Enhancing Recommendation Diversity using Determinantal Point Processes on Knowledge Graphs

Lu Gan
INSA Lyon, University of Lyon
lu.gan@insa-lyon.fr

Diana Nurbakova
INSA Lyon, University of Lyon
diana.nurbakova@insa-lyon.fr

Léa Laporte
INSA Lyon, University of Lyon
lea.laporte@insa-lyon.fr

Sylvie Calabretto
INSA Lyon, University of Lyon
sylvie.calabretto@insa-lyon.fr

ABSTRACT

Top-N recommendations are widely applied in various real life domains and keep attracting intense attention from researchers and industry due to available multi-type information, new advances in AI models and deeper understanding of user satisfaction.While *accuracy* has been the prevailing issue of the recommendation problem for the last decades, other facets of the problem, namely *diversity* and *explainability*, have received much less attention. In this paper, we focus on enhancing diversity of top-N recommendation, while ensuring the trade-off between accuracy and diversity. Thus, we propose an effective framework **DivKG** leveraging knowledge graph embedding and determinantal point processes (DPP). First, we capture different kinds of relations among users, items and additional entities through a knowledge graph structure. Then, we represent both entities and relations as k-dimensional vectors by optimizing a margin-based loss with all kinds of historical interactions. We use these representations to construct kernel matrices of DPP in order to make top-N diversified predictions. We evaluate our framework on MovieLens datasets coupled with IMDb dataset. Our empirical results show substantial improvement over the state-of-the-art regarding both accuracy and diversity metrics.

CCS CONCEPTS

• **Information systems** → **Collaborative filtering**; **Personalization**.

KEYWORDS

Recommender Systems; Knowledge Graph; Diversity; Determinantal Point Processes

ACM Reference Format:
Lu Gan, Diana Nurbakova, Léa Laporte, and Sylvie Calabretto. 2020. Enhancing Recommendation Diversity using Determinantal Point Processes on Knowledge Graphs. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20), July 25–30, 2020, Virtual Event, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3397271.3401213

ACM ISBN 978-1-4503-8016-4/20/07…$15.00
https://doi.org/10.1145/3397271.3401213

1 INTRODUCTION

Top-N recommender systems are widely exploited in real life scenarios and have been intensely studied in the last decades. They aim at selecting a set of N items that represent the highest interest to a user. Among various kinds of recommendation techniques, collaborative filtering methods (CF), in particular model-based matrix factorization methods such as [6, 13], have been widely suggested due to their predictive power in terms of accuracy. They make use of user-item interactions in order to determine user preferences.

However, the use of only this type of relations lacks explicit semantics and implies the search for latent user-item relations. Besides, apart from user-item interactions (among which rating is commonly used), there exist various relations between items and other entities that could be helpful for a better understanding of the users' behaviour. All these relationships can be modeled as a graph structure that provides richer information about the users, items, and their interactions. Mind that this graph may also contain the direct user-item interaction as one of its relations and therefore, can be seen as an extension of CF model. Figure 1 demonstrates such a knowledge graph structure in a movie recommendation domain.

Furthermore, knowledge graph embedding methods [1, 7, 11, 14], naturally capturing and conserving different types of relations among various kinds of entities including users, items and others, can provide a promising model for this purpose. F. Zhang et *al.* propose the framework CKE [18] incorporating one translation-based embedding method Bayesian TransR for recommendation. X. Xin et *al.* [17] propose a two-layer relational collaborative filtering method RCF to exploit knowledge graph embedding for top-N recommendation. Both of them have revealed improved performance of recommendation accuracy due to exploiting structural information on knowledge graphs.

However, accuracy should not be the ultimate goal of the recommendation task as it results in returning to the user highly similar items, ignoring the relations between them, and finally, decreasing user's satisfaction with the provided service. For example, in E-commerce, after detecting a user's interest in laptops, a recommender system returning a list purely composed of laptops is inefficient as the user is very unlikely to purchase more than one model at a time. In movie recommendation, users may get bored after watching several theme-alike movies sequentially. Thus, a returned list of items should be diverse enough, implying both, redundancy reduction and novelty increase. Despite the importance of diversity, it has received much less attention than accuracy. Carbonell and Goldstein provide a diversification method Maximal Marginal

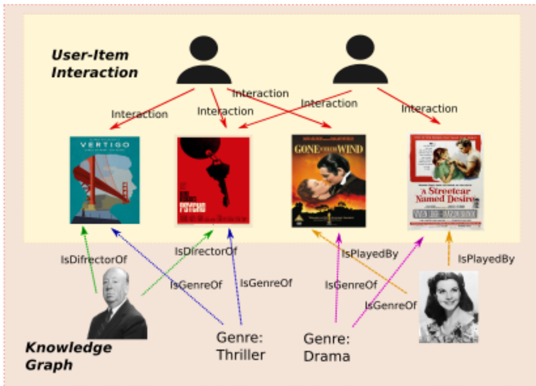

Figure 1: An example of knowledge graph for movie recommendation, reflecting different relations (*Interaction, IsGenreOf,* etc.) between various entities (user, item, genre, etc.).

Relevance (MMR) [6] for re-ordering the items through an iterative process by selecting the most dissimilar item to the existent item list. Further, A. Borodin *et al.* [2] provide its extended version and redefine the problem as max sum diversification problem in order to give a theoretically provable solution. However, these diversification solutions adopt pairwise similarity which tends to be suboptimal by ignoring the correlation within the item list.

A recent emergence of *determinantal point processes* (DPPs) brings new potential to enhancing diversity in multiple machine learning problems, such as extractive summarization [10] and basket completion [5, 15]. DPPs are probabilistic models of sets parameterized with positive semi-definite matrix which characterizes naturally both element-wise relevance towards a query/user and the repulsiveness of subsets on the ground set. DPP captures item similarity in an unified feature space and propose list-wise dissimilar items. Thus, using DPP in recommendation models improves their results in terms of diversity. The challenge here is to find an efficient way to construct the positive semi-definite kernel matrix in order to balance relevance and diversity of items.

In this paper, we address the top-N recommendation problem from diversity perspective, while ensuring a trade-off between accuracy and diversity. To achieve this goal, we propose a general framework that incorporates knowledge graph embedding with DPP. To the best of our knowledge, our approach is the first to combine graph embedding methods with DPP Maximum a Posteriori (MAP) inference to solve top-N diversity-aware recommendation problem. We propose a simplistic but effective way to construct DPP kernel matrix based on knowledge graph embedding results which proves to enhance recommendation performance. We conduct extensive experiments on MovieLens datasets to prove the effectiveness of the combination of knowledge graph embedding and DPP MAP inference for diversity-aware recommendation tasks.

2 DIVERSITY-AWARE RECOMMENDATION ON KNOWLEDGE GRAPH

In this section, we present our two-step framework **DivKG** to make recommendations combining knowledge graph embeddings (step 1) and determinantal point processes (DPP) prediction (step 2).

2.1 Knowledge Graph Embedding for Entity and Relation Representations

To improve the accuracy of recommendation, one can make use of additional information incorporated into collaborative filtering methods. Recently, using knowledge graphs to model this kind of data has been shown to enhance the recommendation [17, 18].

In this paper, we argue that knowledge graph is a robust and meaningful model that helps to blend multiple relations in one data structure. However, different from [17], where solely items are taken as vertices of the knowledge graph, we propose to consider all entities, including users, items and other additional entities (*e.g.* genre, actor, etc.). Moreover, we consider user-item interaction used traditionally for CF-based methods just as a specific relation on the knowledge graph. More formally, we represent every relation instance as a triplet (h, r, t) having semantic interpretation, where h and t denote the head and tail entities linked by one relation r.

To apply embedding on such a knowledge graph, we represent each entity and each relation as a vector, *i.e.* h, r, t are represented as $\mathbf{v_h}$, $\mathbf{v_r}$,$\mathbf{v_t}$, respectively. We use translation-based embedding methods [1?] to interpret the translation semantics among vectors $\mathbf{v_h}, \mathbf{v_r}, \mathbf{v_t}$ which is $translation(\mathbf{v_h}, \mathbf{v_r}) \approx \mathbf{v_t}$. We use a margin-based loss function with margin γ to optimise the vector representation:

$$Loss_{KGE} = \sum_{(h,r,t)} \sum_{(h',r',t')} [f_r(h,t) + \gamma - f_{r'}(h',t')]_+$$

The corrupted triplets (h', r', t') are derived from golden triplets (h, r, t) by (1) keeping the relation unvaried, *i.e.* $r = r'$, and (2) by either keeping unvaried the head entity and randomly selecting the tail entity, *i.e.* $h' = h$, $t' \neq t$, or keeping unvaried the tail entity and randomly selecting the head entity, *i.e.* $t' = t$, $h \neq h'$. And f_r denotes the translation function: TransE [1] takes $f_r(h,t) = \|\mathbf{v_h} + \mathbf{v_r} - \mathbf{v_t}\|_2$, TransH [?] takes $f_r(h,t) = \|(\mathbf{v_h} - \mathbf{w_r}^\top \mathbf{v_h} \mathbf{w_r}) + \mathbf{v_r} - (\mathbf{v_t} - \mathbf{w_r}^\top \mathbf{v_t} \mathbf{w_r})\|_2$, where $\mathbf{w_r}$ is a projection vector.

2.2 Diversity-Aware Recommendation on Determinantal Point Processes

Diversity is considered to be an important factor to improve user satisfaction, and thus, ameliorate overall performance of recommendation. Recently, it has attracted attention of recommender systems community, and various techniques have been proposed. We refer the reader to [16] for a brief survey on diversified recommendation. Here, we propose to exploit determinantal point processes (DPP) models to improve feature representation-based diversity, where feature representations are generated on the previous step, *i.e.* knowledge graph embedding (KGE). Note, that our framework is modular, and allows other quality estimation techniques to be used as input to the current step. However, we argue that the combination of KGE and DPP is the most efficient in terms of diversity-accuracy trade-off.

2.2.1 Construction of DPP Kernel Matrix. DPPs are a group of probability models to reflect the distribution of items from item list X over the set Y, $Y \subseteq 2^X$, where the selection of a subset $S, S \in Y$ of items is proportionate to the indexed determinant of the kernel matrix of DPP [9]. The kernel matrix of DPP is a positive semi-definite matrix which records the inherent affinity of each item appeared in the set Y and the similarities of every two different items. More specifically the diagonal elements of the kernel matrix reflect the inherent affinity of each item and the non-diagonal elements reflect the pairwise similarity of the item set.

In order to construct a kernel matrix $\mathbf{L_u}$ for each user u for top-N diversified recommendation, we define two auxiliary matrices. The first is user's u affinity profile w.r.t. candidate items defined as a diagonal matrix $\mathbf{A}_u = \text{diag}(a_1, ..., a_m)$, where m is the number of candidate items, $a_i = \frac{e^{-(fr(u,i)-\delta)}}{\sum_{j \in X, j \neq i} e^{-(fr(u,j)-\delta)}}$, where $f_r(u, i)$ is the result of the item quality estimation function calculated in the previous step and δ is the average of $f_r(u, i)$ for u. We consider here the embedding translation function for user u, item i and translation type r. Here, we use the softmin function to normalize the affinities of all items for each user. The second matrix reflects item pairwise similarity and is defined as $\mathbf{D}_u = [d_{ij}]^{m \times m}$, whose entries $d_{ij} = \frac{e^{-fr_0(i,j)}}{\sum_{k \in X, k \neq i} e^{-fr_0(i,k)}}$, and $d_{ii} = 0$, where $f_{r_0}(i, j)$ is the result of the embeddings of items i and j and relation r_0, whose vector $v_{r_0} = \overrightarrow{0}$ if items i and j share the same relation value (category).

Finally, the kernel matrix $\mathbf{L_u}$ for user u can be defined as: $\mathbf{L_u} = \alpha\mathbf{A}_u + \mathbf{D}_u$, where α is a parameter to adjust the trade-off between user's affinities and similarities among the items, or in other words, between accuracy and diversity.

2.2.2 MAP Inference for Prediction. After the construction of the kernel matrix for each user u, we aim at selecting a list S of size N of items from total candidate items, s.t. $S_{map} = \underset{S \in Y, |S|=N}{\text{argmax}} \; log\text{det}(\mathbf{L}_S)$ where $\mathbf{L}_S$ is the kernel matrix $\mathbf{L}_u$ indexed by items from S. We recall that the probability of selecting a subset S is proportionate to the determinant of the indexed kernel matrix and DPP promotes a diversified selection of items under its property by definition. Thus, the selected items with the maximum *log* determinant value theoretically determine the best related and diverse top-N items for user u. However, such an optimization problem has been proven to be NP-hard, thus we use the fast greedy algorithm proposed by [4] to retrieve an approximate top-N list as the result. We refer to this DPP model with MAP inference as **FastDPP**.

3 EVALUATION

In this section, we describe the used evaluation procedure to assess our framework and report the obtained results.

3.1 Experimental Settings

3.1.1 Datasets. For the evaluation purpose, we construct multi-relation dataset combining two real-world datasets. Our first dataset is MovieLens-100K (denoted ML-100K) containing 100,000 user ratings ranging from score 1 to 5 from 943 users on 1,682 movies. Each user has rated at least 20 movies. However, the rating matrix of ML-100K is still highly sparse, with a sparsity of 93.70%. We follow traditional idea to binarize explicit rating data by keeping the ratings of four or higher and interpret them as implicit feedback.

Our second dataset is IMDb Dataset which is currently released on IMDb website [1]. It contains information including crews, principals, different releasing versions of more than 947K films. We extract movie genre, director, actor, actress, composer etc. in total 13 categories of information from IMDb to combine with MovieLens dataset for constructing multi-relation datasets, using the extracted categories to determine the relations within our knowledge graph.

3.1.2 Evaluation Protocol. We evaluate our framework regarding the accuracy and diversity of the returned results. The evaluation is performed in two steps. We first assess the quality of the embedding part, and then the results of DPP part.

For measuring accuracy of the recommendation, we use two traditional metrics of information retrieval: (1) normalized discounted cumulative gain *NDCG@N* and (2) hit ratio *hit@N*. We calculate both metrics *hit@N* and *NDCG@N* for each test user and report the average score.

For assessing diversity of the recommendation, we use two list-wise metrics used by [4], where S_{ij} denotes the similarity between i and j: (1) $ILAD = \text{mean}_{u \in U} \text{mean}_{i,j \in R_u, i \neq j}(1 - S_{i,j})$, and (2) $ILMD = \text{mean}_{u \in U} \min_{i,j \in R_u, i \neq j}(1 - S_{i,j})$. We calculate *ILAD* and *ILMD* on for each result list and report the average score.

We apply the following evaluation procedure. To evaluate the accuracy performance of recommendation, we adopt the *leave-one-out* strategy which is widely used in literature [17] in both knowledge graph embedding and DPP processes. Thus, for knowledge graph embedding part (see Section 2.1), for each user, we randomly select one user-item interaction (rating) to constitute our *test set*, and then we randomly split the remaining interactions to *training set*, and *validation set* with ratio 80 : 20, respectively.

And for the diversification part using DPP (see Section 2.2), we randomly hold one user-item rating and mix with other 50 most similar items calculated from knowledge graph embedding results.

3.1.3 Compared Methods. To compare the performance of the first step of our framework, we use the following baseline algorithms:

BPRMF [12]. This is a matrix factorization method optimised by a pairwise ranking loss to learn from implicit feedback. It does not use relational data for learning process.

RCF [17]. This is a knowledge graph based method that considers both user-item interactions and other types of item relations and proposes a double-layer neural model for learning-to-rank top-N recommendation.

FISM [8]. This is an item-based collaborative filtering method that uses latent factor matrices to capture the relations between items. We adopt the implementation provided by [17].

For the sake of a fair comparison, we combine the aforementioned models with two diversification models, namely our FastDPP (see Section 2.2)and the well-acknowledged method MMR to compare recommendation performance both on accuracy and diversity.

Maximal Marginal Relevance (MMR) [3]. This is a re-ranking criterion to reduce redundancy while maintaining document relevance in the field of text summarization. Specifically, MMR iteratively chooses an item satisfying the following requirement: $\omega_i^* = \underset{\omega_i \subseteq X \setminus S}{\text{argmax}}[\lambda r_{\omega_i} - (1 - \lambda) \max_{\omega_j \in S} sim(\omega_i, \omega_j)]$ where r_{ω_i} is the estimated rating of item ω_i, S is the subset of already selected items, λ is the parameter to adjust the trade-off between relevance and diversity and $sim()$ is the similarity function between two items.

[1] IMDb datasets link: https://datasets.imdbws.com/

Table 1: Accuracy results before diversification with dimension=75, learning rate=0.001.

	Hit			NDCG		
Metric	@5	@10	@20	@5	@10	@20
BPRMF	0.1394	0.2200	0.3240	0.0888	0.1150	0.1412
FISM	0.1182	0.2041	0.3160	0.0746	0.1023	0.1304
RCF	0.1442	0.2179	0.3261	0.0888	0.1123	0.1393
TransE	0.1879	0.2842	0.4087	0.1253	**0.1561**	0.1876
TransH	**0.1917**	**0.2861**	**0.4123**	**0.1257**	**0.1561**	**0.1878**

Table 2: Diversified recommendation results with α=0.9

Metric	Hit@10	NDCG@10	*ILAD*	*ILMD*
BPRMF+MMR	0.2821	0.0789	0.9683	0.8867
BPRMF+FastDPP	0.3065	0.0848	0.9922	0.9726
RCF+MMR	0.2842	0.0785	0.9698	0.8886
RCF+FastDPP	0.3001	0.0798	0.9923	0.9729
TransE+MMR	0.2768	0.0774	0.9911	0.9183
DivKG$_E$	0.3160	0.1176	**0.9959**	**0.9768**
TransH+MMR	0.2693	0.0705	0.9898	0.8951
DivKG$_H$	**0.3175**	**0.1178**	0.9956	0.9690

3.2 Experimental Results

Graph Embeddings. Table 1 shows a general accuracy improvement using translation-based knowledge graph embedding regarding user-item interactions as one kind of relations. The two translation-based embedding methods we use here, TransE and TransH, outperform not only classic BPR-based CF method (BPRMF) and item-based matrix factorization method (FISM), but also outperform the start-of-art relation collaborative filtering method RCF which also takes relation information. We attribute this lower performance of RCF to both the separation of user-item interactions and other relations in the knowledge graphs and its fixed types of relations encoded on the knowledge graph. Moreover, TransH generally outperforms TransE due to the projection of entity vectors to a relation-specific hyperplane which enhances the accuracy.

Diversified Recommendation. Table 2 shows diversified recommendation performance w.r.t.accuracy and diversity.It can be seen that our methods DivKG$_E$ and DivKG$_H$ that combine FastDPP with TransE and TransH respectively outperform baselines methods w.r.t. both, accuracy and diversity. Almost all accuracy-based methods combined with FastDPP outperform those combined with MMR.

4 CONCLUSIONS

In this work, we proposed a framework to address diversity-aware top-N recommendation problem. It combines knowledge graph embedding methods (KGE) and DPP models for diversified prediction. The motivation behind using knowledge graphs lies in their ability to capture various relations between items, users, and auxiliary entities, providing a solid basis for understanding user's behaviour and improving recommendation quality. Moreover, knowledge graphs may facilitate the reasoning behind the recommendation process, making it more convincing. We leave this direction for our future work. In order to diversify the results of top-N recommendation, we further proposed to construct DPP kernels over KGE to facilitate diversified predictions. The construction of kernel provides an accuracy-diversity trade-off. Our evaluation results prove that such a combination is beneficial in terms of both accuracy and diversity.

ACKNOWLEDGMENTS

Lu Gan is supported by the China Scholarship Council scholarship for Ph.D. program at INSA Lyon, France. File No. 201801810094.

REFERENCES

[1] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13)*. 2787–2795.

[2] Allan Borodin, Aadhar Jain, Hyun Chul Lee, and Yuli Ye. 2017. Max-Sum Diversification, Monotone Submodular Functions, and Dynamic Updates. *ACM Trans. Algorithms* 13, 3 (2017), 25. https://doi.org/10.1145/3086464

[3] Jaime Carbonell and Jade Goldstein. 1998. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In *Proceedings of the 21st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98)*. 335–336.

[4] Laming Chen, Guoxin Zhang, and Hanning Zhou. 2018. Fast Greedy MAP Inference for Determinantal Point Process to Improve Recommendation Diversity. In *Proceedings of the 32Nd International Conference on Neural Information Processing Systems (NIPS'18)*. 5627–5638.

[5] Mike Gartrell, Ulrich Paquet, and Noam Koenigstein. 2016. Bayesian Low-Rank Determinantal Point Processes. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. 349–356.

[6] Xiangnan He, Hanwang Zhang, Min-Yen Kan, and Tat-Seng Chua. 2016. Fast Matrix Factorization for Online Recommendation with Implicit Feedback. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '16)*. 549–558.

[7] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge Graph Embedding via Dynamic Mapping Matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 687–696.

[8] 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.

[9] Alex Kulesza and Ben Taskar. 2011. k-DPPs: Fixed-size determinantal point processes. In *Proceedings of the 28th International Conference on Machine Learning (ICML'11)*. 1193–1200.

[10] Alex Kulesza and Ben Taskar. 2012. Learning Determinantal Point Processes. arXiv:cs.LG/1202.3738

[11] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*. 2181–2187.

[12] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI '09)*. 452–461.

[13] Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic Matrix Factorization. In *Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS'07)*. 1257–1264.

[14] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI'14)*. 1112–1119.

[15] Romain Warlop, Jérémie Mary, and Mike Gartrell. 2019. Tensorized Determinantal Point Processes for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. 1605–1615.

[16] Qiong Wu, Yong Liu, Chunyan Miao, Yin Zhao, Lu Guan, and Haihong Tang. 2019. Recent Advances in Diversified Recommendation. arXiv:cs.IR/1905.06589

[17] Xin Xin, Xiangnan He, Yongfeng Zhang, Yongdong Zhang, and Joemon Jose. 2019. Relational Collaborative Filtering: Modeling Multiple Item Relations for Recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)*. 125–134.

[18] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative Knowledge Base Embedding for Recommender Systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. 353–362.