879</u>	<u>0.4532</u>	<u>0.4566</u>	<u>0.4590</u>	<u>0.6145</u>	<u>0.6190</u>	<u>0.6200</u>	<u>0.4979</u>	<u>0.5032</u>	<u>0.5064</u>
PGACKG	0.5930	0.5977	0.6006	0.4635	0.4679	0.4710	0.6326	0.6379	0.6424	0.5161	0.5184	0.5109

Table 5. Comparative Performance for Task 3

Methods	AmazonMovie						AmazonMusic					
	HR			NDCG			HR			NDCG		
	5	15	30	5	15	30	5	15	30	5	15	30
NGCF	0.4730	0.4768	0.4821	0.3327	0.3364	0.3418	0.4126	0.4165	0.4190	0.2848	0.2867	0.2900
KGAT	0.4764	0.4790	0.4825	0.3363	0.3387	0.3436	0.4154	0.4189	0.4207	0.2876	0.2924	0.2975
CGAT	0.4817	0.4850	0.4883	0.3389	0.3441	0.3476	0.4188	0.4234	0.4279	0.2921	0.2967	0.3020
KerKT	0.5031	0.5072	0.5118	0.3600	0.3637	0.3685	0.4305	0.4340	0.4376	0.3104	0.3142	0.3179
SemStim	0.5024	0.5060	0.5099	0.3576	0.3613	0.3655	0.4321	0.4346	0.4387	0.3123	0.3170	0.3204
GFM	0.5105	0.5137	0.5186	0.3638	0.3660	0.3728	0.4419	0.4452	0.4487	0.3236	0.3271	0.3330
CD-GNN	<u>0.5164</u>	<u>0.5200</u>	<u>0.5242</u>	<u>0.3687</u>	<u>0.3724</u>	<u>0.3755</u>	<u>0.4439</u>	<u>0.4501</u>	<u>0.4542</u>	<u>0.3306</u>	<u>0.3453</u>	<u>0.3480</u>
PGACKG	0.5305	0.5341	0.5387	0.3832	0.3854	0.3893	0.4716	0.4848	0.4900	0.3764	0.3805	0.3848

these phenomena may be that the graph attention mechanism can better assist the models to obtain the context preference of the target entities in the KG, which can significantly improve the prediction performance of the recommendation system.

- ☆ The CDR approaches (KerKT, SemStim, GFM, CD-GNN, and PGACKG) yield better performance than the SDR methods (NGCF, KGAT and CGAT) in three tasks. Further experimental results show that, compared with the pure KG-free solution (KerKT), other cross-domain models based on KG or graph structure achieve better predictive performance

than the pure cross-domain, especially our proposed PGACKG. Furthermore, our proposed PGACKG model exploits preference-aware entity embeddings within domains and across domains to more accurately capture user preference features, whereas other baseline models employ only aligned entity embeddings. The reason might be that other baselines fail to fully explore the preference information of entities and our proposed approach can better capture preference signals in the collaborative KG based on attentive preference generation mechanisms within domains and across domains.

- ☆ The PGACKG performs consistently better than other baseline approaches in all three tasks. Our PGACKG improves over the best-performing baseline (with results marked in Tables 3 to 5) with regard to HR@5, 15, 30 by 6.24%, 7.71%, 7.88%; NDCG@5, 15, 30 by 13.85%, 10.19%, 10.57%, when the AmazonMusic is used as the target domain in task 3. By stacking multiple preference embedding layers, the PGACKG is capable of capturing the preference features of higher-order entity nodes in the collaborative KG for recommendations, while the CD-GNN explores the preference of only first-order neighbors to guide the embedding learning. Moreover, compared with other baselines, the PGACKG also considers preference-aware graph attention layers to encode more preference signals into the embedding network for training. Therefore, we naturally conclude that our proposed PGACKG outperforms all baseline models in terms of cross-domain rating predictions.

5.4 Interaction Sparsity Study (for RQ2)

The advantage of incorporating the collaborative KG into cross-domain recommendation is that it can help alleviate the data sparsity issue in the target domain, which always limits the improvement of recommendation performance. In scenarios with insufficient interactions, models have difficulty learning strong representations for item predictions. Therefore, we observe the performance of the proposed PGACKG and three other cross-domain baseline approaches with KG (SemStim, GFM, CD-GNN) in alleviating the data sparsity issue.

To this end, we perform experiments over user groups with different sparsity levels and further redefine three specific tasks in the experimental setting section. BC_r, AB_o, and AM_o in the three tasks are sequentially used as the target domains; the other three datasets are used as the source domains. Each target domain is divided into four groups {<10, [10, 40), [40, 80), ≥80} based on interaction number per user in the test set, while keeping the interaction number in the source domain unchanged. Figure 3 exhibits the results measured by NDCG@15 on different user groups under the three tasks.

The PGACKG usually outperforms other KG-based cross-domain baseline models, especially on fairly sparse user groups in all three tasks. These experimental results demonstrate the effectiveness of the PGACKG in alleviating the data sparsity issue. The potential reason is that the proposed PGACKG can not only aggregate the preference features of similar entities within domains but also aggregate the preference features of similar entities across domains to model the preferences of target users, which can adequately leverage the rich knowledge from the source domain and improve the prediction performance for target users.

5.5 Study of PGACKG (for RQ3)

To answer RQ3 and observe the effect of different parameter settings on our proposed PGACKG approach, we investigate its impact on recommendation performance. First, we explore the effect of model depth on network training. Then, we analyze the influence of iteration times for layers on prediction performance. Finally, we study the influence of node dropout.

5.5.1 Effect of Model Depth. To investigate whether the PGACKG can benefit from the preference embedding transformation, the model depth L is changed from 1 to 4. BC_r, AB_o, and AM_o

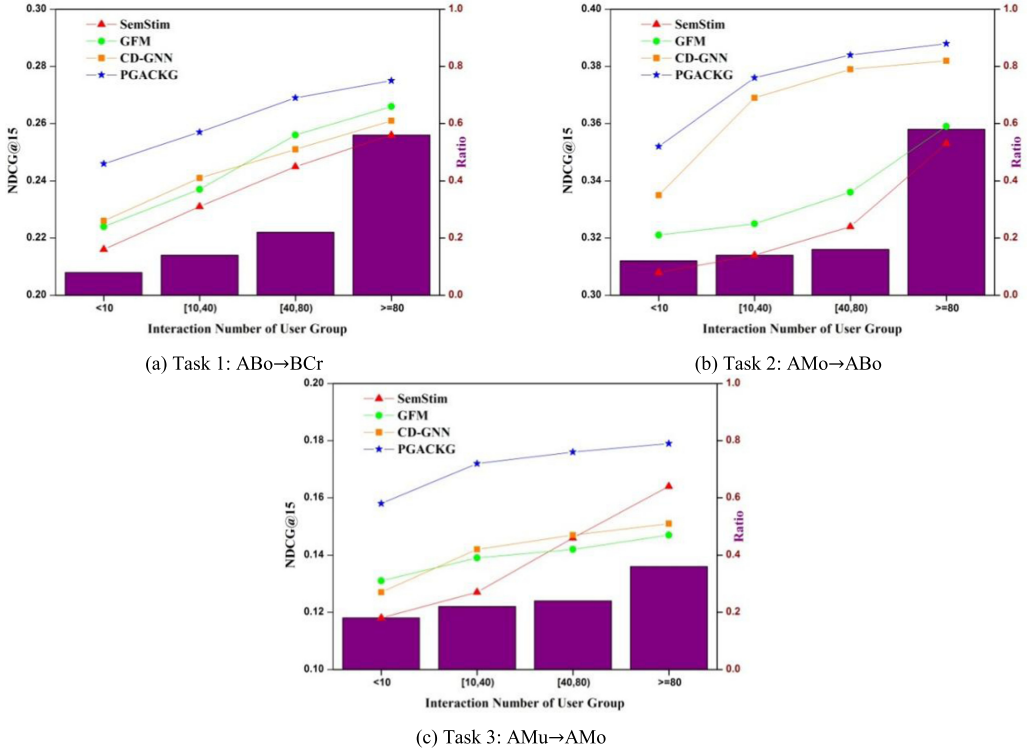


Fig. 3. Performance comparison in different interaction sparsity levels of user groups on three tasks.

Table 6. Effect of Model Depth Under Different Tasks

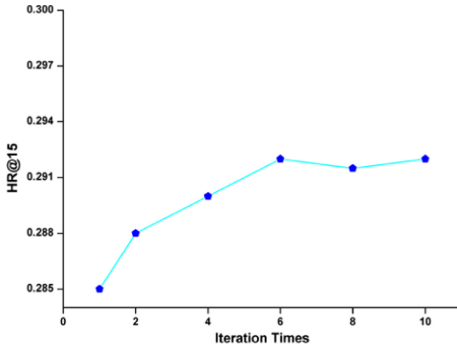
	Task 1 (ABo→BCr)		Task 2 (AMo→ABo)		Task 3 (AMu→AMo)	
	HR@15	NDCG@15	HR@15	NDCG@15	HR@15	NDCG@15
PGACKG-1	0.3108	0.2035	0.4322	0.3109	0.3704	0.2538
PGACKG-2	0.3271	0.2140	0.4374	0.3157	0.3742	0.2580
PGACKG-3	0.3297	0.2209	0.4420	0.3218	0.3801	0.2635
PGACKG-4	0.3316	0.2224	0.4433	0.3186	0.3817	0.2649

in three tasks are sequentially used as the target domains; the other three datasets are used as the source domains. Table 6 shows the experimental results from all three tasks; some conclusions are summarized as follows.

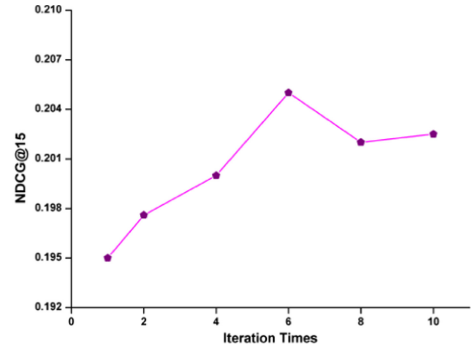
As the model depth increases, our PGACKG can continuously enhance the performance of recommendations. Obviously, the PGACKG-3 and PGACKG-2 consistently perform better than the PGACKG-1 across all three tasks. This may be because multi-order neighbors across domains can capture more entity preference signals for training than only first-order neighbors.

In addition, we can find that the PGACKG-4 achieves marginal improvements only by further stacking one more layer over the PGACKG-3, and it appears to be overfitting in task 2. These results indicate that preference collaborative signals can be adequately captured by considering third-order entity relations, which is consistent with some findings in [71].

5.5.2 Effect of Iteration Times for Layers. To further investigate how the number of iterations affects recommendation performance, we set K to 15 and use AmazonBook as the target domain in



(a) Performance of different iterations under HR@15



(b) Performance of different iterations under NDCG@15

Fig. 4. Effect of different iteration times under AmazonBook used as target domain in task 1.

task 1. Since the performance trends of other tasks are similar to those of task 1, Figure 4 presents only the changing trend of HR@15 and NDCG@15 with different iteration times in task 1. From Figure 4, we have the following observations.

Before the smoothness, more iterations consistently yield better performance, and the proposed PGACKG approach also continues to gain better convergence properties accordingly. The performance of our proposed model becomes gradually stable when the number of iterations is 6. Such an observation demonstrates the efficiency of performing preference-aware embedding propagation and the better model capacity of the PGACKG.

We can also see from Figure 4 and Table 6 that the effect of the number of iterations on the proposed PGACKG approach is slightly smaller compared with the model depth in the previous subsection. This may be that multiple epochs of network training result in our proposed model having a smaller and faster iterative convergence.

5.5.3 Effect of Dropout. To prevent overfitting, message dropout and node dropout techniques are adopted to train our proposed model along with the previous work [62]. For our three tasks, the experimental performance is averaged to observe the influence of message dropout ratio and node dropout ratio on our proposed PGACKG approach, as shown in Figure 5.

For both dropout strategies, Figure 5 shows that node dropout achieves better performance than message dropout in most cases. For example, setting the Dropout Ratio (DR) to 0.2 yields the highest HR@15 of 0.377, which is 3.01% higher than that of message dropout. One possible reason is that dropping out the drifting noise from the high-order propagation process makes the preference-aware embedding capture more preference signals for recommendations. Hence, it can be seen that node dropout is a more effective strategy than message dropout to solve the overfitting in GNNs.

5.5.4 Effect of Random Walks. The proposed PGACKG approach employs the random walk algorithm to aggregate similar nodes across domains in a collaborative KG. To investigate its impact on our proposed method, we observe and analyze the walking length L and the sampling number N for random walks. Tables 7 and 8 exhibit the performance of the PGACKG with respect to different settings of L and N in random walk exploration. NDCG@15 is adopted to evaluate the performance of our proposed approach on three tasks.

As can be seen from Table 7, better performance can be achieved by setting the walking length L around 16. Further increasing L introduces more noise and training complexity, resulting in

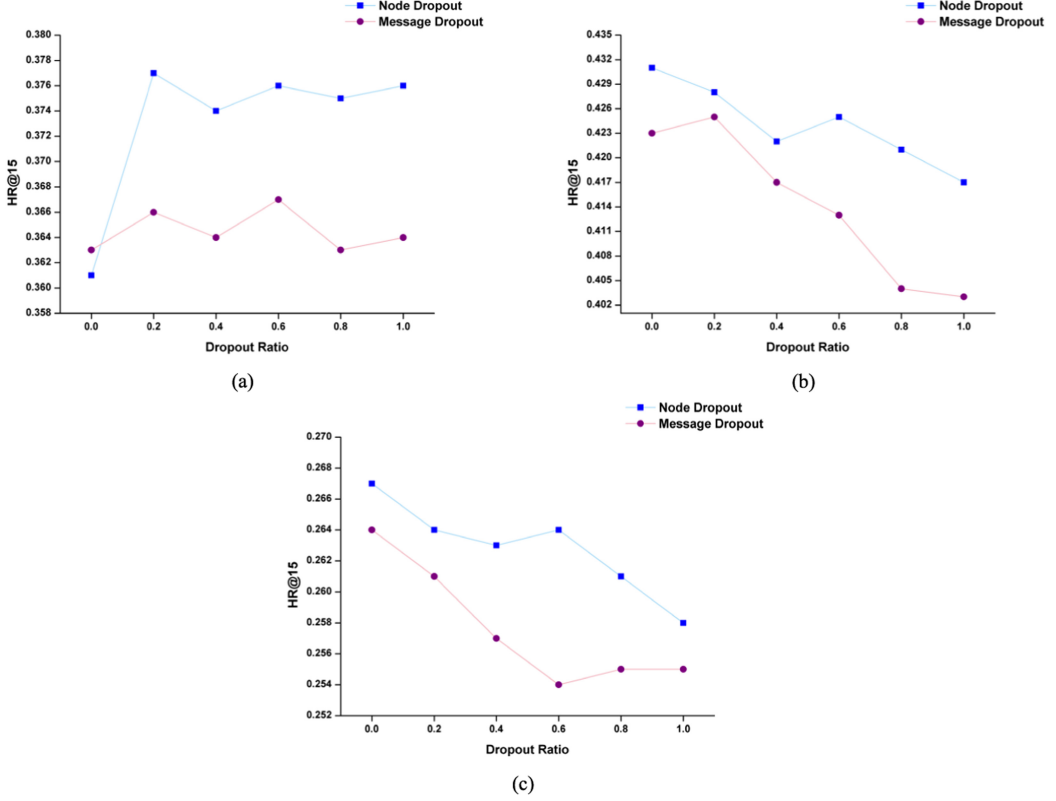


Fig. 5. Effect of node dropout and message dropout ratios under different tasks.

Table 7. Performance of the PGACKG With Respect to Different Walking Lengths L for Random Walks

Auxiliary Datasets	$L = 4$	$L = 8$	$L = 16$	$L = 32$	$L = 64$
ABO (Task 1)	0.3368	0.3419	0.3587	0.3531	0.3475
AMo (Task 2)	0.5210	0.5345	0.5416	0.5362	0.5279
AMu (Task 3)	0.2181	0.2247	0.2309	0.2275	0.2218

Table 8. Performance of the PGACKG With Respect to Different Sampling Numbers N for Random Walks

Auxiliary Datasets	$N = 3$	$N = 6$	$N = 12$	$N = 24$	$N = 48$
ABO (Task 1)	0.3018	0.3083	0.3125	0.3214	0.3150
AMo (Task 2)	0.4739	0.4847	0.4913	0.4982	0.4921
AMu (Task 3)	0.1900	0.1974	0.2028	0.2135	0.2056

lower model performance. In addition, it can be seen from the results in Table 8 that our approach achieves the best performance when performing 24 random walk samplings. This indicates that the most similar nodes across domains can be captured by setting the sampling number N to around 24 for random walks.

Table 9. Recommendation Performance Achieved by PGACKG Variants on Different Tasks

	Metrics	PGACKG _{/GA}	PGACKG _{/WP}	PGACKG _{/AP}	PGACKG
Task 1	HR@15	0.2285	0.1802	0.2896	0.3031
(ABo→BCr)	NDCG@15	0.1379	0.0924	0.1933	0.2154
Task 2	HR@15	0.3226	0.3047	0.3900	0.4126
(AMo→ABo)	NDCG@15	0.2568	0.2381	0.3104	0.2914
Task 3	HR@15	0.2705	0.2569	0.3463	0.3620
(AMu→AMo)	NDCG@15	0.1688	0.1328	0.2401	0.2757

5.6 Ablation Study (for RQ4)

To investigate and expose the importance of each key components of PGACKG, we perform ablation studies to evaluate the performance of the following three variants. Each variant of our proposed approach is described here in detail.

PGACKG_{/GA} deletes the collaborative KG and graph attention mechanism from PGACKG, and considers and aggregates only the preference features of similar neighbors of the target entity to make recommendations based on partially overlapping entities.

PGACKG_{/WP} removes the attentive preference features within domains from the final fusion features, and considers only the attentive preference features across domains to generate cross-domain recommendations.

PGACKG_{/AP} retains the attentive preference features within domains, which is contrary to PGACKG_{/WP}, and deletes the attentive preference features across domains from the final fusion features to generate the final recommendation.

Additionally, we design three tasks to validate the comparative results of ablation studies. In the three tasks, BCr, ABo, and AMo are sequentially used as the target domain datasets, and ABo, AMo, and AMu are sequentially used as the source domain datasets. The number of interactions of each entity used in the target domain is no more than 5 to ensure data sparsity, whereas the number of interactions of each entity used in the source domain is more than 15 to ensure the richness of knowledge. The ablation results are shown in Table 9.

In all comparisons, PGACKG_{/GA} is slightly better than PGACKG_{/WP}, which achieves the worst performance. This indicates that exploiting only similar preferences across domains cannot capture the target entity's preference features well and is even inferior to some previous models that directly aggregate similar preference features.

Compared with PGACKG_{/GA} and PGACKG_{/WP}, PGACKG_{/AP} has greatly improved the recommendation performance, indicating that the attentive preference features within domains are essential for recommendations. Additionally, our PGACKG outperforms PGACKG_{/AP}, indicating that preference-aware embeddings within domains and across domains cooperate to enhance the quality of recommendations.

In summary, PGACKG is consistently superior to other variants on all tasks in terms of HR@15 and NDCG@15 metrics. Such results demonstrate that the collaborative KG can significantly facilitate the improvement of cross-domain recommendation accuracy, thus showing the effectiveness of the graph attention mechanism in our approach.

5.7 Visualization (for RQ5)

In this section, we attempt to investigate preference embedding in the target domain feature space to further analyze the reason why the PGACKG can improve the performance of the model. ABo

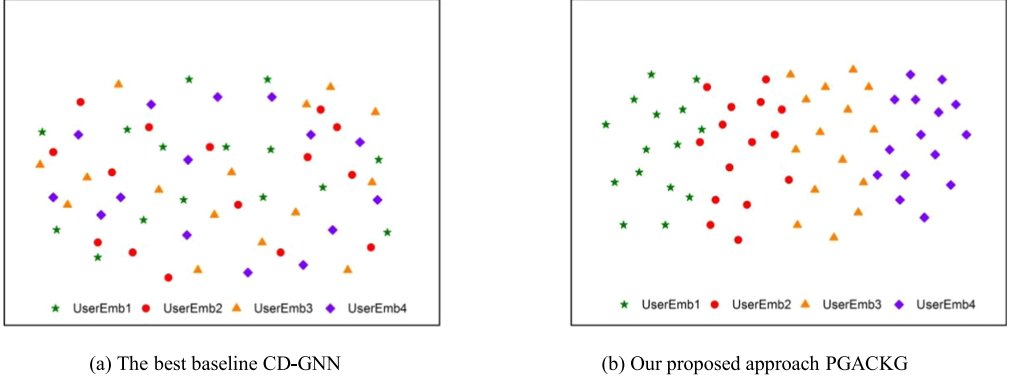


Fig. 6. Visualization of transformed user embeddings derived from a CD-GNN and PGACKG.

is used as the source dataset and BCr is used as the target dataset. Four categories of users are randomly selected to implement visualization experiments, following the default setting of the t-SNE [72, 73] in Scikit-learn. Additionally, the best cross-domain baseline CD-GNN is adopted to compare the performance of our proposed PGACKG.

As can be seen from Figure 6(a), different user embeddings in a CD-GNN do not have discernible boundaries and clusters, and the transformed embeddings are scattered across target domain feature space. However, in Figure 6(b), different user embeddings exhibit distinct clusters, which demonstrates that user embeddings transformed by the PGACKG can inject some preference signals into representation learning. The visual experiments also explain the fundamental reason why the PGACKG can achieve better performance.

6 CONCLUSION

In this work, we devise a novel cross-domain recommendation framework that can employ preference-aware graph attention networks in the CKG to explore high-order semantic preferences among entities, thereby improving cross-domain recommendation accuracy. The PGACKG leverages the graph embedding model to transform and obtain the preference-aware embeddings with rich semantic preference signals. A preference-aware graph attention network model is proposed to aggregate the preferences of similar entities within domains and across domains in the CKG via multi-hop reasoning and frequency visits-based node random walk model, respectively. We fuse entity preference features within domains and across domains to remodel user final preferences, and the CBPR is newly proposed to generate the cross-domain recommendations. Compared with state-of-the-art baselines, the superiority of the PGACKG has been verified by extensive experiments on four real-world datasets.

For future work, we would like to explore the PGACKG on more KG-based cross-domain recommendation scenarios. We also intend to develop different preference aggregation strategies to capture the dynamic interactions of entities in temporal KGs.

REFERENCES

- [1] Zhiqiang Pan, Fei Cai, Wanyu Chen, and Honghui Chen 2021. Graph Co-Attentive session-based recommendation. *ACM Transactions on Information Systems* 40, 4 (2021), 1–31.
- [2] Paolo Cremonesi, Antonio Tripodi, and Roberto Turrin. 2011. Cross-domain recommender systems. In *Proceedings of the International Conference on Data Mining Workshops (ICDMW'11)*. ACM, 496–503.
- [3] Iván Cantador, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Paolo Cremonesi. 2015. Cross-domain recommender systems. In *Recommender Systems Handbook* (2015), 919–959.

- [4] Feng Zhu, Yan Wang, Jun Zhou, Chaochao Chen, Longfei Li, and Guanfeng Liu. 2021. A unified framework for cross-domain and cross-system recommendations. *IEEE Transactions on Knowledge and Data Engineering* (2021).
- [5] Pan Li and Alexander Tuzhilin. 2021. Dual metric learning for effective and efficient cross-domain recommendations. *IEEE Transactions on Knowledge and Data Engineering* (2021).
- [6] Chuan Shi Binbin Hu, Wayne Xin Zhao, and S. Yu Philip. 2018. Heterogeneous information network embedding for recommendation. *IEEE Transactions on Knowledge and Data Engineering* 31, 2 (2018), 357–370.
- [7] Quan Wang, Zhendong Mao, Bin Wang and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29, 12 (2017), 2724–2743.
- [8] 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 (AAAI'19)*. AAAI Press, 5329–5336.
- [9] Rui Sun, Xuezhi Cao, Yan Zhao, Junchen Wan, Kun Zhou, Fuzheng Zhang, Zhongyuan Wang, and Kai Zheng. 2020. Multi-modal knowledge graphs for recommender systems. In *Proceedings of the ACM International Conference on Information & Knowledge Management (CIKM'20)*. ACM, 1405–1414.
- [10] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2020. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering* 34, 8 (2020), 3549–3568.
- [11] Ron Levie, Federico Monti, Xavier Bresson, and Michael M. Bronstein. 2019. CayleyNets: Graph convolutional neural networks with complex rational spectral filters. *IEEE Transactions on Signal Processing* 67, 1 (2019), 97–109.
- [12] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2014. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 (2014).
- [13] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)*. ACM, 3844–3852.
- [14] Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2017).
- [15] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2009. The graph neural network model. *IEEE Transactions on Neural Networks and Learning Systems* 20, 1 (2009), 61–80.
- [16] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. 2017. Neural message passing for quantum chemistry. In *Proceedings of the International Conference on Machine Learning (ICML'17)*. PMLR, 1263–1272.
- [17] Will Hamilton, Zitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NIPS'17)*. ACM, 1024–1034.
- [18] Hongyang Gao, Zhengyang Wang, and Shuiwang Ji. 2018. Large-scale learnable graph convolutional networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'18)*. ACM, 1416–1424.
- [19] Yujia Li, Oriol Vinyals, Chris Dyer, Razvan Pascanu, and Peter Battaglia. 2018. Learning deep generative models of graphs. arXiv preprint arXiv:1803.03324 (2018).
- [20] Aleksandar Bojchevski, Oleksandr Shchur, Daniel Zügner, and Stephan Günnemann. 2018. NetGAN: Generating graphs via random walks. In *Proceedings of the International Conference on Machine Learning (ICML'18)*. PMLR, 610–619.
- [21] Jiaxuan You, Rex Ying, Xiang Ren, William Hamilton, and Jure Leskovec. 2018. GraphRNN: A deep generative model for graphs. In *Proceedings of International Conference on Machine Learning (ICML'18)*.
- [22] Tengfei Ma, Jie Chen, and Cao Xiao. 2018. Constrained generation of semantically valid graphs via regularizing variational autoencoders. In *Advances in Neural Information Processing Systems (NIPS'18)*. ACM, 7110–7121.
- [23] 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 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*. ACM, 353–362.
- [24] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y. Chang. 2018. Improving sequential recommendation with knowledge-enhanced memory networks. In *Proceedings of the International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR'18)*. ACM, 505–514.
- [25] Hongwei Wang, Fuzheng Zhang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2019. Multi-task feature learning for knowledge graph enhanced recommendation. In *Proceedings of the International World Wide Web Conference (WWW'19)*. ACM, 2000–2010.
- [26] Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, and Tat-Seng Chua. 2019. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. In *Proceedings of the International World Wide Web Conference (WWW'19)*. ACM, 151–161.

- [27] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S. Yu. 2018. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'18)*. ACM, 1531–1540.
- [28] Huan Zhao, Quanming Yao, Jianda Li, Yangqiu Song, and Dik Lee. 2017. Meta-graph based recommendation fusion over heterogeneous information networks. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD'17)*. ACM, 635–644.
- [29] Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. 2017. Metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD'17)*. ACM, 135–144.
- [30] Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu, Bradley Stuart, Urvashi Khandelwal, Brandon Norrick, and Jiawei Han. 2014. Personalized entity recommendation: A heterogeneous information network approach. In *Proceedings of the ACM International Conference on Web Search and Data Mining (ICDMW'14)*. ACM, 283–292.
- [31] Chuan Shi, Zhiqiang Zhang, Ping Luo, Philip S. Yu, Yading Yue, and Bin Wu. 2015. Semantic path based personalized recommendation on weighted heterogeneous information networks. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM'15)*. ACM, 453–462.
- [32] 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 ACM Conference on Recommender Systems (RecSys'12)*. ACM, 297–305.
- [33] Kangzhi Zhao, Xiting Wang, Yuren Zhang, Li Zhao, Zheng Liu, Chunxiao Xing, and Xing Xie. 2020. Leveraging demonstrations for reinforcement recommendation reasoning over knowledge graphs. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20)*. ACM, 239–248.
- [34] Sijin Zhou, Xinyi Dai, Haokun Chen, Weinan Zhang, Kan Ren, Ruiming Tang, Xiuqiang He, and Yong Yu. 2020. Interactive recommender system via knowledge graph-enhanced reinforcement learning. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20)*. ACM, 179–188.
- [35] Pengfei Wang, Yu Fan, Long Xia, Wayne Xin Zhao, Shaozhang Niu, and Jimmy Huang. 2020. KERL: A knowledge-guided reinforcement learning model for sequential recommendation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20)*. ACM, 209–218.
- [36] Xiang Wang, Yaokun Xu, Xiangnan He, Yixin Cao, Meng Wang, and Tat-seng Chua. 2020. Reinforced negative sampling over knowledge graph for recommendation. In *Proceedings of the International World Wide Web Conference (WWW'20)*. ACM, 99–109.
- [37] Feng Zhu, Wang Yan, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfang Liu. 2021. Cross-domain recommendation: Challenges, progress, and prospects. *arXiv preprint arXiv:2103.01696*.
- [38] Sibo Gai, Feng Zhao, Yachen Kang, Zhengyu Chen, Donglin Wang, and Ao Tang. 2019. Deep transfer collaborative filtering for recommender systems. In *Proceedings of Pacific Rim International Conference on Artificial Intelligence (PRICAI'19)*. Springer, 515–528.
- [39] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-domain recommendation: An embedding and mapping approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17)*. ACM, 2464–2470.
- [40] Yakun Li, Jiadong Ren, Jiaomin Liu, and Yixin Chang. 2021. Deep sparse autoencoder prediction model based on adversarial learning for cross-domain recommendations. *Knowledge-Based Systems* 220, 5 (2021), 1–14.
- [41] Qian Zhang, Dianshuang Wu, Jie Lu, and Guangquan Zhang. 2018. Cross-domain recommendation with probabilistic knowledge transfer. In *International Conference on Neural Information Processing (NeurIPS'18)*. 208–219.
- [42] Qian Zhang, Jie Lu, Dianshuang Wu, and Guangquan Zhang. 2018. A cross-domain recommender system with kernel-induced knowledge transfer for overlapping entities. *IEEE Transactions on Neural Networks and Learning Systems* 30, 7 (2018), 1998–2012.
- [43] Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, and Aixin Sun. 2020. CATN: Cross-domain recommendation for cold-start users via aspect transfer network. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20)*. ACM, 229–238.
- [44] Lile Li, Quan Do, and Wei Liu. 2019. Cross-domain recommendation via coupled factorization machines. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'19)*. AAAI Press, 9965–9966.
- [45] Ming He, Jiuling Zhang, Peng Yang, and Kaisheng Yao. 2018. Robust transfer learning for cross-domain collaborative filtering using multiple rating patterns approximation. In *Proceedings of the ACM International Conference on Web Search and Data Mining (ICWDM'18)*. ACM, 225–233.
- [46] Hyunwoo Hwangbo and Yangsok Kim. 2017. An empirical study on the effect of data sparsity and data overlap on cross domain collaborative filtering performance. *Expert Systems with Applications* 89 (2017), 254–265.

- [47] Qian Zhang, Peng Hao, Jie Lu, and Guangquan Zhang. 2019. Cross-domain recommendation with semantic correlation in tagging systems. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN'19)*. IEEE, 1–8.
- [48] Shulong Tan, Jiajun Bu, Xuzhen Qin, Chun Chen, and Deng Cai. 2014. Cross domain recommendation based on multi-type media fusion. *Neurocomputing* 127 (2014), 124–134.
- [49] Shaghayegh Sahebi and Trevor Walker. 2014. Content-based cross-domain recommendations using segmented models. In *CBRecSys@ RecSys*. 57–64.
- [50] Heishiro Kanagawa, Hayato Kobayashi, Nobuyuki Shimizu, Yukihiro Tagami, and Taiji Suzuki. 2019. Cross-domain recommendation via deep domain adaptation. In *Proceedings of European Conference on Information Retrieval (ECIR'19)*. 20–29.
- [51] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. MTNet: A neural approach for cross-domain recommendation with unstructured text. *KDD Deep Learning Day (KDD'18)*. 1–10.
- [52] Wenjing Fu, Zhaoxue Peng, Senzhang Wang, Yang Xu, and Jin Li. 2019. Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'19)*. AAAI Press, 94–101.
- [53] Anu Taneja and Anuja Arora. 2018. Cross domain recommendation using multidimensional tensor factorization. *Expert Systems with Applications* 92 (2018), 304–316.
- [54] Shampa Chakraverty and Mala Saraswat. 2017. Review based emotion profiles for cross domain recommendation. *Multimedia Tools and Applications* 76, 24 (2017), 25827–25850.
- [55] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*. ACM, 950–958.
- [56] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL'15)*. 687–696.
- [57] Xi Victoria Lin, Richard Socher, and Caiming Xiong. 2018. Multi-hop knowledge graph reasoning with reward shaping. *arXiv preprint arXiv:1808.10568*.
- [58] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran. 2018. Image transformer. In *Proceedings of the International Conference on Machine Learning (ICML'18)*. PMLR, 4055–4064.
- [59] Yikun Xian, Zuohui Fu, Shan Muthukrishnan, Gerard De Melo, and Yongfeng Zhang. 2019. Reinforcement knowledge graph reasoning for explainable recommendation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)*. ACM, 285–294.
- [60] Jaehun Jung, Jinhong Jung, and U. Kang. 2021. Learning to walk across time for interpretable temporal knowledge graph completion. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD'21)*. ACM, 786–795.
- [61] Yunpeng Weng, Xu Chen, Liang Chen, and Wei Liu. 2020. GAIN: Graph Attention and Interaction Network for inductive semi-supervised learning over large-scale graphs. *IEEE Transactions on Knowledge and Data Engineering* 34, 9 (2020), 4257–4269.
- [62] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural graph collaborative filtering. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)*. ACM, 165–174.
- [63] Ruining He and Julian McAuley. 2016. VBPR: Visual Bayesian personalized ranking from implicit feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'16)*. AAAI Press.
- [64] Yuqi Cui, Dongrui Wu, and Jian Huang. 2020. Optimize TSK fuzzy systems for classification problems: Mini-batch gradient descent with uniform regularization and batch normalization. *IEEE Transactions on Fuzzy Systems* 28, 12 (2020), 3065–3075.
- [65] Shulong Tan, Jiajun Bu, Xuzhen Qin, Chun Chen, and Deng Cai. 2014. Cross domain recommendation based on multi-type media fusion. *Neurocomputing* 127 (2014), 124–134.
- [66] Walid Krichene and Steffen Rendle. 2020. On sampled metrics for item recommendation. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD'20)*. ACM, 75–83.
- [67] Yong Liu, Susen Yang, Yonghui Xu, Chunyan Miao, Min Wu, and Juyong Zhang. 2021. Contextualized graph attention network for recommendation with item knowledge graph. *IEEE Transactions on Knowledge and Data Engineering* 35, 1 (2021), 181–195.
- [68] Benjamin Heitmann and Conor Hayes. 2016. SemStim: Exploiting knowledge graphs for cross-domain recommendation. In *Proceedings of IEEE International Conference on Data Mining Workshops (ICDMW'16)*. ACM, 999–1006.
- [69] Dongbo Xi, Fuzhen Zhuang, Yongchun Zhu, Pengpeng Zhao, Xiangliang Zhang, and Qing He. 2020. Graph factorization machines for cross-domain recommendation. *arXiv preprint arXiv: 2007.05911*.

- [70] Ziqi Liu, Yue Shen, Xiaocheng Cheng, Qiang Li, Jianping Wei, Zhiqiang Zhang, Dong Wang, Xiaodong Zeng, Jinjie Gu, and Jun Zhou. 2021. Learning representations of inactive users: A cross domain approach with graph neural networks. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM'21)*. ACM, 3278–3282.
- [71] Lei Sang, Min Xu, Shengsheng Qian, and Xindong Wu. 2021. Knowledge graph enhanced neural collaborative recommendation. *Expert Systems with Applications* 164, 2 (2021), 1–13.
- [72] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. Decaf: A deep convolutional activation feature for generic visual recognition. In *Proceedings of the International Conference on Machine Learning (ICML'14)*. PMLR, 647–655.
- [73] Yongchun Zhu, Zhenwei Tang, Yudan Liu, Fuzhen Zhuang, Ruobing Xie, Xu Zhang, Leyu Lin, and Qing He. 2022. Personalized transfer of user preferences for cross-domain recommendation. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM'22)*. ACM, 1507–1515.

Received 10 March 2022; revised 13 August 2022; accepted 22 November 2022