2018 IEEE International Conference on Data Mining

Social Recommendation with Missing Not at Random Data

Jiawei Chen[1,2], Can Wang[1,2,†], Martin Ester[3], Qihao Shi[1,2], Yan Feng[1,2], and Chun Chen[1]
[1]*College of Computer Science, Zhejiang University, China*
[2]*Zhejiang University-LianlianPay Joint Research Center*
[3]*School of Computing Science, Simon Fraser University, Canada*
{sleepyhunt,wcan}@zju.edu.cn;ester@cs.sfu.ca;{shiqihao321,fengyan,chenc}@zju.edu.cn

***Abstract*—With the explosive growth of online social networks, many social recommendation methods have been proposed and demonstrated that social information has potential to improve the recommendation performance. However, existing social recommendation methods always assume that the data is missing at random (MAR) but this is rarely the case. In fact, by analysing two real-world social recommendation datasets, we observed the following interesting phenomena: (1) users tend to consume and rate the items that they like and the items that have been consumed by their friends. (2) When the items have been consumed by more friends, the average values of the observed ratings will become smaller, not larger as assumed by the existing models. To model these phenomena, we integrate the missing not at random (MNAR) assumption in social recommendation and propose a new social recommendation method SPMF-MNAR, which models the observation process of rating data based on user's preference and social influence. Extensive experiments conducted on large real-world datasets validate that SPMF-MNAR achieves better performance than existing social recommendation methods and the non-social methods based on MNAR assumption.**

***Index Terms*—MNAR, Social recommendation, Graphic model**

I. Introduction

Nowdays, recommender systems have become a core component of many online services such as Amazon and IMDB. Collaborative filtering (CF), as the most prevalent recommendation model in these systems, infers user's preference and produces recommendations based on user's historical feedback. In practice, most users consume and rate only a small fraction of the available items. Thus, traditional CF algorithms are impeded by the data sparsity problem. To mitigate this problem, many methods have been proposed to integrate social network information into recommender systems. These methods mainly model social influence on user's rating values and show that the social-based CF model has potential to improve recommendation performance [1]–[3].

However, existing social recommendation methods usually assume that the rating data is *missing at random (MAR)*. That is, the process by which users select the items to consume and rate (the *observation process* of rating data) is independent from what rating values users give [4]. In fact, the MAR assumption may not be suitable for the recommendation data [5] since users consume items they like more than ones they dislike [6]–[8]. Accordingly, mostly large-value ratings are observed in rating datasets while the low-value ratings are usually missing. When the MAR assumption is incorrect, these methods will suffer from the biased parameter estimation and prediction [7].

[†]Corresponding author: wcan@zju.edu.cn

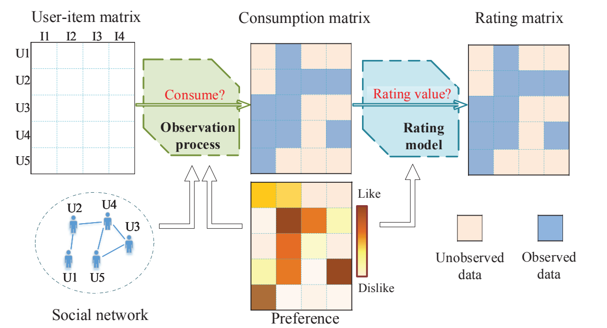


Fig. 1. The generative process of rating data in social recommendation

To better understand the characteristics of the observation process of the rating data in social recommendation, in this paper we conduct a thorough analysis on two well-known social recommendation datasets, Ciao and Epinions. We make the following observations: (1) Users tend to consume (rate) the items that they like. (2) Users tend to consume (rate) the items that have been consumed by their friends. (3) The influence of the friends on user's rating values is smaller than their influence on the user's consumption. In fact, we observed the following interesting phenomenon: When the items have been consumed by more friends, the average values of the observed ratings will become smaller, not larger as assumed by the state-of-the-art models. This phenomenon may be explained as follows. The items, which are popular among our friends, easily attract our attention. Under social influence, we may consume these items even if they do not meet our tastes. Thus, the average rating values of these items will be smaller than those of other consumed items.

Motivated by our analysis, in this paper we integrate the missing not at random (MNAR) assumption in social recommendation and propose a new recommendation method SPMF-MNAR. SPMF-MNAR models both the rating values and the rating observation process. As shown in Fig. 1, on the one hand, SPMF-MNAR generates users' consumption based on social influence and their own preference (Observation

978-1-5386-9159-5/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDM.2018.00018

process). On the other hand, the rating values given by users to his consumed items are generated by classic CF model based on their preference (Rating model). Specifically, different from existing social recommendation methods that exploit social influence on users' rating values, we focus on exploiting social influence on the observation process of rating data. Then, we develop an accelerated inference algorithm for our SPMF-MNAR model based on stochastic variational inference (SVI) to improve scalability. Our comprehensive experimental results clearly demonstrate the effectiveness of our algorithm compared to both existing social recommendation methods and the non-social methods based on MNAR assumption.

The main contributions of this paper are as follows:

- We integrate the missing not at random (MNAR) assumption in social recommendation and conduct substantial analyses on real-world datasets to explore the following questions: (1) Is it beneficial to model the MNAR assumption in the social recommendation? (2) What contributes to user's consumption?
- We propose a novel probabilistic model SPMF-MNAR for social recommendation by integrating the social-based generative model of rating observation process into the classic CF rating model. Besides, we develop an efficient and effective stochastic variational inference method to infer the posterior for our probabilistic model.
- Our experimental evaluation on large real-world datasets show that our method outperforms both state-of-the-art social recommendation methods and the non-social methods based on MNAR assumption.

The rest of this paper is organized as follows. We briefly review related works in section 2. The social empirical analyses were conducted in section 3. In section 4, we present the details of the SPMF-MNAR model. The experiment results and discussions are presented in section 5. Finally, we conclude the paper and present some directions for future work in section 6.

II. Related works

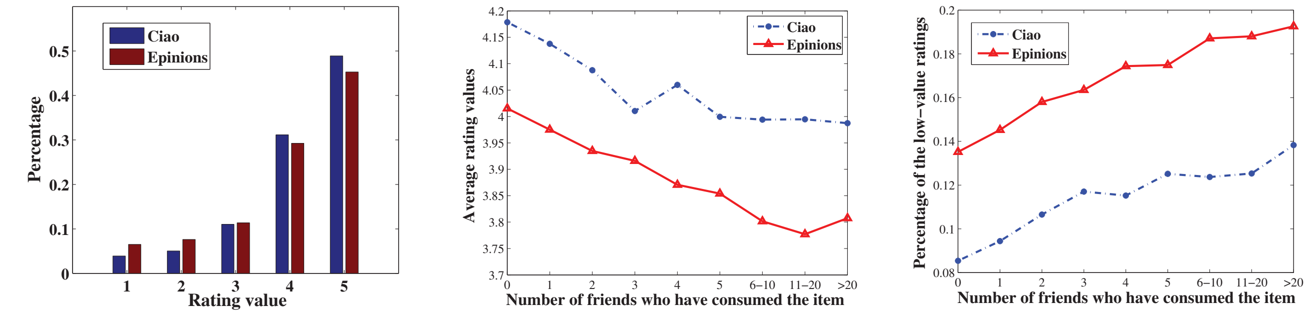
With the exponential growth of information generated on consumer review websites and e-commerce websites, recommender systems are drawing more attention from both academia and industry. For the system with explicit feedback (numerical ratings), substantial works have been done about collaborative filtering (CF) model during the past two decades for its accuracy and scalability [9], [10]. Here, we review the most related works from two perspectives: one on social recommendation and the other on the recommendation with missing not at random assumption (MNAR).

Social recommendation. To address the limitation of the traditional CF methods such as cold-start problem, many methods have been proposed to integrate social information into recommender systems [11]–[15]. These methods mainly assumed that connected users will share similar preference. For example, some social recommendation methods including SoRec [16], TrustSVD [3], PSLF [17], jointly factorized rating matrix and social(trust) matrix by sharing a common latent user space. Yang [18] further extended SoRec to their hybrid method TrustMF that combines both a truster model and a trustee model. They believed that both the users who trust the active user and those who are trusted by the user will influence the user's rating values on the items. Also, some methods utilized a social regularization term [11], [19], [20] to model the social influence on ratings. Specifically, SocialMF [19] introduces an additional regularization term to constrain user's latent preference close to the average of his trusted friends. In some other methods [21]–[24], users' ratings are considered as synthetic results of their preference and social influence. Specifically, in RSTE [21], user's rating values are generated by combing user's own taste and his friends' preference.

However, existing social recommendation methods usually assumed that the rating data is missing at random (MAR). In fact, our analysis in section 3 shows that this is not the case. Thus, without modeling observation process of rating data, inferences can be biased and prediction accuracy suffers [7]. Besides, we observe that the social influence on user's consumption is larger than the influence on user's rating values. It will be interesting and effective to exploit the social influence on the rating observe process (user's consumption), which has not been considered by these works.

Recommendation with MNAR assumption. Most recent CF methods assumed that the CF data is missing at random (MAR). That is, the process that selects the observed data is independent of the value of unobserved rating data. However, as mentioned in [4], [5], [7], [8], when the MAR assumption is incorrect, inferences are biased and predictive performance can suffer. In fact, there are many evidences that CF data is missing not at random (MNAR) [5], [6]. Thus, to deal with this problem, some CF methods with MNAR assumption have been proposed during the past decade. These methods had the generative process of user's consumption (Observation process) to model the dependency between which items a user consumes (rates) and what ratings the user gives. Marlin and Zemel [4] used a mixture of Multinomials (MM) to model user's rating value and generated user's consumption based on this value. Hernandez-Lobato et al. [7] noticed the poor flexibility of MM model and the powerful performance of matrix factorization method. Thus they proposed a probabilistic matrix factorization model to deal with CF data that is missing not at random (MNAR). More recently, Ohsawa et al. [6] further extended PMF [10] to their GPMF by considering the dependency between why a user consumes an item and how that affects the rating value. Thus, users' consumption had been modeled in their GPMF to control the weights on different latent dimensions when they performed probabilistic matrix factorization on rating matrix.

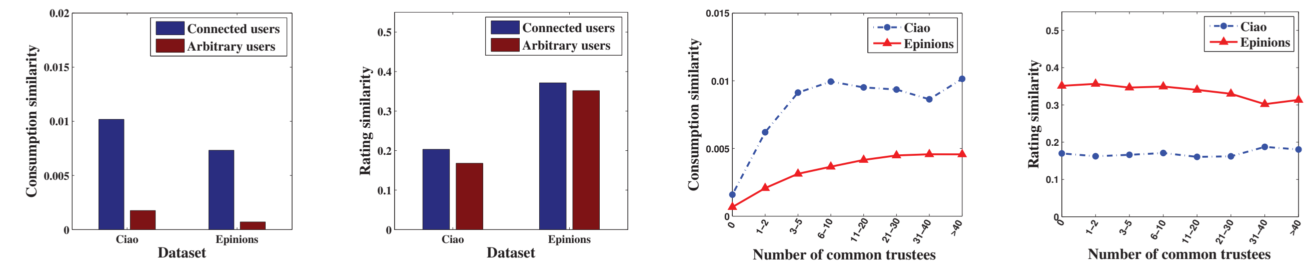
However, all of these methods ignore the social influence on user's behavior. In fact, connected users in a social network often (un)knowingly recommend products to each other and their consumption will inevitably be influenced by their social relations. [25], [26]. Thus, the SPMF-MNAR proposed in this paper, models user's consumption with social influence for better estimating user's preference on items.



(a) The distribution of the rating values of the consumed items

(b) The average rating values with varying number of friends who have consumed the item

(c) The percentage of the low-value ratings with varying number of friends who have consumed the item



(d) The average consumption similarity between the connected users and arbitrary two users

(e) The average rating similarity between the connected users and arbitrary two users

(f) The consumption similarity for user pairs with their varying number of common friends

(g) The rating similarity for user pairs with their varying number of common friends

Fig. 2. The results of statistics analyses on datasets Ciao and Epinions

TABLE I
STATISTICS OF TWO DATASETS

Features	Epinions	Ciao
Number of users	49,289	7,375
Number of items	139,738	106,797
Number of ratings	664,824	280,391
Density of ratings	0.0097%	0.0356%
Number of edges	487,183	111,781
Density of edges	0.0201%	0.2055%

III. USER'S CONSUMPTION IN SOCIAL-BASED RECOMMENDATION DATASETS

In this section, we conducted a thorough analysis of two social-based recommendation datasets, Ciao and Epinions. Our analysis is intended to address the following questions: (1) Is it beneficial to model the MNAR assumption in social recommendation? (2) What contributes to users' consumption?

A. Dataset description

We use two datasets Epinions[1] and Ciao[2] for our analyses and experiments on social recommendation. Epinions and Ciao are two knowledge sharing websites where users can give ratings to their consumed products from 1 to 5. These datasets contain several kinds of information: The consumption information, which denotes the items a user have consumed(rated), can be got from the observation process of rating data since a user must consume an item before he rates it [6]. Also there is rating value information, which denotes the observed rating values that the user gave to his consumed items. Besides, these datasets contain social information since the user in the websites can maintain a "trust" (friend) list which forms a directed social network between users. Thus, Epinions and Ciao are ideal datasets that have been used widely for analyses and experiments on social recommendation. The datasets statistics are presented in Table I.

B. Analyses of user's consumption in social recommendation

To answer above questions, we conducted five statistics analyses: (1) We calculate the distribution of the rating values for those consumed items as shown in Fig. 2(a). (2) We calculate the average rating similarity and consumption similarity for each connected pair. The rating similarity between two users is represented by the Pearson correlation coefficient [15] between their rating values on the common consumed items. The consumption similarity between two users is the Jaccard coefficient [27] between the sets of items consumed by them. In comparison, we calculate the average rating similarity and consumption similarity between arbitrary two users. The results are presented in Fig. 2(d),2(e). (3) We divide pairs of users into several groups according to the number of their common friends. We then calculate the average rating similarity and consumption similarity between every user pairs in each group. The results are shown in Fig. 2(f),2(g). (4)

[1]http://www.trustlet.org/epinions
[2]http://www.cse.msu.edu/~tangjili/trust

We divide observed ratings into several parts according to the number of friends have consumed the item. Then we calculate the average rating values and the percent of the low-value ratings (less than 3) in each part. The results are presented in Fig. 2(b),2(c).

Four important observations are concluded from these results.

***Observation** 1:* Users tend to consume (rate) the items that they like.

As shown in Fig. 2(a), more than 70% ratings are larger than 3 and just 15% ratings are less than 3. High-value ratings are usually given by the users to their consumed items. That is, users tend to consume the items that they think they will like. This observation is consistent with the finding in [4]. They conduct online survey to collect users' ratings for randomly selected items and find that the average rating values for these random items are much smaller than the rating values for the user-selected (consumed) items.

***Observation** 2:* Users tend to consume (rate) the items that have been consumed by their friends.

***Observation** 3:* The influence of the friends on user's ratings is smaller than their influence on the user's consumption.

As shown in Fig. 2(d),2(e), connected users tend to consume much more items in common, but their rating values does not exhibit so much more similarity than ordinary user pairs as their consumption. Meanwhile, as shown in Fig. 2(f),2(g), as the number of common friends increases, users tend to have more commonality in their consumption while their rating values remain mostly unaffected by the number of their shared social relations. Thus, comparing with users' rating values, their consumption is more sensitive to the social influence.

***Observation** 4:* The rating value of a user decreases with increasing number of friends that have consumed the item.

As shown in Fig. 2(c),2(b), when the items are consumed by more friends, the average rating values will become smaller and the percent of the low-value ratings will become larger.

Based on these observations, we can answer above two questions:

***Conclusion** 1:* It is beneficial to model MNAR assumption in the social recommendation.

On the one hand, based on our observations, we can conclude that the rating data is missing not at random. Specifically, users tend to consume the items that they like and the low-value ratings are usually missing (Observation 1). When we deal with MNAR data without MNAR assumption, inferences can be biased and prediction accuracy suffers [7]. On the other hand, as mentioned in our observation 3, comparing with users' rating values their consumption is more strongly influenced by their friends. It seems effective to further exploit social influence on users' consumption for better inferring users true preference.

***Conclusion** 2:* Users' consumption can be considered as contributions of users' preference and social influence.

On the one hand, preference precedes choices because choices are made to maximize preference [28]. When users make the decisions on the consumption, they will observe themselves and judge whether they will like these items [7]. On the other hand, in the social network, each of us belongs to some content-shared communities [29]. Users are explicitly linked and often (un)knowingly recommend products/services to friends [17]. The items consumed by our friends are more likely brought to our attention and stored in episodic memory [25], which contributes to our consumption [26]. Also, the interesting Observation 4 can be explained by this conclusion. Comparing with the popular items among user's friends, user's consumption on these unpopular items was more attributed to his preference. Thus, when the items have been consumed by more friends, the average rating values will become smaller.

IV. Problem and algorithm

In this section, we present our SPMF-MNAR model. we will start with the problem definition.

A. Problem definition

In recommender systems, we are given a set of users U (including n users) and a set of items I (including m items) as well as an observed rating set $R^o = \{r_{ij} : i \in U, j \in I, (i,j) \in \mathcal{O}\}$, where the entries r_{ij} denote the rating value of item j given by user i and $\mathcal{O}$ is the set of user-item pairs for which a rating is observed. In practice, R^o is usually a small subset of the entries of a complete $n \times m$ rating matrix R. Note that the observation process of rating data indicates user's consumption since a user must consume an item before he rates it [6], [7]. Thus, we model the location of the entries included in R^o using a $n \times m$ consumption matrix X. For each entry in X, $x_{ij} = 1$ denotes user i have consumed (and rated) item j ($r_{ij} \in R^o$ is observed) and $x_{ij} = 0$ means not ($r_{ij} \notin R^o$). When it comes to social recommendation, we also have social network information, which indicates the connection between users. Let $g^i = [g^i_1, g^i_2, \dots g^i_{|g^i|}]$ be a vector, which we refer to as the friends of user i. The k-th element of g^i (g^i_k) denotes the id of k-th friend of user i. The length of vector g^i ($|g^i|$) denotes the number of friends of user i. The task of social recommendation is to predict user's preference(ratings) on items precisely so that the recommendations will meet user's taste.

B. Social Probabilistic Matrix Factorization with missing not at random data assumption

Different from existing social recommendation methods, we integrate MNAR assumption into social recommendation and model the observation process of rating data. As shown in Fig. 1, on the one hand, we generate users' consumption based on social influence and their own preference. On the other hand, the rating values given by users to their consumed items are generated by the classic CF model. Here we start with the overall generative process behind SPMF-MNAR, and then describe the details step by step. For each consumption $x_{ij} \in X$ and rating value $r_{ij} \in R$, SPMF-MNAR assumes the following generative process, corresponding to the graphical model presented in Fig. 3.

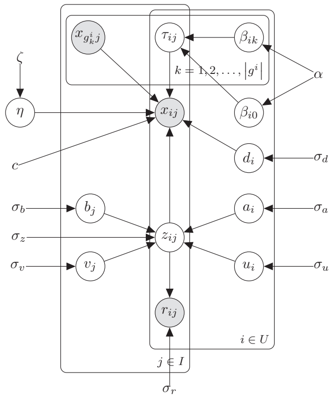


Fig. 3. The graphical model of SPMF-MNAR.

1) Draw the influence of friend's consumption: $\eta_0 \sim Beta(\zeta)$, $\eta_1 \sim Beta(\zeta)$.
2) For each user i:
   I. Draw the bias: $a_i \sim \mathcal{N}(0, \sigma_a^2)$.
   II. Draw the latent preference: $u_i \sim \mathcal{N}(0, \sigma_u^2 \mathbf{I})$.
   III. Draw the acceptive threshold: $d_i \sim \mathcal{N}(0, \sigma_d^2)$.
   IV. Draw the importance of different components on the consumption: $\beta_i = Dir(\alpha)$.
3) For each item j:
   I. Draw the bias: $b_j \sim \mathcal{N}(0, \sigma_b^2)$.
   II. Draw the latent attribute: $v_j \sim \mathcal{N}(0, \sigma_v^2 \mathbf{I})$.
4) For each user-item pair (i, j):
   I. Draw user's preference on the item: $z_{ij} \sim \mathcal{N}(u_i^\top v_j + a_i + b_j, \sigma_z^2)$.
   II. Draw the component contributing to the consumption: $\tau_{ij} \sim Multinomial(\beta_i)$.
   III. When $\tau_{ij0} == 1$, draw the consumption based on user's preference: $x_{ij} \sim Bernoulli(x_{ij}|\sigma(c(z_{ij} - d_i)))$.
   IV. When $\tau_{ijk} == 1 (1 \leq k \leq |g^i|)$, draw the consumption based on the influence of user's k-th friend: $x_{ij} \sim Bernoulli(x_{ij}|\eta_1 x_{g_k^i j} + \eta_0(1 - x_{g_k^i j}))$.
5) For each observed rating value r_{ij}:
   I. Draw the rating value based on user's preference on the item: $r_{ij} \sim \mathcal{N}(z_{ij}, \sigma_r^2)$.

Several assumptions are made in the SPMF-MNAR model.

First, users' preference on the items can be generated by probabilistic matrix factorization [10]. For each user-item pair (i, j), we generate a latent continuous variable z_{ij}, which characterizes the preference of user i on item j as follow:

$$p(z_{ij}|a_i, b_j, u_i, v_j) = \mathcal{N}(z_{ij}|a_i + b_j + u_i^\top v_j, \sigma_z^2) \tag{1}$$

where $\mathcal{N}(.)$ denotes Gaussian distribution; u_i denotes the D-dimensional latent preference vector of user i; v_j denotes the D-dimensional latent attribute vector of item j; a_i and b_j denote the biases of user i and item j. Here we use a_i and b_j to capture the common phenomenon that some users tend to be generous and have a wide range of interest while some items tend to have high quality [3]. Then, we can generate user's rating values based on his preference as follow:

$$p(r_{ij}|z_{ij}) = \mathcal{N}(r_{ij}|z_{ij}, \sigma_r^2) \tag{2}$$

Here we remark that our simple rating model can be replaced by some sophisticated social-based recommended models such as SocialMF [19], TrustMF [18]. Also, we do agree that modeling social influence on ratings is potential to further improve recommendation performance. However, in this paper we focus on exploiting social influence on the observation process (user's consumption) and exploring how it improves recommendation performance, while the social-based rating model has been explored by substantial related literatures and is also orthogonal to our proposed model. Thus, we try to isolate the effect of these sophisticated rating models to set off the effect of our social-based model of the observation process.

Second, motivated by the analyses in section 3, user's consumption is modeled as contributions of user's preference and social influence. For each element $x_{ij} \in X$, with probability β_{i0}, x_{ij} is generated based on user's own preference $x_{ij} \sim p(x_{ij}|z_{ij}, d_i, c)$. With probability β_{ik} $(k = 1, 2, \ldots |g^i|)$, x_{ij} is generated based on the influence of his k-th friends $x_{ij} \sim p(x_{ij}|x_{g_k^i j}, \eta)$. Here β_{ik} for $k = 0, 1, \ldots |g^i|$ denotes the mixing coefficients for each user i, which captures the importance of social influence and user's own preference on his consumption. $\beta_i = [\beta_{i0}, \beta_{i1}, \ldots \beta_{i|g^i|}]$ satisfies $\sum_{k=0}^{|g^i|} \beta_{ik} = 1$ as it follows a Dirichlet distribution with parameter α. Based on this generative process, x_{ij} can be represented as random mixtures based on users' own preference and social influence:

$$p(x_{ij}) = \beta_{i0} p(x_{ij}|z_{ij}, d_i, c) + \sum_{k=1}^{|g^i|} \beta_{ik} p(x_{ij}|x_{g_k^i j}, \eta) \tag{3}$$

To facilitate the derivation and description, we introduce Multinomial auxiliary variables $\tau_{ij} \sim Mult(\beta_i)$ to indicate which component contributes to the consumption x_{ij}. $\tau_{ij} = [\tau_{ij0}, \tau_{ij1}, \ldots \tau_{ij|g^i|}]$ is a $|g^i| + 1$ dimensional binary random variable having a 1-of-K representation in which a particular element is equal to 1 and all other elements are equal to 0.

Third, when x_{ij} was generated based on user's own preference ($\tau_{ij0} = 1$), Bernoulli-logit model could be used to capture the phenomenon that users tend to consume the items that they like (our observation 1):

$$p(x_{ij}|z_{ij}, d_i, c, \tau_{ij0} = 1) = Bernoulli(x_{ij}|\sigma(c(z_{ij} - d_i))) \tag{4}$$

where $\sigma(.)$ denotes logistic function. We introduce offset variable d_i, which is defined as acceptive threshold for each user i, to capture the following intuition: When user's preference on the item is larger than the acceptive threshold ($z_{ij} > d_i$), users tend to consume the item; When the preference is smaller than the acceptive threshold ($z_{ij} < d_i$), users tend to reject it. Also we introduce a scale parameter $c > 0$ to control the influence of user's preference z_{ij} on his consumption x_{ij}. When the scale parameter c has larger value, x_{ij} will be more sensitive to z_{ij}. In other words, a slight disturbance on z_{ij} will contribute to the more change on the probability distribution of x_{ij}. In summary, we can find that the larger z_{ij} will bring the larger values of $p(x_{ij} = 1)$, indicating the user is more likely to consume the item. It coincides with our observation 1.

Fourth, when x_{ij} was generated based on the influence of his k-th friend ($\tau_{ijk} = 1$), we model x_{ij} by Bernoulli distribution and the Bernoulli parameter depends on the consumption of his k-th friend ($x_{g_k^i j}$):

$$p(x_{ij}|x_{g_k^i j}, \eta, \tau_{ijk} = 1) = Bernoulli(x_{ij}|\eta_1 x_{g_k^i j} + \eta_0(1 - x_{g_k^i j})) \tag{5}$$

The parameters $\eta = [\eta_0, \eta_1]$ indicate the conditional dependency between user's consumption and the selected friend's consumption. $\eta_1 = p(x_{ij} = 1|x_{kj} = 1, \tau_{ijk} = 1)$ denotes the probability that the user will follow his friend's consumption to consume the item. $\eta_0 = p(x_{ij} = 1|x_{kj} = 0, \tau_{ijk} = 1)$ denotes the probability that the user will consume the item when his friend didn't consume it. Typically, users tend to consume the items that have been consumed by their friends and thus η_0 should be different from η_1. Our experiment results presented in section 5 validate this point and show that the posterior expectation of η_1 is much larger than η_0.

C. Approximate Inference

Considering that the exact posterior probability of SPMF-MNAR is not tractable to compute, so we develop an efficient approximate method to compute posterior based on variational inference. The mean field theory drives us to partition latent variables into disjoint groups and these variables are governed by their own variational parameters. Besides, we can specify the form of the factored variational distribution of each variable as same as its corresponding conditional [30]. That is, we define variational distribution as follow:

$$\begin{aligned}
&q(u, v, a, b, d, \beta, \tau, \eta, z) = \prod_{i \in U} \Big(\mathcal{N}(a_i|\mu_i^a, (\Lambda_i^a)^{-1}) \\
&\times \mathcal{N}(u_i|\mu_i^u, (\Lambda_i^u)^{-1}) \mathcal{N}(d_i|\mu_i^d, (\Lambda_i^d)^{-1}) Dir(\beta_i|\kappa_i^\beta) \Big) \\
&\times \prod_{j \in I} \mathcal{N}(b_j|\mu_j^b, (\Lambda_j^b)^{-1}) \mathcal{N}(v_j|\mu_j^v, (\Lambda_j^v)^{-1}) \prod_{l=0,1} Beta(\eta_l|\kappa_l^\eta) \\
&\times \prod_{i \in U, j \in I} Mult(\tau_{ij}|\varphi_{ij}) \mathcal{N}(z_{ij}|\mu_{ij}^z, (\Lambda_{ij}^z)^{-1})
\end{aligned} \tag{6}$$

Note that Bernoulli-logit likelihood does not admit a conjugate prior. To address this problem we employ the Gaussian lower bound on the logistic function [31] as follow:

$$\begin{aligned}
&\sigma(x) \geqslant h(x, \xi) = \sigma(\xi) \exp((x - \xi)/2 - \lambda(\xi)(x^2 - \xi^2)) \\
&\lambda(\xi) = \frac{1}{2\xi}(\sigma(\xi) - \frac{1}{2})
\end{aligned} \tag{7}$$

which is a tight lower bound on the logistic function, with an additional parameter ξ. Thus, we have the following objective function, a bound on the log likelihood of the observations.

$$\begin{aligned}
&L(q) = \ln p(R^o, X) \geqslant E_q[\ln p(R^o, X, \Theta)] - E_q[\ln q(\Theta)] \\
&\geqslant \tilde{L}(q, \xi) = \sum_{i \in U, j \in I} E_q[\tau_{ij0} \ln(h((2x_{ij} - 1)\gamma_{ij}, \xi_{ij}))] \\
&+ \sum_{i \in U, j \in I} \sum_{k=1}^{|g_i|} E_q[\tau_{ijk} \ln p(x_{ij}|x_{g_k^i j}, \eta)] \\
&+ \sum_{i \in U, j \in I} E_q[\ln p(\tau_{ij}|\beta_i) + \ln p(z_{ij}|a_i, b_j, u_i, v_j, \sigma_z)] \\
&+ \sum_{i \in U} E_q[\ln p(u_i|\sigma_u) + \ln p(a_i|\sigma_a) + \ln p(\beta_i|\alpha) + \ln p(d_i|\sigma_d)] \\
&+ \sum_{j \in I} E_q[\ln p(v_i|\sigma_v) + \ln p(b_i|\sigma_b)] + \sum_{ij \in R^o} E_q[\ln p(r_{ij}|z_{ij})] \\
&+ \sum_{l=0,1} E_q[\ln p(\eta_l|\zeta)] - E_q[\ln q(\Theta)]
\end{aligned} \tag{8}$$

where $\Theta \equiv \{u, v, a, b, d, \beta, \tau, \eta, z\}$ denotes latent variables and $\gamma_{ij} = c(z_{ij} - d_i)$. We use the coordinate ascent method to optimize variational parameters in turns by optimizing the lower bound $\tilde{L}(q, \xi)$. Note that there are $n \times m$ entries for latent local variables $z_{ij}, \tau_{ij}, \xi_{ij}$. It is inefficient to learn each entry of those local variables in one iteration. In fact, we pay more attention to learn the global variables $u, v, a, b, d, \beta, \eta$. Thus stochastic variational inference (SVI) [30] can be used to optimize global variational parameters quickly. We can learn something about global variational parameters from only a subset of the data in each step. In Algorithm 1, we present the pseudo-code of our stochastic variational inference algorithm, with details in Appendix.

Algorithm 1 Stochastic Variational Inference of SPMF-MNAR

```
Initialize global variational parameters randomly;
while not converge do
  Subsample a set S of user-item pairs
  Local step:
  for each user-item pair (i, j) ∈ S do
    while not converge do
      Update the parameter ξ_ij based on variational parameters
      of global and local variables. {Equation (15),(16)}
      Update the variational parameters of local variables
      z_ij, τ_ij based on global variational parameters and the
      parameter ξ_ij. {Equation (17)-(19)}
    end while
  end for
  Global step:
  Update the variational parameters of global variables
  u, v, a, b, d, β, η with an appropriate step-size ρ_t based on
  local variational parameters. {Equation (20)-(31)}.
end while
```

Complexity Analysis. The computational time of inference is mainly taken by updating variational parameters. In the local step, we just update the local parameters for the user-item pair (i, j) in the set S. The time to update these local variational parameters is $O(t_l(|S| + |G|))$ (Equation (15)-(19)), where $|S|$ denotes the number of user-item pairs in the set S, $|G|$ denotes the number of edges in the datasets, t_l denotes the number of local iterations in the local step. In the global step, we just sum over the entries in the set S. The computational complexities for updating these global variational parameters are $O((n+m)D^3 + |S| + |G|)$ (Equation (20)-(31)), where n and m denote the number of users and items in the dataset. D denotes the dimension of the latent vector u_i, v_j. Hence, the overall computational complexity in one step is $O((n + m)D^3 + t_l(|S| + |G|))$. In fact, we usually let D, t_l be the fixed small numbers and the number of sampled user-item pairs ($|S|$) be twice as large as the number of observed ratings. Thus, our algorithm is efficient on sparse recommendation data.

D. Prediction

Once the posterior is fit, we can predict the rating value of item j given by user i based on those global variational parameters and the specific value of x_{ij}. we calculate the predicted rating by the expectation of r_{ij} as follow:

$$\hat{r}_{ij} = E[r_{ij}|x_{ij}, \Theta] = E[z_{ij}|x_{ij}, \Theta] \approx E_{q(z_{ij}|x_{ij},\Theta)}[z_{ij}] \quad (9)$$

Here we can perform an additional local step of user-item pair (i, j) to get the posterior expectation of z_{ij}.

Note that MNAR assumption has been modeled in SPMF-MNAR. Thus, we can also make recommendations based on the predicted consumption, which combines user's preference and social influence:

$$\begin{aligned}\hat{x}_{ij} &= E_q[p(x_{ij} = 1)] = \beta_{i0} E_q[p(x_{ij} = 1|z_{ij}, d_i, c)] \\ &\quad + \sum_{k=1}^{|g^i|} \beta_{ik} E_q[p(x_{ij} = 1|\eta_1 x_{g^i_k j} + \eta_0(1 - x_{g^i_k j})] \\ &= \beta_{i0} E_q[\sigma(c(z_{ij} - d_i))] + \sum_{k=1}^{|g^i|} \beta_{ik} E_q[\eta_1 x_{g^i_k j} + \eta_0(1 - x_{g^i_k j})] \\ &\approx \beta_{i0}\sigma(\varphi(s^2_{ij})\alpha_{ij}) + \sum_{k=1}^{|g^i|} \beta_{ik} E_q[\eta_1 x_{g^i_k j} + \eta_0(1 - x_{g^i_k j})] \quad (10)\end{aligned}$$

where we approximate the logistic function with a rescaled probit function that has the same slope at the origin as the logistic function $\sigma(.)$ [32], and $\varphi(x) = (1+\pi x/8)^{-1/2}$, $s^2_{ij} = D_q[c(z_{ij} - d_i)]$, $\alpha_{ij} = E_q[c(z_{ij} - d_i)]$.

V. Experiments and results

In this section, we evaluate our SPMF-MNAR on real-world datasets. We start with the description of the experimental protocol.

A. Experimental protocol

The two datasets, Epinions, Ciao presented in section 3 are used in our experiments. These datasets contain user's consumption, ratings and social relations. A 5-fold cross-validation for learning and testing is used in our experiment where the observed ratings are divided into 5 folds. Specially, when an r_{ij} with $x_{ij} = 1$ is added to the test set, we fix that x_{ij} to zero in the train phase to indicate that the entry r_{ij} is not observed as recent works [7].

The optimal experimental settings for our methods are determined either by the grid search in our experiments or suggested by previous works. Specifically, for hyperparameters, we set $\sigma^2_r = 0.1$,$\sigma^2_d = 0.2$, $D = 10$ and others to 1 accross both datasets, while the performance with varying values of parameter c was presented in Fig. 5. For the parameters of step-size, we refer to [30] and set $\rho_t = (t_0 + t)^{-\vartheta}$,$t_0 = 0, \vartheta = 0.8$. For sampled set S, in each step, we choose all the observed entries with $x_{ij} = 1$ and randomly sample the same number of the entries with $x_{ij} = 0$ to train our model [3].

[3]Source code is available at https://github.com/jiawei-chen/spmf-mnar

B. Evaluation metrics

We adopt the following metrics to evaluate the predictive accuracy and recommendation performance:

Mean Absolute Error (MAE), Root Mean Square Error (RMSE): Two well-known metrics to evaluate the predictive accuracy which are defined as follows:

$$\text{MAE} = \frac{\sum_{i,j} |\hat{r}_{ij} - r_{ij}|}{N} \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i,j} (\hat{r}_{ij} - r_{ij})^2}{N}} \quad (12)$$

where N is the number of observed ratings in the test set, r_{ij} is the observed rating values of item j given by user i and $\hat{r}_{ij}$ is the predicted rating value.

NDCG (Normalized Discounted Cumulative Gain): This is widely used in information retrieval and it measures the quality of ranking through discounted importance based on positions. In recommendation, NDCG is computed as follow:

$$NDCG = \frac{1}{|U|} \sum_{i \in U} \frac{DCG_i}{IDCG_i} \quad (13)$$

where DCG_i is defined as follow and $IDCG_i$ is the ideal value of DCG_i coming from the best ranking.

$$DCG_i = \sum_{i \in Fav(i)} \frac{1}{\log_2(rank_{ij} + 1)} \quad (14)$$

where $rank_{ij}$ represents the rank of the item j in the recommended list of the user i, and $Fav(i)$ denotes the set of favorite items in test data for user i. As recent works [13], highly rated items in the test set will be considered as favorite items (larger than 3 in a 5-star system). NDCG can be interpreted as the ease of finding all favorite items, as higher numbers indicate the favorite items are higher in the list.

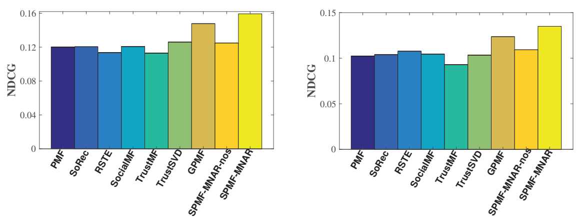
C. Compared methods

To demonstrate the effectiveness of our SPMF-MNAR model, we compared it with the following methods:

- Two baseline Useravg, PMF [7]: Useravg is a non-social simple baseline, which makes use of the average observed ratings of users to predict missing rating values. PMF is a non-social classic probabilistic matrix factorization model on rating data.
- Social recommendation methods without MNAR assumption: SoRec [16], RSTE [21], SocialMF [19], TrustMF [18], TrustSVD [3] are the state-of-the-art methods in social recommendation. Note that MNAR assumption has not been considered by these social recommendation methods.
- Non-social recommendation methods with MNAR assumption: GPMF [6] is a state-of-the-art method with missing not at random data assumption. We also design another method SPMF-MNAR-nos as a comparison to show the effect of modeling social influence on the rating observation process. SPMF-MNAR-nos, a simple version of SPMF-MNAR, leaves out the social influence and generates user's consumption just based on user's preference.

TABLE II
THE CHARACTERISTICS AND PERFORMANCE COMPARISON OF THE COMPARED METHODS. THE BOLDFACE FONT DENOTES THE WINNER IN THAT COLUMN. ALSO WE PRESENT THE RELATIVE IMPROVEMENT ACHIEVED BY THE WINNER OVER THE COMPARISONS.

Methods	Characteristics		Performance on Ciao				Performance on Epinions			
	Social?	MNAR?	MAE	Impv	RMSE	Impv	MAE	Impv	RMSE	Impv
Useravg	\	\	0.7790	10.32%	1.0266	8.12%	0.9293	18.41%	1.2024	14.77%
PMF	\	\	0.7672	8.65%	1.0081	6.17%	0.8503	8.34%	1.0869	3.75%
SoRec	√	\	0.7743	9.65%	0.9965	4.95%	0.8565	9.13%	1.0792	3.01%
RSTE	√	\	0.7679	8.75%	1.0323	8.72%	0.8732	11.26%	1.1093	5.88%
SocialMF	√	\	0.7583	7.39%	0.9924	4.52%	0.8537	8.77%	1.0764	2.74%
TrustMF	√	\	0.7635	8.12%	0.9781	3.01%	0.8519	8.54%	1.0676	1.90%
TrustSVD	√	\	0.7285	3.17%	0.9598	1.08%	0.8132	3.61%	1.0584	1.03%
GPMF	\	√	0.7511	6.37%	0.9830	3.53%	0.8321	6.02%	1.0712	2.25%
SPMF-MNAR-nos	\	√	0.7231	2.40%	0.9639	1.52%	0.8130	3.58%	1.0671	1.86%
SPMF-MNAR	√	√	**0.7061**	-	**0.9495**	-	**0.7848**	-	**1.0476**	-



(a) On dataset Ciao (b) On dataset Epinions

Fig. 4. NDCG of the compared methods.

TABLE III
THE OPTIMAL VARIATIONAL DISTRIBUTIONS OF PARAMETERS η_0 AND η_1, WHERE $B(.)$ DENOTES BETA DISTRIBUTION.

Dataset	$q(\eta_1)$	$q(\eta_0)$
Ciao	$B(7.4 \times 10^3, 1.0 \times 10^0)$	$B(2.2 \times 10^0, 1.8 \times 10^5)$
Epinions	$B(3.1 \times 10^4, 1.0 \times 10^0)$	$B(2.5 \times 10^1, 4.0 \times 10^5)$

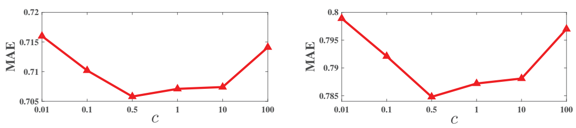
D. Experimental results and analyses

Table II presents the performances of all ten recommendation methods on both datasets, in terms of MAE and RMSE. We observe that SPMF-MNAR consistently outperforms all the other compared methods on both datasets. The improvement of SPMF-MNAR over other methods can be attributed to two aspects: (1) Rating data in social recommendation is missing not at random. If the methods keep the MAR (missing at random) assumption and ignore the dependency between which items a user consumes (rates) and what ratings the user gives, their predictive performance will suffer. This leads to better performance of our method than these methods without MNAR assumption. (2) We integrate social information into the recommendation. Specially, instead of exploiting social influence on users' rating values which has been discussed by substantial social recommendation literatures, our SPMF-MNAR model attempt to exploit social influence on the rating observation process (user's consumption) and observe how it improves the recommendation performance. In fact, our analyses on datasets suggest that in the content-shared social network, connected users often (un)knowingly recommend products to each other and thus their consumption will be heavily influenced by the social relations (Observation 2). As a result, SPMF-MNAR achieves better performance than these non-social methods including the SPMF-MNAR-nos, the special non-social case of SPMF-MNAR.

NDCG comparison. Fig. 4 shows the performance of the compared methods in terms of NDCG. NDCG measures the quality of ranking and the larger values indicate the better performance. We observe that SPMF-MNAR again outperforms other methods, consistent with the findings in Fig. 4. Especially, the relative improvement of SPMF-MNAR over its non-social special case SPMF-MNAR-nos is apparent: 22.3% on dataset Ciao and 19.3% on dataset Epinions in terms of NDCG. These results clearly demonstrate the effectiveness of modeling social influence on user's consumption.

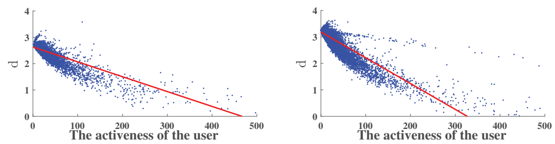
Impact of parameter c. Another experiment is conducted to investigate how parameter c affects the performance of SPMF-MNAR, where c is a scale parameter controlling the influence of user's preference (z_{ij}) on his consumption (x_{ij}). The results in terms of MAE with varying c are presented in Fig. 5. As we can see, as c becomes larger, the performance becomes better first. This is because in social recommendation user's consumption is not independent with user's preference. User's consumption also reflect user's preference on the items. Thus, when we let user's consumption (x_{ij}) be more sensitive to user's preference (z_{ij}) in our model, we can infer more precise user's preference (z_{ij}) reversely based on his consumption information. However, when c surpasses a threshold ($c > 0.5$), the performance becomes worse with further increase of c. As c becomes larger, the inference of user's preference will depend more on his consumption. Other important information including user's real rating values becomes unimportant on the inference of user's preference, which brings the worse results. Thus, when c is set to a appropriate value ($c = 0.5$), when c can best balance the importance of these information, SPMF-MNAR achieves best performance.

Parameter η_0 VS parameter η_1. We also present the optimal variational distribution of parameter η_0 and parameter η_1 as shown in Table III. parameters η_0, η_1 indicate the conditional dependency between user's consumption and the selected friend's consumption. $\eta_1 = p(x_{ij} = 1 | x_{kj} = 1, \tau_{ijk} = 1)$ denotes the probability that the user will follow this selected friend's consumption to consume the item.



(a) On dataset Ciao (b) On dataset Epinions

Fig. 5. Impact of parameter c.



(a) On dataset Ciao (b) On dataset Epinions

Fig. 6. The posterior expectation of acceptive threshold d for different user with varying activeness, where the activeness of the user denotes the number of items consumed by the user. Also, we presented the fitting linear in red.

$\eta_0 = p(x_{ij} = 1|x_{kj} = 0, \tau_{ijk} = 1)$ denotes the probability that the user will consume the item when the friend didn't consume it. From Table III, we can find that the expectation of η_1 is larger than 0.99, while the expectation of η_0 is less than 0.01. This result validates the social conformity phenomenon that the user will follow friend's consumption when he makes the decision on the consumption. Thus, when we model the generative process of user's consumption, we need consider social influence to undo the effect caused by conformity for better inferring user's true preference on the items.

Parameter d for different users. It will be interesting to explore the correlation between the acceptive threshold d and the activeness of the user. Fig. 6 presents the posterior expectation of parameter d for each user with varying activeness, where the activeness denotes the number of items consumed by the user. We can find the more active user corresponds to the smaller d. It can be explained as follow: the users with large acceptive threshold d, tend to be fastidious about the items. These users usually don't accept the items from friends or systems. On the contrary, the generous users with small d, tend to be tolerant towards varied kinds of items. Thus, more items will be consumed by them.

VI. Conclusions

In this paper, we conduct a thorough analysis of real-world social recommendation datasets and find that the observation process of rating data depends on user's preference and social relations. Therefore, we integrate the missing not at random assumption into social recommendation and propose a new recommendation method SPMF-MNAR with a probabilistic generative model of the rating observation process. Different from existing social recommendation methods that exploit social influence on users' rating values, we focus on exploiting the social influence on the rating observation process. Our comprehensive experimental results clearly demonstrate the effectiveness of our proposed method compared to both existing social recommendation methods and the non-social methods based on MNAR.

In the future, it will be interesting to exploit more factors which contribute to user's consumption, such as item's popularity and user's location. We can extend our SPMF-MNAR to model these factors for better recommendation performance.

Acknowledgments

This work is supported by Zhejiang University-LianlianPay Joint Research Center and Zhejiang Provincial Key Research and Development Plan (Grant no. 2017C01012).

Appendix

Here we describe the details of how we optimize the lower bound $\tilde{L}(q, \xi)$ and how we update the variational parameters in the global step and local step.

Local step. As is shown in algorithm 1, in each step we firstly subsample a set S of user-item pairs. Then, we iteratively update the optimal parameter ξ_{ij} and variational parameters of local variables z_{ij}, τ_{ij} for each user-item pair $(i, j) \in S$ as follows:

$$\xi_{ij} \leftarrow c\sqrt{(\Lambda^z_{ij})^{-1} + (\mu^z_{ij})^2 - 2\mu^d_i\mu^z_{ij} + (\Lambda^d_i)^{-1} + (\mu^d_i)^2} \quad (15)$$

$$\lambda_{ij} \leftarrow \frac{1}{2\xi_{ij}}(\sigma(\xi_{ij}) - \frac{1}{2}) \quad (16)$$

$$\Lambda^z_{ij} \leftarrow \frac{1}{\sigma^2_z} + \frac{1}{\sigma^2_r} + 2\lambda_{ij}c^2\varphi_{ij0} \quad (17)$$

$$\mu^z_{ij} \leftarrow (\Lambda^z_{ij})^{-1}(\frac{\Omega_{ij}}{\sigma^2_z} + \frac{r_{ij}}{\sigma^2_r} + \varphi_{ij0}(c(2x_{ij} - 1)/2 + 2\lambda_{ij}c^2\mu^d_i)) \quad (18)$$

$$\begin{aligned} &\varphi_{ijk} \propto \exp(\Psi(\kappa^\beta_{ik}) + \theta_{ijk}) \\ &\theta_{ij0} = \ln(\sigma(\xi_{ij})) + \frac{(2x_{ij} - 1)c(\mu^z_{ij} - \mu^d_i) - \xi_{ij}}{2} \\ &\theta_{ijk} = \Psi({e_{ij}}^\top\kappa^\eta f_{ij}) - \Psi(\|\kappa^\eta f_{ij}\|_1), k = 1, 2, \ldots \left|g^i\right| \end{aligned} \quad (19)$$

where Ω_{ij} denotes the predicted average rating based on latent features of users and items $\Omega_{ij} = (\mu^u_i)^\top(\mu^v_j) + \mu^a_i + \mu^b_j$; $e_{ij} = [x_{ij}, 1-x_{ij}]$, $f_{ij} = [1-x_{g^i_k j}, x_{g^i_k j}]$ are two-dimensional binary vector to indicate which part of κ^η contributes to the evaluation of parameter φ_{ij}; $\|.\|_1$ is a $l1$ norm; $\Psi(.)$ is Digamma function, the first derivative of the log Gamma function. We find the good convergence of the local iteration in local step and fix the number of the local iteration at 10.

Global step. Here we first compute the natural gradient [33] of the $\tilde{L}$ on the global variational parameters $\Xi \equiv \{\mu^u_i, \mu^v_j, \mu^d_i, \mu^a_i, \mu^b_j, \Lambda^u_i, \Lambda^v_j, \Lambda^d_i, \Lambda^a_i, \Lambda^b_j, \kappa^\beta_i, \kappa^\eta_l\}$ to get the intermediate optimal global parameter $\hat{\Xi}$ as follow:

$$\hat{\Lambda}^u_i \leftarrow \frac{1}{\sigma^2_z}\sum_{j\in I,(i,j)\in S}((\Lambda^v_j)^{-1} + \mu^v_j(\mu^v_j)^\top) + \frac{1}{\sigma^2_u}\mathbf{I} \quad (20)$$

$$\hat{\mu}^u_i \leftarrow (\Lambda^u_i)^{-1}\frac{1}{\sigma^2_z}\sum_{j\in I,(i,j)\in S}(\mu^z_{ij} - \mu^a_i - \mu^b_j)\mu^v_j \quad (21)$$

$$\hat{\Lambda}^v_j \leftarrow \frac{1}{\sigma^2_z}\sum_{i\in U,(i,j)\in S}((\Lambda^u_i)^{-1} + \mu^u_i(\mu^u_i)^\top) + \frac{1}{\sigma^2_v}\mathbf{I} \quad (22)$$

$$\hat{\mu}^v_j \leftarrow (\Lambda^v_j)^{-1}\frac{1}{\sigma^2_z}\sum_{i\in U,(i,j)\in S}(\mu^z_{ij} - \mu^a_i - \mu^b_j)\mu^u_i \quad (23)$$

$$\hat{\Lambda}^a_i \leftarrow \frac{1}{\sigma^2_z}\sum_{j\in I,(i,j)\in S}1 + \frac{1}{\sigma^2_u} \quad (24)$$

$$\hat{\mu}_i^a \leftarrow (\Lambda_i^a)^{-1} \frac{1}{\sigma_z^2} \sum_{j \in I, (i,j) \in S} (\mu_{ij}^z - (\mu_i^u)^\top (\mu_j^v) - \mu_j^b) \tag{25}$$

$$\hat{\Lambda}_j^b \leftarrow \frac{1}{\sigma_z^2} \sum_{i \in U, (i,j) \in S} 1 + \frac{1}{\sigma_v^2} \tag{26}$$

$$\hat{\mu}_j^b \leftarrow (\Lambda_j^b)^{-1} \frac{1}{\sigma_z^2} \sum_{i \in U, (i,j) \in S} (\mu_{ij}^z - \mu_i^a - (\mu_i^u)^\top (\mu_j^v)) \tag{27}$$

$$\hat{\kappa}_i^\beta \leftarrow \alpha + \sum_{j \in I, (i,j) \in S} \varphi_{ij} \tag{28}$$

$$\hat{\kappa}_l^\eta \leftarrow \zeta + \sum_{(i,j) \in S, k=1,2\ldots|g_i|} e_{ij} \varphi_{ijk} \mathrm{I}[x_{g_k^i j} = l] \tag{29}$$

$$\hat{\Lambda}_i^d \leftarrow \frac{1}{\sigma_d^2} + \sum_{j \in I, (i,j) \in S} 2\varphi_{ij0} \lambda_{ij} c^2 \tag{30}$$

$$\mu_i^d \leftarrow (\Lambda_i^d)^{-1} \Big(\frac{m_d}{\sigma_d^2} + \sum_{i \in U, (i,j) \in S} \varphi_{ij0} \Big(2\lambda_{ij} c^2 \mu_{ij}^z - \frac{c(2x_{ij} - 1)}{2}\Big)\Big) \tag{31}$$

where I[.] is the binary indicator function. Having these intermediate optimal global parameter $\hat{\Xi}$, we can update the global parameters Ξ with an appropriate step-size ρ_t as follow:

$$\Xi \leftarrow (1 - \rho_t)\Xi + \rho_t \hat{\Xi} \tag{32}$$

REFERENCES

[1] S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha, "Like like alike: joint friendship and interest propagation in social networks," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 537–546.

[2] T. Zhao, J. McAuley, and I. King, "Leveraging social connections to improve personalized ranking for collaborative filtering," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 2014, pp. 261–270.

[3] G. Guo, J. Zhang, and N. Yorke-Smith, "Trustsvd: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings," in *AAAI*, 2015.

[4] B. M. Marlin and R. S. Zemel, "Collaborative prediction and ranking with non-random missing data," in *Proceedings of the third ACM conference on Recommender systems*. ACM, 2009, pp. 5–12.

[5] B. M. Marlin, R. S. Zemel, S. Roweis, and M. Slaney, "Collaborative filtering and the missing at random assumption," in *UAI*. AUAI Press, 2007, pp. 267–275.

[6] S. Ohsawa, Y. Obara, and T. Osogami, "Gated Probabilistic Matrix Factorization : Learning Users ' Attention from Missing Values," *Ijcai*, pp. 1888–1894, 2016.

[7] J. M. Hernández-Lobato, N. Houlsby, and Z. Ghahramani, "Probabilistic matrix factorization with non-random missing data." in *ICML*, 2014, pp. 1512–1520.

[8] B. Pradel, N. Usunier, and P. Gallinari, "Ranking with non-random missing ratings: influence of popularity and positivity on evaluation metrics," in *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 2012, pp. 147–154.

[9] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, 2009.

[10] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *NIPS*, 2008, pp. 1257–1264.

[11] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 287–296.

[12] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1267–1275.

[13] X. Wang, W. Lu, M. Ester, C. Wang, and C. Chen, "Social recommendation with strong and weak ties," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 5–14.

[14] L. Xiao, Z. Min, Z. Yongfeng, L. Yiqun, and M. Shaoping, "Learning and transferring social and item visibilities for personalized recommendation," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2017, pp. 337–346.

[15] M. Jamali and M. Ester, "Trustwalker: a random walk model for combining trust-based and item-based recommendation," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 397–406.

[16] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: social recommendation using probabilistic matrix factorization," in *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008, pp. 931–940.

[17] Y. Shen and R. Jin, "Learning personal+ social latent factor model for social recommendation," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1303–1311.

[18] B. Yang, Y. Lei, J. Liu, and W. Li, "Social collaborative filtering by trust," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 8, pp. 1633–1647, 2017.

[19] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 2010, pp. 135–142.

[20] X. Wang, S. C. Hoi, M. Ester, J. Bu, and C. Chen, "Learning personalized preference of strong and weak ties for social recommendation," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 1601–1610.

[21] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2009, pp. 203–210.

[22] J. Tang, H. Gao, and H. Liu, "mtrust: discerning multi-faceted trust in a connected world," in *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 2012, pp. 93–102.

[23] A. J. Chaney, D. M. Blei, and T. Eliassi-Rad, "A probabilistic model for using social networks in personalized item recommendation," in *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 2015, pp. 43–50.

[24] Y. Bao, H. Fang, and J. Zhang, "Leveraging decomposed trust in probabilistic matrix factorization for effective recommendation," in *AAAI*, 2014, p. 350.

[25] E. Tulving, "Episodic memory: From mind to brain," *Annual review of psychology*, vol. 53, no. 1, pp. 1–25, 2002.

[26] U. M. Dholakia and D. Talukdar, "How social influence affects consumption trends in emerging markets: An empirical investigation of the consumption convergence hypothesis," *Psychology & Marketing*, vol. 21, no. 10, pp. 775–797, 2004.

[27] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of jaccard coefficient for keywords similarity," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, no. 6, 2013.

[28] J. J. Louviere and G. Woodworth, "Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data," *Journal of marketing research*, pp. 350–367, 1983.

[29] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[30] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 1303–1347, 2013.

[31] T. S. Jaakkola and M. I. Jordan, "Bayesian parameter estimation via variational methods," *Statistics and Computing*, vol. 10, no. 1, pp. 25–37, 2000.

[32] J. M. Hernández-Lobato, N. Houlsby, and Z. Ghahramani, "Stochastic inference for scalable probabilistic modeling of binary matrices," in *International Conference on Machine Learning*, 2014, pp. 379–387.

[33] A. Honkela, M. Tornio, T. Raiko, and J. Karhunen, "Natural conjugate gradient in variational inference," in *International Conference on Neural Information Processing*. Springer, 2007, pp. 305–314.