Knowledge-Based Systems 140 (2018) 120–133

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Improving user recommendation by extracting social topics and interest topics of users in uni-directional social networks☆

Ke Xu[a], Xushen Zheng[b], Yi Cai[a,*], Huaqing Min[a], Zhen Gao[c], Benjin Zhu[a], Haoran Xie[d], Tak-Lam Wong[e]

[a] South China University of Technology, Guangzhou, China
[b] Carnegie Mellon University, Pittsburgh, United States
[c] Alibaba Group Holding Limited, Guangzhou, China
[d] The Education University of Hong Kong, Hong Kong, China
[e] Douglas College, New Westminster, British Columbia, Canada

A R T I C L E I N F O

Article history:
Received 25 April 2017
Revised 23 October 2017
Accepted 26 October 2017
Available online 27 October 2017

Keywords:
User recommendation
Topic modeling
Generalized Pólya Urn model
Matrix factorization

A B S T R A C T

With the rapid growth of population on social networks, people are confronted with information overload problem. This clearly makes filtering the targeted users a demanding and key research task. Uni-directional social networks are the scenarios where users provide limited *follow or not* binary features. Related works prefer to utilize these *follower-followee* relations for recommendation. However, a major problem of these methods is that they assume every *follower-followee* user pairs are equally likely, and this leads to the coarse user following preferences inferring. Intuitively, a user's adoption of others as followees may be motivated by her interests as well as social connections, hence a good recommender should be able to separate the two situations and take both factors into account for better recommendation results. In this regard, we propose a new user recommendation framework namely *UIS-MF* in this work. *UIS-MF* can well capture user preferences by involving both interest and social factors in prediction, and targeted to recommend Top-*N* followees who have similar interest and close social connection relevant to a target user. Specifically, we first present a unified probabilistic topic model on *follower-followee* relations, namely *UIS-LDA*, and it employs Generalized Pólya Urn (GPU) models on *mutual-following* relations for discovering interest topics and social topics of users. Next we propose a community-based method for user recommendation, it organizes social communities and interest communities based on the estimation of topics obtained from *UIS-LDA*, and then performs Matrix Factorization (MF) method on each community to generate *N* most likely followees for individual user. Systematic experiments on Twitter, Sina Weibo and Epinions datasets have not only revealed the significant effect of our *UIS-LDA* model for the extraction of interest and social topics of users in improving recommending accuracy, but also demonstrated the advantage of our proposed recommendation framework over competitive baselines by large margins.

1. Introduction

Nowadays advances in uni-directional social networks, such as Twitter[1] and Sina weibo,[2] which enable social interaction and information exchanging, have sharply increased. They allow users to follow any other user without approval and post tweets substance-in 140 characters or less. The visibilities of tweets are triggered by *follower-followee* relations, that is a follower will receive all the tweets from her followees. As a result of the rapidly increasing number of population, more than 300 million monthly active users generate 500 million tweets per day, most Twitter users encounter a serious problem of information overload. Therefore, recommending relevant users so as to alleviate the flooding of information brought by irrelevant followees is a significant challenge and also what we concerned with in this paper.

Collaborative Filtering (CF) and topic modeling algorithms (topic models) have proven to be effective for user recommendation tasks

☆ This article is an extended work of a conference proceeding, which has been published in the 2017 International Conference on Web Intelligence [1].

* Corresponding author.
E-mail addresses: kexu@scut.edu.cn (K. Xu), xushenz@andrew.cmu.edu (X. Zheng), ycai@scut.edu.cn (Y. Cai), hqmin@scut.edu.cn (H. Min), condtend@aliyun.com (Z. Gao), sepoodar@mail.scut.edu.cn (B. Zhu), hxie@eduhk.hk (H. Xie), wongi5@douglascollege.ca (T.-L. Wong).

[1] http://www.twitter.com.
[2] http://www.weibo.com.

https://doi.org/10.1016/j.knosys.2017.10.031

and receiving a lot of attention. They are methods of making predictions about a user's preferences commonly base on utilizing *follower-followee* relations. However, a major problem of these methods is that they assume every user pair with *follower-followee* relations are equally likely. For example, they cannot clearly separate the two situations: you are following someone who isn't following you (namely *uni-following*), and you are *mutual-following* with someone. In fact, each situation can be viewed as an indicator for a user's preferences at a different granularity. The former situation has a higher probability to indicate individual interest. It allows a user collect information or to be entertained through following anyone whose interest is aligned with her, and usually they may not know each other offline; The latter situation is more likely to infer the social connection of users [2]. People prefer to establish *mutual-following* relations among their social circle since they'd like to maintain and strengthen existing social ties online. In these regards, a user's interests and social connections can both provide valuable clues about her preferences. Therefore, it is worthwhile to take both factors into account when making recommendations. Failure to distinguish whether one follows the other based on her interest or social connection by the roughly equal use of every *follower-followee* relations, results in the poor quality of topics discovered from topic models or unreasonable similarity calculations from CF, and thus leading to suboptimal recommendation.

In this paper, we try to improve user recommending task through the identification of a user's preferences in a fine granularity level via both interest and social connection. That is, we recommend candidate users to a target user base on similarity in interest and close social connection. Specifically, we rely purely on binary *follow or not* relevance data from uni-directional social networks. There are three main questions: (1) What relations can benefit us for indicating a user's interest and social connection? (2) How to model a user's preferences with these critical relations? (3) How to enhance the accuracy of state-of-the-art user recommendation works by using proposed model? Providing solutions to these questions results in a novel recommendation framework *UIS-MF*. As its core, a topic model is designed to jointly model a user's interest and social connection, and a community-based method is devised for making user recommendation bases on this topic level extraction. The contributions are summarized below:

- In view of the limitations of previous works that they roughly model a user's preferences, we propose a new user recommendation framework called *UIS-MF* in this work. It can involve both social connection factors and interest factors in predicting a user's preferences, and thus offering analytical benefits and improving user recommendation accuracy on uni-directional social networks. Particularly, since only *follow or not* information is required, proposed *UIS-MF* is more practical in lots of recommendation scenarios.
- In *UIS-MF*, we present a novel probabilistic topic model *UIS-LDA*, which is built with *follower-followee* and *mutual-following* relations. This model has the advantage of organizing a user's preferences related to social topics and interest topics respectively. In particular, it enables a *Bernoulli* switch to decide the kind of topics, and leverages Generative Pólya Urn (short for GPU) model [3] in sampling process to promote *mutual-following* users to be observed under social topics. Our evaluation shows that the *UIS-LDA* model significantly improves user recommending effectiveness, and outperforms topic models that discovery interest topics or social topics alone. To the best of our knowledge, this is the first time to incorporate users' social connection and interest into a coherent model, as well as introduce GPU model into social networks. Further, the use of GPU model shed some light on how we can build topic level models to depict the different relations of users on social networks.
- After topic modeling, we propose a community-based method for our top-*N* user recommendation task. This method uses the variables of *UIS-LDA* as community membership measures. The memberships are fed into the community formation equations in order to select qualified users to separately build related interest communities of users and social communities. After that, Matrix Factorization (MF) algorithm is performed on each community to obtain a ranked list of followees for each follower. Our evaluation shows the data sparsity is alleviated by forming communities and the ability of MF for recommendation is remarkable.
- We conduct extensive experiments using real-world Twitter, Sina Weibo and Epinions datasets to verify the effectiveness of our recommendation framework. Our evaluation demonstrates that the proposed *UIS-MF* produces considerable improvements in F1 Score, Conversation Rate, Precision, Recall, NDCG over the baseline methods. In particular, our experiments on the three datasets suggest that the proposed framework is very general, it not only works well with user follow relationships, but also performs well with user trust relationships.

The remainder of the paper is structured as follows. We review related work in Section 2, give the motivation and basic assumption of our work in Section 3, detail our approach in Section 4, evaluations are represented in Section 5, describe and discuss the result of our experimentation in Section 6, and conclude this paper in Section 7.

2. Related work

The boom of social networks fosters the research on user preferences analysis and recommendation. User preferences data that is used for recommendation can be partitioned into two types: explicit feedback (typically ratings) and implicit feedback. Most of the available implicit data is binary (e.g. follow or not, consume or not) where users conducts the action on a given item/user. In this section, we introduce a series of related work ranging from MF, topic models, GPU model, community-based recommendation methods and discuss the relation between proposed *UIS-MF* and *CB-MF* [4].

2.1. MF-based approaches

CF approach utilizes the wisdom of crowds and has achieved great success in recommending area [5,6]. It can be broadly divided into two categories: neighbor based (NB) algorithms [7] and latent factor models (LFMs). NB algorithms calculate user preferences similarities whereas LFMs learn latent factors [1]. MF method is one of the most successful realizations of LFMs for recommender systems and it's superior to NB algorithms [8]. The basic idea of MF is it factorizes the user-item rating matrix into two low rank user-specific and item-specific matrices, and making predictions for unknown ratings heavily relying on these two matrices. Probabilistic Matrix Factorization (PMF) [9] is the pioneer algorithm of MF which is suitable for rating predictions with observed ratings.

While a lot of the MF literature has been devoted to recommendation scenarios with explicit feedback, MF has also shown to be very valuable in scenarios with implicit feedback [10–13]. *IF-MF* [10] and *BPR-MF* [11] have emerged as two state-of-the-art MF extensions for implicit feedback scenarios. *IF-MF* algorithm optimizes to predict if an item is selected or not. Within *IF-MF*, each user-item preference is coupled with a confidence level and its latent factors are computed by a cost function. And the confidence levels are considered as weighting variables. *BPR-MF* deals with binary relevance data and uses MF as the learning model. It utilizes a Bayesian probabilistic optimization method which takes into account all pairwise preferences among relevant and irrelevant items partially ordered relations to reconstructs the training data. *CLiMF* [12] is also tailored the scenarios for binary relevance data. Unlike *BPR-MF, CLiMF* optimizes Mean Reciprocal Rank (MRR optimization) [14] instead of Area Under Curve (AUC). This difference especially benefits *CLiMF* on ranking of top items. Work [13] proposes to weight missing data based on the popularity of items and designs a new learning algorithm for efficiently optimizing the MF model. However, data sparsity is one of the most serious limitations of these aforementioned MF approaches.

Researchers have started to exploit various information to better shape the MF's latent spaces, alleviate the data sparsity problem and hence improve recommendation accuracy. For instance, Authors in [15] fuse user-item ratings, explicit social relations and common neighbors information together based on PMF to exploit rating predictions. Ref. [16] studies a research problem on how to improve recommender systems using implicit social information. It defines the implicit user/item social information as the similar or dissimilar users/items, where user-item ratings are used to measure the similarity/dissimilarity between users/items. And a very general MF framework is employed to incorporate different implicit social information. UIContexRank [17] is a pairwise learning to recommendation model. It integrates users' friendship or trust relation and items' concurrent rated information to predict users' preferences between item pairs. A two phase trust-aware recommendation process [18] is proposed to utilize deep learning in MF's initialization and to synthesize the users' interests and their trust friends' interests together with the impact of community effect based on MF for recommendations. In [19], authors exploit locations, chek-in information, etc. and intergrades them into a MF model for better rating prediction. Volkovs and Yu [20] enriches the binary feedback matrix with neighbor-based similarly, followed by applying SVD. However, as mentioned, the above recommendation methods rely heavily on additional information, which may limit their impact and utilization.

In this subsection, we final mention some sophisticated and thrilling related MF work [21–23]. For example, The RSNMF recommender [21] is a recent promising model. It investigates a non-negative update process depending on each involved feature parameter rather than on the whole feature matrices (SNMF). With SNMF, it subsequently integrate the Tikhonov regularizing terms. The integration of the Tikhonov regularizing terms focus on developing an non-negative update process depending on each involved feature. The experiments well demonstrate that RSNMF can obtain high efficiency and accuracy under constraint of non-negative. Authors in [22] implement a highly efficient CF-based approach to binary interactome mapping. They first propose a CF framework for it. Under this framework, they model the given data into an interactome weight matrix, where the feature-vectors of involved proteins are extracted. With them, they design the rescaled cosine coefficient to model the inter-neighborhood similarity among involved proteins, for taking the mapping process. A novel UOCFF [24] approach based on *CLiMF* and PMF is developed that simultaneously optimizes both the ratings and rank of the recommended items.

2.2. Topic model and GPU model approaches

In another direction, various topic models, especially the adaption of the original LDA [25] models have been proposed and applied for social networks. *CB-MF* [4] utilizes LDA for clustering users into communities to enhance the existing MF-based user recommendation. A LDA-based model [26] is designed to group users based on *follower-followee* relations for handling popular users. Work in [27] presents a topic model to discover user-oriented and community-oriented topics simultaneously for recommending users and communities. Recommendation system [28] adopts topic models at the user-level where documents are replaced by users' streams and recommends users that have a distribution highly similar to a target user. LDA is used in [29] to mine interests of users based on ratings and tags and generate recommendation based on the clustering results. A Bayesian model developed in [30] benefits from both the analysis of the latent structure of signed social networks and the propagation of distrust for link prediction and rating prediction. In [31], LDA is applied on tweet content and extract topic-level distributions of items. Ref. [32] first uses topic model to analysis users' repost behaviors and measure the topic-specific relationship strength, then incorporates the relationship factor into the MF framework, their task is for microblog recommendation. However, none of these works consider integrating a user's interest and social connection together for making user recommendation.

Pólya Urn model [33] is a famous statistic model and GPU is an effective variation. Our work is also related to works [3,34–37] since we introduce GPU model into our topic modeling algorithm. In work [3], GPU is used to smooth the probabilities of words with co-document frequencies for generating coherent topics. Authors in [34] propose to use GPU on word to phase and phase to word level. Paper [35] is an enhanced work of [36], it utilizes GPU to promote words that sharing some similar semantic meaning. Work in [37] propose a topic model for short texts, named GPU-DMM. GPU-DMM promotes the semantically related words by using the GPU model. Apparently, these GPU efforts are all applied in the domain of text processing. In this paper, GPU is applied in social networks for the first time and is used to improve state-of-the-art topic models. Interest and social connection of users are respectively seized as interest topics and social topics which can be effectively generated by proposed *UIS-LDA* topic model.

Also, except for topic model, there are other methods can be used to class users with their preferences. Work on [38] models user behaviors to detect segments of users to target and to whom address ads. It presents an approach to segment the users by analyzing the items positively evaluated by them, in order to consider reliable user preferences. These items are analyzed by extracting the word embeddings and by building from Neural Class Embedding. Ref. [39] presents a sentiment classification system for social media (such as Twitter) genres. It introduces an approach to hybridise a general purpose lexicon, SentiWordNet, with genre-specific vocabulary and sentiment to improve classification accuracy.

2.3. Community-based recommendation approaches

Another line of related approaches is community-based recommendation. *CB-MF* employs a community-based approach to user recommendation in Twitter-style social networks. Ref. [4] is a typical work for such approach. The work suggested that forming communities benefit us to reduce data sparsity and focus on discovering the latent characteristics of communities instead of individuals. Work on [40] proposes a community-based user domain model collaborative recommendation algorithm (CUCRA). This algorithm maps a user-item data set to a user-user social network based only on user-item preference data. It then finds user similar preference communities to define a user domain model by detecting communities on a user-user social network. Finally it makes memory-based recommendations in the community-based user domain model. This algorithm needs rating and tag information. Research [41] applies K-means algorithm to cluster similar users. The similarity of every user pair is calculated by Euclidean distance on ratings matrix. The new user belongs to the most similar cluster and the average ratings of the cluster are recommended to the new user. Jindal and Anjali [41] also proves the superiority to CUCRA experimentally. Ref[23]. extends topology based approach and Friends of Friends (FOF) algorithm to recommend users by using the information related to users (user actions and mentions) and by applying the classifier method.

2.4. Relation to existing work

We differs from the previous work because (1) we focus on recommending targeted users on uses only the binary relevance data of *follow or not* relevance data. While most of existing work probing into the text of tweets, user-item ratings, or other information like tag, mention, retweet, etc. Obviously, due to the follow or not feedback that ease to be obtained, our method is more practical in lots of scenarios. (2) we focus on rank top-*N* candidate followees rather than rating, while ratings have been so far mainly used to improve item recommendation [30] not user recommendation.

CB-MF is the most similar work to our *UIS-MF* in terms of recommending process and application scenarios, but we are substantially different. First of all, *UIS-MF* is a framework for recommending users based on their similarity in interest and social connection, whereas *CB-MF* recommends users with similar influence and interest; apart from this, while *CB-MF* applies LDA on coarse *follower-followee* relations, we propose a novel probabilistic topic model *UIS-LDA* on *follower-followee* and *mutual-following* relations to organize a user's preferences related to social topics and interest topics respectively. Particularly, GPU is used in *UIS-LDA*'s sampling process; Furthermore, in contrast to *CB-MF*, which maps both followers and followees into the same latent space, *UIS-LDA* only maps followers into two type of topics, this leads to the different strategy of communities construction. Also we demonstrate superiority of our *UIS-MF* over *CB-MF* experimentally.

3. Motivation and basic assumption

To improve user recommendation, we initially need to know what spurs us to follow someone on uni-directional social networks. A large number sociology and psychology research have investigated into the nature of the social networks that matter to people. For example, work in [42] investigates two motivation in twitter usage, one is social-relational motivation that centers on building and maintaining interpersonal networks, the other is less social that gravitated towards personal interest such as acquisition of information and entertainment. Similarly, work in [43] explicates two functions of Twitter: one is helping strengthen existing social circles offline, because users would be more inclined to replicate social relationship offline in network platform; the other is helping users establish new *follower-followee* relations probably rely common interest. Moreover, authors in Ref. [44] investigate the correlations between social friend and user interest similarity in the context of recommender systems. They find social friend relationships generally cannot represent user interest similarity, in other words, users may not share similar taste with most of their friends. Following the same idea, we seek to estimate user preferences in a fine-grained level, involving both interest and social connection, to obtain better recommending performance. This becomes the major motivation of our work. Noted that social connection refers to the cooperative and mutual relationships that surround us [44], such as friends, colleagues, classmates, relatives or neighbors, etc.

As discussed, recommending users to users requires not only what we understand their interest, but also that we understand their social connection with others. Related recommender works [4,26,45,46] prefer to explore *follower-followee* relations for recommendation and find that they are dominant features to indicate a user's preferences [46]. No matter what techniques are developed, the basic assumption employed in these work is that every *follower-followee* user pairs are equally likely. However, this hypothesis often don't hold. Intuitively, the *follower-followee* relations cover two organized states: one is for *uni-following* state, which depicts users who are being followed by other users do not reciprocate by following them back; the other is for *mutual-following* state, which represents mutual connection between users. Authors in [2] find that *mutual-following* is a signal for real life social connection. Information can flow in both directions between two users, this indeed promotes offline social connectors share more direct two-way follow links with each other. Whereas, in *uni-following* situation, where a one-way information flow from the followee to the follower, this do not offer strong evidence for offline connections. Work in [47] suggests mutuality affords users the ability to engage in conversation, which may also be a proxy for existing relational strength. Papers [48,49] figure that mutual directed communications are useful for maintaining existing ties and for encouraging the growth of new ones. Yoo et al. [50] also suggest that close friends use Twitter to share personal information or feelings by following each other. Encouraged by these research, we may perceive that *mutual-following* may help to enhance more intimacy and credibility relations than *uni-following* relations, which in turn attracts more social connected people. Therefore, we explore *mutual-following* relations to give us a hand as we take the following assumption:

Assumption 1. People socially connected offline are tend to follow each other online.

Based on this assumption, we can consider that the *mutual-following* relation creates a better social connection inferring to us. That is, *mutual-following* users are more likely motivated by their social ties, whereas, *uni-following* users are more likely motivated by their common interest. The advantage of utilizing *mutual-following* relations can not only help us enrich our model to infer a user's social connection, but also purify individual interest that inferred by rough *follower-followee* relations.

4. Our recommendation framework

Our intuition in this work is that by treating every *follower-followee* relation equally likely, we would obtain rough preferences inferring and inaccurate personalized user recommendation. To motivate this work, we define the scope of our recommendation framework. Given a target user and *follow or not* information of users, our goal is to rank candidate users matching her interest and social connection at top-*N* positions. To achieve this goal, we need to model interest as well as social connection of users from these limited binary relevance data and propose user recommendation rules. Technically, the proposed recommendation framework *UIS-MF* comprises two main phases. Phase 1 (See Section 4.1) is for topics extraction. We present a *UIS-LDA* model to discover interest topics of users and social topics of users, where both *follower-followee* relations and *mutual-following* relations as explored as user preferences indications. In Phase 2 (See Section 4.2), a community-based method for user recommendation is proposed.

4.1. Phase 1: topics extraction

In this work, we learn to know each user's preferences with respect to the set of latent interest topics and social topics, and how to extract both types of topics is the fundamental and critical problem to be solved. Topic modeling algorithms is a common approach for issues alike, which works like a cluster that categorizes documents to latent topics. LDA is one of the most

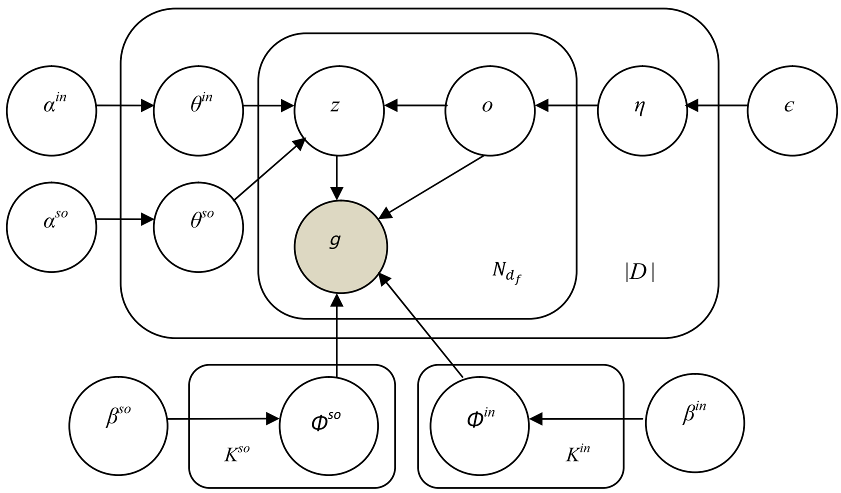


Fig. 1. The plate notation for *UIS-LDA*.

advanced algorithms for topic modeling. Though a bunch of LDA variants [4,26–28] have been successfully applied on social networks, none of them can reach our goal. The major reason is for the limitation of LDA itself that disregarding the relations of words. This line of thinking leads to the roughly equal treatment of *mutual-following* users with *uni-following* users in our scenario. To overcome this weakness, inspired by work [3], we incorporate GPU model into our topic model's sampling process for appropriately depicting users with *mutual-following* relations.

In this subsection, we first describe our proposed topic model *UIS-LDA* and its generative scheme. To learn variables, we then train the *UIS-LDA* model with GPU model and Collapsed Gibbs sampling [51].

4.1.1. UIS-LDA *model*

Each document in LDA model is represented as a random mixture over a given number of topics, and each topic is characterized by a distribution over bag-of-words. As presented in our earlier work [1], we regard each followee $g \in G$ as a word, every follower $f \in F$ as a document containing all her followees, and the graphic model for *UIS-LDA* in plate notation is showed in Fig. 1. The predefined K topics are classified into K^{in} interest topics and K^{so} social topics. Z^{so} is the set of all social topics. Z^{in} is the set of all interest topics. Followers set F , followees set G , a document d_f and documents corpus D are formulated as follows:

$$F = \{f | f \in U \wedge \exists (g \in U \wedge e(f, g) \in E)\} \tag{1}$$

$$G = \{g | g \in U \wedge \exists (f \in U \wedge e(f, g) \in E)\} \tag{2}$$

$$d_f = \{g | g \in G \wedge \exists e(f, g) \in E\} \tag{3}$$

$$D = \bigcup_{f \in F} d_f \tag{4}$$

where U is the set of users, E is the set of user pairs, and each $e(f, g) \in E$ indicates a *follower-followee* relation from follower f to followee g. We also denote that $|D|$ is the number of documents, $|G|$ is the number of followees, and each d_f has N_{d_f} followees. Particularly, we add a latent *Bernoulli* variable o as a switch to indicate the corresponding kind of topics, a social topic z^{so} given $o = 0$ and otherwise an interest topic z^{in} given $o = 1$. θ^{in} with *Dirichlet* prior α^{in} depicts the distribution of per-document on K^{in} interest topics; ϕ^{so} with *Dirichlet* prior β^{so} captures the proportion of per-followee is assigned from social topics Z^{so}; η with *Beta* prior ε represents the distribution of per-document over o. Z is the set of all topics, where $Z = Z^{so} \cup Z^{in}$. Given hyper-parameters α^{so}, α^{in}, β^{so}, β^{in}, and ε, the generative scheme is shown in Algorithm 1.

Algorithm 1: Generative scheme of *UIS-LDA* model.

input: Document D; Topic number K, and the ratio of K^{in} to K^{so} denoted by $\hat{R}$; Hyper parameters α^{in}, α^{so}, β^{in}, β^{so}, ε; The set of *mutual-following* users by user g denoted by T_g and the social promotion matrix $\tilde{A}$
output: θ^{in}, θ^{so}, ϕ^{in}, and ϕ^{so}.

```
for each topic z ∈ Z do
    if z ∈ Z^so then
        Draw φ^so ~ Dirichlet(β^so)
    else
        Draw φ^in ~ Dirichlet(β^in)
    end if
end for
for each document d_f ∈ D do
    Draw η ~ Beta(ε)
    Draw θ^so ~ Dirichlet(α^so)
    Draw θ^in ~ Dirichlet(α^in)
    for each followee g in d_f do
        Draw a switch o ~ Bernouli(η)
        if o == 0 then
            Draw a social topic assignmentz^so ~ Multinomial(θ^so)
            Employ GPU model to promote users with
            mutual-following relations(Detailed in Section 4.1.2).
            Draw a followee g ~ Multinomial(φ^so, z^so)
        else
            Draw an interest topic assignmentz^in ~
            Multinomial(θ^in)
            Draw a followee g ~ Multinomial(φ^in, z^in)
        end if
    end for
end for
```

4.1.2. Sampling and variables inference

To estimate the model, we use the Collapsed Gibbs sampler. Particularly, we employ GPU model in the sampling process given $o = 0$ to encode the characteristic implied by Assumption 1, that is, if a user is being observed under a social topic, it is reasonable to expect a higher likelihood of observing any of her *mutual-following* users under the same topic.

Let T be the set of all *mutual-following* user pairs. Each $t(g, g') \in T$ is referring to a *mutual-following* relation between followees g and followee g', formally, $\exists(e(g, g') \in E \wedge e(g', g) \in E)$. We use T_g to denote the set of g's *mutual-following* users, as:

$$T_g = \{g' | g' \in G \wedge t(g, g') \in T\} \tag{5}$$

We also fix the amount of δ for each *mutual-following* users g' when working on user g. Here, δ is called a social promotion weight. The promotion matrix $\tilde{A}$ is a $|G| \times |G|$ dimensional real-valued matrix. Each entry in $\tilde{A}$ is denoted by $a_{g,g'}$ defined as:

$$a_{g,g'} = \begin{cases} 1 & g' = g \\ \delta & g' \in T_g \wedge g' \neq g \\ 0 & otherwise \end{cases} \tag{6}$$

Under the GPU hood, the urn contains various colored balls, when a ball is drawn, that ball is put back along with a certain number of similar colored balls, thus increasing all their probabilities of being observed. Similarly, we can consider each followee is a ball, and balls with *mutual-following* relations are painted with similar color. Once a ball is drawn from a social topic, our GPU process is triggered. That is, if a followee g is observed under a social topic z^{so}, her *mutual-following* users $g' \in T_g$ will be added to this z^{so} with a small count δ, thus guiding the sampler has higher probability to cluster them to the same topic. Therefore, we can expect to associate *mutual-following* users with more accurate social topics as they are given adequate promotion values.

Comparatively, under the normal Collapsed Gibbs sampling process, the probability of drawing a user under an interest topic z^{in} given $o = 1$ would only raise every time itself are drawn under z^{in}. Obviously, incorporating GPU to *UIS-LDA* distinguishes the generative process of interest topics and social topics.

In our *UIS-LDA*, z^{in}, z^{so}, θ^{in}, θ^{so}, o are random variables and g is the observed variable. We denote the variables $\{z, o\}$ by singular subscripts $\{z_i, o_i\}$, where i corresponding to the ith followee in a document d_f, z_i is the topic allocation for this followee, and o_i is the switch assignment for this followee. Symbol $\mathtt{z}, \mathtt{h}$ respectively refers to topics allocation and all hyper parameters in a general sense. $\mathtt{z}^{\neg i}$ represents all topics allocation except for z_i. During our sampling process, we jointly sample variable o_i with z_i for each iteration via Eqs. (7) and (8), which are shown as follows:

$$\begin{aligned} Pr(z_i{=}z^{in}, o_i{=}1 | \mathtt{z}^{\neg i}, D, \mathtt{h}, g) \propto & \frac{n^{\neg i}_{d_f, o=1} + \varepsilon}{\sum_{o' \in \{0,1\}} n^{\neg i}_{d_f, o'} + 2\varepsilon} \\ & \times \frac{n^{\neg i}_{d_f, z^{in}} + \alpha^{in}}{\sum_{z' \in Z^{in}} n^{\neg i}_{d_f, z'} + K^{in} \times \alpha^{in}} \times \frac{n^{\neg i}_{z^{in}, g} + \beta^{in}}{\sum_{g' \in G} n^{\neg i}_{z^{in}, g'} + |G| \times \beta^{in}} \end{aligned} \tag{7}$$

$$\begin{aligned} Pr(z_i{=}z^{so}, o_i{=}0 | \mathtt{z}^{\neg i}, D, \mathtt{h}, g, \tilde{A}) \propto & \frac{n^{\neg i}_{d_f, o=0} + \varepsilon}{\sum_{o' \in \{0,1\}} n^{\neg i}_{d_f, o'} + 2\varepsilon} \\ & \times \frac{n^{\neg i}_{d_f, z^{so}} + \alpha^{so}}{\sum_{z' \in Z^{so}} n^{\neg i}_{d_f, z'} + K^{so} \times \alpha^{so}} \times \frac{\sum_{g' \in G} a_{g,g'} n^{\neg i}_{z^{so}, g'} + \beta^{so}}{\sum_{\nu' \in G} \sum_{g' \in G} a_{\nu',g'} n^{\neg i}_{z^{so}, g'} + |G| \times \beta^{so}} \end{aligned} \tag{8}$$

In these formulas, $n^{\neg i}_{z^{so}, g}$ denotes the number of times that an observed followee g under topic z^{so} excluding z_i; $n^{\neg i}_{d_f, z^{in}}$ refers to the count of a document d_f was assigned to topic z^{in} except for z_i; $n^{\neg i}_{d_f, o=1}$ describes the number of times a document d_f was allocated to $o = 1$ excluding o_i. After sampling is complete, we infer the latent variables $\theta^{in}_{d_f}$ and $\theta^{so}_{d_f}$ with the following Eqs. (9) and (10).

$$\theta^{in}_{d_f} = \frac{n_{z^{in}, g} + \alpha^{in}}{\sum_{z' \in Z^{in}} n_{d_f, z'} + K^{in} \times \alpha^{in}} \tag{9}$$

$$\theta^{so}_{d_f} = \frac{n_{z^{so}, g} + \alpha^{so}}{\sum_{z' \in Z^{so}} n_{d_f, z'} + K^{so} \times \alpha^{so}} \tag{10}$$

We can see that by introducing GPU in sampling process, it differentiates latent topics on social and interest factors. Our *UIS-LDA* model is actually an improved topic level model, it allows us to depict *mutual-following* users in this work to facilitate the recommendation better. More importantly, other user relations, such as *repost* and *like*, can also be depicted alike in the future work, showing large flexibility and promising potential.

4.1.3. UIS-LDA model complexity

We now analyze the time complexity of *UIS-LDA* with references to LDA and the topic model used by *CB-MF* [4]. The time complexity of LDA in an iteration is $O(K|D|\overline{\ell})$. $|D|$ is the number of document, K is the number of all topics and $\overline{\ell}$ is the average document length. Within *UIS-LDA*, every document contains followees for each follower. For each user g, her *mutual-following* users are organized in set T_g. As GPU is employed, in each iteration of Algorithm 1, *UIS-LDA* has a time complexity of $O(|D|K\overline{\ell} + \tau(|D||\overline{T}|\overline{\ell}))$, where τ is the average probability that social topic is assigned, $|\overline{T}|$ is the average size of *mutual-following* set. In this case, $|D||\overline{T}|\overline{\ell}$ is linear to $|\overline{T}|$.

CB-MF employees LDA on both followers document and followees document to generate pre-defined number of K topics. For convenient, we refer to its topic model the *CB-LDA*. Similar to LDA, *CB-LDA* has a time complexity of $O(K|D|\overline{\ell'})$, where $\overline{\ell'}$ equals the average size of followers and followees for each user. Obviously, since the documental formation of *UIS-LDA* and *CB-LDA* is different, the average document length of *UIS-LDA* $\overline{\ell}$ is less than $\overline{\ell'}$.

In order to lower the time complexity, we can limit upper bound of $|\overline{T}|$ to a constant. In experiments, we set the max size of the *mutual-following* users to 25 for every user. Hence at most 25 random users with the *mutual-following* relations are chosen into a *mutual-following* set. Therefore, it is expected that $O(|D|K\overline{\ell} + \tau(|D||\overline{T}|\overline{\ell})) < O(|D|K\overline{\ell} + 25|D|\overline{\ell}) < O(|D|K\overline{\ell'} + 25|D|\overline{\ell'}) = O(|D|(K + 25)\overline{\ell'})$, which is identical to add constant topics to *CB-LDA*. Suppose that $\overline{\ell'}$ is two times of $\overline{\ell}$, the number of topics K is larger than 25, the time complexity of our algorithm actually is less than *CB-LDA*. Hence, although the GPU process is included, we can conclude that our *UIS-LDA* do not increase much of computation expense. Noted that we also empirically verify this argument in Section 6.5.

4.2. Phase 2: user recommendation

Recently MF approaches gained popularity due to attractive accuracy. However, data sparsity is their major challenge [52,53]. One promising approach is the community-based method. It forms communities before applying MF algorithms, whereby a community is a subset of original dataset including both representative followers and followees. Forming communities can not only reduce the sparsity in MF but also save time since MF on each community can be carried out in parallel [4]. Therefore, this work adopts community-based method for user recommendation. We first group users into social communities and interest communities whose memberships are over some thresholds, whereby the memberships are defined by users' probabilities of a given topic. Thereafter, we perform MF algorithm on each community to obtain the maximum score that a follower will follow a user in the community, and then sort these scores to output the final Top-*N* candidate followees list for a target user.

4.2.1. Community formations

Since a user follows another user based on individual interest and social connection, interest communities and social communities exist on uni-directional social networks. Within a community, users might have different degrees of memberships [54]. The higher the degree of a membership, the better a user can reflect the essence of a community. In our scenario, we group followers and followees as their memberships are greater than or equal to some thresholds.

For each interest topic z^{in}, we form a corresponding interest community c^{in}. It includes followers in $c^{in}.F$ and followees in $c^{in}.G$, which are given by follows:

$$c^{in}.F = \{f | f \in F \wedge \exists (Pr(z^{in}|d_f) \geq \gamma)\} \tag{11}$$

$$c^{in}.G = \{g | g \in G \wedge \exists (Pr(z^{in}|d_g) \geq \zeta)\} \tag{12}$$

where both γ, ζ are thresholds. Like d_f, each followee $g \in G$ can be regarded as a document d_g and its content is expressed as:

$$d_g = \{f | f \in F \wedge \exists e(f, g) \in E\} \tag{13}$$

Since a higher $Pr(z^{in}|d_f)$ or $Pr(z^{in}|d_g)$ indicates the user is more strongly associated with the topic, for the corresponding community, we regard $Pr(z^{in}|d_f)$ as the followers membership and $Pr(z^{in}|d_g)$ as the followees membership. Among them, $Pr(z^{in}|d_f)$ is defined as:

$$Pr(z^{in}|d_f) = \frac{Pr(z^{in}|d_f)}{\sum_{z' \in Z} Pr(z'|d_f)} \tag{14}$$

The numerator $Pr(z^{in}|d_f)$ can be obtained from $\theta^{in}_{d_f}$, and the denominator is solely for normalization. As contrast with $Pr(z^{in}|d_f)$, $Pr(z^{in}|d_g)$ cannot directly obtained from *UIS-LDA*. Alternatively, it can be achieved with equation Eq. (15) since we take the following assumption:

Assumption 2. A higher probability of a user's followers exists under a topic indicates a higher chance of himself exists under this same topic.

$$Pr(z^{in}|d_g) = \frac{\sum_{f \in d_g} Pr(z^{in}|d_f)}{\sum_{z' \in Z^{in}} \sum_{f \in d_g} Pr(z'|d_f)} \tag{15}$$

Noted that when inferring the communities, in Eq. (14), summation over all topics is used while in Eq. (15), summation is done only over the interest topics. The reason for such a distinction is that we'd like to infer the communities as accurately as possible since a followee is assigned to a topic according to the number of times her followers have been previously assigned to this topic. That is, if some assignments are made in Eq. (14), they affect assignments in Eq. (15).

The edge in a community c^{in} denoted as $c^{in}.E$ is given by:

$$c^{in}.E = \{e(f, g) | e(f, g) \in E \wedge f \in c^{in}.F \wedge g \in c^{in}.G\} \tag{16}$$

Analogously, we can form a bunch of social communities $c^{so} \in C^{so}$ with the same thresholds γ and ζ. Therefore, the output for this section is *K* numbers of interest communities C^{in} with social communities C^{so}, where $|C^{in}| = K^{in}$ and $|C^{so}| = K^{so}$.

Table 1
Statistics of three experimental datasets.

Statistic	Twitter size	Sina Weibo size	Epinions size
$\|U\|$	163,048	169,750	14,256
$\|F\|$	156,048	168,561	14,165
$\|G\|$	135,774	150,761	14,130
$\|E\|$	14,909,700	40,358,104	399,298
$\|T\|$	8,589,402	14,291,214	175,559
$Avg(\|E(*, g)\|)$	109.81	267.70	28.25
$Avg(\|E(f, *)\|)$	95.56	239.43	28.19
$Avg(\|T(*, g)\|)$	56.49	84.79	12.39
$Avg(\frac{\|T(f,*)\|}{\|E(f,*)\|})$	51.8%	35.3%	43.9%
Sparsity	99.93%	99.83%	99.80%

*4.2.2. Top-*N *followee recommendation*

MF-based approaches prove to be highly accurate and scalable in addressing CF problems [21]. To perform user recommendation, we conduct MF method over these separate communities. For each community *c* where $c \in C^{in} \cup C^{so}$, we regard $f \in c.F$ as users and $g \in c.G$ as items for *IF-MF* method. Let $\tilde{M}_c$ is a $|c.F| \times |c.G|$ dimensional matrix, each entry is given by p_{fg}, derived by follow:

$$p_{fg} = \begin{cases} 1 & \exists (e(f, g) \in c.E) \\ 0 & otherwise \end{cases} \tag{17}$$

After performing *IF-MF* on each $\tilde{M}_c$, we can obtain two latent matrixes: $\tilde{X}^{|c.F| \times L}$ and $\tilde{Y}^{L \times |c.G|}$. They denote the mappings of followers and followees into the reduced latent space of *L* respectively. We then calculate the predict score $C\_score(f, g, c)$ by Eq. (18). It indicates how much a follower *f* prefers to follow a followee *g* under a community *c*. x_f are latent feature vectors in $\tilde{X}$ and y_g are latent feature vectors in $\tilde{Y}$, where $f \in c.F$ and $g \in c.G$.

$$C\_score(f, g, c) = x_f \cdot y_g \tag{18}$$

Following that, we take the maximum score of $C\_score(f, g, c)$ among *K* communities as the final score, denoted by $F\_score(f, g)$.

$$F\_score(f, g) = \underset{c \in C}{Maximum}(C\_score(f, g, c)) \tag{19}$$

It is worth noticing that we have tried other operations, such as a summation or weighted combination, max over all scores by Eq. (19) provides best results. Finally, for each follower *f*, we sort the highest *N* final scores and recommend them as candidates.

4.2.3. User recommendation complexity

After *UIS-LDA* is employed, the computational complexity for *UIS-MF* is affected by the complexity of community formations and the complexity of recommending users. In the process of community formations, as indicated in Ref. [4], the time complexity of *CB-MF* is $O(|D|K)$. Our algorithm's time complexity is $O(|D|K + |G||D|K)$, where $|G|$ is the total followees within the dataset. The first part of $O(|D|K)$ represents the computation expense of Eq. (14) for all documentations. In the second part, for each followee, the computation for Eq. (15) is required, which in total has time complexity of $O(|G||D|K)$. When recommending users, both of *CB-MF* and *UIS-MF* applied matrix factorization on communities. The time complexity of MF is highly dependent on community size. When the community size is identical, the time complexity is approximate. Since communities are independently, we should emphasize that parallel implementation of matrix factorization phases can both speed up *UIS-MF* and *CB-MF*.

5. Evaluations

In this section, we preface our evaluation including: (1) Experimental Datasets, (2) Evaluation Metrics, (3) Comparative baselines.

Table 2
Example performance tuning results on three datasets for varying K and $R\hat{}$.

	F1 Score on Twitter where $N = 5$				F1 Score on Sina Weibo where $N = 5$				F1 Score on Epinions where $N = 5$			
	$K = 130$	$K = 140$	$K = 150$	$K = 160$	$K = 110$	$K = 120$	$K = 130$	$K = 140$	$K = 10$	$K = 20$	$K = 30$	$K = 40$
$R\hat{} = 1/4$	0.0777	0.0779	**0.0803**	0.0798	0.0681	**0.0685**	0.0657	0.0650	0.0324	0.0292	0.0263	0.0238
$R\hat{} = 2/3$	0.0758	0.0767	0.0775	0.0783	0.0679	0.0676	0.0652	0.0642	0.0285	0.0311	0.0312	0.0260
$R\hat{} = 1/1$	0.0735	0.0740	0.0723	0.0732	0.0672	0.0666	0.0642	0.0633	0.0318	**0.0327**	0.0299	0.0279
$R\hat{} = 3/2$	0.0691	0.0701	0.0680	0.0665	0.0661	0.0651	0.0631	0.0621	0.0289	0.0314	0.0301	0.0299
$R\hat{} = 4/1$	0.0653	0.0667	0.0632	0.0637	0.0611	0.0598	0.0574	0.0567	0.0299	0.0302	0.0307	0.0275

5.1. Experimental datasets

We conduct our experiments on three real world datasets.

1. **Twitter dataset and Sina Weibo dataset**. The Twitter dataset is obtained from [55], we randomly sampled users from this dataset and discard all users with less than 10 followers and followees. The Sina Weibo dataset we choose is used in [4] which has already been pre-processed. These two datasets consist a list of *follower-followee* relations. Table 1 gives the statistics of these two datasets after pre-processing. On average, each user has 95.56 (239.43) followers, 109.81 (267.7) followees, and maintains *mutual-following* relations with 56.49 (94.79) users on Twitter (Sina Weibo). To each follower, the average ratio of *mutual-following* relations to *follower-followee* relations is more than 51% on Twitter and over 35% on Sina Weibo. This indicates that *mutual-following* relations are widely exist among users on both datasets, making them worth our attention for in-depth research.
2. **Epinions dataset**. A concern for this work is whether the framework is generalizable, or can translate to other networks. Therefore, we try to extrapolate our results further on another application scenario. Epinions.com[3] is a famous product review site where users can specify whom to trust and build a social trust-network. The trust-network of Epinions shares similar uni-directional structure as Twitter and Sina Weibo, i.e., if user f is a truster of user g, user g is not necessarily a truster of user f. Our task becomes to generate trustee recommendations for individual users. The selected Epinions dataset is publicly available.[4] As also showed on Table 1, it contains 14,256 users, 399,298 explicit *truster-trustee* relations, 43.9% are *mutual-trust* relations, and the sparsity is 99.8%. Noted that ratings and users with less than five trusters/trustees have been excluded.

Moreover, we separate each dataset into a training set and a test set. For each follower, we randomly choose 90% followees she has followed as training set data for training the latent topics and latent factors, and the remaining 10% followees are used as testing set data.

5.2. Evaluation metrics

In order to evaluate our recommendation framework, we employ five standard metrics in the area of information retrieval. Each metric is calculated at a given cut-off rank, denoted as @N, indicates only the first N ranked followees are regarded by evaluation metrics.

1. **Recall**. This metric measures the proportion of users from the test sets that appears among the Top-N ranked list, for some given N.
2. **Precision**. This metric measures the proportion of recommended users that are true followees.
3. **F1 Score**. This score is the harmonic mean of recall and precision, defined as:

$$F1@N = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{20}$$

4. **Conversion rate**. This metric is to confirm if a user has obtained at least one ground-truth followees from recommended list L.

$$Conversion \quad rate@N = \begin{cases} 1 & if|L \cap L'| > 0 \\ 0 & otherwise \end{cases} \tag{21}$$

where L' is the list of N users actually followed by the user in test set.

5. **NDCG** [56]. This metric is to evaluate ranked list as it gives higher reward for the top users in a recommended list, which is defined as:

$$NDCG@N = \frac{DCG@N}{IDCG@N} \tag{22}$$

where $DCG@N = \sum_{i=1}^{N} \frac{2^{y_i}-1}{\log_2(1+i)}$, y_i is 1 if the recommended user at position i in the ranking is a hit followee, and 0 otherwise. $IDCG@N$ is the $DCG@N$ of the sorted optimal ranked list.

Noted that we compare above metrics in various algorithms by taking the average of values computed for each test user, and all the metrics should be as high as possible for better performance.

5.3. Comparative baselines

Since the task is to generate user recommendations for individual users, we compare *UIS-MF* against the six following user recommendation baselines to evaluate the effectiveness of our proposed framework.

1. ***CB-MF*** [4]. A community-based user recommendation method on uni-directional social networks. Considering that *CB-MF* shares a close relationship with *UIS-MF* as presenting a topic model and forming communities before applying *IF-MF* method for making user recommendation, we choose *CB-MF* as the main baseline to compare against.
2. ***IF-MF*** [10]. A state-of-the-art MF technique for implicit feedback data.
3. ***BPR-MF*** [11]. A state-of-the-art Bayesian Personalized Ranking MF method for binary relevance data.
4. ***LDA-Based***. An implication of the LDA-based model proposed in [26], and we recommend the N followees with the highest score using the equation as follow:

$$Pr(g|f) = \sum_{z \in Z} Pr(g|z) Pr(z|d_f) \tag{23}$$

5. ***UIS-Based***. After *UIS-LDA* modeling, we recommend N most likely users using Eq. (23).
6. ***PopRec***. A naive baseline and non-personalized recommendation approach. For any target user, it generates the same recommendation list containing most popular users.

[3] http://www.epinions.com.
[4] http://www.trustlet.org/wiki/Downloaded_Epinions_dataset.

On the other hand, to evaluate how the unified extraction of social topics and interest topics of users does affect the recommend performance, we make following internal comparisons with our *UIS-LDA* model:

1. ***In-LDA***. This model discovers interest topics is also proposed in [26]. In practice, we remove the social factors by setting $K^{so} = 0$ for *UIS-LDA*.
2. ***So-GPU***. This model takes the GPU model for encouraging *mutual-following* relations so as to extract social topics, this is achieved by $K^{in} = 0$ for *UIS-LDA*.

The recommendation with *In-LDA* and *So-GPU* model can be performed by using the strategies similar to Section 4.2.

6. Experiments and discussion

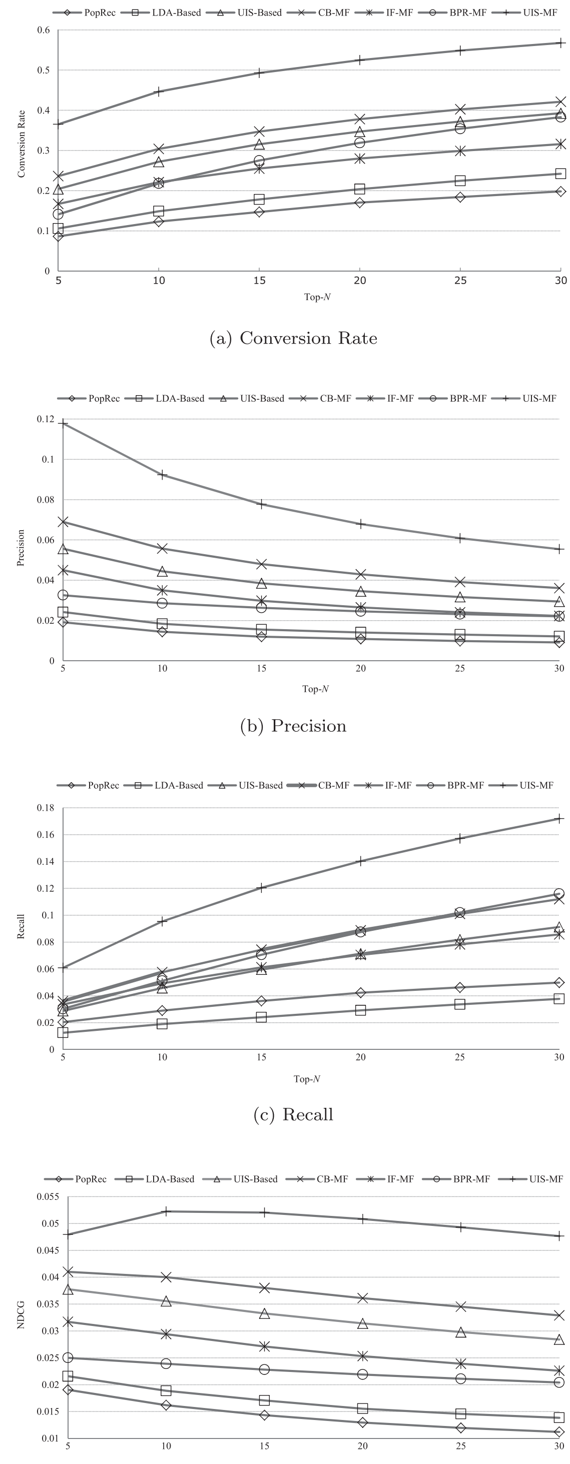
We present the parameters setting and report the results of the extensive experiments we have carried out to evaluate the advantage of our *UIS-MF* in this section. We code the *UIS-LDA* model and use the Java implementation provided in software MyMediaLite [57] for *PopRec, IF-MF* and *BPR-MF*. All the experiments are carried out on an Intel Xeon CPU E5606 with 2.13GHz, 128GB RAM, 64 bit, Linux Ubuntu 14.04 operating system.

6.1. Parameters setting

For *UIS-LDA* model, we set Dirichlet prior hyper-parameters as $\alpha^{in} = \alpha^{so} = \beta^{in} = \beta^{so} = 0.01$ since they are not very sensitive. Meanwhile, we set *Bernoulli* prior hyper-parameter $\varepsilon = 0.5$ to make a social topic and an interest topic have equally chance to be chosen by a follower at the initial time of sampling. Besides, social promotion weight δ is set to 0.01, it results in a higher weight to *mutual-following* user pairs compared to *uni-following* user pairs, meaning that the probability of *mutual-following* users under social topics will all get promoted. For the community formation process, we empirically set thresholds $\gamma = 2 \times \frac{1}{K}$, $\zeta = 0.05$. In order to compare performances fairly, for all the MF models, we set the number of latent factors $L = 16$ as suggested in work [4].

6.2. Parameters sensitivity experiments

The N values, the number of latent topics K and the ratio of K^{in} to K^{so}, denoted by $R\hat{}$ are important parameters affecting the performance of our method. The higher value of $R\hat{}$ indicates more interest topics generated by a given K topic numbers. In the first set of experiments, we focus on the exploration of our recommendation accuracy evaluated with F1 Score by varying K, $R\hat{}$, and N. Concretely, we vary K and $R\hat{}$, whereby $R\hat{}$ is separately set to 1/1, 2/3, 1/4, 3/2, 4/1 on three datasets. For each fixed pair of K and $R\hat{}$, we change the number of N and obtain the optimal parameters setting on F1 Score: (N = 15, K = 150, $R\hat{}$ = 1/4) for Twitter dataset, (N = 30, K = 120, $R\hat{}$ = 1/4) for Sina Weibo dataset and (N = 15, K = 20, $R\hat{}$ = 1/1) for Epinions dataset. Due to limitation of pages, we just present some example results of the parameters tuning with F1 Score that K is tuned from 130/110/10 to 160/140/40 on Twitter/Sina Weibo/Epinions datasets where N=5 (See Table 2). Noted that (1) since a user's satisfaction is dominated by items at the top of the recommendation list and a large value of N is usually ignored for a typical Top-N recommendation task [24,29], we experimentally change N from 5 to 50 and show only the performance N is in the range of {5,10,15,20,25,30} for the rest of experiments. (2) the highest F1 Score on Twitter/Epinions dataset appears at $N = 15$ whereas the highest F1 Score on Sina Weibo appears in much longer $N = 30$ list. However, for drawing a fair comparison, we also record @20, @25, @30 evaluations on Twitter and Epinions datasets.



(a) Conversion Rate

(b) Precision

(c) Recall

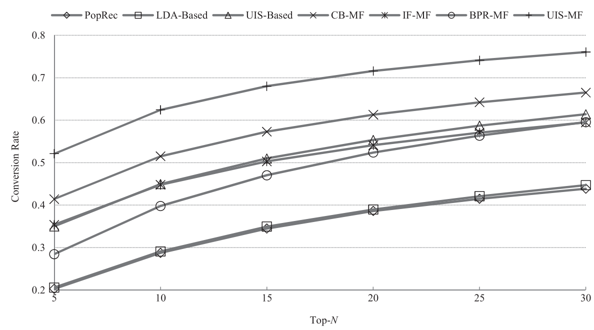
(d) NDCG

Fig. 2. Comparison of user recommendation on Twitter dataset.

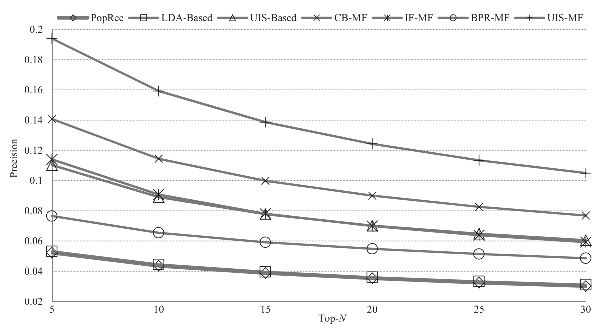
6.3. Method comparisons

In the second set of experiments, we compare the performance of the seven user recommendation methods with a series of N values on three datasets.

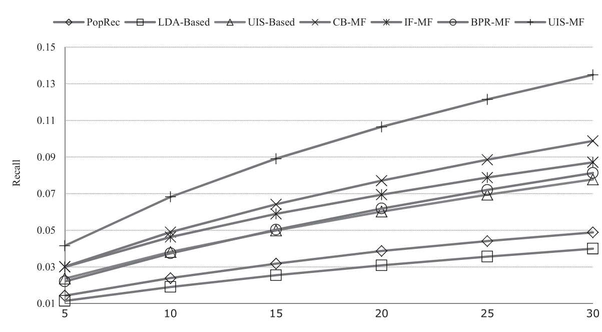
Figs. 2 and 3 give the results for various estimations on Twitter and Sina Weibo datasets. We look into Figs. 2(a, c) and 3(a, c) at first, which show that Conversion Rate and Recall of the seven methods improve as the number of N increases. Compared to



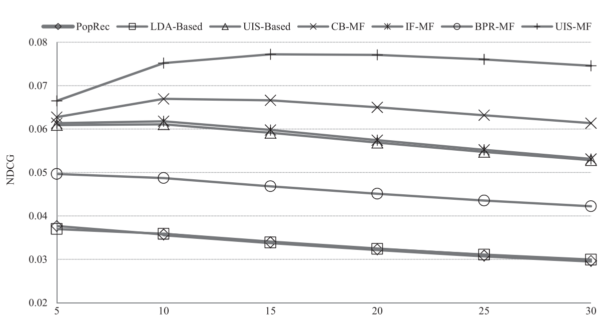
(a) Conversion Rate



(b) Precision

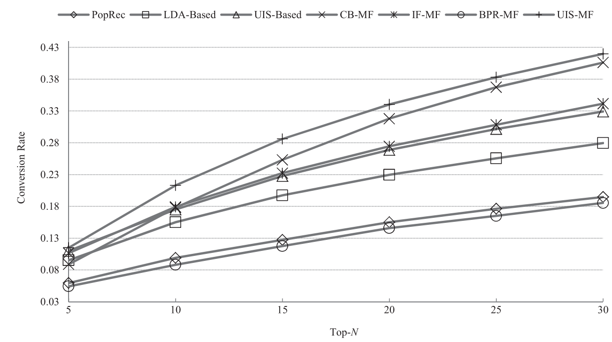


(c) Recall

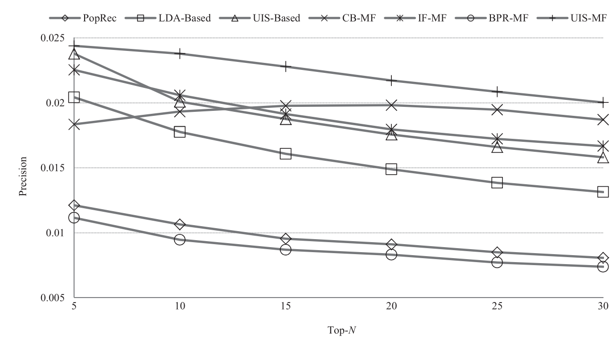


(d) NDCG

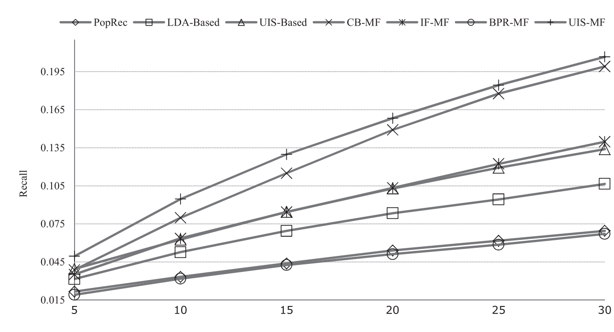
Fig. 3. Comparison of user recommendation on Sina Weibo dataset.



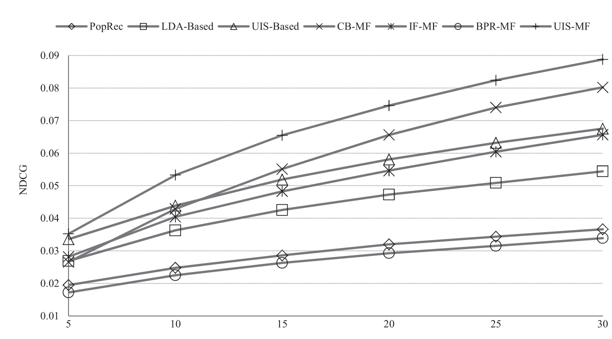
(a) Conversion Rate



(b) Precision



(c) Recall



(d) NDCG

Fig. 4. Comparison of user recommendation on Epinions dataset.

the best performance of baseline methods *CB-MF*, *UIS-MF* increases the Conversion Rate@30 by 34.91% and significantly raises the Recall@30 by 53.57% in Twitter dataset, as well as, improves the Conversion Rate@30 by 14.44%, and raises the Recall@30 by 36.54% in Sina Weibo dataset. Figs. 2b and 3b illustrate that all the methods degrade on Precision for longer recommending. Although the precision of *UIS-MF* decreases to 0.055 and 0.105 for Precision@30 in two datasets respectively, we believe that they are acceptable results since it gives the best performance as well. In Figs. 2d and 3 d, again, they confirm the NDCG results of *UIS-MF* are reasonable and better than other baselines.

Table 3
Internal comparisons of *UIS-LDA* model on three datasets.

Twitter dataset		$N=5$	$N=10$	$N=15$	$N=20$	$N=25$	$N=30$
Method	$R\hat{}$	F1 Score	F1 Score	F1 Score	F1 Score	F1 Score	F1 Score
So-GPU	0/150	0.0797	0.0927	0.0932	0.0900	0.0860	0.0820
UIS-LDA	30/120	**0.0803**	**0.0938**	**0.0945**	**0.0915**	**0.0877**	**0.0839**
In-LDA	150/0	0.0551	0.0627	0.0624	0.0602	0.0576	0.0550
Sina Weibo dataset		$N=5$	$N=10$	$N=15$	$N=20$	$N=25$	$N=30$
Method	$R\hat{}$	F1 Score	F1 Score	F1 Score	F1 Score	F1 Score	F1 Score
So-GPU	0/120	0.0637	0.0888	0.1007	0.1064	0.1086	0.1090
UIS-LDA	24/96	**0.0685**	**0.0955**	**0.1085**	**0.1147**	**0.1173**	**0.1181**
In-LDA	120/0	0.0502	0.0680	0.0760	0.0800	0.0811	0.0813
Epinions dataset		$N=5$	$N=10$	$N=15$	$N=20$	$N=25$	$N=30$
Method	$R\hat{}$	F1 Score	F1 Score	F1 Score	F1 Score	F1 Score	F1 Score
So-GPU	0/20	0.0296	0.0352	0.0370	0.0371	0.0365	0.0357
UIS-LDA	10/10	**0.0327**	**0.0380**	**0.0388**	**0.0382**	**0.0375**	**0.0365**
In-LDA	20/0	0.0306	0.0364	0.0378	0.0373	0.0358	0.0345

Table 4
Average running time comparisons on three datasets.

Dataset	Per iteration consumed for topic models				Running time consumption	
	UIS-LDA	*In-LDA*	*So-GPU*	*CB-LDA*	*UIS-MF*	*CB-MF*
Twitter	22,023 ms	10,483 ms	23,614 ms	15,942 ms	251 m	235 m
Sina Weibo	43,451 ms	21,898 ms	45,623 ms	35,893 ms	228 m	212 m
Epinions	272 ms	220 ms	356 ms	353 ms	37 m	34 m

Detailed results for various estimations on Epinions dataset can be found in Fig. 4. We have the following findings: (1) Among all the baselines, *UIS-MF* is most effective, *CB-MF* takes the second place, and both of the *BPR-MF* and *PopRec* perform poorly on this dataset. (2) Compared with *CB-MF, UIS-MF* increase the average prediction accuracy by 8.4% on Recall, 15.7% on Precision and 9.04% on Conversion Rate. (3) When employing NDCG as the evaluation metric, *UIS-MF* can also obtain obvious advantages compared with *CB-MF*, as shown in Fig. 4d. For instance, the highest NDCG of *UIS-MF* is at 0.089 when $N=30$, which is about 10.7% higher than that of *CB-MF*. In addition, although their curves seem to be very close to each other on Recall, Precision and Conversion Rate, *UIS-Based* outperforms *IF-MF* on different NDCG@N points, which shows the ranking priority.

These results prove that our *UIS-MF* gains significant performance improvements among all the baselines in regard of all evaluation metrics on three datasets. Specifically, we can derive several insights. First of all, *UIS-MF* outperforms MF based methods (*BPR-MF, IF-MF* and *CB-MF*) without *UIS-LDA* model, and we attribute this results to the advantage of *UIS-LDA* model in extracting sufficient topics properly captures the user preferences. That is, integrating both interest topics and social topics to learn user preferences can significantly improve the recommendation performance. On the other hand, between the two *UIS-LDA* based method (*UIS-MF* and *UIS-Based*), *UIS-MF* gives better results, owing to the contribution of our community-based recommendation, which utilizes the hidden characteristics of users via MF method. Additionally, except for Recall, *UIS-Based* performs better than *IF-MF, BPR-MF, PopRec* and *LDA-Based* methods on all the evaluation metrics in Twitter dataset. In Sina Weibo and Epinions dataset, *UIS-Based* exhibits competitive performance with *IF-MF* in most of the conditions, which demonstrates that our *UIS-LDA* model is indeed useful. Furthermore, *UIS-MF* is a general framework since it works pretty well with user follow relationships (verified on Twitter and Sina Weibo datasets) and user trust relationships (verified on Epinions dataset). In a word, our *UIS-MF* with *UIS-LDA* model could give a better explanation for the followees/trustees selection process and thus move a step closer to persuasive recommendations.

6.4. Model internal comparisons

The focus of the third set of experiments is on in-depth investigating the respective roles of social topics and interest topics in augmenting the recommending results. We apply the same matrix factorization approach IF-MF with $L=16$ on three different models: *UIS-LDA, In-LDA* and *So-GPU* described in Section 5.3, for user recommendation. Table 3 lists their results for F1 Score on three datasets. Based on the comparison, three main observations can be drawn from Twitter and Weibo datasets: First, the performance of *In-LDA* is the worst, suggesting that the poor quality of interest topics it has captured. The main reason is due to its utilization of coarse *follower-followee* relations that cannot make a distinction between a user's interest and social connection. Second, *So-GPU* gives impressive accurate results, highlighting the predominant role of social topics captured from *follower-followee* relations promoted by our GPU process. Third, our *UIS-LDA* model always obtains the best performance, indicating that combining the extraction of social topics and interest topics improves user recommendation over extracting either of the two separately. This suggests that *mutual-following* relations promoted by GPU process is a good complement to the traditional use of *follower-followee* relations under the normal Collapsed Gibbs sampling process that treat users with *mutual-following* relations as same as with *uni-following* relations. These results can also be explained by the fact that our model includes both interest and social connection of users as preferences indicators to infer more reasonable user topics, and this helps boost the recommending performance.

However, on Epinions dataset, we find that *In-LDA* wins *So-GPU* in most cases, while this result is in contrary in Twitter and Weibo datasets. One possible reason may be that Epinions is a consumer rating website, thus users using these ratings to help them making right choices from trustees. Whereas, Twitter is a famous social-oriented platform. thus users are more inclined to connect with each other. Therefore, a user chooses her trustees probably according to her trust. Since Ref. [44] have confirmed that trust information is a very idea source to represent users' interest, we can refer trust as a proxy for interest in this work. As Epinions is not a social-oriented platform, if we put more focus on interest (like

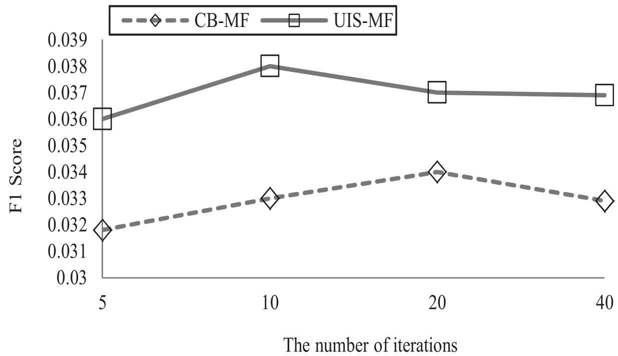


Fig. 5. F1 scores as the number of iterations varies on Epinions dataset.

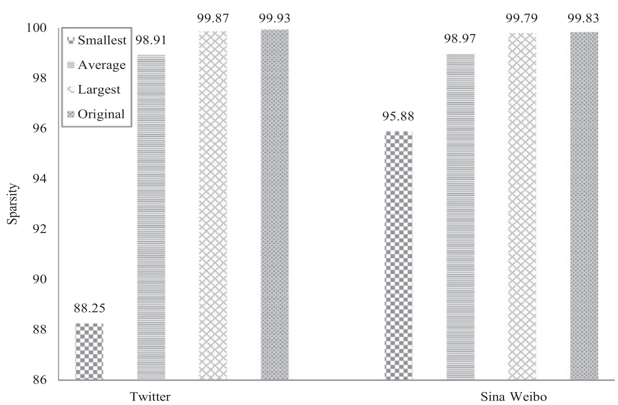


Fig. 6. Sparsity on Twitter and Sina weibo dataset.

In-LDA) rather than social connection (like *So-GPU*), we can get better trustee recommendation results. Furthermore, from the final results, if we combine user interest with social connection as *UIS-LDA*, the recommendation accuracy is the best. In short, social factors of users is still indispensable on Epinions dataset to leveling up recommendation performance. We also believe that this will give us another point of view on understanding the correlations between trust-network and user preferences inferring on Epinions.

6.5. Run time comparisons and analysis

The fourth set of experiments leads us to benchmark the running time for our *UIS-MF* and the best existing approach *CB-MF*. In particular, we also analyze the running time of *UIS-LDA* compared to related topic models.

Fig. 5 shows how dose *UIS-MF*/*CB-MF*'s accuracy change with number of *UIS-LDA*/*CB-LDA*'s learning iterations χ on Epinions dataset. For *UIS-MF*, F1 Score increases as χ increases, when χ is larger than 10, lower F1 Scores are obtained. While *CB-MF* requires larger $\chi = 20$ to obtain better F1 Score. Noted that, 1) to avoid heavy time burden, we directly set $\chi = 200$ on Twitter and Sina Weibo dataset. 2) we apply the same matrix factorization approach *IF-MF* for *UIS-MF* and *CB-MF* with latent factors $L = 16$ on three datasets.

Table 4 shows the average run time of per iteration for four topic models (*UIS-LDA, In-LDA, So-GPU* and *CB-LDA*). Noted that *CB-LDA* is the topic model employed by *CB-MF*. The execution times are measured in milliseconds. Our findings are as follows: (1) On Epinions dataset, the consumed iteration time of *UIS-LDA* is slightly shorter than *CB-LDA*, whereas, its time is a bit more than *CB-LDA* on Twitter and Sina Weibo datasets. However, their actual running time are in the same magnitude. These results are consistence with the analytical time complexity presented in 4.1.3. (2) The *In-LDA* is the least time consuming. That is because *In-LDA* deletes the GPU part from *UIS-LDA*, and degenerated to a basic LDA model. Thus, it does not require additional cost for GPU process. (3) While the mechanism of *So-GPU* is treating all the topics of *UIS-LDA* as social topics, therefore, GPU process would be triggered in every iteration, making it be the slowest of the four topic models. Table 4 also records *UIS-MF* to compare its running speed with *CB-MF*. The execution times are measured in minutes. We find that *CB-MF*'s recommendation gets faster than *UIS-MF*. The increased time cost of *UIS-MF* is mainly caused by its community formations process, which slows down the run time (Detailed in Section 4.2.3). Admittedly, some implementation details, such as caches used on Twitter and Sina Weibo, are also affect the actual running time.

6.6. Data sparsity analysis

In the last set of experiments, we are also interested in comparing the sparsity of the original dataset and the communities we obtained. Since the original dataset is divided into K smaller communities, Fig. 6 reports that the average sparsity of a community is improved to 98.91% on Twitter and 98.97% on Sina Weibo. Fig. 7 gives the detailed sparsity information on Epinions dataset where C0 to C19 refers to 20 discovered communities. These observations allow us to confirm that forming communities alleviates the sparsity problem suffered by original dataset positively, hence help us to improve the MF performance.

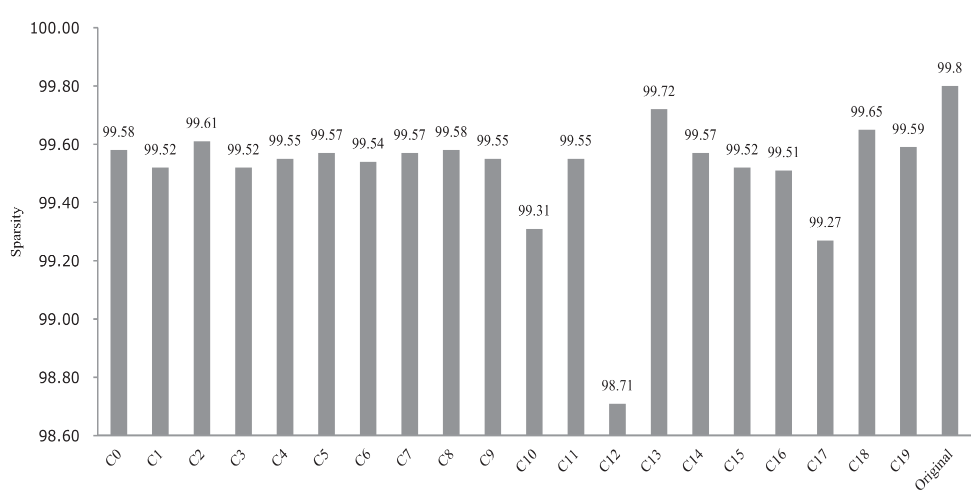


Fig. 7. Sparsity of original dataset vs. discovered communities on Epinions datasets.

7. Conclusion

Modeling user preferences involving interest factors and social factors in uni-directional networks introduces complexity but offers significant accuracy improvements and analytical benefits. In this paper, we have proposed a framework *UIS-MF*, which integrates both types of preferences in a principled manner and targeted to recommend Top-*N* followees relevant to a target user. To be concrete, we have first designed a novel topic model on *follower-followee* and *mutual-following* relations to discover latent interest topics and social topics of users, and then proposed a community-based method for making user recommendation. According to our analyses and experiments on Twitter, Sina Weibo and Epinions datasets, the distinguished effectiveness of *UIS-MF* was demonstrated. Moreover, we found that the unified extraction of social topics and interest topics not only reflecting the nature of a user's following preferences better but also is essential to improving the accuracy of user recommendation. More notably, the significant role of *mutual-following* relations in improving the quality of social topic distributions was identified. On the other hand, in the sparse data condition, our community-based method can alleviate the sparsity problem troubled by MF method. Furthermore, we believe that our GPU process has improved existing topic models substantially as to depicting social connection among users. Besides, our work is especially useful when only follow or not data is available. In conclusion, these findings illustrate meaningful insights for our recommendation work on uni-directional social networks, and corroborate the significant impact of our topic model on augmenting the performance of the recommendation. In particular, a series of experiments on Epinions' trust-network also demonstrates that our recommendation framework may generalizes to other domains. For our upcoming research, we shall continue our study along the following directions: (1) Social promotion weight δ will be deeply discussed. (2) When forming communities, the proportion of interest or social factors affecting a person should be considered. (3) We will apply our model for other challenging domains, such as protein interactive mapping [22], etc.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, SCUT (No. 2017ZD048), the Tiptop Scientific and Technical Innovative Youth Talents of Guangdong special support program (No.2015TQ01X633), the Science and Technology Planning Project of Guangdong Province (No.2016A030310423, 2017B050506004), the Science and Technology Program of Guangzhou International Science & Technology Cooperation Program (No. 201704030076), the Science and Technology Planning Major Project of Guangdong Province (No. 2015A070711001), and the Internal Research Grant (RG 66/2016-2017) of The Education University of Hong Kong. We would like to express our sincere appreciation to the anonymous reviewers for their valuable and constructive comments.

References

[1] K. Xu, Y. Cai, H. Min, X. Zheng, H. Xie, T.-L. Wong, Uis-lda: a user recommendation based on social connections and interests of users in uni-directional social networks, in: ACM Conference on Web Intelligence, WI, 2017, pp. 260–265, doi:10.1145/3106426.3106494.

[2] W. Xie, C. Li, F. Zhu, E.-P. Lim, X. Gong, When a friend in twitter is a friend in life, in: ACM Web Science Conference, WebSci, 2012, pp. 344–347, doi:10.1145/2380718.2380763.

[3] D. Mimno, H.M. Wallach, E. Talley, M. Leenders, A. McCallum, Optimizing semantic coherence in topic models, in: Conference on Empirical Methods in Natural Language Processing, EMNLP, 2011, pp. 262–272.

[4] G. Zhao, M.L. Lee, W. Hsu, W. Chen, H. Hu, Community-based user recommendation in uni-directional social networks, in: ACM Conference on Information & Knowledge Management, CIKM, 2013, pp. 189–198, doi:10.1145/2505515.2505533.

[5] X. Su, T.M. Khoshgoftaar, A survey of collaborative filtering techniques, Adv. Artif. Intell. 2009 (12) (2009) 1–19, doi:10.1155/2009/421425.

[6] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, S. Ma, Explicit factor models for explainable recommendation based on phrase-level sentiment analysis, in: ACM Conference on Research and Development in Information Retrieval, SIGIR, 2014, pp. 83–92, doi:10.1145/2600428.2609579.

[7] B. Shams, S. Haratizadeh, Iterank: an iterative network-oriented approach to neighbor-based collaborative ranking, Knowl. Based Syst. 128 (2017) 102–114, doi:10.1016/j.knosys.2017.05.002.

[8] G. Takács, I. Pilászy, B. Németh, D. Tikk, Matrix factorization and neighbor based algorithms for the netflix prize problem, in: ACM Conference on Recommender Systems, RecSys, 2008, pp. 267–274, doi:10.1145/1454008.1454049.

[9] R. Salakhutdinov, A. Mnih, Probabilistic matrix factorization, in: International Conference on Neural Information Processing Systems, NIPS, 2007, pp. 1257–1264.

[10] Y. Hu, Y. Koren, C. Volinsky, Collaborative filtering for implicit feedback datasets, in: IEEE International Conference on Data Mining, ICDM, 2008, pp. 263–272, doi:10.1109/ICDM.2008.22.

[11] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, Bpr: Bayesian personalized ranking from implicit feedback, in: Conference on Uncertainty in Artificial Intelligence, UAI, 2009, pp. 452–461.

[12] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver, A. Hanjalic, Climf:learning to maximize reciprocal rank with collaborative less-is-more filtering, in: ACM Recommender Systems, RecSys, 2012, pp. 139–146, doi:10.1145/2365952.2365981.

[13] X. He, H. Zhang, M.-Y. Kan, T.-S. Chua, Fast matrix factorization for online recommendation with implicit feedback, in: ACM Conference on Research and Development in Information Retrieval, SIGIR, 2016, pp. 549–558, doi:10.1145/2911451.2911489.

[14] E.M. Voorhees, D.M. Tice, The trec-8 question answering track evaluation, Trec National Inst. Stand. Technol. 7 (1999) 77–82.

[15] W. Reafee, N. Salim, A. Khan, The power of implicit social relation in rating prediction of social recommender systems, PLoS ONE 11 (5) (2016) 1–20, doi:10.1371/journal.pone.0154848.

[16] H. Ma, An experimental study on implicit social recommendation, in: ACM Conference on Research and Development in Information Retrieval, SIGIR, 2013, pp. 73–82, doi:10.1145/2484028.2484059.

[17] Z. Huang, S. E, J. Zhang, B. Zhang, Z. Ji, Pairwise learning to recommend with both users' and items' contextual information, IET Commun. 10 (16) (2016) 2084–2090, doi:10.1049/iet-com.2016.0112.

[18] S. Deng, L. Huang, G. Xu, X. Wu, Z. Wu, On deep learning for trust-aware recommendations in social networks, IEEE Trans. Neural Netw. Learn. Syst. 28 (5) (2017) 1164–1177, doi:10.1109/TNNLS.2016.2514368.

[19] K. Lu, Y. Zhang, L. Zhang, S. Wang, Exploiting user and business attributes for personalized business recommendation, in: ACM Conference on Research and Development in Information Retrieval, SIGIR, 2015, pp. 891–894, doi:10.1145/2766462.2767806.

[20] M.N. Volkovs, G.W. Yu, Effective latent models for binary feedback in recommender systems, in: ACM Conference on Research and Development in Information Retrieval, SIGIR, 2015, pp. 313–322, doi:10.1145/2766462.2767716.

[21] X. Luo, M. Zhou, Y. Xia, Q. Zhu, An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems, IEEE Trans. Ind. Inf. 10 (2) (2014) 1273–1284, doi:10.1109/TII.2014.2308433.

[22] X. Luo, Z. You, M. Zhou, S. Li, H. Leung, Y. Xia, Q. Zhu, A highly efficient approach to protein interactome mapping based on collaborative filtering framework, Sci. Rep. 5 (2015) 1–10, doi:10.1038/srep07702.

[23] M. Yigit, B.E. Bilgin, A. Karahoca, Extended topology based recommendation system for unidirectional social networks, Expert Syst. Appl. 42 (2015) 3653–3661, doi:10.1016/j.eswa.2014.12.043.

[24] G. Li, Z. Zhang, L. Wang, Q. Chen, J. Pan, One-class collaborative filtering based on rating prediction and ranking prediction, Knowl Based Syst. 124 (2017) 46–54, doi:10.1016/j.knosys.2017.02.034.

[25] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.

[26] Y. Cha, J. Cho, Social-network analysis using topic models, in: ACM Conference on Research and Development in Information Retrieval, SIGIR, 2012, pp. 565–574, doi:10.1145/2348283.2348360.

[27] L. Li, W. Peng, S. Kataria, T. Sun, T. Li, Frec: a novel framework of recommending users and communities in social media, in: ACM Conference on Information & Knowledge Management, CIKM, 2013, pp. 1765–1770, doi:10.1145/2505515.2505645.

[28] M. Pennacchiotti, S. Gurumurthy, Investigating topic models for social media user recommendation, in: International Conference on World Wide Web, WWW, 2011, pp. 101–102, doi:10.1145/1963192.1963244.

[29] S. Wang, M. Gong, H. Li, J. Yang, Y. Wu, Memetic algorithm based location and topic aware recommender system, Knowl. Based Syst. 131 (2017) 125–134, doi:10.1016/j.knosys.2017.05.030.

[30] G. COSTA, R. ORTALE, Model-based collaborative personalized recommendation on signed social rating networks, ACM Trans. Internet Technol. 16 (3) (2016) 1–21, doi:10.1145/2934681.

[31] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, S. Yang, Social contextual recommendation, in: ACM Conference on Information & Knowledge Management, CIKM, 2012, pp. 45–54, doi:10.1145/2396761.2396771.

[32] X. Lu, P. Li, H. Ma, S. Wang, A. Xu, B. Wang, Computing and applying topic-level user interaction in microblog recommendation, in: ACM Conference on

Research and Development in Information Retrieval, SIGIR, 2014, pp. 843–846, doi:10.1145/2600428.2609455.
[33] H. Mahmoud, Pólya Urn Models, Chapman & Hall/CRC Texts in Statistical Science, 2008. 43(2)
[34] G. Fei, Z. Chen, B. Liu, Review topic discovery with phrases using the Pólya urn model, in: International Conference on Computational Linguistics: Technical Papers, COLING, 2014, pp. 667–676.
[35] Z. Chen, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, R. Ghosh, Leveraging multi-domain prior knowledge in topic models, in: International Joint Conference on Artificial Intelligence, IJCAI, 2013, pp. 2071–2077.
[36] Z. Chen, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, R. Ghosh, Discovering coherent topics using general knowledge, in: ACM Conference on Information & Knowledge Management, CIKM, 2013, pp. 209–218, doi:10.1145/2505515.2505519.
[37] C. Li, H. Wang, Z. Zhang, A. Sun, Z. Ma, Topic modeling for short texts with auxiliary word embeddings, in: ACM Conference on Research and Development in Information Retrieval, SIGIR, 2016, pp. 165–174, doi:10.1145/2911451.2911499.
[38] L. Boratto, S. Carta, G. Fenu, R. Saia, Using neural word embeddings to model user behavior and detect user segments, Knowl. Based Syst. 108 (2016) 5–14, doi:10.1016/j.knosys.2016.05.002.
[39] A. Muhammad, N. Wiratunga, R. Lothian, Contextual sentiment analysis for social media genres, Knowl. Based Syst. 108 (2016) 92–101, doi:10.1016/j.knosys.2016.05.032.
[40] F. Qian, Y. Zhang, Y. Zhang, Z. Duan, Community-based user domain model collaborative recommendation algorithm, Tsinghua Sci. Technol. 18 (4) (2013) 353–359.
[41] H. Jindal, Anjali, Graph based recommendation system in social networks, Int. J. Comput. Appl. 113 (2) (2015) 36–40, doi:10.5120/19801-1583.
[42] E.-J. Lee, Y.W. Kim, How social is twitter use? Affliative tendency and communication competence as predictors, Comput. Human Behav. 39 (2014) 296–305, doi:10.1016/j.chb.2014.07.034.
[43] S.J. Kwon, E. Park, K.J. Kim, What drives successful social networking services? a comparative analysis of user acceptance of facebook and twitter, Soc. Sci. J. 51 (2014) 534–544, doi:10.1016/j.soscij.2014.04.005.
[44] H. Ma, On measuring social friend interest similarities in recommender systems, in: ACM Conference on Research and Development in Information Retrieval, SIGIR, 2014, pp. 465–474, doi:10.1145/2600428.2609635.
[45] M.G. Armentano, D. Godoy, A. Amandi, Topology-based recommendation of users in micro-blogging communities, J. Comput. Sci. Technol. 27 (3) (2012) 624–634, doi:10.1007/s11390-012-1249-5.
[46] J. Hannon, M. Bennett, B. Smyth, Recommending twitter users to follow using content and collaborative filtering approaches, in: ACM Conference on Recommender Systems, RecSys, 2010, pp. 199–206, doi:10.1145/1864708.1864746.
[47] S.A. Golder, S. Yardi, Structural predictors of tie formation in twitter: transitivity and mutuality, in: IEEE Second International Conference on Social Computing, SOCIALCOM, 2010, pp. 88–95, doi:10.1109/SocialCom.2010.22.
[48] D. Zhao, M.B. Rosson, How and why people twitter: the role that micro-blogging plays in informal communication at work, in: ACM International Conference on Supporting Group Work, GROUP, 2009, pp. 243–252, doi:10.1145/1531674.1531710.
[49] C. Hutto, S. Yardi, E. Gilbert, A longitudinal study of follow predictors on twitter, in: Conference on Human Factors in Computing Systems, CHI, 2013, pp. 821–830, doi:10.1145/2470654.2470771.
[50] J. Yoo, S. Choi, M. Choi, J. Rho, Why people use twitter: social conformity and social value perspectives, Online Inf. Rev. 38 (2) (2014) 265–283, doi:10.1108/OIR-11-2012-0210.
[51] T.L. Griffiths, M. Steyvers, Finding scientific topics, Proc. Natl. Acad. Sci. U.S.A. 101 (Suppl 1 (1)) (2010).
[52] Q. Yuan, L. Chen, S. Zhao, Factorization vs. regularization: fusing heterogeneous social relationships in top-n recommendation, in: ACM Conference on Recommender Systems, RecSys, 2011, pp. 245–252, doi:10.1145/2043932.2043975.
[53] M. Kunaver, T. Požrl, Diversity in recommender systems - a survey, Knowl. Based Syst. 123 (2017) 154–162, doi:10.1016/j.knosys.2017.02.009.
[54] H. Xie, Q. Li, Y. Cai, Community-aware resource profiling for personalized search in folksonomy, J. Comput. Sci. Technol. 27 (3) (2012) 599–610, doi:10.1007/s11390-012-1247-7.
[55] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media? in: International Conference on World Wide Web, WWW, 2010, pp. 591–600, doi:10.1145/1772690.1772751.
[56] K. Järvelin, J. Kekäläinen, Ir evaluation methods for retrieving highly relevant documents, in: ACM Conference on Research and Development in Information Retrieval, SIGIR, 2000, pp. 41–48.
[57] Z. Gantner, S. Rendle, C. Freudenthaler, L. Schmidt-Thieme, Mymedialite:a free recommender system library, in: ACM Conference on Recommender Systems, RecSys, 2011, pp. 305–308, doi:10.1145/2043932.2043989.