

# Not All Links Are Created Equal: An Adaptive Embedding Approach for Social Personalized Ranking

Qing Zhang, Houfeng Wang\*

Key Laboratory of Computational Linguistics (Peking University) Ministry of Education, China  
{zqicl,wanghf}@pku.edu.cn

## ABSTRACT

With a large amount of complex network data available, most existing recommendation models consider exploiting rich user social relations for better interest targeting. In these approaches, the underlying assumption is that similar users in social networks would prefer similar items. However, in practical scenarios, social link may not be formed by common interest. For example, one general collected social network might be used for various specific recommendation scenarios. The problem of noisy social relations without interest relevance will arise to hurt the performance. Moreover, the sparsity problem of social network makes it much more challenging, due to the two-fold problem needed to be solved simultaneously, for effectively incorporating social information to benefit recommendation. To address this challenge, we propose an adaptive embedding approach to solve the both jointly for better recommendation in real world setting. Experiments conducted on real world datasets show that our approach outperforms current methods.

## 1. INTRODUCTION

With a large amount of complex network data available, how to effectively incorporate rich social information for recommendation is an important research topic [1]. Collaborative Filtering (CF) with implicit feedback [7] which is also called personalized ranking [7] is the consideration in this paper, since in most real applications, the collected data of user behaviors, e.g., assigning tags, purchasing products and watching videos, are usually implicit [5] without explicit feedback. To incorporate rich social information, existing approaches utilize social connections in different ways, but all of those hold the same underlying assumption [1] that given social networks, *similar users would prefer similar items*. In this paper, we argue that this assumption is too strong, and even not reasonable in many real applications, because social link may not be formed by common interest. For example, one general collected social network might be used for various specific recommendation scenarios. To investigate the issue, we first show the following two problems observed from the widely used real-world datasets.

\*Corresponding author.

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.

SIGIR '16, July 17-21, 2016, Pisa, Italy

© 2016 ACM. ISBN 978-1-4503-4069-4/16/07...\$15.00

DOI: <http://dx.doi.org/10.1145/2911451.2914740>

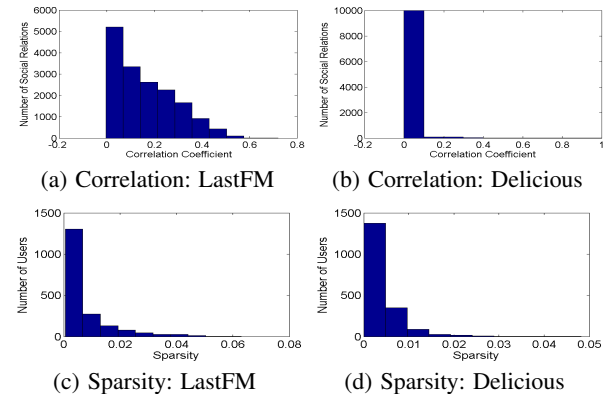


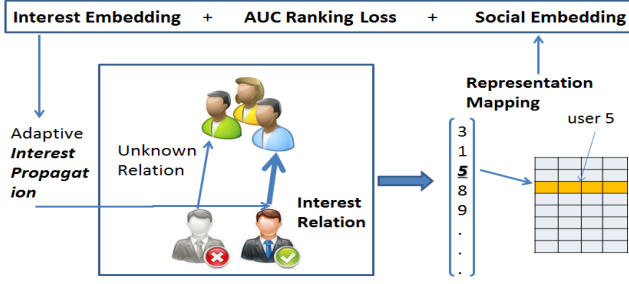
Figure 1: Subfigure(a,b): The distribution of interest correlation for links. Subfigure(c,d): The distribution of link sparsity.

**Problem 1: Noisy Social Link.** From Figure (1.a,1.b), we can see that the interest correlation, calculated by Pearson's correlation coefficient, is not high enough to support the assumption in the conventional social regularization based approaches, that the similar user in social networks leads to similar interest. In fact, this is the fundamental reason why social regularization based approach may lead to inconsistent results for different datasets, due to the noisy social patterns, violating the model assumption.

**Problem 2: Sparse Social Link.** From Figure (1.c,1.d), we can see that the sparsity of the social relation is quite low. A large number of users have only a small fraction of social connections in the entire network. Thus, directly using the social information is hard to achieve the best performance, and even may hurt the performance, due to the underlying noisy links in real applications.

Intuitively, to solve the problem 1, we would like to eliminate the noisy social links, but this will cause the sparse social link problem. Likewise, to solve the problem 2, we would like to create more links, which may also incorporate extra noisy social links causing the problem 1, even though the underlying noisy patterns that may have existed in the sparse links as shown in Figure 1, are not considered. Thus, we are in a dilemma, while only considering one specific problem of the two. In fact, the two tasks are mutually beneficial. If we can incorporate interest information to control social link selection, the noisy link problem could be alleviated, and if we can utilize more clean and higher order social relations, the sparsity problem could be alleviated for benefiting interest targeting. Thus, joint considering the two problems is an intuitive solution.

However, in previous widely adopted Matrix Factorization (MF) framework, the matrix factorizing process is a global process with



**Figure 2: Model Framework. Joint learning interest and social embeddings with adaptive interest propagation.**

low-rank constraint. It is hard to incorporate local interest information to control the global factorization process. To solve this challenge, inspired by the recent revealed equivalency [2] between word embedding (word2vec) [3] and matrix factorization, we propose a novel solution from an embedding perspective instead of the original matrix factorization view, as illustrated in Fig.(2), which allows a more flexible way to model user interest with complex social interactions.

## 2. A UNIFIED EMBEDDING APPROACH

The goal of the joint embedding is to learn user interest embedding  $e_u^i$ , user social embedding  $e_u^s$  and item embedding  $e_i^r$ , by maximizing a posterior of AUC ranking based criterion [7] for interest embedding, and maximizing a criterion for social embedding,

$$\begin{aligned} \mathcal{L} &= \ln p(\Theta | >_u, G) \propto \ln p(>_u, G | \Theta) \cdot p(\Theta), \\ &= \ln \left( \underbrace{p(>_u | \Theta)}_{\text{Interest Embedding}} \cdot \underbrace{p(G | \Theta)}_{\text{Social Embedding}} \cdot \underbrace{p(\Theta)}_{\text{Prior}} \right) \end{aligned} \quad (1)$$

where  $\Theta$  the model parameters,  $G$  the user social information, and  $>_u$  the ranking-based user interest information.

### 2.1 Interest Embedding

**BPR-opt Criterion.** For personalized ranking, a widely employed optimization criterion is Bayesian Personalized Ranking criterion denoted as BPR-opt as shown in Eq.(2). It has a strong theoretical underpinning as a surrogate objective, for optimizing the ranking-based AUC loss [7].

$$\begin{aligned} \text{BPR-opt} &:= \ln p(\Theta | >_u) = \ln p(>_u | \Theta) p(\Theta), \\ &= \ln \prod_{(u,i,j)} \sigma(\hat{x}_{uij}) p(\Theta) \end{aligned} \quad (2)$$

where  $\sigma$  is a sigmoid function  $\sigma(x) = \frac{1}{1+\exp(-x)}$ ,  $p(\Theta)$  is a normal distribution  $N(0, \Sigma_\Theta)$  with zero mean and a diagonal covariance matrix  $\Sigma_\Theta = (1/\sigma)I$ ,  $I$  identity matrix,  $x_{uij}$  suggests that user  $u$  is assumed to prefer item  $i$  over item  $j$ . In the following, we show how to formulate the ranking score  $\hat{x}_{uij}$  as a function of the model parameters  $\Theta$ . In Eq.(2), we define the ranking score as

$$\hat{x}_{uij} = e_u e_i^T - e_u e_j^T \quad (3)$$

where  $e_u$  is the user  $u$  embedding;  $e_i^r$  and  $e_j^r$  are the item  $i$  and item  $j$  embeddings. All these embeddings are the model parameters which can be learned by our model.

**Interest Embedding.** We model item  $i$  embedding  $e_i^r$  in Eq.(3) as a vector representation with real values. We model user  $u$  embedding  $e_u$  in Eq.(3) as a linear combination of  $e_u^i$  and  $e_u^s$ ,  $\lambda_s \in [0, 1]$ ,

$$e_u = (1 - \lambda_s) e_u^i + \lambda_s e_u^s \quad (4)$$

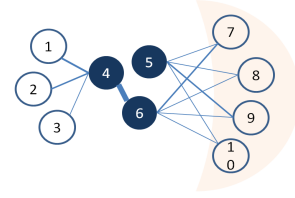
where  $e_u^i$  is the user  $i$  interest-specific embedding, and  $e_u^s$  is the corresponding social-specific embedding as a structured bias, which will be discussed in the next section.

### 2.2 Social Embedding

The goal of social embedding is to maximize the likelihood of a specific sequence of vertexes appearing in a user neighbour context, with the fundamental assumption that *similar context leads to similar representation*. To achieve the goal, formally, given a sequence of vertexes  $V^n = (v_0, v_1, \dots, v_n)$ , where  $v \in V$  the vertex set in a social network, in general, we would like to maximize the Eq.(5) over all the training corpus with the window size  $w$ ,

$$\sum_i \log P(e_i^s | e_{i-w}^s, \dots, e_{i+w}^s \setminus e_i^s) \quad (5)$$

where  $\setminus e_i^s$  excluding  $e_i^s$ . For our  $p(G | \Theta)$ , we present a novel so-



**Figure 3: A toy example of social network.**

lution as a variant of Deep-walk (Perozzi et al. 2014) [6] and LINE (Tang et al. 2015) [9], the state-of-the-art network embedding models, to incorporate both first-order and second-order proximities simultaneously in a unified framework with interest propagation.

**Incorporating First-order Proximity.** Many existing graph embedding algorithms, have the objective to preserve the first-order proximity [9] directly. However, in real world, the links observed are only a small proportion, and even with much noise. To solve the problem, we do not directly enforce the pairwise constraint on the two connected users. Instead, we add a virtual common node, representing the neighbourhood relation, for each user and its adjacent neighbours. The effectiveness of this modeling approach with the similar idea has been proved in NLP community [8].

**DEFINITION 1 (LOCAL: FIRST-ORDER PROXIMITY).** *The first-order proximity in a network is the local pairwise proximity between two vertices, such as nodes 4,6 in Fig.(3). For each pair of vertices linked by an edge  $(u, v)$ , the weight on that edge,  $w_{uv}$ , indicates the first-order proximity between  $u$  and  $v$ . If no edge is observed between  $u$  and  $v$ , their first-order proximity is 0.*

The objective function of the first-order modeling is the log likelihood of all users,

$$l_1 = \sum_{n=1}^N \sum_{u_i \in A_n} \log(P(e_{u_i}^s | e_{A_n}^s)) \quad (6)$$

where  $A_n$  denotes the neighbour set of user  $n$  as user context;  $e_{u_i}^s$  denote the user  $i$  embedding in context  $A_n$ , and  $e_{A_n}^s$  denotes the user context embedding  $e_{A_n}^s$ . We use a softmax function to define the probability,  $e_U^s$  denotes the entire set of user social embeddings,

$$P(e_{u_i}^s | e_{A_n}^s) = \frac{\exp(e_{u_i}^s \cdot e_{A_n}^s)}{\sum_{e_u^s \in e_U^s \setminus A_n + u_i} \exp(e_u^s \cdot e_{A_n}^s)}. \quad (7)$$

**Incorporating Second-order Proximity.** Following previous study [6], we assume nodes occurring in similar contexts tend to have similar meanings. This assumption leads to a more robust

way of measuring the similarity between two users, compared with the direct pair-wise link assumption, which may cause overfitting due to the link sparsity and noise problems.

**DEFINITION 2 (GLOBAL: SECOND-ORDER PROXIMITY).**

The second-order proximity between two vertices  $(u, v)$  in a network is the similarity between their neighborhood network structures, such as nodes 5,6 in Fig.(3). Let  $p_u = (w_{u,1}, \dots, w_{u,|V|})$  denote the first-order proximity of  $u$  with all other vertices, then the second-order proximity between  $u$  and  $v$  is determined by the similarity between  $p_u$  and  $p_v$ . If no vertex is linked from/to both  $u$  and  $v$ , the second-order proximity between  $u$  and  $v$  is 0.

The objective function of the second-order modeling is as follows,

$$l_2 = \sum_{i=1}^N \log(P(e_{u_i}^s | e_{c_i}^s)) \quad (8)$$

where  $P(e_{u_i}^s | e_{c_i}^s)$  is similarly defined by using a softmax function as in the first-order. We define the user 2-order context embedding,

$$e_{c_i}^s = f(c_{u_{i_1} \in A(u_i)}, \dots, c_{u_{i_k} \in A(u_i)}), \quad (9)$$

where  $f()$  can be sum, average, concatenate or max pooling of context user vectors  $c_{u_{i_k}}$  in set  $A(u_i)$ . In this paper, we use **weighted average** on user social neighbour set  $A(u_i)$  (SCHEME-1), which is different from that of [3] and [6]. The weights here are learnt from data automatically for adaptive interest propagation. We define the weight  $p(u_{i_k} \in A(u_i) | u_i)$  for  $e_{u_{i_k}}$  in Eq.(10), a random-walk like transition probability for each user to propagate interest,

$$p(u_{i_k} \in A(u_i) | u_i) = \frac{1}{Z} \cdot \frac{e_{i_k} e_i^T}{\|e_{i_k}\| \|e_i\|}, \quad (10)$$

where  $e_u, e_j$  are defined in Eq.(4);  $p(u_{i_k} \in A(u_i) | u_i)$  is calculated by the percentage of weights, i.e., the interest embedding based cosine similarity by this edge against the total edge weight  $Z$  of a user. Note that in the above objective, the interest strength is naturally incorporated for different users by Eq.(10) adaptively.

Moreover, we also introduce **an alternative scheme** to generate  $e_{c_i}^s$  (SCHEME-2), which is similar to DeepWalk. The difference is that we adaptively generate the *higher order* user specific context set  $A'(u_i)$  by an interest-aware random walk process [6] using Eq.(10). Then we use an average  $f()$  for those generated vectors in  $A'(u_i)$ , instead of using the social neighbour set  $A(u_i)$  for Eq.(9).

**Combination: First-order and Second-order Proximities.** To embed the networks by preserving both the first-order and second-order proximities with interest propagation, we linearly combine the two objectives for joint learning as  $\ln(p(G|\Theta))$  in Eq.(1),

$$l = \alpha l_1 + (1 - \alpha) l_2 \quad (11)$$

where  $\alpha \in [0, 1]$  is the strength weight;  $l_1$  and  $l_2$  are the first-order objective and second-order objective respectively. The social regularization parameter  $\lambda_l$  can be added  $\lambda_l * l$  for controlling the overall social network strength, as in the traditional MF approaches.

## 2.3 Parameter Learning

We employ the widely used Stochastic Gradient Descent (SGD) algorithm to optimize the objective function in Eq.(1). The main process of SGD is to randomly select a pair of positive and negative feedbacks, and iteratively update model parameters. Specifically, for each training instance (user, positive-item, negative-item), we calculate the derivative and update the corresponding parameters by walking along the ascending gradient direction,

$$\Theta^{t+1} = \Theta^t + \eta \frac{\partial \mathcal{L}}{\partial \Theta}. \quad (12)$$

where the model parameters  $\Theta$  are user interest and user social embeddings  $e_u^i, e_u^s$ ; item embedding  $e_i^i$  and user context embeddings  $e_{A_n}^s, e_{u_{i_k}}^s$ . In the above process, to optimize Eq.(11), we adopt the negative sampling technique (Mikolov et al., 2013) [3] for efficient learning since it is generally intractable for direct optimization.

## 3. RELATED WORK

Social recommendation methods focus on using the social network to improve the recommendation performance. Most existing methods are based on Matrix Factorization (MF) with regularization approaches. The social information is typically used as two ways. The first way is to directly use the **first-order proximity** to construct Laplacian matrix for regularizing the latent user. The limitation is that it only works well if the social is dense and noiseless. To overcome the limitation, recent approaches [1] employ the twice matrix factorization approach, i.e., not only factorizing the original rating matrix but also the social adjacent matrix, and then employ the latent social representation as regularization for users. This approach can be seen as a way of utilizing the **second-order proximity** due to the equivalency between the word2vec objective [3] and matrix factorization as shown in [2]. While the sparse social link problem may be solved, the noisy link problem is seldom addressed in previous work successfully. The most similar work to this paper is [10], which proposed a model with joint friendship and interest propagation in social network. However, it mainly focus on modeling the explicit feedback using regression-based modeling approach, and did not consider social representation modeling with ranking-based optimization criterion. Different from the above approaches, this work can incorporate *both first-order and second-order proximities* simultaneously in a unified embedding framework jointly with *local interest propagation*.

## 4. EXPERIMENTS

### 4.1 Experiment Setup

**Dataset and Evaluation Metrics.** We evaluate our method on two public real-world datasets<sup>1</sup>, LastFM and Delicious as shown in Table 1. We use three popular metrics [5], Recall@K (R@K), NDCG@K (Normalized Discounted Cumulative Gain@K) and Area Under the Curve (AUC), to measure the recommendation quality of our proposed approach in comparison to baseline methods.

$Recall@k = \frac{\#relevant \text{ in Top}K}{\#total \text{ relevant}}$ ,  $NDCG@k = \frac{\sum_{i=1}^k \frac{2^{rel_i} - 1}{\log_2(1+i)}}{IDCG}$ , where  $rel_i$  denotes the relevant degree which is binary in our task, i.e., 1 denotes it is relevant, 0 denotes it is not relevant and IDCG is the optimal score computed using the same form in numerator but with optimal ranking result known in advance.  $AUC = \frac{1}{|U|} \sum_u \frac{1}{|E(u)|} \sum_{(i,j) \in E(u)} \delta(\hat{x}_{ui} > \hat{x}_{uj})$ , where  $U$  user set; the evaluation pairs per user  $u$  are  $E(u) := \{(u, i) \in S_{test} \wedge (u, j) \notin (S_{test} \cup S_{train})\}$ ,  $S$  denotes observed Set.

Dataset	#User	#Item	Sparsity	#Ratings	#Avg.R
LastFM	1892	18745	0.28%	92834	49
Delicious	1867	69223	0.08%	104799	56

**Table 1: Datasets Statistics. R denotes the #rated per user. The statistics of social relations can be found in Fig.(1).**

**Settings.** We conducted experiments on one validation set and one test set, by randomly drawing new train/test (90%/10%) splits of the observed data, for users with more than 20 items. Our

<sup>1</sup>Data available at <http://grouplens.org/datasets/hetrec-2011/>

Dataset	Metric	WRMF	BPR-MF	GBPR	GBPR-1	MR-BPR	EBPR-U	EBPR
LastFM $\lambda_s = 0.2, \alpha = 0.7$	R@30	0.2221	0.2383	0.2418	0.2497	0.2501	0.2545	<b>0.2636</b>
	NDCG@30	0.1312	0.1472	0.1518	0.1515	0.1597	0.1613	<b>0.1694</b>
	R@100	0.4414	0.4622	0.4791	0.4742	0.5036	0.5247	<b>0.5428</b>
	NDCG@100	0.2008	0.2119	0.2292	0.2286	0.2353	0.2416	<b>0.2494</b>
	AUC	0.9772	0.9848	0.9866	0.9855	0.9886	0.9902	<b>0.9924</b>
Delicious $\lambda_s = 0.5, \alpha = 0.4$	R@30	0.1533	0.1622	0.1815	0.1785	0.1933	0.2047	<b>0.2280</b>
	NDCG@30	0.1031	0.1119	0.1342	0.1304	0.1419	0.1531	<b>0.1814</b>
	R@100	0.2215	0.2303	0.2441	0.2432	0.2520	0.2597	<b>0.2789</b>
	NDCG@100	0.1212	0.1329	0.1519	0.1441	0.1604	0.1701	<b>0.1974</b>
	AUC	0.6813	0.6962	0.7091	0.7041	0.7101	0.7106	<b>0.7143</b>

**Table 2: Model Comparison.** Our best results are achieved while adopting the Scheme-2 on LastFM dataset, and the Scheme-1 on Delicious dataset respectively, for modeling  $e_{c_i}^s$  as shown in Eq.(9). For simplicity, we do not consider item bias fairly for all methods.

datasets are more challenging than [5], since we have a much larger number of item candidates for each user while predicting. The hyperparameters for baseline methods are optimized via grid search following the original works in the validation set, and afterwards are kept constant in the test set. For the optimization, we iteratively update the parameters until the likelihood does not increase by 0.01 or the maximum iteration limit is reached, i.e., 250 on LastFM; 400 on Delicious. In each iteration, we randomly select instances  $2*|U|$  for training. The latent dimension is fixed as 100. For our method, the parameters are randomly initialized within  $[0, 1]$ . The learning rate  $\eta$  for SGD is set to 0.01. The optimal parameters  $\sigma$  in  $p(\Theta)$  for user embedding  $e_u$  and item embedding  $e_i$  are 0.1, 0.1 on LastFM, and 0.01, 0.1 on Delicious respectively. The best social strength parameters  $\lambda_l$  are 1 on LastFM, and 0.001 on Delicious respectively.

**Baselines.** We compare the following popular and the state-of-the-art social regularization based models for personalized ranking. **WRMF** [4] is a weighted matrix factorization method, which uses a point-wise strategy for one-class recommendation. **BPR-MF** [7] is a well known BPR method with pairwise assumption for item ranking. **GBPR** [5] is a state-of-the-art BPR method which relaxes BPR’s assumption to a group pairwise preference assumption. Here we also introduce additional first-order social connection to construct groups, denoted as **GBPR-1**. **MR-BPR** [1] is a state-of-the-art social regularization based BPR method, as a second-order regularization approach which combines BPR-MF with social matrix factorization. **EBPR** is our proposed method. In addition, we also test a variant, **EBPR-U**. **EBPR-U** denotes using a uniform distribution instead of the weighting in Eq.(10).

## 4.2 Result and Analysis

From Table 2, we can see that the proposed method consistently performs better than other baselines in different datasets. It shows the increased recommendation performance in the presence of link sparsity and noise, getting improvements of Recall, NDCG and AUC, with strong baselines. It could be explained that the adaptive interest propagation for social link modeling is important for achieving the good results in real applications. In addition, we show that an appropriate linear combination in Eq.(4) with  $\lambda_s$  can benefit the performance for different datasets while using different second order proximities. In fact, different social networks may have different properties. Empirically, in the process of finding the optimal  $\alpha$  in Eq.(11), to balance the strength of different order proximities, we found that our model is more relying on the first-order proximity on LastFM, and slightly more relying the second-order proximity on Delicious, but adding the complementary one is also a necessary which can improve the performance. In contrast, purely using one specific of the two did not achieve the best results, which verifies our joint proximity modeling is a necessity.

## 5. CONCLUSION

In this paper, we studied the social personalized ranking problem. To relax the conventional assumption, in practical scenarios we argue that not all links are created equal, and we consider to do automatic social link selection jointly with learning user interest representation for social recommendation. The challenge is that solving the noisy social link problem solely may aggravate the sparsity problem and vice versa. To tackle this challenge, we developed a unified embedding framework to present a new viewpoint of social network regularization, which provides an avenue to allow a more flexible way of incorporating user social information with complex interactions. Experiment results demonstrate the superiority of the proposed method for real world applications.

## 6. ACKNOWLEDGMENTS

Our work is supported by National High Technology Research and Development Program of China(863 Program)(No.2015AA015402), National Natural Science Foundation of China(No.61370117) and Major National Social Science Fund of China(No.12&ZD227).

## 7. REFERENCES

- [1] A. Krohn-Grimberghe, L. Drumond, C. Freudenthaler, and L. Schmidt-Thieme. Multi-relational matrix factorization using bayesian personalized ranking for social network data. In *Proceedings of WSDM*, pages 173–182, 2012.
- [2] Y. Li, L. Xu, F. Tian, L. Jiang, X. Zhong, and E. Chen. Word embedding revisited: A new representation learning and explicit matrix factorization perspective. In *Proceedings of IJCAI*, pages 3650–3656, 2015.
- [3] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.
- [4] R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. M. Lukose, M. Scholz, and Q. Yang. One-class collaborative filtering. In *Proceedings of ICDM*, pages 502–511, 2008.
- [5] W. Pan and L. Chen. GBPR: group preference based bayesian personalized ranking for one-class collaborative filtering. In *Proceedings of IJCAI*, 2013.
- [6] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: online learning of social representations. In *Proceedings of KDD*, pages 701–710, 2014.
- [7] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. BPR: bayesian personalized ranking from implicit feedback. In *Proceedings of UAI*, pages 452–461, 2009.
- [8] F. Sun, J. Guo, Y. Lan, J. Xu, and X. Cheng. Learning word representations by jointly modeling syntagmatic and paradigmatic relations. In *Proceedings of ACL*, pages 136–145, 2015.
- [9] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. LINE: large-scale information network embedding. In *Proceedings of WWW*, pages 1067–1077, 2015.
- [10] S. Yang, B. Long, A. J. Smola, N. Sadagopan, Z. Zheng, and H. Zha. Like like alike: joint friendship and interest propagation in social networks. In *Proceedings of WWW*, pages 537–546, 2011.