Social Recommendation with Optimal Limited Attention

Xin Wang
xin_wang@tsinghua.edu.cn
Tsinghua University

Wenwu Zhu*
wwzhu@tsinghua.edu.cn
Tsinghua University

Chenghao Liu*
chliu@smu.edu.sg
Singapore Management University

ABSTRACT

Social recommendation has been playing an important role in suggesting items to users through utilizing information from social connections. However, most existing approaches do not consider the attention factor causing the constraint that people can only accept a limited amount of information due to the limited strength of mind, which has been discovered as an intrinsic physiological property of human by social science. We address this issue by resorting to the concept of limited attention in social science and combining it with machine learning techniques in an elegant way. When introducing the idea of limited attention into social recommendation, two challenges that fail to be solved by existing methods appear: i) how to develop a mathematical model which can *optimally* choose a subset of friends for each user such that these friends' preferences can best influence the target user, and ii) how can the model learn an *optimal* attention for each of these selected friends. To tackle these challenges, we first propose to formulate the problem of optimal limited attention in social recommendation. We then develop a novel algorithm through employing an EM-style strategy to jointly optimize users' latent preferences, optimal number of their best influential friends and the corresponding attentions. We also give a rigorous proof to guarantee the algorithm's optimality. The proposed model is capable of efficiently finding an optimal number of friends whose preferences have the best impact on target user as well as adaptively learning an optimal personalized attention towards every selected friend w.r.t. the best recommendation accuracy. Extensive experiments on real- world datasets demonstrate the superiority of our proposed model over several state-of-the-art algorithms.

CCS CONCEPTS

• **Information systems → Social recommendation**.

KEYWORDS

Recommendation, User Behavior Modeling, Limited Attention

ACM Reference Format:
Xin Wang, Wenwu Zhu, and Chenghao Liu. 2019. Social Recommendation with Optimal Limited Attention. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19), August 4–8, 2019, Anchorage, AK, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3292500.3330939

*Corresponding Authors

KDD '19, August 4–8, 2019, Anchorage, AK, USA

ACM ISBN 978-1-4503-6201-6/19/08...$15.00
https://doi.org/10.1145/3292500.3330939

1 INTRODUCTION

Being capable of efficiently filtering the exploding information on Internet, recommender systems have become an indispensable tool in recommending relevant items that may potentially be attractive to users. As a hot research topic, recommendation with no doubt has received a lot of attention from both academy and industry [1, 28]. Nevertheless, traditional recommender systems suffer from *data sparsity* which is caused by the fact that the number of items is normally very huge while users commonly consume only a very small portion of these items. In addition, traditional recommendation approaches have a deteriorative performance on new users without any historical behaviours, resulting in the *cold start* problem. This brings the idea of *social recommendation* which utilizes social information from social connections (such as friends) to mitigate the above two problems [11, 19].

Although there have been a lot of works on social recommendation, most of them ignore the attention factor which results in the constraint that only a small portion of information can be processed in real time by each individual due to her limited mind strength and brain capacity [13, 27]. Recent works [4, 8, 10, 37] have also confirmed the important role this factor plays in affecting people's behaviours and their interactions in social media. Actually people now with online social networks are easier to get connected, especially for those who are not close enough to become friends off-line, causing the fact that many of our friends on social networks may produce noisy/useless information. This being the case, bringing the concept of attention factor into social recommendation becomes very necessary. The only two works [14, 15] considering the attention factor in recommendation simply assign non-zero weights to all social connections. On the one hand, this fails to simulate the real-world scenario where people only take information from a small number of friends into consideration and ignore useless information produced by all other friends who have noisy influence on the target user. On the other hand, aggregation of preferences from all social connections is computationally expensive and time-consuming when making recommendation to users (especially for those who have a huge number of connections). Overall, none of existing works in social recommendation could handle the problem related to attention factor appropriately or efficiently.

To address the above problem in social recommendation, we borrow the idea of *limited attention*, a well-documented psychological and cognitive concept from social science that can affect user behaviours. The insights from social science and computer science inspire us to incorporate the notion of limited attention into social recommendation in a more appropriate and efficient way, considering that people may only take information from a limited number of their social connections into account. That is to say, each individual should be influenced by only a limited number of her social connections and thus her preference should also depend only on the preferences of these social connections. Besides, it is shown that social connections normally receive non-uniform distributed attentions from the target user and thus have different influence on her [7].

Therefore, two challenges exist: i) How can we develop an algorithm capable of filtering out an *optimal* group of friends for each user such that these friends' preferences can best influence the target user and ii) How to learn an *optimal* personalized attention towards each of these selected friends for every target user. Obviously none of previous works are able to resolve the first challenge. One naive way of implementing such subset selection on the set of friends may be to find the distance between users and their friends, then fill up the set of useful friends by adding friends sequentially in order beginning from the closest one first, imposing a threshold at a level where, say, the distance is equivalent to the distance to people that are not explicit friends. This seems to be quite straightforward and intuitive, but how to calculate the distance between users may have a significant influence on the selection outcome and there are too many of them (PCC, Cosine, etc.) that we can choose. More importantly, some commonly adopted distance calculation metric such as PCC even fails to optimally capture the similarities between two users (discussed in Section 3.2). As for the second challenge, existing work [20] calculates the weight for each friend through the *Pearson Correlation Coefficient* (PCC) between her and the target user, which is suboptimal because PCC is static and independent of user latent feature vectors (more details will be discussed in Section 3.2).

To handle these two challenges, we elegantly combine social science concepts with machine learning techniques and formulate the problem of *optimal limited attention* in the context of social recommendation. We then propose a novel social recommendation model capable of i) selecting an optimal number of social connections for each individual efficiently such that the preferences of these chosen friends are able to best influence the target user, and ii) learning the optimal attentions from the target user towards these chosen friends adaptively as well. To be more concrete, we first employ latent feature factors obtained through matrix factorization to express the latent user and item preferences. Then we develop a novel algorithm to simultaneously learn the optimal number of influential social connections, their corresponding optimal attentions from each target individual and other model parameters including the latent feature vector for each user and each item. The proposed algorithm has an advantage in a joint optimization of finding the 'optimal' combination of influential social connections and the corresponding attentions as well as other parameters rather than a two-stage procedure. Experiments on real-world datasets demonstrate the improvement of our proposed model against state-of-the-art approaches.

To recapitulate, the highlight of this paper is that inspired by the sociological discoveries, we develop a model which combines social science concepts and mathematical formulations in an elegant way. We address the challenges raised in social science by means of machine learning techniques in the context of social recommendation. We believe our elegant combination of machine learning with social science can help to achieve a performance boost in terms of social recommendation accuracy. The contributions of this paper are summarized as follows.

- Motivated by the challenges discovered in social science, we propose to combine machine learning techniques with social science concepts, and formulate the problem of *optimal limited attention* in social recommendation.
- We develop a novel algorithm that is able to pick up a group of social connections (friends) for each individual efficiently such that the preferences of these chosen friends can best influence the target user, and then learn the optimal attentions from the target user towards these chosen friends adaptively with respect to the best recommendation accuracy.
- We show the optimality of our proposed model in selecting the optimal number of social connections as well as adaptively learning the attentions.
- We conduct extensive experiments on real-world datasets to show that our proposed algorithm can clearly beat existing approaches in various evaluation metrics.

2 RELATED WORK

Collaborative Filtering. Being capable of predicting user preferences through uncovering complex and unexpected patterns hidden in users' past behaviours without any domain knowledge, collaborative filtering has become one of the most popular methods in recommender systems. Collaborative filtering methods generally can be classified into memory-based methods and model-based methods [31]. Memory-based approach can be further categorized as user-based and item-based approaches, according to which of similar users and similar items will be taken into account. User-based methods predict an unknown rating from a target user on a target item through calculating the weighted average of all the ratings on the target item from users similar to the target user, while item-based methods obtain the rating from a target user on a target item via computing the average ratings by the same user on items similar to the target item. In contrast to the memory-based methods, model-based approaches operate the observed ratings scores with the help of machine learning techniques to train a predefined learning model which will later be used to predict unknown ratings. Matrix factorization, as one of the most widely used model-based collaborative filtering methods, has achieved a promising success in both academia and industry [16, 33]. Among the literature of matrix factorization, Salakhutdinov and Mnih [23] propose a probabilistic version of matrix factorization whose time complexity is linear in the number of observations and is more resistant the overfitting problem. We refer readers to a general treatment [17] for detailed introduction on matrix factorization in recommendation. Recent works on collaborative filtering [23, 26, 29, 30] take notice of the fact that only a small number of factors are important (sparseness) and one user's preference vector is determined by how each factor applies to that user. Therefore, these methods focus on factorizing the user-item rating matrix with low-rank representations which will then be utilized to make further predictions.

Social Recommendation. Since users become connected through online social networks, their preferences are no longer independently and identically distributed. Therefore, correlated relationships among connected users that provide social information have drawn more and more interests from researchers [36]. For instance, Weng et al. [38] find that users with social connections are more likely to share similar interests in various topics than two randomly chosen users, and Tang et al. [32] show that users with trust relations are more likely to have similar preferences in item ratings. These phenomena can be observed in most online social networks and explained by social influence [21] as well as homophily [22] in social correlation theories. This motivates the advent of social recommendation [3, 5, 9, 12, 18, 20, 24, 25, 39–45]. Particularly, Ma et al. [19] propose a probabilistic matrix factorization model through factorizing user-item rating matrix and user-user linkage matrix simultaneously. Later another matrix factorization model aggregating a user's own rating and her friends' ratings to predict the target user's final rating on an item is introduced by them as well [18]. Jamali and Ester [12] formulate another matrix factorization model based on the assumption that users' latent feature vectors are dependent on their social ties'. There also exist several works trying to incorporate the concept of strong and weak ties into social recommendation. Wang et al. [35] integrate the concepts of strong and weak ties documented in social science into social recommendation through presenting a more fine-grained categorization of user-item feedback for Bayesian Personalized Ranking by leveraging the knowledge of tie strength and tie types. They later enable the learning of personalized tie type preference for each individual in probabilistic matrix factorization [34].

Limited Attention. Limited attention is a concept widely discussed in social science [13, 27] and Kang et al. [14, 15] are the first to consider it in recommendation. However, they simply assign non-zero attentions (weights) to all social connections (friends), failing to simulate the real-world scenario where people with limited attention only take information from a small number of friends into consideration and ignore useless information produced by all other friends who have noisy influence on the target user. Moreover, aggregation of preferences from all social connections is computationally expensive and time-consuming when making recommendation to users (especially those who have a huge number of social connections). Our proposed optimal limited attention social recommendation model in this paper overcomes previous works' disadvantages through finding an optimal subset of best influential friends for each individual and calculating the optimal corresponding attentions for these selected friends.

3 SOCIAL RECOMMENDATION WITH OPTIMAL LIMITED ATTENTION

In this section, we first give a problem definition on the application of matrix factorization in recommendation as prior knowledge. We then define the problem of *optimal limited attention* in which limited attention is elegantly brought to social recommendation. Finally, we detailedly explain our proposed LA-Rec model which incorporates the concept of optimal limited attention into social recommendation.

3.1 Problem Definition

In recommender systems, we are given a set of users $\mathbb{U}$ and a set of items $\mathbb{I}$, as well as a $|\mathbb{U}| \times |\mathbb{I}|$ rating matrix R whose non-empty (observed) entries R_{ij} represent the feedbacks (e.g., ratings, clicks etc.) of user $i \in \mathbb{U}$ for item $j \in \mathbb{I}$. When it comes to social recommendation, another $|\mathbb{U}| \times |\mathbb{U}|$ social tie matrix T whose non-empty entries T_{iu} denote $i \in \mathbb{U}$ and $u \in \mathbb{U}$ are ties, may also be necessary. The task is to predict the missing values in R, i.e., given a user $u \in \mathbb{U}$ and an item $p \in \mathbb{I}$ for which R_{up} is unknown, we predict the rating of u for p using observed values in R and T (if available).

A matrix factorization model assumes the rating matrix R can be approximated by a multiplication of d-rank factors,

$$R \approx U^T V, \tag{1}$$

where $U \in \mathbb{R}^{d\times|\mathbb{U}|}$ and $V \in \mathbb{R}^{d\times|\mathbb{I}|}$. Normally d is far less than both $|\mathbb{U}|$ and $|\mathbb{I}|$. Thus given a user i and an item j, the rating R_{ij} of i for j can be approximated by the dot product of user latent feature vector U_i and item latent feature V_j,

$$R_{ij} \approx U_i^T V_j, \tag{2}$$

where $U_i \in \mathbb{R}^{d\times 1}$ is the i_{th} column of U and $V_j \in \mathbb{R}^{d\times 1}$ is the j_{th} column of V. For ease of notation, we let $|\mathbb{U}| = M$ and $|\mathbb{I}| = N$ in the remaining of the paper.

3.2 Considering Optimal Limited Attention

We define the *optimal limited attention* problem in Problem 1.

Problem 1. **Optimal Limited Attention (OLA)** Given a set of users, their social linkage information, a set of items as well as a subset of user-item ratings as input in the context of social recommendation, for each user select an optimal subset of her friends such that these friends' preferences can best influence this user and learn an optimal attention for each of these selected friends.

Existing social recommendation models either simply treat different social connections equally or employ Pearson Correlation Coefficient (PCC) to calculate similarities between users. On the one hand, giving each social connection equal attention is not optimal because of the non-uniform distributions among friendships [7]. On the other hand, the calculation of PPC between two users i and f is based on those common items that these two users both rate, as is shown in (3).

$$S(i,f) = \frac{\sum\limits_{j\in I(i)\cap I(f)} (R_{ij} - \overline{R}_i)(R_{fj} - \overline{R}_f)}{\sqrt{\sum\limits_{j\in I(i)\cap I(f)} (R_{ij} - \overline{R}_i)^2}\sqrt{\sum\limits_{j\in I(i)\cap I(f)} (R_{fj} - \overline{R}_f)}}, \tag{3}$$

where $I(\cdot)$ denotes the set of items rated by the corresponding user and $\overline{R}$ denotes the average rating score of the corresponding user. We give an instance in which PCC fails to optimally capture the similarities between two users as follows. Suppose user i and u have exactly the same latent feature vector $(0, 0, 0, 0.9, 0.1, 0, 3)$, and item j and k also have the same latent feature vector $(0, 0, 0, 0.5, 4, 0, 0.2)$.

Let us consider the scenario where i only rates j and u only rates k. The PCC similarity between user i and u is apparently 0, which is reasonable according to the definition of PCC but obviously not realistic. Furthermore, we can also observe from (3) that PCC is static and independent of user latent feature vectors. Therefore, applying PCC similarity to the calculation of attentions between users will result in a suboptimal result.

As a conclusion, all the existing approaches fail to solve the *optimal limited attention* problem.

3.3 OLA-Rec: Social Recommendation with Optimal Limited Attention

To solve Problem 1, we propose a novel algorithm OLA-Rec which is capable of finding an optimal number of best influential friends and their corresponding attentions from each target user. We begin by introducing a new $d \times 1$ vector ϕ_i for each user i, such that

$$\phi_i = \sum_{u \in F(i)} \alpha_{iu} U_u, \tag{4}$$

where $F(i)$ is the set of user i's friends and α_{iu} is the attention from i to u. We further constrain $\alpha_{iu} \geq 0$ and $\sum_{u=1}^{|F(i)|} \alpha_{iu} = 1$ so that all variables are in a comparable magnitude. We denote ϕ_i as the social factor which is an aggregation of the influence of user i's friends, weighted by attentions ($\alpha_{i\cdot}$) from i. Larger α_{iu} indicates user u receives more attention from user i and has more impact on i. Similarly, smaller α_{iu} means user u receives less attention from user i and is less important in influencing i. We minimize the absolute difference between ϕ_i and $\alpha_{iu} U_u$ so that they are close to each other:

$$\min_{\boldsymbol{\alpha}_i} |\sum_{u \in F(i)} \alpha_{iu} U_u - \phi_i|. \tag{5}$$

As we discussed, the challenge in Problem 1 is to find an optimal number of best influential friends for each individual and learn the optimal attention for them with respect to the best recommendation accuracy. We start to tackle this challenge by considering the following inequality based on (5):

$$\begin{aligned} &|\sum_{u \in F(i)} \alpha_{iu} U_u - \phi_i| = |\sum_{u \in F(i)} \alpha_{iu}(U_u - \phi_u + \phi_u) - \phi_i| \\ \leq &|\sum_{u \in F(i)} \alpha_{iu}(U_u - \phi_u)| + |\sum_{u \in F(i)} \alpha_{iu}(\phi_u - \phi_i)| \\ \leq &|\sum_{u \in F(i)} \alpha_{iu}\epsilon_u| + L\sum_{u \in F(i)} \alpha_{iu} d(U_u, U_i), \end{aligned} \tag{6}$$

where L is the Lipschitz constant of ϕ_i given its Lipschitz continuity and $d(\cdot,\cdot)$ is a preset distance function (Euclidean distance in this paper). By Hoeffding's inequality, with probability at least $1-\delta$, we have :

$$|\sum_{j \in F(i)} \alpha_{ij}\epsilon_j| \leq C\|\boldsymbol{\alpha}_i\|_2, \quad s.t. \quad C = b \times \sqrt{2log(\frac{2}{\delta})}, \tag{7}$$

whose proof will be given in Appendix B. Here we assume that $|\epsilon_j| \leq b$ for some given $b > 0$ to bound ϵ. In addition, it is also assumed that $\{\epsilon_i\}_{i=1}^n$ are independent so that we are able to apply Hoeffding's inequality and bound the so-called variance. The assumption of Lipschitz continuous function, on the other hand, is required to bound the so-called bias term. Thus there comes another optimization problem from (6) such that solving it could obtain a guarantee for (5) with high probability:

$$\begin{aligned} &\min_{\boldsymbol{\alpha}_i} C\|\boldsymbol{\alpha}_i\|_2 + L\sum_{u \in F(i)} \alpha_{iu} d(U_u, U_i), \quad or \\ &\min_{\boldsymbol{\alpha}_i} C(\|\boldsymbol{\alpha}_i\|_2 + \boldsymbol{\alpha}_i^T \boldsymbol{\beta}_i), \end{aligned} \tag{8}$$

where $\boldsymbol{\beta}_i \in \mathbb{R}^{|F(i)|}$ such that:

$$\beta_{iu} = L \cdot d(U_u, U_i)/C, \tag{9}$$

and user i's friends $u \in F(i)$ are assumed to be in an ascending order with respect to $d(U_u, U_i)$. In fact, we only need to set the value of L_C ratio which equals to L/C in (9) rather than setting ϵ and b during practical implementation. We will discuss the impact of L_C ratio later in the experimental section. L_C ratio indirectly determines the attention weights $\boldsymbol{\alpha}$ and the value of optimal k^* which will in turn influence the selected subsets of social connections. Therefore, we test different settings of L_C ratio and examine the corresponding effects on the performances in later experiments.

Inspired by Anava's work [2], we come out with the following theorem and corollary:

THEOREM 3.1. *The optimal $\boldsymbol{\alpha}_i$ of (8) for each user i, denoted as $\boldsymbol{\alpha}_i^*$, can be written in the following form:*

$$\alpha_{iu}^* = \frac{(\lambda - \beta_{iu})}{\sum_u^{|F(i)|}(\lambda - \beta_{iu})}, \tag{10}$$

where we require $\beta_{iu} < \lambda$ for some $\lambda > 0$ in (10).

PROOF. Consider the alternative expression in (8), i.e., $\min_{\boldsymbol{\alpha}_i} C(\|\boldsymbol{\alpha}_i\|_2 + \boldsymbol{\alpha}_i^T \beta)$, by ignoring C and introducing the Lagrange Multipliers, we have:

$$L(\boldsymbol{\alpha}_i, \lambda, \boldsymbol{\theta}_i) = \|\boldsymbol{\alpha}_i\|_2 + \boldsymbol{\alpha}_i^T \beta + \lambda(1 - \sum_{u=1}^{|F(i)|} \alpha_{iu}) - \sum_{u=1}^{|F(i)|} \theta_{iu}\alpha_{iu}.$$

Given the convexity of (8), a global optimum is guaranteed for any solution satisfying the KKT conditions. Take the partial derivative of Lagrangian with respect to $\boldsymbol{\alpha}_i$, set it to 0:

$$\begin{aligned} \frac{\partial L}{\partial \alpha_{iu}} &= \alpha_{iu} - \|\boldsymbol{\alpha}_i\|_2 \times (\lambda - \beta_{iu} + \theta_{iu}) = 0, \\ \frac{\alpha_{iu}}{\|\boldsymbol{\alpha}_i\|_2} &= \lambda - \beta_{iu} + \theta_{iu}, \end{aligned} \tag{11}$$

where $\forall \alpha_{iu} > 0$, $\theta_{iu} = 0$ (i.e., $\beta_{iu} < \lambda$) and $\forall \alpha_{iu} = 0$, $\theta_{iu} \geq 0$ (i.e., $\beta_{iu} \geq \lambda$) by KKT conditions. Thus for any optimal attention $\alpha_{iu}^* > 0$, we have:

$$\frac{\alpha_{iu}^*}{\|\boldsymbol{\alpha}_i^*\|_2} = \lambda - \beta_{iu}. \tag{12}$$

Further combining (12) with the constraint that $\sum_{u=1}^{|F(i)|} \alpha_{iu} = 1$, any $\alpha_{iu}^* > 0$ can be calculated as follows:

$$\alpha^*_{iu} = \frac{\lambda - \beta_{iu}}{\sum_{\alpha_{iu}>0}(\lambda - \beta_{iu})}, \tag{13}$$

which completes the proof. □

A direct statement from Theorem 3.1 is as follows.

Corollary 3.2. *There exists* $1 \le k^*_i \le |F(i)|$ *whose relation to* $\boldsymbol{\alpha}^*_i$ *in Theorem 3.1 is as follows:* $\forall u > k^*_i$, $\alpha^*_{iu} = 0$ *and* $\forall u \le k^*_i$, $\alpha^*_{iu} > 0$.

Theorem 3.1 and Corollary 3.2 confirm the existence of optimal solution for Problem 1, which is that for each target user i, k^*_i is the optimal number of best influential friends needed whose attentions from i should be non-zero and whose attentions correspond to the k^*_i smallest values of $\boldsymbol{\beta}_i$. We will show how the optimal solution $\boldsymbol{\alpha}^*_i$ can be efficiently found and be incorporated in social recommendation. The following equation can be obtained by squaring and summing both sides in (12) over all non-zero $\boldsymbol{\alpha}^*_i$:

$$\sum_{\alpha^*_{iu}>0} \frac{(\alpha^*_{iu})^2}{\|\boldsymbol{\alpha}^*_i\|_2^2} \quad = \quad \sum_{\alpha^*_i>0}(\lambda - \beta_{iu})^2 = 1. \tag{14}$$

Through rewriting (14) in a quadratic form, we have (15):

$$k^*_i\lambda^2 - 2\lambda \sum_{u=1}^{k^*_i} \beta_{iu} + \Big(\sum_{u=1}^{k^*_i} \beta^2_{iu} - 1\Big) = 0. \tag{15}$$

Thus, λ for user i can be calculated in (16) through solving (15). We note that we only keep the solution satisfying $\alpha_{iu} \ge 0, \forall u \in F(i)$.

$$\lambda = \frac{1}{k^*_i}\left(\sum_{u=1}^{k^*_i}\beta_{iu} + \sqrt{\left(k^*_i + \Big(\sum_{u=1}^{k^*_i}\beta_{iu}\Big)^2 - k^*_i\sum_{u=1}^{k^*_i}\beta^2_{iu}\right)}\right). \tag{16}$$

Therefore given k^*_i, the optimal attention $\boldsymbol{\alpha}^*_i$ can be obtained through substituting the computed value of λ by (16) into (10). Algorithm 1 presents the details for finding the optimal number k^*_i and optimal attention $\boldsymbol{\alpha}^*_i$ for target user i.

Algorithm 1 Optimal Limited Attention

Require: target user i, i's social connections (friends) $u \in F(i)$, $\beta_{iu} \in \mathbb{R}$ sorted in ascending order.
Initialization: $\lambda_0 = \beta_{i,1} + 1$, $k = 0$
while $\lambda_k > \beta_{i,k+1}$ and $k \le |F(i)|$ **do**
  $k \leftarrow k + 1$

$$\lambda_k = \frac{1}{k}\left(\sum_{u=1}^{k}\beta_{iu} + \sqrt{\left(k + \Big(\sum_{u=1}^{k}\beta_{iu}\Big)^2 - k\sum_{u=1}^{k}\beta^2_{iu}\right)}\right).$$

end while
return k and $\boldsymbol{\alpha}_i$ (whose u-th element $\alpha_{iu} = \frac{\lambda_k - \beta_{iu}}{\sum_{\alpha_{iu}>0}(\lambda_k - \beta_{iu})}$);

Next we incorporate the concept of optimal limited attention into social recommendation through combining the optimal k^* and $\boldsymbol{\alpha}^*$ with matrix factorization. Generally, we estimate user i's rating on item j, R_{ij}, through the dot product of social factor ϕ_i and item j's latent feature vector V_j:

$$R_{ij} = \phi_i^T V_j. \tag{17}$$

Thus, we keep R_{ij} and $\phi_i^T V_j$ close to each other through minimizing the square loss shown in (18):

$$\min \sum_{i=1}^{M}\sum_{j=1}^{N}(R_{ij} - \phi_i^T V_j)^2. \tag{18}$$

Besides, given the additional social information for user i, we also hope that U_i is close to ϕ_i and ϕ_i in turn is close to $\sum_{u\in F(i)_{k^*}} \alpha_{iu} U_u$ as well:

$$\min \sum_{i=1}^{M}(U_i - \phi_i)^T(U_i - \phi_i). \tag{19}$$

$$\min \sum_{i=1}^{M}\Big(\phi_i - \sum_{u\in F(i)_{k^*}} \alpha^*_{iu}U_u\Big)^T\Big(\phi_i - \sum_{u\in F(i)_{k^*}} \alpha^*_{iu}U_u\Big), \tag{20}$$

where we denote $F(i)_{k^*}$ as the set of user i's k^*_i best influential friends and α^*_{iu} as the optimal attention from i to u. Putting (18) (19) and (20) together, our objective function is:

$$\begin{aligned}\mathcal{L} \quad = \quad & \min\Bigg[\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N} I^R_{ij}(R_{ij} - \phi_i^T V_j)^2 \\ + \quad & \frac{\delta_\phi}{2}\sum_{i=1}^{M}\Big(\phi_i - \sum_{u\in F(i)_{k^*}}\alpha^*_{iu}U_u\Big)^T\Big(\phi_i - \sum_{u\in F(i)_{k^*}}\alpha^*_{iu}U_u\Big) \\ + \quad & \frac{\delta_\phi}{2}\sum_{i=1}^{M}(U_i - \phi_i)^T(U_i - \phi_i) \\ + \quad & \frac{\delta_U}{2}\sum_{i=1}^{M} U_i^T U_i + \frac{\delta_V}{2}\sum_{j=1}^{N} V_j^T V_j\Bigg],\end{aligned} \tag{21}$$

where $\sum U_i^T U_i$ and $\sum V_j^T V_j$ are regularization terms preventing overfitting. I^R_{ij} is the indicator function that equals to 1 if user i has rated item j and equals to 0 otherwise. Assuming the optimal k^*_i and attention α^*_{iu} for user i are known, a local minimum of (21) can be found by taking the derivative and performing gradient descent on U_i, V_j, ϕ_i separately. The corresponding partial derivatives are shown as follows:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\phi_i} \quad &= \quad -\sum_{j=1}^{N} I^R_{ij}(R_{ij} - \phi_i^T V_j)V_j \\ &+ \quad \delta_\phi\Big(\phi_i - \sum_{u\in F(i)_{k^*}}\alpha_{iu}U_j\Big) - \delta_\phi(U_i - \phi_i), \end{aligned}\tag{22}$$

$$\frac{\partial\mathcal{L}}{\partial U_i} \quad = \quad \delta_\phi(U_i - \phi_i) + \delta_U U_i, \tag{23}$$

$$\frac{\partial\mathcal{L}}{\partial V_j} \quad = \quad -\sum_{i=1}^{M} I^R_{ij}(R_{ij} - \phi_i^T V_j)\phi_i + \delta_V V_j. \tag{24}$$

We close this section by presenting the whole picture of our proposed OLA-Rec model. We employ an *Expectation-Maximization* (EM) [6] style optimization strategy to alternatively learn the parameters $\boldsymbol{k}^*$, $\boldsymbol{\alpha}^*$, ϕ, U, V that minimize $\mathcal{L}$.

E-step. In each iteration, the optimal number $\boldsymbol{k}^*$ and optimal attention $\boldsymbol{\alpha}^*$ for each user are calculated based on the current ϕ and U through employing Algorithm 1.

M-step. Given the optimal $\boldsymbol{k}^*$ and $\boldsymbol{\alpha}^*$ obtained from E-step, $\boldsymbol{\phi}$, U, V are updated using standard gradient descent:

$$x^{(t+1)} = x^{(t)} - \eta^{(t)} \cdot \frac{\partial \mathcal{L}}{\partial x}(x^{(t)}), \tag{25}$$

where η is the learning rate and $x \in \{U, V, \phi\}$ denotes any model parameter.

Finally, the whole procedure terminates when the absolute difference between the losses in two consecutive iterations is less than 10^{-5}.

We close this section by pointing out that the concept of limited attention is a well-studied cognitive factor in social science which claims only a small portion of information can be processed in real time by each individual due to her limited mind strength. People nowadays with online social networks are much easier to get connected than before, especially for those who are not close enough to become friends off-line. This results in the problem that many of our friends on social networks may produce noisy/useless information. By elegantly introducing the concept of limited attention in social recommendation through a mathematical model with theoretical analyses, we optimally find a subset of most useful friends as well as their corresponding attention weights to solve the above problems simultaneously. In the following section, we will show that our solution is adequate to help boost recommendation performance through extensive experiments.

4 EMPIRICAL EVALUATION

In this section, we compare our proposed algorithm (OLA-Rec) with several state-of-the-art methods on four real-world datasets to demonstrate the superiority of OLA-Rec model over the others with respect to various evaluation metrics.

4.1 Experimental Setup

Evaluation Metrics. The following metrics are used to measure the recommendation accuracy.

- *Root Mean Square Error (RMSE).*
$$\text{RMSE} = \sqrt{\frac{\sum_{i,j}(R_{ij} - \hat{R}_{ij})^2}{N}}.$$
- *Mean Absolute Error (MAE).*
$$\text{MAE} = \frac{\sum_{i,j} |R_{ij} - \hat{R}_{ij}|}{N},$$
where R_{ij}, $\hat{R}_{ij}$ and N are the original rating, predictive rating and the number of ratings in test set.
- *Recall@K*. This metric quantifies the fraction of consumed items that are in the top-K ranking list sorted by their estimated rankings. For each user u we define $S(K; u)$ as the set of already-consumed items in the test set that appear in the top-K list and $S(u)$ as the set of all items consumed by this user in the test set. Then, we have
$$Recall@K(u) = \frac{|S(K; u)|}{|S(u)|}.$$
- *Precision@K*. This measures the fraction of the top-K items that are indeed consumed by the user (test set):
$$Precision@K(u) = \frac{|S(K; u)|}{K}.$$

	Douban	*CiaoDVD*	*Epinions*	*Flixster*
#users	64, 812	480	12, 319	85, 899
#items	56, 005	9, 623	119, 995	48, 602
#ratings	10, 555, 299	17, 684	459, 619	7, 680, 974
#connections	1, 397, 833	9, 451	355, 310	1, 322, 912

Table 1: Summary of datasets

Datasets. Our experiments are performed on four real-world datasets whose detailed filtering information will be presented in Appendix A. Table 1 gives a summary about their basic statistics.

Comparable Approaches. The following seven recommendation methods including our proposed OLA-Rec model are compared.

- *OLA-Rec.* The proposed OLA-Rec model.
- *TrustMF.* The method capable of handling trust propagation, originally proposed by Yang et al. [40].
- *SoReg.* The individual-based regularization model with Pearson Correlation Coefficient (PCC) proposed in [20], which outperforms its other variants.
- *SMF.* This model [12] assumes that users' latent feature factors are dependent on their ties'.
- *STE.* The model proposed by Ma et al. [18] which aggregates a user's own rating and her friends' ratings to predict the target user's final rating on an item.
- *SoRec.* The probabilistic matrix factorization model proposed by Ma et al. [19] which factorizes user-item rating matrix and user-user linkage matrix simultaneously.
- *PMF.* The classic probabilistic matrix factorization model first proposed in [23].

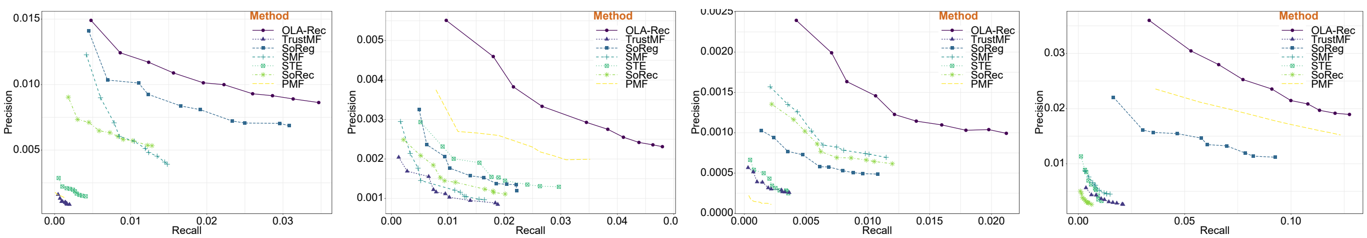
4.2 Experimental Results

In Table 2, we show the performances of the above seven comparative models on four datasets, in terms of RMSE, MAE, Precision@5 (Pre@5) and Recall@5 (Rec@5). We conduct paired difference tests for two ranking metrics, Pre@5 and Rec@5, and $*$ indicates the significance of testing results at $p < 0.05$ with degree of freedom as # users - 1 on each dataset.

RMSE and MAE. As for RMSE and MAE, we observe that social recommendation models including SoRec, STE, SMF, SoReg and TrustMF benefit from taking extra social network information into account and therefore outperform vanilla matrix factorization based collaborative filtering model such as PMF. This result confirms the assumption in social recommendation literature that social information does help boost the accuracy of traditional recommendation methods. Besides, we can also observe from Table 2 that our proposed OLA-Rec model clearly beats all other methods on all datasets for both metrics, demonstrating the advantage of joint optimization in learning users' latent preferences and finding the optimal number of their best influential social connections as well as obtaining the optimal corresponding attentions towards these chosen

		PMF	SoRec	STE	SMF	SoReg	TrustMF	OLA-Rec
Douban	RMSE	0.737926	0.719282	0.716398	0.716762	0.701026	0.720205	**0.691787**
	MAE	0.589502	0.570079	0.564863	0.573150	0.574231	0.567358	**0.553230**
	Rec@5	0.000456	0.001806	0.000633	0.004179	0.004460	0.000562	**0.004785** ∗
	Pre@5	0.001767	0.009037	0.002864	0.012273	0.014106	0.001908	**0.014901** ∗
CiaoDVD	RMSE	1.166616	1.090378	1.089358	1.105861	1.121661	1.089063	**1.037851**
	MAE	0.876568	0.831397	0.835996	0.830795	0.863786	0.830732	**0.792926**
	Rec@5	0.007885	0.002143	0.005151	0.001609	0.004908	0.001290	**0.009710** ∗
	Pre@5	0.003751	0.002488	0.002934	0.002948	0.003251	0.002043	**0.005510** ∗
Epinions	RMSE	1.184620	1.126539	1.120531	1.116998	1.121887	1.138754	**1.063084**
	MAE	0.948665	0.900854	0.883898	0.871509	0.895373	0.879354	**0.835751**
	Rec@5	0.000655	0.002217	0.000770	0.002091	0.001360	0.000601	**0.004207** ∗
	Pre@5	0.000223	0.001354	0.000664	0.001570	0.001030	0.000567	**0.002393** ∗
Flixster	RMSE	1.020013	1.009463	0.974493	0.962956	0.969857	1.002425	**0.923399**
	MAE	0.813994	0.794524	0.773342	0.768965	0.771018	0.790036	**0.739452**
	Rec@5	**0.034244**	0.000951	0.001236	0.003093	0.016525	0.003579	0.033312
	Pre@5	0.023560	0.005045	0.011332	0.008725	0.022068	0.005647	**0.035993** ∗

Table 2: Performances of comparative algorithms in terms of RMSE, MAE, Rec@5 and Pre@5 on different datasets (boldface font denotes the winner in that row).



(a) Douban (b) CiaoDVD (c) Epinions (d) Flixster

Figure 1: Precision@K vs Recall@K, where K ranges from 5 to 50

social connections. We note that due to the randomness in data splitting, model initialization and even data preprocessing, our results for some baselines may not be exactly the same as reported in the original works, though given our best efforts to diminish the variances.

Pre@5 and Rec@5. As for Pre@5 and Rec@5, we observe from Table 2 that our proposed OLA-Rec model significantly outperforms several baselines on almost every dataset, demonstrating its superiority over other state-of-the-art methods. For instance, the performance of OLA-Rec is roughly 9 times and 7 times better than PMF in terms of Rec@5 and Pre@5 on Douban, 6 times better and 3 times better than TrustMF in terms of Rec@5 and Pre@5 on Epinions.

Precision v.s. Recall. In Figure 1, we draw the Precision (Y-axis) vs. Recall (X-axis) curves of all seven recommendation methods for comparison. Data points from left to right on each line were calculated at different values of K, ranging from 5 to 50. The closer the line is to the top right corner, the better the algorithm is, indicating that both precision and recall are high. We observe from Figure 1 that OLA-Rec with no doubt achieves better performances than all other methods. Further, Figure 1 also confirms the trade-off between precision and recall —- as K increases, precision tends to go down while recall moves toward the opposite direction.

Impact of L_C Ratio. We next discuss the impact of L_C ratio ($L_C = L/C$) in (9) on the performances of OLA-Rec. We can tell that L_C controls the value of β_{iu} in (9), which in turn affects the calculation of λ in (16) and therefore indirectly influences the value of k. When quite a small L_C is adopted, k tends to be very large, indicating that information from a large number of the target user's friends (all of her friends in the extreme case) is considered to infer the target user's preference. Extremely, we may simply take every friend of the target user into account, ignoring the factor of limited attention. When L_C is very large, the target user's taste will merely depend on very few friends (none of her friends in the extreme case). Figure 2 displays the performances of OLA-Rec with different L_C ratio values. We observe that for all of the four datasets, as L_C increases, RMSE and MAE first decrease (prediction accuracy increases) and then increase (prediction accuracy decreases) after L_C goes beyond a certain threshold whose value is 100 for Douban, CiaoDVD, Epinions and 10 for Flixster. This confirms our intuition that by finding an optimal number of best influential friends for each user and learning their optimal attentions received from the target user, OLA-Rec is able to achieve a performance boost over models purely utilizing information from every friend without considering limited attention and models simply ignoring any social information.

Histograms on $\frac{k^*}{n}$. It is obvious that the upper bound of k_i^* is $|F(i)|$ which is the size of user i's total social connections. Therefore it is worth comparing the optimal number of selected social ties (denoted as k^*) with the exact number of total social ties (denoted as n) in the experiments. Figure 3 presents the histograms over $\frac{k^*}{n}$ for users on each dataset. We observe a skewed distribution on each of the four datasets, showing that the majority of the population

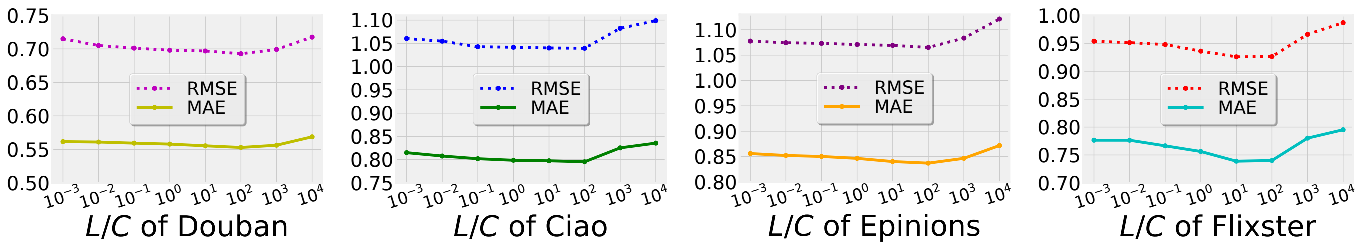


Figure 2: Impact of different L_C ratio values in OLA-Rec on RMSE and MAE for all four datasets

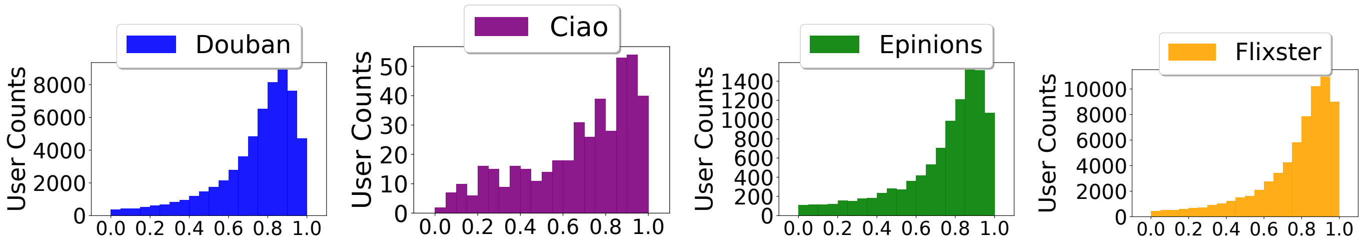


Figure 3: Histograms of users over $\frac{k^*}{n}$ (horizontal axis) on all four datasets

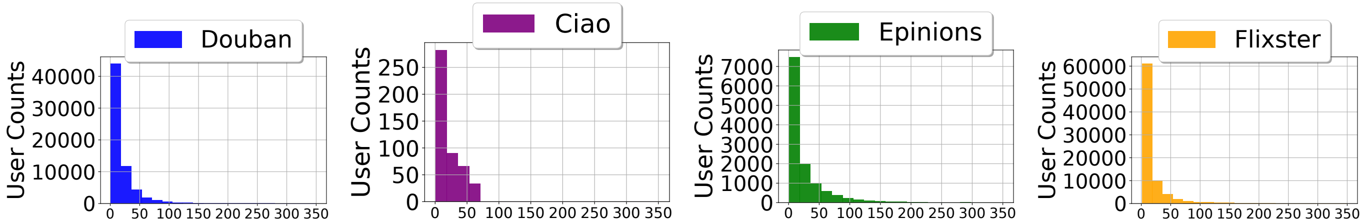


Figure 4: Histograms of users over number of total social connections (horizontal axis) on all four datasets

select between 70% and 100% of their social connections as social information sources. Furthermore, the histograms over total number of social connections are given in Figure 4, demonstrating a skewed distribution on each dataset as well. The combination of Figure 3 and Figure 4 provides us with two indications:

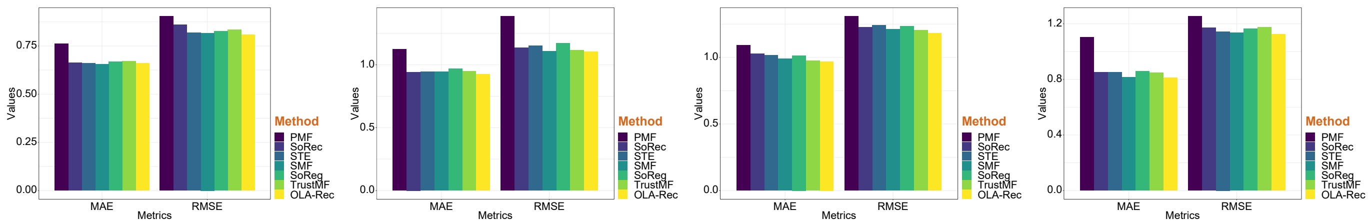
(1) Most users in practice only need social information from a subset of their total social connections (i.e., $\frac{k^*}{n} < 1$) when making recommendations to them.
(2) Users with a larger number of social connections tend to have a smaller $\frac{k^*}{n}$ and vice versa. This actually makes sense because having a large number of social connections may produce noisy information that has negative impact on inferring the preferences of these target users.

We remark that it is possible to use all the social connections when necessary, which actually is not a "bad" choice. The contribution of our work is that for each user we can find an optimal subset (size $k^* \in [1, n]$, with theoretical guarantee) of her social ties who contribute in affecting her preference without any useful social information loss through an elegant combination of motivation from social science and formulation from math. In fact,our model is expected to reduce useless/noisy social information when $k^* < n$, which happens for over 80% of the users having more than 15 social ties.

Cold Start Problem. Last but not least, we drill down to the performances of different algorithms on *cold-start* users. As is common practice, we define users rating less than five items as cold-start users. Figure 5 depicts the performances of various methods on cold start users. Our observations that social recommendation methods (including SoRec, STE, SMF, SoReg, TrustMF and OLA-Rec) significantly outperform PMF (a non-social algorithm) in terms of both MAE and RMSE in Figure 5 confirm the fact that social recommendation methods are superior to their non-social competitors particularly for cold-start users. We also observe that our OLA-Rec model have similar accuracies in terms of both MAE and RMSE to the other social recommendation baselines. This may due to the reason that when the target user processes a very small number of social connections, the diversity of these social ties reduces dramatically, making the proposed OLA-Rec model simply take all social connections into account to obtain as much useful information as possible. In this case, given exactly the same amount of social information shared by all social recommendation methods, the credit of slightly better performances obtained by OLA-Rec may go to learning the optimal attentions (i.e., $\boldsymbol{\alpha}_i^*$) for each target user i.

5 CONCLUSIONS

Limited attention is very important to social recommendation as it has been proved to have significant impact on users' online behaviours. Therefore, we propose to incorporate limited attention, a well-studied social science notion into social recommendation in an appropriate way. We first formulate the optimal limited attention problem, aiming to optimally bring the concept of limited attention into social recommendation. Then we develop a novel model which efficiently finds an optimal number of friends whose preferences have the best impact on the target user and adaptively learns an



(a) Douban (b) CiaoDVD (c) Epinions (d) Flixster

Figure 5: Performances of seven methods on cold start users in terms of MAE and RMSE for all four datasets

optimal personalized attention towards every selected friend, as well as the latent preference for each user. We also provide a proof on the optimality of the proposed algorithm. Extensive experiments on four real-world datasets demonstrate the improvement of our proposed method over existing approaches.

ACKNOWLEDGMENTS

This research is supported by China Postdoctoral Science Foundation No. BX201700136 , National Program on Key Basic Research Project No. 2015CB352300, National Natural Science Foundation of China Major Project No. U1611461.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *TKDE* 17, 6 (2005), 734–749.
[2] Oren Anava and Kfir Levy. 2016. k*-Nearest Neighbors: From Global to Local. In *Advances in Neural Information Processing Systems*. 4916–4924.
[3] Jiawei Chen, Yan Feng, Martin Ester, Sheng Zhou, Chun Chen, and Can Wang. 2018. Modeling Users' Exposure with Social Knowledge Influence and Consumption Influence for Recommendation. In *CIKM*. ACM, 953–962.
[4] Scott Counts and Kristie Fisher. 2011. Taking It All In? Visual Attention in Microblog Consumption. *ICWSM* 11 (2011), 97–104.
[5] Lin Cui, Caiyin Wang, Jia Wu, Jian Yang, and Michael Sheng. 2018. Individual Interest and Trust Driving Collective Intelligence Awareness for Social Recommendation. In *IJCNN*. IEEE, 1–6.
[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)* 39, 1 (1977), 1–38.
[7] Eric Gilbert and Karrie Karahalios. 2009. Predicting tie strength with social media. In *SIGCHI*. ACM, 211–220.
[8] Bruno Goncalves, Nicola Perra, and Alessandro Vespignani. 2011. Validation of Dunbar's number in Twitter conversations. *arXiv:1105.5170* (2011).
[9] Lesly Alejandra Gonzalez Camacho and Solange Nice Alves-Souza. 2018. Social network data to alleviate cold-start in recommender system. *Information Processing and Management: an International Journal* 54, 4 (2018), 529–544.
[10] Nathan Oken Hodas and Kristina Lerman. 2012. How visibility and divided attention constrain social contagion. In *PASSAT, SocialCom*. IEEE, 249–257.
[11] Mohsen Jamali and Martin Ester. 2009. Trustwalker: a random walk model for combining trust-based and item-based recommendation. In *KDD*. ACM, 397–406.
[12] Mohsen Jamali and Martin Ester. 2010. A matrix factorization technique with trust propagation for recommendation in social networks. In *Recsys*. ACM, 135–142.
[13] Daniel Kahneman. 1973. *Attention and effort*. Vol. 1063. Prentice-Hall Englewood Cliffs, NJ.
[14] Jeon-Hyung Kang and Kristina Lerman. 2013. LA-CTR: A Limited Attention Collaborative Topic Regression for Social Media.. In *AAAI*.
[15] Jeon-Hyung Kang, Kristina Lerman, and Lise Getoor. 2013. LA-LDA: a limited attention topic model for social recommendation. In *SBP-BRiMS*. Springer, 211–220.
[16] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *KDD*. 426–434.
[17] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.
[18] Hao Ma, Irwin King, and Michael R Lyu. 2009. Learning to recommend with social trust ensemble. In *SIGIR*. ACM, 203–210.
[19] Hao Ma, Haixuan Yang, Michael R Lyu, and Irwin King. 2008. Sorec: social recommendation using probabilistic matrix factorization. In *CIKM*. ACM, 931–940.
[20] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. 2011. Recommender systems with social regularization. In *WSDM*. ACM, 287–296.
[21] Peter V Marsden and Noah E Friedkin. 1994. Network studies of social influence. *Sociological Methods 'I&' Research* 22, 1 (1994), 127–151.
[22] Miller Mcpherson, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27, 1 (2001), 415–444.
[23] Andriy Mnih and Ruslan Salakhutdinov. 2007. Probabilistic matrix factorization. In *NIPS*. 1257–1264.
[24] Sanjay Purushotham, Yan Liu, and C-C Jay Kuo. 2012. Collaborative Topic Regression with Social Matrix Factorization for Recommendation Systems. (2012).
[25] Xueming Qian, He Feng, Guoshuai Zhao, and Tao Mei. 2014. Personalized recommendation combining user interest and social circle. *TKDE* 26, 7 (2014), 1763–1777.
[26] Jasson DM Rennie and Nathan Srebro. 2005. Fast maximum margin matrix factorization for collaborative prediction. In *ICML*. ACM, 713–719.
[27] Ronald A Rensink, J Kevin O'Regan, and James J Clark. 1997. To see or not to see: The need for attention to perceive changes in scenes. *Psychological science* 8, 5 (1997), 368–373.
[28] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). 2011. *Recommender Systems Handbook*. Springer.
[29] Ruslan Salakhutdinov and Andriy Mnih. 2008. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *ICML*. ACM, 880–887.
[30] Nathan Srebro and Tommi Jaakkola. 2003. Weighted low-rank approximations. In *ICML-03*. 720–727.
[31] Xiaoyuan Su and Taghi M Khoshgoftaar. 2009. A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009 (2009).
[32] Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. 2013. Exploiting homophily effect for trust prediction. In *WSDM*. 53–62.
[33] Xin Wang, Roger Donaldson, Christopher Nell, Peter Gorniak, Martin Ester, and Jiajun Bu. 2016. Recommending groups to users using user-group engagement and time-dependent matrix factorization. In *AAAI*.
[34] Xin Wang, Steven CH Hoi, Martin Ester, Jiajun Bu, and Chun Chen. 2017. Learning personalized preference of strong and weak ties for social recommendation. In *WWW*. 1601–1610.
[35] Xin Wang, Wei Lu, Martin Ester, Can Wang, and Chun Chen. 2016. Social Recommendation with Strong and Weak Ties. In *CIKM*. ACM, 5–14.
[36] Xin Wang, Wenwu Zhu, Chun Chen, and Martin Ester. 2018. Joint User-and Event-Driven Stable Social Event Organization.. In *WWW*. 1513–1522.
[37] Lillian Weng, Alessandro Flammini, Alessandro Vespignani, and Fillipo Menczer. 2012. Competition among memes in a world with limited attention. *Scientific reports* 2 (2012), 335.
[38] Li Weng and Filippo Menczer. 2010. GiveALink tagging game:an incentive for social annotation. In *ACM SIGKDD Workshop on Human Computation*. 26–29.
[39] Le Wu, Peijie Sun, Richang Hong, Yong Ge, and Meng Wang. 2018. Collaborative Neural Social Recommendation. *TSMC: Systems* (2018).
[40] Bo Yang, Yu Lei, Dayou Liu, and Jiming Liu. 2013. Social collaborative filtering by trust. In *IJCAI*. AAAI Press, 2747–2753.
[41] Shuang-Hong Yang, Bo Long, Alex Smola, Narayanan Sadagopan, Zhaohui Zheng, and Hongyuan Zha. 2011. Like like alike: joint friendship and interest propagation in social networks. In *WWW*. 537–546.
[42] Mao Ye, Xingjie Liu, and Wang-Chien Lee. 2012. Exploring social influence for recommendation: a generative model approach. In *SIGIR*. ACM, 671–680.
[43] Yijia Zhang, Wanli Zuo, Zhenkun Shi, Lin Yue, and Shining Liang. 2018. Social Bayesian Personal Ranking for Missing Data in Implicit Feedback Recommendation. In *KSEM*. Springer, 299–310.
[44] Tong Zhao, Julian McAuley, and Irwin King. 2014. Leveraging social connections to improve personalized ranking for collaborative filtering. In *CIKM*. ACM, 261–270.
[45] Zhou Zhao, Hanqing Lu, Deng Cai, Xiaofei He, and Yueting Zhuang. 2016. User preference learning for online social recommendation. *TKDE* 28, 9 (2016), 2522–2534.

A SUPPLEMENT

Datasets.

- *Douban.* This is a public dataset from a Chinese movie forum (http://movie.douban.com/), containing user-user friendships and user-movie ratings, and is publicly available from (https://www.cse.cuhk.edu.hk/irwin.king.new/pub/data/douban).
- *CiaoDVD.* The trust relationships among users from CiaoDVD as well as their ratings on DVDs are included. It was crawled from the entire category of DVDs of a UK DVD community website (http://dvd.ciao.co.uk) in December, 2013.
- *Epinions.* This dataset comes from an American website and consists of trust relationships and user-item ratings. This dataset (http://www.trustlet.org/wiki/Epinions_dataset) is extracted from the consumer review website Epinions (http://www.epinions.com/), which contains user-user trust relationships and numerical ratings.
- *Flixster.* This dataset (http://www.cs.ubc.ca/~jamalim/datasets/) contains the information of user-movie ratings as well as user-user friendships from Flixster, an American social movie site for discovering new movies (http://www.flixster.com/).

We remove users with less than 2 ratings and select 80% of each user's ratings at random for training, leaving the remainder as test set.

B PROOF OF EQUATION (7)

THEOREM B.1. *(Hoeffding's Inequality). Let $\{x_j\}_{j=1}^n \in [L_j, U_j]^n$ be a sequence of independent random variables, such that $\mathbb{E}[x_j] = \mu_j$. Then, it holds that:*

$$\mathbb{P}\Big(\Big|\sum_{j=1}^{n} x_j - \sum_{j=1}^{n} \mu_j\Big| \geq x\Big) \leq 2e^{-\frac{2x^2}{\sum_{j=1}^{n}(U_j - L_j)^2}}.$$

We want to prove that by Hoeffding's inequality, with probability at least $1-\delta$, we have :

$$\Big|\sum\nolimits_{j\in F(i)} \alpha_{ij}\epsilon_j\Big| \leq C\|\boldsymbol{\alpha}_i\|_2, \quad s.t. \quad C = b \times \sqrt{(2log(\frac{2}{\delta}))}, \tag{26}$$

where $|\epsilon_j| \leq b$ for some $b \geq 0$ and $\mathbb{E}[\epsilon_j] = 0$. We further constrain $\alpha_{ij} \geq 0$ and $\sum_{j=1}^{|F(i)|} \alpha_{ij} = 1$.

PROOF. Given $x_j = \alpha_{ij}\epsilon_j$, we have:

$$\mu_j = \mathbb{E}[x_j] = \mathbb{E}[\alpha_{ij}\epsilon_j] = \mathbb{E}[\alpha_{ij}]\mathbb{E}[\epsilon_j] = 0. \tag{27}$$

and thus:

$$\mathbb{P}\Big(\Big|\sum_{j=1}^{n} x_j\Big| \geq x\Big) \leq 2e^{-\frac{2x^2}{\sum_{j=1}^{n}(U_j - L_j)^2}}. \tag{28}$$

Let $n = |F(i)|$ and $x = C\|\boldsymbol{\alpha}_i\|_2$, we have:

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\sum\nolimits_{j\in F(i)} \alpha_{ij}\epsilon_j\Big| \geq C\|\boldsymbol{\alpha}_i\|_2\Big) \\
&\leq 2e^{-\frac{2C^2\|\boldsymbol{\alpha}_i\|_2^2}{\sum_{j=1}^{n}(U_j-L_j)^2}} \\
&= 2e^{-\frac{2*b^2*2log(\frac{2}{\delta})\|\boldsymbol{\alpha}_i\|_2^2}{\sum_{j=1}^{n}(U_j-L_j)^2}}.
\end{aligned} \tag{29}$$

Recall that $-b \leq \epsilon_j \leq b$ and therefore $L_j = -b * \alpha_{ij}$ and $U_j = b * \alpha_{ij}$. Thus we have:

$$\sum_{j=1}^{n}(U_j - L_j)^2 = \sum_{j=1}^{n} 4b^2\alpha_{ij} = 4b^2\|\boldsymbol{\alpha}_i\|_2^2. \tag{30}$$

Substitute Eq (30) into Eq (29), we have:

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\sum\nolimits_{j\in F(i)} \alpha_{ij}\epsilon_j\Big| \geq C\|\boldsymbol{\alpha}_i\|_2\Big) \\
&\leq 2(\frac{\delta}{2})^{\frac{4b^2\|\boldsymbol{\alpha}_i\|_2^2}{\sum_{j=1}^{n}(U_j-L_j)^2}} \\
&= 2 * \frac{\delta}{2} = \delta.
\end{aligned} \tag{31}$$

By rewriting Eq (31), we finally get:

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\sum\nolimits_{j\in F(i)} \alpha_{ij}\epsilon_j\Big| \leq C\|\boldsymbol{\alpha}_i\|_2\Big) \\
&\geq 1-\delta,
\end{aligned} \tag{32}$$

which completes the proof.

□