



# EMDKG: Improving Accuracy-Diversity Trade-Off in Recommendation with EM-based Model and Knowledge Graph Embedding

Lu Gan, Diana Nurbakova, Léa Laporte, Sylvie Calabretto

## ► To cite this version:

Lu Gan, Diana Nurbakova, Léa Laporte, Sylvie Calabretto. EMDKG: Improving Accuracy-Diversity Trade-Off in Recommendation with EM-based Model and Knowledge Graph Embedding. The 20th IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Dec 2021, Melbourne, Australia. 10.1145/3486622.3493925 . hal-03518867

**HAL Id: hal-03518867**

**<https://hal.science/hal-03518867v1>**

Submitted on 10 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# EMDKG: Improving Accuracy-Diversity Trade-Off in Recommendation with EM-based Model and Knowledge Graph Embedding

Lu Gan

INSA Lyon, University of Lyon  
Villeurbanne, France  
lu.gan@insa-lyon.fr

Léa Laporte

INSA Lyon, University of Lyon  
Villeurbanne, France  
lea.laporte@insa-lyon.fr

Diana Nurbakova

INSA Lyon, University of Lyon  
Villeurbanne, France  
diana.nurbakova@insa-lyon.fr

Sylvie Calabretto

INSA Lyon, University of Lyon  
Villeurbanne, France  
sylvie.calabretto@insa-lyon.fr

## ABSTRACT

To maintain attractiveness and reduce redundancy of recommendation, the concept of diversity has been brought up in recommender systems (RS). Thus, advanced RS aim at achieving both better accuracy and diversity facing a trade-off issue between the two aspects. Recently, knowledge graphs embedding methods have been widely used in RS for achieving better accuracy provided with auxiliary information along with historical user-item interactions. However, little work has been done to investigate what effects of diversity it brings along with higher accuracy results and how to achieve the best accuracy-diversity trade-off under such circumstances. In this paper, we propose an EM-model capable of incorporating a generalized concept of diversity for a diversity-encoded knowledge graph embedding based recommendation. Our EM-model alternates between a general item diversity learning and knowledge graph embedding learning for user and item representation, which helps to achieve better results in terms of both accuracy and diversity compared to the state-of-art baselines on datasets MovieLens and Anime. Moreover, extensive experiments prove our model outperforms the baseline with existing diversification methods (MMR and DPP) achieving a better accuracy-diversity trade-off.

## CCS CONCEPTS

• **Information systems** → **Recommender systems; Personalization; Information retrieval diversity.**

## KEYWORDS

Recommender Systems; Knowledge Graph Embedding; Diversity; Accuracy; Trade-off

## ACM Reference Format:

Lu Gan, Diana Nurbakova, Léa Laporte, and Sylvie Calabretto. 2021. EMDKG: Improving Accuracy-Diversity Trade-Off in Recommendation with EM-based Model and Knowledge Graph Embedding. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '21)*, December 14–17, 2021, ESSENDON, VIC, Australia. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3486622.3493925>

## 1 INTRODUCTION

Recommender systems (RS) have been intensively studied over the last two decades and have reached a remarkable effectiveness. Amazon, Netflix, Facebook...: all these applications and e-commerce sites make a very intensive use of recommender systems. Top- $N$  recommender systems exploiting user-item interactions achieve high-level accuracy. Besides, recent works [3, 15] have shown that incorporating multiple relations among items and their semantic information into recommendation task, e.g. via knowledge graphs, can even improve the accuracy in recommendation results. Knowledge graph embeddings have been shown to be highly effective methods to represent user, items and their structural relations [32]. But despite their effectiveness, most recommender systems still suffer from a number of drawbacks and limitations which recently raised the scientific community interest, among which *diversity*.

A bunch of work have been done recently to address the problem of diversity in recommendation (e.g. [4, 5, 9, 20, 22, 25, 27, 31]). However, achieving a good accuracy-diversity trade-off is still an open challenge. Thus, most of existing works confront with this trade-off assuming submodular feature of optimisation function. A few works [6, 24, 31] try to find solutions both diverse and accurate. Thus, newly proposed diversity-aware recommendation models (e.g. [5, 9, 16, 30, 31]) make use of Determinantal Point Processes (DPPs) [14], an elegant probabilistic model that has been actively conquering the ML and IR communities for the last years. However, there is still lacking of the understanding of a good trade-off between accuracy and diversity and when and how a better trade-off can be achieved. In our opinion, a diversity-aware recommendation algorithm should not only achieve both high accuracy and diversity, but also be robust under different parameter settings.

In this paper, we address the top- $N$  recommendation problem from diversity perspective, while **ensuring a trade-off between**

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 the author(s) 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](mailto:permissions@acm.org).

WI-IAT '21, December 14–17, 2021, ESSENDON, VIC, Australia

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9115-3/21/12...\$15.00

<https://doi.org/10.1145/3486622.3493925>

**accuracy and diversity and making use of rich auxiliary semantic information and relations about items** (if available). Another challenge that we respond to lies in the fact that vector representations widely used in modern RS thanks to the popularity of matrix factorisation and various embedding techniques do not account for diversity directly. We argue that it should be **encoded in vector representations of items** in order to provide a more systematic view on the diversity of user preferences.

To achieve this goal, we propose a general framework called **EMDKG** that incorporates knowledge graph embedding with Determinantal Point Process (DPP). We make the following **contributions**. First, we propose an item diversity learning (IDL) framework to learn the vector representations of items based on ground-truth item sets generated given a certain chosen diversity metric. We then propose an EM scheme to co-learn IDL representations with knowledge graph embedding for diversity-aware vector representations of users and items. To the best of our knowledge, we are the first to exploit such combination. Based on these representations we can further generate top- $N$  recommendations to leverage better accuracy and diversity. We perform extensive experiments on two datasets (MovieLens-100k and Anime) to evaluate our framework while comparing it against multiple state-of-the-art algorithms.

## 2 RELATED WORKS

Recent advances in RS focus on key objectives other than accuracy that contribute to the overall satisfaction of the users from the underlying service [12]. Diversity is one of such factors. Thus, more diversified lists of recommendation results bring more satisfaction to users, even at cost of some loss of accuracy [25].

In the context of recommendation, two levels of diversity are usually distinguished: aggregate and individual [8]. *Individual diversity* reflects item dissimilarity in each list of recommendations for an individual user, thus helping to reduce the ‘*filter bubble*’ effect. Most of the existing works deal with it [6, 24]. *Aggregate diversity* (or catalog diversity) [6, 8, 26] reflects the ability of the system to create a more balanced recommendation over all items, thus reducing the long-tail effect and mitigating the popularity bias. It is often measured as a ratio of recommended items to the set of all available items. In this work, we also focus on individual diversity and use the term ‘*diversity*’ to refer to it, if not precised otherwise.

Diversification techniques can be categorised into two groups: re-ranking and diversity modelling. Most of the existing works (e.g. [4, 22, 25]) approach the diversified recommendation task in two steps, and are based on *re-ranking*. First, a candidate list is generated typically by choosing the most relevant items (i.e. targeting the accuracy objective). Second, the re-ranking procedure is applied in post-processing to produce the resulting list of recommendations by maximising an objective function. Such function is usually defined as a combination of a candidate’s relevance and its dissimilarity (diversity or distance) with the items already added to the result list (greedy strategy). The methods then differ in the way of selecting the candidate items and computing the dissimilarity measure. A typical example of such a greedy re-ranking technique is **Maximal Marginal Relevance (MMR)** [4]. It introduces a relevance-diversity trade-off via the the notion of marginal

relevance that combines two metrics: relevance and diversity. Probabilistic IR re-ranking models such as IA-Select [1] have also been adapted for diversified recommendation using the latent item feature space (e.g. [27]). In contrast to greedy re-ranking strategy, some of the proposed methods (e.g. [20]) incorporate diversity along with relevance as optimisation objectives and/or constraints to find an optimal item ranking for each user (e.g. [33]). The main advantage of the re-ranking techniques lies in its modular nature allowing to easily incorporate the diversification step into the existing recommendation process and explicitly control the accuracy-diversity trade-off. At the same time, the latter represents a limitation of this group of techniques, as an extensive hyperparameter tuning may be required for a better performance. Moreover, such post-processing strategies may result in a significant loss in terms of accuracy. Another important limitation is the adoption of pairwise measures of diversity to characterise the list that do not account for some complex relationships between items [5].

New recommendation algorithms tend to consider diversity directly when generating recommendations, thus developing *diversity models* and finding a more sophisticated accuracy-diversity trade-off (e.g. [6, 9, 16, 23, 31]). Some works approach accuracy-diversity trade-off as exploration-exploitation trade-off and propose bandit-based models (e.g. [18]). Recently, **Determinantal Point Processes (DPPs)** [13, 14] have been used to improve recommendation diversity (e.g. [5, 9, 16, 30, 31]). This probabilistic model allows to select relevant yet diverse items without losing importantly the item relevance to a user. A key component of a DPP is its kernel matrix that models item feature space and is indexed with candidate set of items. Its diagonal elements reflect items ‘quality’ (in RS context, relevance), while off-diagonal elements measure their similarity. The methods then differ in the way they determine this kernel matrix and the sampling techniques applied to generate the result list of items. The kernel matrix can be learnt from data (e.g. [17, 30, 31]) or constructed heuristically (e.g. [5, 9]). Chen *et al.* [5] proposed a fast greedy **Maximum A Posteriori (MAP)** inference for effective sampling of the result list from candidate items. This method was further adopted by other algorithms, including **DivKG** [9] that uses it for the prediction part. **DivKG** constructs the DPP kernel matrix empirically based on the relevance and similarity measures issued from knowledge graph embedding model. Recent advances in generative adversarial networks (GAN) frameworks have been employed for diversified recommendations (e.g. [31]). For instance, **PD-GAN (Personalized Diversity-promoting GAN)** [31] combines DPP model [5] used as the generator with the discriminator aiming to distinguish between generated lists and the ground truth. **PD-GAN** allows to capture individual preferences towards diversity, and generated diversified yet relevant items accordingly.

The work closest to ours is **DivKG** [9]. Similar to DivKG [9], we exploit the knowledge graph embedding (KGE) techniques for representing users and items and capturing auxiliary information and relations that can improve the recommendation. However, DivKG assumes the learned item vectors represent the similarity/dissimilarity of item lists with semantic information, which is not always the case. In contrast to that, our proposal EMDKG explicitly propose an Item Diversity Learning module to distill the semantic diversity into item vector representations. Also EMDKG bridges this Item Diversity Learning with KGE for learning both

an accurate and diversity-encoded representations for users, items and their affinity relation, prompting to enable a better-off between accuracy and diversity when applied with diversification methods.

### 3 BACKGROUND

The goal of our recommendation task is two-fold: given a set of users  $U$ , a set of items  $I$ , user-item interactions, provide each user  $u \in U$  with **relevant** yet **diverse** recommendations based on historical user-item interactions and auxiliary information for items. Our solutions lies on two ideas: Determinantal Point Processes and Translation-Based Knowledge Graph Embedding, which we present briefly in this Section to provide the background for our solution.

#### 3.1 Determinantal Point Processes

A determinantal point process (DPP) is a probabilistic model over selection of points. Originating from quantum physics, this model is characterized by its repulsiveness, which means a higher probability of a subset selection associates with more repelling points to each other in the subset [14]. DPP  $\mathcal{P}$  over a discrete point set  $\Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$  is determined by a  $M \times M$  positive semi-definite matrix  $L$  indexed by the elements from  $\Omega$  and defining the probability of point selection. In our case,  $\Omega$  is the set of items.

Given a discrete point set  $\Omega$ , a *determinantal point process*  $\mathcal{P}$  is a probability measure defined on  $2^\Omega$ , the set of all subsets of  $\Omega$ , such that if  $A \sim \mathcal{P}$  is a random subset, then we get:

$$\mathcal{P}(A = A) \propto \det(L_A) \quad (1)$$

where  $L_A \equiv [L_{ij}]_{\omega_i, \omega_j \in A}$ . The diagonal elements of  $L$  provide the probabilities of selecting individual items from  $\Omega$  ( $\mathcal{P}(\omega_i) \in W, i = 1, \dots, M$ ), while the off-diagonal elements of  $L$  reflect the negative correlations between item pairs. The larger the values of  $L_{ij}$ , the smaller the tendency of  $\omega_i$  and  $\omega_j$  to co-occur. The determinants of entries  $L_{ij}$  can be viewed as measurements of the similarity between  $\omega_i$  and  $\omega_j$ . Therefore, more similar items are less likely to get selected together. In the RS context, the diagonal elements  $L_{ii}$  can be seen as user's affinities towards an item  $\omega_i$ .

As we mentioned in Section 2, various sampling procedures can be applied to DPP for generating diversified item lists. For instance, MAP inference [5] suggests to iteratively add items  $j$  to the list  $A$  by maximising the following objective function:

$$\mathcal{L}_{MAP} = \log \det(L_{A \cup \{j\}}) - \log \det(L_A) \quad (2)$$

Here, we also consider a pairwise loss function between selected and all remaining items, and uniform sampling under the framework of Bayesian Personalized Ranking (BPR) [19]. For instance, given a user  $u \in U$ , the set of items  $I$ , the BPR loss function for the parameter vector of an arbitrary model class  $\Theta$  is defined as:

$$\mathcal{L}_{BPR} = \sum_{(u,i,j) \in D_T} \ln \sigma(\hat{x}_{uij}) - \lambda_\Theta \|\Theta\|^2 \quad (3)$$

where  $\sigma(x) = \frac{1}{1+e^{-x}}$  is the logistic sigmoid function,  $\hat{x}_{uij}(\Theta)$  is a model specific function estimating a real value of preferences of user  $u$  and items  $i$  and  $j$ ,  $\lambda_\Theta$  are model specific regulation parameters,  $D_T : U \times I \times I$  s.t.  $D_S = \{(u, i, j) | i \in I_u^+ \wedge j \in I \setminus I_u^+\}$ , and  $(u, i, j) \in D_S$  denotes that the user  $u$  prefers the item  $i$  over  $j$ ,  $T \subseteq U \times I$  being

the available user-item interactions and  $I_u^+ = \{i \in I : (u, i) \in T\}$ . The estimator  $\hat{x}_{uij}$  is decomposed as  $\hat{x}_{uij} = \hat{x}_{ui} - \hat{x}_{uj}$ . We will specify the function used for optimisation in the next Section.

#### 3.2 Translation-Based Knowledge Graph Embedding for Recommendation

A *knowledge graph* (KG) is a multi-relational graph  $(V, Edges, R)$  consisting of nodes  $v \in V$ , i.e. *entities* such as users, items, genres, actors, etc., and edges  $e \in E$  defining a *relation* between them  $r \in R$  (e.g. user-item interaction, belonging to a category, being directed by a certain person, etc.). Thus, an edge  $e \in Edges$  is a triplet of the head (or left) entity  $h \in V$ , the relation  $r \in R$ , and the tail (or right) entity  $t \in V$ , i.e.  $e = (h, r, t)$ . We denote by  $r_0$  the affinity relation between a user and an item (user-item interaction), and by  $r_j$  any other relation between entities.

The main idea of translation-based embedding is to project the entities and the relations of the knowledge graph into the *embedding space*, i.e. a  $d$ -dimensional vector space  $\mathbb{R}^d$ . Thus, to each entity  $h \in V$  (resp.  $t \in V$ ) corresponds a vector  $\mathbf{h} \in \mathbb{R}^d$  (resp.  $\mathbf{t} \in \mathbb{R}^d$ ). A *score function*  $f_r(\mathbf{h}, \mathbf{t})$  is defined to measure the plausibility that the triplet is incorrect. In other words, the relation should correspond to a translation of the embeddings. Various score functions have been proposed for achieving better accuracy for tasks as link prediction, node classification and recommendation [10]. Here we use two forms of score functions from TransE and TransH as examples of KGE for recommendation. More advanced and complex forms of score functions are believed to provide a better performance. The score function of TransE [2] is given by:  $f_r^E(\mathbf{h}, \mathbf{t}) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_p$ ,  $\mathbf{r} \in \mathbb{R}^d$ . TransH [29] projects the embeddings  $\mathbf{h}$  and  $\mathbf{t}$  to a relation-specific hyperplane  $\mathbf{w}_r$  and considers the relation-specific translation vector  $\mathbf{d}_r$  in  $\mathbf{w}_r$ . Its score function is defined as:  $f_r^H(\mathbf{h}, \mathbf{t}) = \left\| \underbrace{(\mathbf{h} - \mathbf{w}_r^T \mathbf{h} \mathbf{w}_r)}_{\text{projection of } \mathbf{h} \text{ to } \mathbf{w}_r} + \mathbf{d}_r - \underbrace{(\mathbf{t} - \mathbf{w}_r^T \mathbf{t} \mathbf{w}_r)}_{\text{projection of } \mathbf{t} \text{ to } \mathbf{w}_r} \right\|_2^2$ ,  $\mathbf{w}_r, \mathbf{d}_r \in \mathbb{R}^d$ . For training, the following margin-based ranking loss function [29] can be used:

$$\mathcal{L}_{KGE} = \sum_{(h,r,t) \in \Delta} \sum_{(h',r',t') \in \Delta'} [f_r(\mathbf{h}, \mathbf{t}) + \gamma - f_{r'}(\mathbf{h}', \mathbf{t}') ]_+ \quad (4)$$

where  $[x]_+ \triangleq \max(0, x)$ ,  $\gamma$  is the margin between positive and negative triplets,  $\Delta$  and  $\Delta'$  denote the sets of positive and negative triplets, respectively. The negative triplets  $(h', r', t')$  are the results of the corruption of  $(h, r, t)$ .

### 4 EMDKG MODEL

In this Section, we describe our solution **EMDKG** which targets at optimizing both item diversity representations and knowledge graph embedding for top- $N$  recommendations. In the following of this paper, EMDKG<sub>E</sub> (resp. EMDKG<sub>H</sub>) denotes our solution EMDKG incorporating TransE (resp. TransH) knowledge-graph embedding.

#### 4.1 General Overview

We propose an EM-schemed representation learning for recommendation. Corresponding to the two objectives for diversity-aware translation-based recommendation, the E-step aims at optimising translation-based knowledge graph representations for users, items

and entities with a modified margin-based ranking loss for taking into account the diversity of item vectors. And the M-step aims at learning a diversity-encoded item representations with a pairwise BPR-based loss. The general model alternates the learning processes of these two parts until it converges.

## 4.2 Item Diversity Learning.

Diversity can be considered in terms of categories. In this case, one-hot encoding [34] can be used to represent item features. However, such approach becomes infeasible when the definition of diversity goes wider or the category number increases significantly. To overcome these limitations, vector representations can be used. Indeed, with the growing popularity of matrix factorisation and embedding techniques, modern RS largely rely on user and item representations in continuous vector space. We argue that diversity should be encoded in vector representations of items. To do so, we propose an item diversity learning framework with negative sampling.

We first need to demarcate the concepts of (a) *semantic diversity* based on the available information about the items (take categories/movie genres as example), and (b) *vectorial diversity* based on item vector representations and calculated using item vectors. In the **Item Diversity Learning (IDL)** module, we make use of both concepts. Our motivation behind that is as follows. In traditional RS, item vectors are often learnt by optimising item relevance to a given user profile (user vector). At the same time, a general principle is that similar users tend to like similar items. Thus, there exists a correlation between the similarity of items and their vector representations. However, it is not always the case.

A diversity measure  $div(\cdot)$ , whether on discrete or continuous space, should be defined to characterise the diversity for a given list. For instance, one can use *pairwise vector dissimilarity metrics* like intra-list average distance (ILAD) [33], intra-list minimal distance (ILMD), or *set-level metrics* like category coverage (CC),  $\alpha$ -NDCG and log-determinant of item-indexed kernel matrix employed in DPP [5] defining diversity in the vector space of the entire list of items. As semantic information is handled in one-hot encoding on discrete space, only ILAD, ILMD, CC and  $\alpha$ -NDCG can be used to calculate this information. In contrast, learned vector representations are in continuous vector space, and metrics as determinant, ILAD and ILMD are the possible diversity measures here.

For a given user  $u \in U$ , the IDL-module aims at learning the representations of items which can reflect the semantic diversity based on the available information (features, relations). As input, it takes the set of ground-truth item sets, denoted  $\{T_+\}$ , each of the element  $T_+$  (one individual item set) having the same size (length) consisting of the items the user had interactions with<sup>1</sup>. We consider the item sets in this set to be the most diverse for a given user.

Based on each element from the ground-truth set  $T_+$  and the remaining items  $I \setminus T_+$ , the *negative sampling* is performed by randomly replacing all the items from  $T_+$  except for one with the items from  $I \setminus T_+$ . Note that the size of the item sets is the same, i.e.  $|T_-| = |T_+|$ . The items for substitution are selected so that the overall semantic diversity of these negative item sets is inferior to the one of the ground-truth. In other words, if  $div(\cdot)$  is a list diversity

measure, then  $div(T_+) > div(T_-)$ . At this step, we consider semantic diversity and suggest to apply dissimilarity measure on discrete space to calculate it. Thus, negative item sets  $T_-$  are generated.

Once the negative items sets are constructed, the IDL-module learns item vector representations by optimising the following loss function:  $\mathcal{L}_{IDL} = -\log \left( 1 + e^{-div(T_+) + div(T_-)} \right)$ . Here, we make use of vectorial diversity and apply log-determinant measure to calculate it. Thus, the loss function gets the following form:

$$\mathcal{L}_{IDL} = \sum_{A_+ \in \{T_+\}} \sum_{A_- \in \{T_-\}} -\log (\sigma(\log \det(L_{A_+}) - \log \det(L_{A_-})) \quad (5)$$

where  $A_+$  is the ground truth diverse item set, and  $A_-$  is one item set from all negative item sets,  $L_{A_+}$  and  $L_{A_-}$  are the kernel matrix of DPP indexed with the elements from  $A_+$  and  $A_-$ , respectively.

This kernel matrix is built as follows. Given an item list  $A \subset I$ , we notate with  $\mathbf{v}_A$  a vector representation of the item list  $A$ . Then  $L_A = \mathbf{v}_A \mathbf{v}_A^\top$  is a positive semidefinite matrix. The larger the value  $\det(L_A)$  is, the more diverse the item set  $A$  is [14].

As the result of the IDL-module, we obtain vector representations of items that reckon with semantic similarity and vector similarity.

## 4.3 Modified Translation-Based Knowledge Graph Embedding for Recommendation.

Auxiliary semantic information related to items such as categories, film directors, actors, etc. can provide valuable assets for enhancing recommendation accuracy [3, 15]. Such auxiliary information can contain categorical information and relational knowledge with other entities which do not engage in the recommendation directly. It has been shown [10] that a translation-based knowledge graph embedding for recommendation can efficiently take into account (1) various entities that may affect user's preferences/choices for items and (2) different types of relations between them.

In this work, when constructing a knowledge graph, in terms of modelled relations, we distinguish between user-item interactions and any other relation between entities. We refer to the latter as auxiliary relations. Let  $r_0$  be the relation between users and items reflecting a user-item interaction. As described in Section 3.2, for the triplet  $(u, r_0, i)$  we can define a translation-based score function  $f_{r_0}(u, i)$  to measure the affinity value between the user  $u$  and the item  $i$ . The smaller the value of  $f_{r_0}(u, i)$  is, the larger the affinity between the user  $u$  and item  $i$  is. Similarly, we can define the score function  $f_{r_j}(i, e)$  for any auxiliary relation  $r_j$ ,  $j \in \{1, 2, \dots\}$  between the item  $i$  and the auxiliary entity  $e \in E = V \setminus \{U \cup I\}$ .

To learn the embedding accounting for both types of relations (user-item interactions  $r_0$  and auxiliary relations  $r_j, j \neq 0$ ), we can rewrite the margin-based ranking loss function from eq. 4 as:

$$\mathcal{L}_{KGE} = \sum_{(u, r_0, i)} \sum_{(u, r_0, i')} [-f_{r_0}(u, i) + f_{r_0}(u, i'), 0]_{++} + \sum_{(i, r_j, e)} \sum_{(i, r_j, e')} [-f_{r_j}(i, e) + f_{r_j}(i, e'), 0]_{++}. \quad (6)$$

As stated above, such loss function aims at separating the golden triplets (both, historical user-item interactions and existent item-entity relations) from the negative triplets.

<sup>1</sup>The generation of the ground-truth item sets of a given size is out of the focus of our model. In Section 5, we will precise the procedure used for evaluation.

However, conventional KGE only considers the existent relations of entities in the graph and does not explicitly optimize the diversity representations of item vectors as we do in Section 4.2. Thus we propose a modified knowledge graph embedding loss function to bridge KGE and Item Diversity Learning:

$$\mathcal{L}_{xKGE} = \mathcal{L}_{KGE} + KLDivergence(\mathbf{v}_I^{KGE}, \mathbf{v}_I^{IDL}) \quad (7)$$

where  $\mathbf{v}_I^{KGE}$  and  $\mathbf{v}_I^{IDL}$  represent correspondingly the item vectors in KGE and Item Diversity Learning. We minimize KL-divergence of the vector representations of items in two modules in order to resemble the two item representations. Finally, we alternate the learning of knowledge graph embedding and item diversity learning by alternating optimisation of functions (eq. 5) and (eq. 7) until the the learning converges. Thus, we can formalize our proposal EMDKG as a dual-goal optimization problem of the Diversity-Aware Translation-Based Recommendation as follows.

**(Diversity-Aware Translation-Based Recommendation.)** Given a set of users  $U$ , items  $I$ , other entities  $E$ , historical user-item interactions  $H_{u,r_0,i}$  and item-side relation information triplets  $H_{i,r_j,e}$ , the diversity-aware translation-based recommendation aims at minimising two loss functions 5 and 7 simultaneously.

The process of co-learning is shown in Algorithm 1.

---

**Algorithm 1** Co-learning of KGE and IDL

---

**Require:**  $\mathbf{T} = \{T_+\}$ ,  $\mathbf{P} = \{(u, i)\}$ ,  $\mathbf{Q} = \{(i, r, e)\}$ ,  $k$

**Ensure:**  $\mathbf{v}_I^{IDL}, \mathbf{v}_I^{KGE}, \mathbf{v}_U, \mathbf{v}_R$

- 1: Initialize  $\mathbf{v}_I^{IDL}, \mathbf{v}_I^{KGE}, \mathbf{v}_U, \mathbf{v}_R$
  - 2: **while** not converge **do**
  - 3:   **for**  $k$  times **do**
  - 4:     Get batch  $\mathbf{p} \subseteq \mathbf{P}$  and  $\mathbf{q} \subseteq \mathbf{Q}$
  - 5:     Optimize Eq.(7) with  $\mathbf{p}$  and  $\mathbf{q}$
  - 6:   Get batch  $\mathbf{t} \subseteq \mathbf{T}$
  - 7:   Using negative sampling to obtain  $\mathbf{t}_-$  from  $\mathbf{t}$
  - 8:   Optimize Eq.(5) with  $\mathbf{t}$  and  $\mathbf{t}_-$
- 

#### 4.4 Prediction

To make a top- $N$  recommendation, we take the learned vectors of users and items and affinity relation  $r_0$  from KGE and calculate the affinity score for each user  $u$  with any item  $i$  using the score function  $f_{r_0}(\mathbf{u}, \mathbf{i})$ . For each user  $u \in U$ , we sort the items by ascending score function values, and return the top- $N$  items as the result.

To further adjust the recommendation list, we can also apply diversification methods to balance the accuracy-diversity trade-off of the list. In diversification methods, such as MMR [4], XQuAD [22], a threshold parameter that we denote  $\alpha$  is used to adjust the trade-off between accuracy and diversity. For instance, MMR optimisation function is given by:  $\max\{\alpha \text{Sim}_1(u, i) - (1 - \alpha) \max(\text{Sim}_2(i, j))\}$ , where  $\text{Sim}_1(u, i)$  reflects the relevance of the item  $i$  to the user  $u$ , and  $\text{Sim}_2(\cdot, \cdot)$  is a similarity measure between two items.

We propose the following MMR-like determinant-based trade-off equation:  $\log \det(L_A) \propto \alpha \sum Q(u, i) + (1 - \alpha) \cdot \log \det(L_A)$ , which is equivalent to the log-determinant of submatrix  $A$  on a new kernel  $L_u = \text{Diag}(\exp(\beta Q(u))) \cdot L \text{Diag}(\exp(\beta Q(u)))$ . The change of operations from  $-$  to  $+$  is due to the semantics of  $\log \det(\cdot)$ ,

interpreted as the dissimilarity of the item list  $\cdot$ , contrary to the  $\text{Sim}_2(\cdot, \cdot)$  being the similarity between any item pair.  $Q(u, i)$  is the affinity measure between any user-item pair  $(u, i)$ . Higher value of  $Q(u, i)$  corresponds to closer affinity of the pair  $(u, i)$ . However, in KGE setting, the lower the value of  $f_{r_0}(\mathbf{u}, \mathbf{i})$  is, the higher the affinity is. Thus, we apply a monotonically decreasing function, i.e.  $\exp(-x)$  to satisfy the requirements of  $Q(u, i)$ .

## 5 EVALUATION

### 5.1 Experiment Settings

**5.1.1 Datasets & ground truth.** We consider two publicly available datasets for recommendation: **MovieLens-100k**<sup>2</sup>, also used in [28, 31, 32] and **Anime**<sup>3</sup>, also used in [31]. In both cases, we keep only users with at least 20 ratings. Thus, the **MovieLens-100k** dataset, denoted as ML-100K, contains 100K of 5-grades ratings from 943 users on 1,682 movies (with at least 20 ratings per user). Following common practice in the field, we consider ratings of four and higher as positive implicit feedback, leading to 82K positive interactions. Movies are categorized into 18 non-exclusive genres (e.g. action, adventure, comedy, fantasy, etc), so any movie belongs to at least one category. Similar to [9], we combined ML-100K and the IMDb dataset<sup>4</sup> to construct the related knowledge graph. We extract the ratings and 12 categories of information, including movie genre, director, actor, actress, composer, writer, production designer, archive footage, archive sound, cinematographer, producer and editor, that are used to define the entities and relations within our knowledge graph. The **Anime** dataset contains 1 million 10-grades ratings from 73,516 users on 12,294 animes (with at least 20 ratings per user). Ratings of 6/10 and higher are treated as positive feedback, leading to more than 1.8 million of positive interactions. Each anime may belong to one or more of the 44 non-exclusive genres available in the dataset, e.g. drama, romance, slice of life, action, etc. We used these categories to construct the knowledge graph. We use *leave-one-out* strategy to divide datasets into training/ validation/ test datasets. For generating ground-truth item sets for a given user, we use items from historical user-item interactions as candidates and randomly generate item sets of length 10 and keep the top-100 item sets with the highest semantic diversity scores.

**5.1.2 Baselines.** We consider both diversified and not diversified baselines to evaluate EMDKG in terms of diversity, accuracy and diversity-accuracy trade-off:

**BPRMF** [19] is a MF approach that uses a pairwise ranking loss to provide recommendations. Similar to EMDKG, it considers implicit feedback. However, it does not focus on diversity, nor it uses relational information from knowledge graph.

**FISM** [11] is an item-based CF method that exploits user-item relations through low-dimensional latent factor matrices. It is a strong baseline for highly sparse datasets.

**TransKG**<sub>[.]</sub> [9] models users, items and all the associated entities in a knowledge graph, then uses the embedded vectors obtained with translation-based KG embedding, TransE [2] (TransKG<sub>E</sub>) or TransH [29] (TransKG<sub>H</sub>), to give the result.

<sup>2</sup><https://grouplens.org/datasets/movielens/100k/>

<sup>3</sup><https://www.kaggle.com/CooperUnion/anime-recommendations-database>

<sup>4</sup><https://datasets.imdbws.com/>

**Table 1: Accuracy and diversity results before diversification method for ml100k extended datasets. Bold results are significantly higher than other results in the same column with  $p = 0.01$ .**

| Metrics              | Hit (%)      |              |              | NDCG (%)     |              |              | CC (%)       |       |       | $\alpha$ -NDCG (%) |              |              |
|----------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------|-------|--------------------|--------------|--------------|
|                      | @5           | @10          | @20          | @5           | @10          | @20          | @5           | @10   | @20   | @5                 | @10          | @20          |
| BPRMF                | 8.91         | 17.71        | 32.03        | 5.51         | 8.34         | 11.93        | 34.96        | 49.51 | 65.32 | 39.78              | 51.61        | 62.96        |
| FISM                 | 22.87        | 31.44        | 42.82        | 15.47        | 18.23        | 21.09        | 36.57        | 50.21 | 63.47 | 39.73              | 51.73        | 62.31        |
| IRGAN                | 10.55        | 16.50        | 23.45        | 6.99         | 8.91         | 10.65        | 37.17        | 53.26 | 69.18 | 37.55              | 51.92        | 64.74        |
| RCF                  | 12.20        | 19.51        | 29.59        | 7.51         | 9.86         | 12.39        | 37.14        | 53.40 | 69.88 | 38.60              | 52.71        | 65.86        |
| TransKG <sub>E</sub> | 22.07        | 32.43        | 46.49        | 14.25        | 17.57        | 21.13        | 36.18        | 50.87 | 65.73 | 41.83              | 53.85        | 65.14        |
| TransKG <sub>H</sub> | <b>23.38</b> | <b>34.36</b> | <b>47.01</b> | <b>15.70</b> | <b>19.24</b> | <b>22.44</b> | 36.05        | 50.93 | 65.65 | 41.50              | 53.77        | 64.95        |
| EMDKG-E              | 22.12        | 33.65        | 46.03        | 14.59        | 17.97        | 21.32        | 36.45        | 51.22 | 66.24 | 42.00              | 54.23        | 65.58        |
| EMDKG-H              | 22.08        | 33.01        | 46.34        | 14.08        | 17.60        | 20.95        | <b>37.34</b> | 52.15 | 66.67 | <b>43.27</b>       | <b>55.50</b> | <b>66.66</b> |

**Table 2: Accuracy and diversity results before diversification method for anime datasets. Bold results are significantly higher than other results in the same column with  $p = 0.01$ .**

| Metrics              | Hit (%) |              |              | NDCG (%) |              |              | CC (%)       |              |              | $\alpha$ -NDCG(%) |              |       |
|----------------------|---------|--------------|--------------|----------|--------------|--------------|--------------|--------------|--------------|-------------------|--------------|-------|
|                      | @5      | @10          | @20          | @5       | @10          | @20          | @5           | @10          | @20          | @5                | @10          | @20   |
| BPRMF                | 8.13    | 11.56        | 17.81        | 5.13     | 6.25         | 7.83         | 22.95        | 34.23        | 46.46        | 23.95             | 32.06        | 40.21 |
| FISM                 | 11.72   | 17.03        | 22.97        | 7.29     | 8.98         | 10.48        | 24.95        | 37.31        | 50.43        | 25.63             | 34.26        | 42.97 |
| IRGAN                | 14.84   | 19.38        | 24.06        | 9.79     | 11.30        | 12.46        | 26.28        | 35.63        | <b>52.43</b> | 27.97             | 34.89        | 44.95 |
| RCF                  | 13.02   | 16.37        | 19.22        | 7.46     | 9.01         | 11.20        | 26.80        | 38.99        | 51.97        | 28.10             | 34.94        | 45.33 |
| TransKG <sub>E</sub> | 14.38   | 20.00        | 25.63        | 9.37     | 11.00        | 12.54        | 26.64        | 39.56        | 52.09        | 29.41             | 38.47        | 47.30 |
| TransKG <sub>H</sub> | 13.44   | 20.00        | 26.88        | 9.06     | 11.02        | 12.89        | 26.70        | 39.77        | 52.21        | 29.52             | 38.66        | 47.31 |
| EMDKG-E              | 14.84   | <b>20.16</b> | 26.25        | 9.72     | <b>11.42</b> | 12.89        | 26.65        | 39.40        | 51.50        | 29.14             | 38.08        | 46.77 |
| EMDKG-H              | 14.84   | 19.22        | <b>27.34</b> | 9.64     | 11.05        | <b>13.08</b> | <b>27.00</b> | <b>40.66</b> | 51.60        | 29.49             | <b>38.94</b> | 47.14 |

**IRGAN** [28] is a framework that makes compete a discriminative and a generative model in an adversarial way to solve several IR problems, including top- $N$  recommendation.

**RCF** [32] is a CF approach that exploits user-item relations and other item relations through knowledge graph embedding.

These baselines have not been designed to promote diversity. We use them to evaluate the accuracy of our approach before applying diversification to the recommended list. To ensure a fair comparison when evaluating diversity ability, we adopt a similar approach as [9]. We thus combine each algorithm with two baselines diversification techniques, namely **MMR** [4] and the method from [5] that we denote **FastDPP**. MMR is a well-known re-ranking approach to promote diversity in recommendation, that iteratively re-rank items to adjust the trade-off between accuracy and diversity. FastDPP promotes diversity by re-ranking the results using DPP and MAP inference. In total, we obtain 12 diversified baselines.

Finally, we also compare EMDKG with two recent baselines that consider DPP to provide diversity-promoting recommendations:

**DivKG** [9] is a re-ranking approach exploiting multiple relations of items using KGE to provide recommendations. It combines TransKG with FastDPP to enhance the diversity of the result list, but it does not explicitly encode diversity into the representations.

**PD-GAN** [31] combines DPP with GAN to generate personalized and diverse recommendations without using the relations between items and other entities or knowledge graphs.

For reproducibility sake, we provide our source code<sup>5</sup>.

**5.1.3 Performance Metrics.** To assess the accuracy of the proposed method and the baselines, we consider standard metrics widely used in the field (e.g. [11, 28]), namely Hit Rate (Hit@ $n$ ) and Normalized Discounted Cumulative Gain (NDCG@ $n$ ). We take  $n = \{5, 10, 20\}$ .

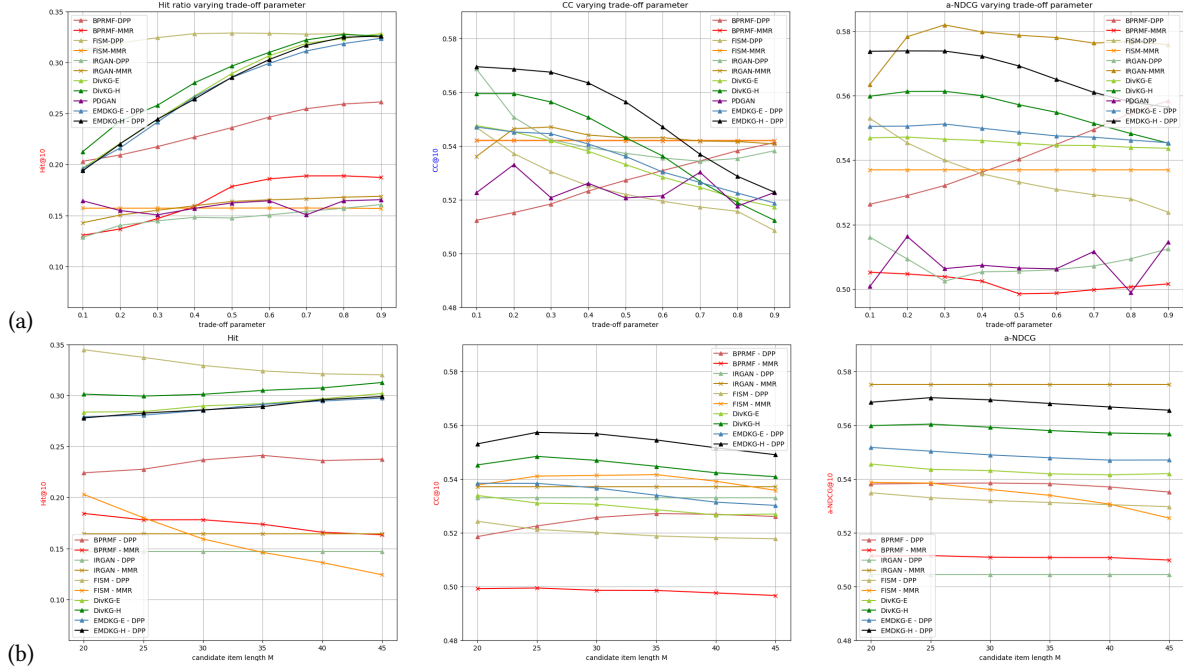
Similar to [6, 21, 31], we evaluate the diversity through: (1) Category Coverage  $CC@n = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{|C_u^n|}{|C|}$ , and (2)  $\alpha$ -NDCG@ $n$  [7]  $\alpha$ -NDCG@ $n = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{\alpha DCG_u@n}{\alpha IDCG_u@n}$  in which  $\alpha DCG_u@n = \sum_{k=1}^n \frac{\sum_{l=1}^L J_{kl}^u (1-\alpha)^{q_{l,k-1}^u}}{\log_2(1+k)}$ , with  $J_{kl}^u$  equals to the rating of the  $k$ th item in the list for user  $u$  if  $k$ th item belongs to the genre  $l$  otherwise 0.  $q_{l,k-1}^u$  counts the number of items belonging to genre  $l$  up to the  $k-1$  position in the list, which accompanying the constant  $\alpha$  to modify the redundancy in the recommendation list.  $\alpha IDCG_u@n$  denotes the largest value of  $\alpha DCG_u@n$  which achieves the ideal diversification of recommendation lists. We take  $n = \{5, 10, 20\}$ .

**5.1.4 Hyperparameter Settings.** For hyperparameters in KGE we use the same variations of values as in TransKG. In IDL we generate 99 negative item sets for each ground-truth diverse item set.

## 5.2 Results

**5.2.1 General results.** First, we evaluate EMDKG before diversification. Table 1 shows the results of accuracy and diversity for each

<sup>5</sup>[https://github.com/LGanShare/EMDKG\\_WIgit](https://github.com/LGanShare/EMDKG_WIgit)



**Figure 1: Comparison of EMDGKG against state-of-the-art approaches in terms on Hit Ratio (left), Coverage (center),  $\alpha$ -NDCG (right) when varying: (a) the trade-off parameter, (b) the candidate item list length (trade-off parameter: 0.5)**

method on MovieLens. We can see that EMDKG-E and EMDKG-H perform much better for all accuracy and diversity metrics compared to BPRMF. Compared to FISM, although its performance in term of accuracy shows a slight advantage (not statistically significant), EMDKG-E and EMDKG-H bring a significant improvement in diversity. Although IRGAN and RCF show higher values in terms of  $CC@10$  and  $CC@20$ , EMDKG-E and EMDKG-H still defeats these two methods in other diversity metrics and largely in accuracy. Compared to TransKGs, it is observed that  $TransKG_H$  outperforms EMDKG-E and EMDKG-H in terms of accuracy but EMDKG-H compensates it by an obvious gain in diversity w.r.t. all diversity metrics. Besides, EMDKG-E shows both improvements in accuracy and diversity comparing to  $TransKG_E$ .

Table 2 shows the results of accuracy and diversity on dataset Anime. Our proposals EMDKG-E and EMDKG-H outperform BPRMF and FISM in all metrics of accuracy and diversity. Compared to IRGAN, although IRGAN show comparable results or slightly better results in terms of accuracy, EMDKG-E and EMDKG-H win with a margin for most diversity metrics except for  $CC@20$ . Compared to  $TransKG_E$  and  $TransKG_H$ , EMDKG-E and EMDKG-H respectively have outperform on accuracy and diversity. We have conducted pairwise t-test to confirm the result difference with confidence 99%.

**5.2.2 Impact of trade-off parameter  $\alpha$ .** We show the impact of parameter  $\alpha$  in Fig. 1 (a). We choose  $Hit@10$  as accuracy metric and  $CC@10$  and  $\alpha$ -NDCG@10 as diversity metrics for each method combined with one of diversification methods *MMR* or *FastDPP*. In each chart, we present the results of the corresponding metric for every method by varying the trade-off parameter  $\alpha$ . We can see that by increasing the value  $\alpha$ , accuracy metrics tend to increase for most

methods combined with any of the diversification method while diversity metrics have tendency of increasing in most diversification combinations. In terms of accuracy, EMDKG demonstrates a huge advantage under various  $\alpha$  compared to other baseline methods, except for DivKG and FISM. Besides, we can tell that the combinations of EMDKG and FastDPP maintain better their accuracy while decreasing the  $\alpha$  compared to those with MMR. While EMDKG and DivKG have comparable results in accuracy, EMDKG improves diversity in terms of both  $CC$  and  $\alpha$ -NDCG for both diversification methods. Compared to FISM, EMDKG wins with a large margin in terms of both  $CC@10$  and  $\alpha$ -NDCG.

In terms of diversity, EMDKG-H shows competitive results in both diversity and accuracy results compared to all other methods. While the combinations of IRGAN+MMR gains advantage of  $\alpha$ -NDCG, it also presents much lower accuracy for  $Hit@10$  and lower diversity for  $CC@10$ . And compared to DivKG (noted as  $TransH+DPP$ ) and  $TransH+MMR$ , although our proposal does not gain a huge margin in terms of accuracy (still in general no worse than both of them), EMDKG-H +DPP and EMDKG-H +MMR bring obvious improvements for both diversity metrics when varying trade-off parameter  $\alpha$  correspondingly.

**5.2.3 Impact of candidate item set length  $M$ .** We show the impact of candidate item set length  $M$  for applying diversification methods in Fig. 1 (b). To speed up the re-ranking and keep the accuracy performance, we select the top- $M$  ( $M > N$ ) items from the prediction of each recommendation before diversification as candidate items for re-ranking. We can tell from Fig. 1 (b) that EMDKG combined with FastDPP achieves stable results for both diversity and accuracy, while the combination with MMR suffers from a loss of accuracy

when increasing the size  $M$  of candidate item sets Besides, EMDKG shows better stability compared to DivKG and TransKG+MMR.

## 6 CONCLUSION

In this paper, we aim at learning representations for top- $N$  recommendation to achieve better accuracy-diversity trade-off. In our perspective, the latter is not limited to only achieve both high accuracy and diversity, but should also be robust under different parameter settings. Thus, we propose a novel EM-schemed diversity-encoded knowledge graph embedding model EMDKG which incorporates Item Diversity Learning and Knowledge Graph Embedding for this purpose. We compare EMDKG with multiple state-of-art baseline methods before and after applying two diversification methods MMR and DPP. The results show that before diversification EMDKG can adjust accuracy and diversity to a better trade-off and after diversification EMDKG can outperform the baselines when varying the trade-off parameters and the candidate item set length. In all, EMDKG demonstrates better performance in terms of accuracy-diversity trade-off compared to competitive state-of-art works.

## ACKNOWLEDGMENTS

Lu Gan is supported by the China Scholarship Council for Ph.D. program (No. 201801810094) at INSA Lyon, France. Experiments presented in this paper were carried out using the Grid'5000 testbed.

## REFERENCES

- [1] Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Jeong. 2009. Diversifying search results. In *Proc. of the Second ACM International Conference on Web Search and Data Mining - WSDM '09*. 5.
- [2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Proc. of the 26th International Conference on Neural Information Processing Systems - Volume 2*. 2787–2795.
- [3] 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 *The World Wide Web Conference*. 151–161.
- [4] Jaime Carbonell and Jade Goldstein. 1998. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In *Proc. of the 21st International ACM SIGIR Conference on Research and Development in Information Retrieval*. 335–336.
- [5] Laming Chen, Guoxin Zhang, and Hanning Zhou. 2018. Fast Greedy MAP Inference for Determinantal Point Process to Improve Recommendation Diversity. In *Proc. of the 32nd International Conference on Neural Information Processing Systems*. 5627–5638.
- [6] Peizhe Cheng and Shuaiqiang Wang. 2017. Learning to Recommend Accurate and Diverse Items. In *Proc. of the 26th International Conference on World Wide Web*. 183–192.
- [7] Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and Diversity in Information Retrieval Evaluation. In *Proc. of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 659–666.
- [8] Farzad Eskandarian and Bamshad Mobasher. 2020. Using Stable Matching to Optimize the Balance between Accuracy and Diversity in Recommendation. In *Proc. of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. 71–79.
- [9] Lu Gan, Diana Nurbakova, Léa Laporte, and Sylvie Calabretto. 2020. Enhancing Recommendation Diversity Using Determinantal Point Processes on Knowledge Graphs. In *Proc. of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2001–2004.
- [10] Ruining He, Wang-Cheng Kang, and Julian McAuley. 2017. Translation-Based Recommendation. In *Proc. of the Eleventh ACM Conference on Recommender Systems*. 161–169.
- [11] Santosh Kabbur, Xia Ning, and George Karypis. 2013. FISM: Factored Item Similarity Models for Top-N Recommender Systems. In *Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 659–667.
- [12] Marius Kaminskis and Derek Bridge. 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Trans. on Interactive Intelligent Systems* 7, 1 (2016), 2:1–2:42.
- [13] Alex Kulesza and Ben Taskar. 2011. k-DPPs: Fixed-size determinantal point processes. In *Proc. of the 28th International Conference on Machine Learning*. 1193–1200.
- [14] Alex Kulesza and Ben Taskar. 2012. Determinantal Point Processes for Machine Learning. *Found. Trends Mach. Learn.* 5, 2-3 (2012), 123–286.
- [15] Dongho Lee, Byungkook Oh, Seungmin Seo, and Kyong-Ho Lee. 2020. News Recommendation with Topic-Enriched Knowledge Graphs. In *Proc. of the 29th ACM International Conference on Information and Knowledge Management*. 695–704.
- [16] Yong Liu, Yingtai Xiao, Qiong Wu, Chunyan Miao, Juyong Zhang, Binqiang Zhao, and Haihong Tang. 2020. Diversified Interactive Recommendation with Implicit Feedback. In *The 34th AAAI Conference on Artificial Intelligence, AAAI 2020, The 32nd Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The 10th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020*. 4932–4939.
- [17] Zelda Mariet, Mike Gartrell, and Suvrit Sra. 2019. Learning Determinantal Point Processes by Corrective Negative Sampling. 89 (2019), 2251–2260.
- [18] Lijing Qin, Shouyuan Chen, and Xiaoyan Zhu. 2014. Contextual Combinatorial Bandit and its Application on Diversified Online Recommendation. In *Proc. of the 2014 SIAM International Conference on Data Mining*. 461–469.
- [19] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proc. of the 25th Conference on Uncertainty in Artificial Intelligence*. 452–461.
- [20] Marco Tulio Ribeiro, Anisio Lacerda, Adriano Veloso, and Nivio Ziviani. 2012. Pareto-Efficient Hybridization for Multi-Objective Recommender Systems. In *Proc. of the 6th ACM Conference on Recommender Systems*. 19–26.
- [21] Rodrygo L.T. Santos, Craig Macdonald, and Iadh Ounis. 2010. Selectively Diversifying Web Search Results. In *Proc. of the 19th ACM International Conference on Information and Knowledge Management*. 1179–1188.
- [22] Rodrygo L. T. Santos, Jie Peng, Craig Macdonald, and Iadh Ounis. 2010. Explicit Search Result Diversification through Sub-Queries. In *Proc. of the 32nd European Conference on Advances in Information Retrieval*. 87–99.
- [23] Yue Shi, Xiaoxue Zhao, Jun Wang, Martha Larson, and Alan Hanjalic. 2012. Adaptive Diversification of Recommendation Results via Latent Factor Portfolio. In *Proc. of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 175–184.
- [24] Jianing Sun, Wei Guo, Dengcheng Zhang, Yingxue Zhang, Florence Regol, Yaochen Hu, Huifeng Guo, Ruiming Tang, Han Yuan, Xiquiang He, and Mark Coates. 2020. A Framework for Recommending Accurate and Diverse Items Using Bayesian Graph Convolutional Neural Networks. In *Proc. of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2030–2039.
- [25] Saúl Vargas, Linas Baltrunas, Alexandros Karatzoglou, and Pablo Castells. 2014. Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *Proc. of the 8th ACM Conference on Recommender systems*. 209–216.
- [26] Saúl Vargas and Pablo Castells. 2014. Improving Sales Diversity by Recommending Users to Items. In *Proc. of the 8th ACM Conference on Recommender Systems*. 145–152.
- [27] Saul Vargas, Pablo Castells, and David Vallet. 2011. Intent-Oriented Diversity in Recommender Systems. In *Proc. of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1211–1212.
- [28] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. 2017. IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models. In *Proc. of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 515–524.
- [29] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. In *Proc. of the 28th AAAI Conference on Artificial Intelligence*. 1112–1119.
- [30] Mark Wilhelm, Ajith Ramanathan, Alexander Bonomo, Sagar Jain, Ed H. Chi, and Jennifer Gillenwater. 2018. Practical Diversified Recommendations on YouTube with Determinantal Point Processes. In *Proc. of the 27th ACM International Conference on Information and Knowledge Management*. 2165–2173.
- [31] Qiong Wu, Yong Liu, Chunyan Miao, Binqiang Zhao, Yin Zhao, and Lu Guan. 2019. PD-GAN: Adversarial Learning for Personalized Diversity-Promoting Recommendation. In *Proc. of the 28th International Joint Conference on Artificial Intelligence, IJCAI-19*. 3870–3876.
- [32] Xin Xin, Xiangnan He, Yongfeng Zhang, Yongdong Zhang, and Joemon Jose. 2019. Relational Collaborative Filtering: Modeling Multiple Item Relations for Recommendation. In *Proc. of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 125–134.
- [33] Mi Zhang and Neil Hurley. 2008. Avoiding Monotony: Improving the Diversity of Recommendation Lists. In *Proc. of the 2008 ACM Conference on Recommender Systems*. 123–130.
- [34] A. Zheng and A. Casari. 2018. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly. <https://books.google.fr/books?id=Ho0UvGAAACAJ>