Summer 6-27-2016

INDIVIDUALITY OR CONFORMITY: RECOMMENDATION EXPLOITING COMMUNITY-LEVEL SOCIAL INFLUENCE

Jiangning He
Tsinghua University, hejn.13@sem.tsinghua.edu.cn

Hongyan Liu
Tsinghua University, hyliu@tsinghua.edu.cn

Jun He
Renmin University of China, hejun@ruc.edu.cn

Sanpu Han
Beijing Qihoo Technology Company Limited, hansanpu@360.cn

INDIVIDUALITY OR CONFORMITY: RECOMMENDATION EXPLOITING COMMUNITY-LEVEL SOCIAL INFLUENCE

Jiangning He, School of Economics and Management, Tsinghua University, Beijing, China, hejn.13@sem.tsinghua.edu.cn

Hongyan Liu, School of Economics and Management, Tsinghua University, Beijing, China, hyliu@tsinghua.edu.cn

Jun He, School of Information, Renmin University of China, Beijing, China, hejun@ruc.edu.cn

Sanpu Han, Beijing Qihoo Technology Company Limited, Beijing, China, hansanpu@360.cn

Abstract

With the increasing prevalence of online businesses and social networking services, a huge volume of data about transaction records and social connections between users is accumulated at an unprecedented speed, which enables us to take advantage of electronic word-of-mouth effect embedded in social networks for precision marketing and social recommendations. Different from existing works on social recommendations, our research focuses on discriminating the community-level social influence of different friend groups to enhance the quality of recommendation. To this end, we propose a novel probabilistic topic model integrating community detection with topic discovery to model user behaviors. Based on this model, a recommendation method taking both individual interests and conformity influence into consideration is developed. To evaluate the performance of the proposed model and method, experiments are conducted on two real recommendation applications, and the results demonstrate that the proposed recommendation method exhibits superior performance compared with the state-of-art recommendation methods, and the proposed topic model exhibits good explainablibity of topic semantics and community interests. Furthermore, as some people are more individual interest oriented and some are more conformity oriented demonstrated by the experiments, we explore factors that influence each individual's conformity tendency, and obtain some meaningful findings.

Keywords: Social recommendation, Community-level social influence, Topic model, Community detection.

1 INTRODUCTION

In the fast emerging business environment, customers are inundated with a large variety of choices among products or services, facing the problem of information overload. At the same time, they are embedded in various kinds of social networks, building social connections anytime and anywhere. The availability of data collected in online businesses and social networking platforms provides opportunities and challenges for us to take advantage of electronic word-of-mouth effect for effective personalized recommendations and for better solutions to the information overload problem. According to a Morpace Omnibus Report[1], 68% of Facebook users say that a recommendation from a Facebook friend would make them more likely to buy a specific product or visit a certain retailer. Under this condition, social recommendation systems (social recommenders) are developed and become prevalent, which aim to identify patterns of user interests by analyzing data about users' purchase behaviors and social relations, such as trust relations in Epinions[2], friendships in Facebook[3], and following relations in Twitter[4] (King et al. 2010; Tang et al. 2013).

Existing social recommenders assume that users' preferences are similar to or influenced by their friends' and the rationale behind can be explained by social correlation theories such as homophily (Mcpherson et al. 2001) and social influence (Marsden and Friedkin 1993). Based on this assumption, some representative recommendation systems such as SocialMF (Jamali and Ester 2010) , SoReg (Ma et al. 2011) and SBPR (Zhao et al. 2014) were developed. These systems largely assume the homogeneity of social connections. For instance, SocialMF formulates a user's latent preference vector as the weighted average of the latent preference vectors of all its neighbors (Jamali and Ester 2010). However, the assumptions in existing social recommendation methods are too strict to be realistic, because the social relationships between people are indeed heterogeneous, and the heterogeneity of social relationships leads to diverse social groups in both physical and virtual world. As illustrated in Figure 1, a person like Alice usually has many friends, which can be divided into multiple groups (communities) with possible overlaps among them. Alice may share different interests with different groups of friends, for instance, hiking and camping with one group, and visiting art museums and attending concerts with another group. In this regard, it would be beneficial to consider appropriate friend groups' influence, instead of the influence of all socially connected neighbors as a whole or very sparse individuals' behavior influence, when we predict users' preferences on items (i.e. products or services). For instance, to recommend a song to Alice, it would be better to consider the preferences of music fans, rather than the preferences of outdoor amateurs or the average preferences of all friends. Therefore, in this paper, we focus on discriminate heterogeneous interest-based social groups and exploit community-level social influence to improve the quality of recommendation. Noting that the words "community" and "social group" are used interchangeably in this paper for the same meaning.

How to model the community-level social influence for item recommendation is a great challenge.

---

1 http://www.prnewswire.com/news-releases/morpace-reports-facebooks-impact-on-retailers-89590997.html

2 http://www.epinions.com

3 https://www.facebook.com

4 https://www.twitter.com

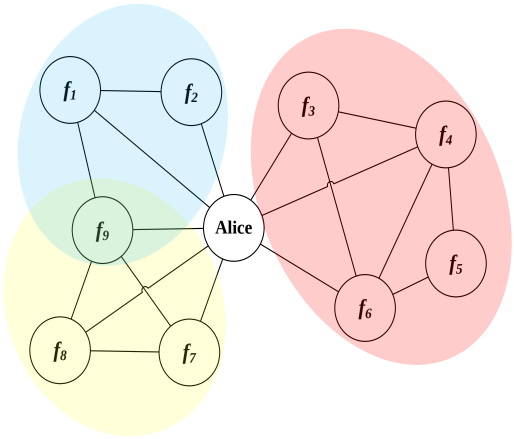

Figure 1. An illustration of diversified social groups in an ego-network

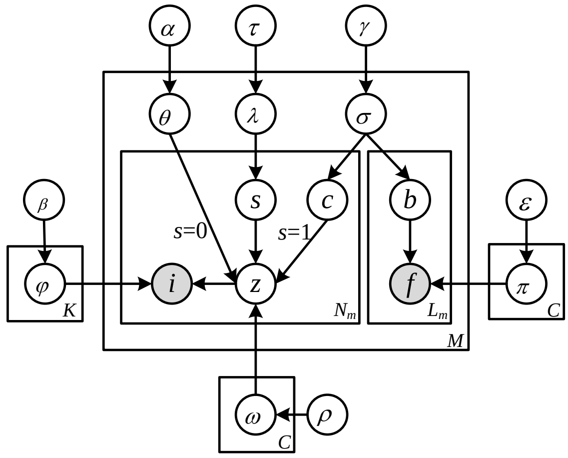

Figure 2. The probabilistic graphical model of ICTM

First, how can we automatically discover interest-based social groups? Second, how can we model the mechanism of users' selection behaviors on items? Third, how can we discriminate heterogeneous social influence of different groups to facilitate item recommendation? Next, we summarize the main work of this paper to answer the above questions.

First, we utilize the information of both social network structures and item selection behaviors for social group discovery or community detection (Fortunato 2009; Zhou et al. 2006) and group interest analysis. The combination guarantees that the discovered communities are more stable to avoid the disturbance of noise from one data source. Moreover, the structural information guarantees that users in a community are densely connected and interact a lot, and the behavioral information makes it possible to detect interest-based groups, namely users in a group have some common interests and different groups have different interests.

Then, we model the mechanism of item selection behaviors according to utility functions based on conformity theory (Bernheim 1994; J. Zhang et al. 2014). Utility is achieved in two channels: the first relies on the fulfillment of individual intrinsic preferences, called *individuality*, and the other comes from the *conformity* that people change behavior, belief or thinking to align with others or group norms (Cialdini and Goldstein 2004). Psychologically speaking, people conform to others in order to be correct in uncertain situations or to be liked by a social group, both of which are prevalent in the scenario of recommendation. Thus, we take both individuality and conformity into consideration when modeling users' selection behaviors on items, and employ a user-specific variable to determine the level of conformity tendency for each individual.

We propose a unified model, called Individual-level and Community-level Topic Model *(ICTM*) for both topic analysis and community detection. Based on this model, an item recommendation approach is proposed, taking both individual intrinsic preference and community-level conformity influence into consideration. *ICTM* discriminates social influence of different social groups by mapping items and community-level influence into a same topic space and then predicts users' preferences on items according to the preferences of social groups. In this way, social influence is exploited at the community level. Finally, we conduct experiments on two real datasets and show that our model is able to detect meaningful social groups and to achive better recommendation performance compared with the state-of-art related recommendation methods. Furthermore, we analyse factors that influence each individual's conformity tendency, and some meaningful findings are obtained.

In the rest of the paper, we first define the problem of social recommendation in Section 2. Then in Section 3 we introduce our model, describe the inference method and give the recommendation approach. To test the proposed model, we conduct experiments for evaluation in Section 4 and give a brief review of related work in Section 5. Finally, we conclude the paper in Section 6.

2 PROBLEM DEFINITION

In this section, we first formulate the input of the problem and then define the problem of social recommendation with implicit feedback.

The input of the problem includes a *user-item selection matrix* and *a user-user social connection matrix*. Let U=$\{u_1, u_2, \dots, u_M\}$ denote a set of users and I=$\{i_1, i_2, \dots, i_V\}$ denote a set of items, where M is the number of users and V is the number of items. We use $R \in \mathcal{R}^{M\times V}$ to denote the *user-item selection matrix* where $R_{ui} = 1$ denotes a selection behavior that user u has selected item i and $R_{ui} = 0$ denotes missing values. In different scenarios, selection behaviors have different meanings. Noting that in this paper we focus on datasets with implicit feedback (Hu et al. 2008; Rendle et al. 2012), rather than datasets with rating scores. The real world contains more implicit data, such as purchase transactions of products, download records of apps, and check-in trajectories of locations, etc. Accordingly, in these scenarios, user selection behaviors mean purchasing, downloading and checking-in, respectively. We can obtain *individual selected items* from the user-item selection matrix. For each user u, *individual selected items* I_u refers to the set of items selected by user u, namely $I_u = \{i_j | i_j \in I \ and \ R_{uj} = 1\}$.

The *user-user social connection matrix* is denoted as $T \in \mathcal{R}^{M\times M}$ where $T_{uv} = 1$ means user u has a social connection with user v, and $T_{uv} = 0$ otherwise. T is a symmetrical matrix if the social connections are undirected like friendships, or unsymmetrical if the social connections are directed like follow relations. And the proposed model can handle both situations. Likewise, we can obtain *socially connected neighbors* from the user-user social connection matrix T. For each user u, *socially connected neighbors* S_u refers to the set of users socially connected with user u, namely $S_u = \{u_v | u_v \in U \ and \ T_{uv} = 1\}$.

The task of social recommendation with implicit feedback: Given the user-item selection matrix R and user-user social connection matrix T, the goal is to predict score R_{ui} for each missing value of user-item pair and to generate a *Top-N* recommendation list L_u for each user u, formulated as follows.

$$f: (u, R, T) \rightarrow list\ L_u =< i_{r_1}, i_{r_2}, \dots, i_{r_N} > \ s.t.: R_{ur_1} \geq R_{ur_2} \geq \cdots \geq R_{ur_N}$$

3 OUR MODEL *ICTM*

In this section, we propose a probabilistic topic model called *ICTM* to discriminate community-level social influence for item recommendation. In the following, we first give a detailed description of the proposed model, describe the model inference method, and finally give a personalized recommendation approach based on the proposed model.

3.1 Model Description

In order to discriminate social influence of different communities, we propose a probabilistic topic model named *ICTM* integrating community detection with topic discovery. The proposed model aims to model the process of user selection behaviors, taking both individuality effect and conformity effect into consideration. Individuality means that users choose an item to achieve the fulfillment of their individual intrinsic preferences. Therefore, each user's individual interests are represented as a user-specific topic distribution following the idea of LDA (Blei et al. 2003). Besides, conformity means that users select an item due to the social influence of a friend group. Thus, to consider the effect of conformity influence, the proposed model needs to automatically detect social groups and discover interests for each group. In this regard, we introduce a hidden variable "community" for characterizing memberships of users in social networks and assigning users into different social groups with possible overlaps. Consequently, each user is described by a multinomial probability distribution over social groups (communities). Moreover, we employ a community-level topic distribution to discover distinct interests of each group, which plays the role of linking the results of community detection and selection behavior modeling into a unified model.

The generative process of the proposed model is illustrated in Figure 2. Here, *C* denotes the number of communities and *K* denotes the number of topics. N_m and L_m denote the number of items and the number of socially connected neighbors of the m^{th} user, respectively. And $\alpha, \beta, \tau, \gamma, \varepsilon, \rho$ are all hyper-parameters of Dirichlet or Beta priors for multinomial or Bernoulli distribution.

In Figure 2, the right part models the generation of *socially connected neighbors*. For the m^{th} user, first a user-specific community distribution σ_m characterizing a user's membership is drawn from its symmetric Dirichlet prior γ. Then for the l^{th} friend of the m^{th} user, a community $b_{m,l}$ is drawn from σ_m and the friend is generated from a multinomial community-level user distribution $\pi_{b_{m,l}}$. In this way, we employ a hidden variable b to cluster users' friends into different communities. The rational behind is that, similar to LDA (Blei et al. 2003) which clusters items into topics based on the co-occurrences of words, the proposed model clusters densely connected users in a social network into communities based on the co-occurrences of socially connected neighbors (H. Zhang et al. 2007). Thus, our model fulfills the task of community detection ensuring that users in a community are densely connected while the connections among communities are sparse (Zhou et al. 2006).

For the generation of *individual item selections*, as shown in the left part of Figure 2, the process is more complex. First of all, each user has a personal tendency of conformity in selecting items based on individual intrinsic preference or conformity influence of a community. The user-specific conformity distribution λ_m characterizes this tendency and is drawn from its asymmetric beta prior $\tau = (\tau_0, \tau_1)$. Thus, for the n^{th} item selected by the m^{th} user, the value of the binary variable $s_{m,n}$ is determined by the binomial user-specific conformity distribution λ_m. If $s_{m,n} = 0$, a topic $z_{m,n}$ is sampled from the user's topic distribution θ_m, and if $s_{m,n} = 1$, a community $c_{m,n}$ is sampled from the user's community distribution σ_m to obtain a influencing social group, and then a topic $z_{m,n}$ is drawn from the community's topic distribution $w_{c_{m,n}}$. Finally, the item is generated from the topic-specific item distribution $\varphi_{z_{m,n}}$.

Notably, we can see three advantages of the proposed model described above. The first is the integration of topic discovery with overlap community detection into a unified model. The second is the consideration of both conformity and individuality in modeling item selection behaviors. And the third is the combination of behavioral and structural information for interest-based community detection. This lies in the fact that variable c integrates behavioral information of item selection and variable b collects structural information of social network, both contributing to the inference of user-specific community distribution σ , as shown in Figure 2.

3.2 Model Inference

While the inference of models in the LDA family cannot be solved with closed-form solutions, a variety of algorithms have been developed to estimate the parameters of these models. In this paper, we exploit the collapsed Gibbs sampling method (Porteous et al. 2008), a form of Markov chain Monte Carlo, which is easy to implement and efficient for inference and learning. To infer *ICTM* model, we need to sample hidden variables $b_{m,l}$, $c_{m,n}$, $s_{m,n}$ and $z_{m,n}$. The posterior distributions for inference are as follows and the detailed derivation of these equations are omitted for space limit.

$$p(b_{m,l}|b_{-(m,l),}f,c,s,z,i,\dots) \propto (\gamma_{b_{m,l}} + p_{m,b_{m,l},*,*,*} + q_{m,b_{m,l},*}^{-(m,l)}) \times \frac{\varepsilon_{f_{m,l}} + q_{*,b_{m,l},f_{m,l}}^{-(m,l)}}{\sum_{f=1}^{M} \varepsilon_f + q_{*,b_{m,l},f}^{-(m,l)}} \quad (1)$$

$$p(c_{m,n}|c_{-(m,n)},b,f,s,z,i,\dots) \propto (\gamma_{c_{m,n}} + p_{m,c_{m,n},*,*,*}^{-(m,n)} + q_{m,c_{m,n},*}) \times \frac{\rho_{z_{m,n}} + p_{*,c_{m,n},1,z_{m,n},*}^{-(m,n)}}{\sum_{k=1}^{K} \rho_k + p_{*,c_{m,n},1,k,*}^{-(m,n)}} \quad (2)$$

$$p(z_{m,n},s_{m,n}=1|z_{-(m,n)},s_{-(m,n)},b,f,c,i,\dots) \propto (\tau_1 + p_{m,*,1,*,*}^{-(m,n)}) \times \frac{\rho_{z_{m,n}} + p_{*,c_{m,n},1,z_{m,n},*}^{-(m,n)}}{\sum_{k=1}^{K} \rho_k + p_{*,c_{m,n},1,k,*}^{-(m,n)}} \times \frac{\beta_{i_{m,n}} + p_{*,*,*,z_{m,n},i_{m,n}}^{-(m,n)}}{\sum_{i=1}^{V} \beta_i + p_{*,*,*,z_{m,n},i}^{-(m,n)}} \quad (3)$$

$$p(z_{m,n},s_{m,n}=0|z_{-(m,n)},s_{-(m,n)},b,f,c,i,\dots) \propto (\tau_0 + p_{m,*,0,*,*}^{-(m,n)}) \times \frac{\alpha_{z_{m,n}} + p_{m,*,0,z_{m,n},*}^{-(m,n)}}{\sum_{k=1}^{K} \alpha_k + p_{m,*,0,k,*}^{-(m,n)}} \times \frac{\beta_{i_{m,n}} + p_{*,*,*,z_{m,n},i_{m,n}}^{-(m,n)}}{\sum_{i=1}^{V} \beta_i + p_{*,*,*,z_{m,n},i}^{-(m,n)}} \quad (4)$$

In the above equations, symbol "…" denotes the omitted hyper-parameters in the model. Let $p_{m,c,s,z,i}$ denotes the count that item i is assigned to topic z for the m^{th} user due to conformity influence of community c (s=1) or due to individual intrinsic preference (s=0), and $q_{m,b,f}$ denotes the count that friend f of the m^{th} user is assigned to community b. And we use * to denote an aggregation on the corresponding dimension and -(m,n) means the exclusion of the n^{th} item for the m^{th} user.

After the sampling algorithm has run a number of iterations till the burn-in time (Porteous et al. 2008), we collect samples to obtain the expectation estimations of parameters in *ICTM* as follows

$$\sigma_m^c = \frac{p_{m,c,*,*,*} + q_{m,c,*} + \gamma_c}{p_{m,*,*,*,*} + q_{m,*,*} + \sum_{i=1}^{C} \gamma_i} \quad (5) \qquad \lambda_m^s = \frac{p_{m,*,s,*,*} + \tau_s}{p_{m,*,*,*,*} + \tau_0 + \tau_1} \quad (6)$$

$$\theta_m^z = \frac{p_{m,*,0,z,*} + \alpha_z}{p_{m,*,0,*,*} + \sum_{i=1}^{K} \alpha_i} \quad (7) \qquad \omega_c^z = \frac{p_{*,c,1,z,*} + \rho_z}{p_{*,c,1,*,*} + \sum_{i=1}^{K} \rho_i} \quad (8)$$

$$\pi_b^f = \frac{q_{*,b,f} + \pi_f}{q_{*,b,*} + \sum_{i=1}^{M} \pi_i} \quad (9) \qquad \varphi_z^i = \frac{p_{*,*,*,z,i} + \beta_i}{p_{*,*,*,z,*} + \sum_{i=1}^{V} \beta_i} \quad (10)$$

The above equations give us a better understanding of the advantages of *ICTM* model. For instance, the user-specific community distribution σ_m^c aggregates $p_{m,c,*,*,*}$ and $q_{m,c,*}$, thus satisfying the need of combining behavioral and structural information for interest-based community detection. In addition, the user-specific conformity distribution λ_m^s quantifies the level of conformity tendency for each individual and θ_m^z quantifies the user-specific individual interests while ω_c^z quantifies the community-level interests. And topic-specific item distribution φ_z^i clusters items into topics to model users' interests.

Complexity Analysis The main cost of *ICTM* lies in the collapsed Gibbs sampling of model inference, including the inference of latent variables *b*, *c*, *z* and *s* in Equations (1-4). The time complexity of inferring *b* is $\mathcal{O}(M\bar{L}C)$, where $\bar{L}$ is the average number of socially connected neighbors of users and does not scale with M. Similarly, the time complexity of inferring *c* is $\mathcal{O}(M\bar{N}C)$ and the complexity of inferring *z* and *s* is $\mathcal{O}(M\bar{N}K)$, where $\bar{N}$ is the average number of items selected by users and is usually very small due to the sparseness of datasets. Thus, the total time complexity of inference in each iteration is $\mathcal{O}(M(\bar{L}C+\bar{N}C+\bar{N}K))$, which is linearly correlated with the number of users M. Therefore, the proposed model is efficient and can scale to very large datasets.

3.3 Recommendation Approach

In order to recommend a list of items to user *u*, we need to predict recommendation scores R_{ui} for all items that user *u* has not selected and recommend top ranking items according to the descending order of recommendation score. Based on the proposed model, we employ $P(i|u)$, the probability that user *u* selects item *i* as the recommendation score R_{ui}, as defined in Equation (11).

$$P(i|u) = (1-\lambda_u^1)\sum_{z=1}^{K}\theta_u^z\varphi_z^i + \lambda_u^1\sum_{c=1}^{C}\sigma_u^c\sum_{z=1}^{K}w_c^z\,\varphi_z^i \tag{11}$$

The above equation takes both intrinsic preference and social influence into account when recommending items and employs the level of conformity λ_u^1 (i.e. λ_u^s given *s*=1) to balance the two effects. Specifically, the first part $\sum_{z=1}^{K}\theta_u^z\varphi_z^i$ denotes the probability that user *u* will choose item *i* based on individual intrinsic preference whereas the second part $\sum_{c=1}^{C}\sigma_u^c\sum_{z=1}^{K}w_c^z\,\varphi_z^i$ denotes the probability that user *u* will choose item *i* due to the conformity influence of social groups.

4 EXPERIMENTAL EVALUATION

In this section, we conduct experiments to test the ability of *ICTM* in item recommendation and give some related empirical analysis. Specifically, we are intended to explore the following aspects: 1) recommendation performances of our model compared with benchmark methods; 2) A case study of our model for topic and community detection; 3) regression analysis to explore factors influencing the level of individual conformity tendency.

4.1 Experimental Setup

Datasets Description. We conduct experiments on a travel package dataset and a mobile app dataset. The travel package dataset comes from a China tourism company, which operates through an e-commerce platform embedded with a social network of travelers. The dataset consists of users' purchase histories of travel packages and friendship relations between users in the social network. Here items refer to travel packages (Liu et al. 2014) and the user-user social connection matrix extracted from the friendship relations is symmetric. We aim to generate personalized recommendation lists of travel packages based on users' purchase histories and friend lists.

The mobile app dataset comes from a popular app downloading platform in China. The dataset consists of a list of apps installed on users' mobile phones and their contact lists in mobile phones. Here items are apps and the user-user social connection matrix extracted from the contact lists is asymmetric, since it is possible that user *u* can adds user *v* in its contact list but user *v* does not add user *u*. We aim to generate personalized recommendation lists of mobile apps based on users' install histories and contact lists.

	Travel package dataset	Mobile app dataset
The number of users	10,812	316,835
The number of distinct items	2,369	11,246
The average number of selected items	6.36	30.54
The average number of social neighbors	24.05	21.99

Table 1. Basic statistics of the travel package dataset and the mobile app dataset

Table 1 shows basic statistics of the two datasets and some differences between them. For example, the quantity and category of social information in the two datasets are distinct. The travel package dataset contains much more social information but very limited behavior information while this is contrary for the mobile app dataset. Another distinct characteristic is that social influence in the travel package dataset is explicit while that in the mobile app dataset is implicit. More specifically, users in the travel platform will be informed once their friends purchase a travel package and can scan the travel histories of their friends through the social networking platforms. However, the contact lists are static in the mobile app dataset and users cannot know what apps their friends install instantly.

Benchmark Methods. Since the proposed model aims to handle the task of item recommendation with implicit feedback, we select some representative recommendation methods for implicit feedback as benchmarks for comparison. The selected benchmarks include SBPR, LinkLDA, LDA, and itemKNN. A short description of the benchmark methods is given as follows.

- SBPR: Social Bayesian Personalized Ranking aims to leverage social connections to better estimate users' preferences in the form of ranking. It is proposed based on an assumption of pairwise ranking that users' preferences on their selected items are higher than items selected by their friends, which are higher than the rest of items (Zhao et al. 2014).
- LinkLDA: The method was proposed for modeling the generation of words in documents considering citation relationships between documents. Here for item recommendation, we treat items selected by a user as words in a document and socially connected neighbors of a user as papers cited by a document. The underlying assumption is that the generation of items and social neighbors share a common user-topic distribution (Erosheva et al. 2004).
- LDA: This is the classical form of topic models for modeling documents (Blei et al. 2003), and also suitable for item recommendation with implicit feedback. LDA treats users as documents and only takes behavioral selection of items into consideration.
- ItemKNN: This is a classical recommendation method of item-based collaborative filtering. It

computes item similarity according to users' co-selections of items and recommends similar items to users based on their selection histories (Sarwar et al. 2001).

Parameter Setting. For experiments, we randomly divide the datasets into a train set, a validation set, and a test set with the ratio of 6:2:2. We use the train set to learn models and the validation set to tune parameters and the recommendation performances of all methods are evaluated on the test set. For *ICTM*, the hyper parameters are set as $\alpha = 0.1, \beta = 0.01, \rho = 0.01, \tau_0 = 2.0, \tau_1 = 0.1, \gamma = 0.1, \varepsilon = 0.01$. For the travel package dataset, *K*=50, *C*=30; and for the mobile app dataset, *K*=100, *C*=50.

Evaluation Metrics. In our experiments, we recommend top-*N* (*N*=5, 10, 15, 20, 25, and 30) items for each user and choose three popular evaluation metrics to measure the recommendation quality. The metrics include *NDCG*, *Recall@N* and *Precision@N*, defined as follows.

NDCG measures the quality of ranking and is defined as the ratio of *DCG* value to the ideal *DCG* value (*IDCG*). For user *u*, we define DCG_u and $IDCG_u$ in Equations (12-13), where $|L_u|$ equals the length of ranked list for user *u*, $|Y_u|$ equals the number of items selected by user *u*, and rel_i equals 1 if item *i* is selected by user *u* and 0 otherwise.

$$DCG_u = \sum_{i=1}^{|L_u|} (2^{rel_i} - 1)/log_2(i + 1) \tag{12}$$

$$IDCG_u = \sum_{i=1}^{|Y_u|} 1/log_2(i + 1) \tag{13}$$

Recall@N measures the fraction of all truly selected items that are returned in the top-*N* recommendation lists. We define $Recall@N_u$ for user *u* in Equation (14), where L_u^N refers to the set of top-*N* recommended items for user *u* and Y_u refers to the set of items truly selected by user *u*.

$$Recall@N_u = |L_u^N \cap Y_u|/|Y_u| \tag{14}$$

Precision@N measures the fraction of top-*N* recommended items that are truly selected by users. We define $Precision@N_u$ for user *u* in Equation (15), where *N* is the size of the recommendation lists.

$$Precision@N_u = |L_u^N \cap Y_u|/N \tag{15}$$

4.2 Recommendation Performances

***NDCG*.** Table 2 lists the average results of *NDCG* of the proposed model in comparison with benchmark methods. It indicates that *ICTM* outperforms all the benchmarks with an average advantage of 20.21% on the travel package dataset and of 8.25% on the mobile app dataset. Moreover, compared with the mobile app recommendation, *ICTM* gains more advantage on the travel package recommendation, probably due to the relatively abundant social information in the travel dataset and the prominent effect of explicit social influence through social networking platforms. In addition, it may seem counterintuitive that some social recommendation methods, such as SBPR and LinkLDA, perform even worse than methods without utilizing social information such as LDA. However, this can be explained by the fact that social relationships are diversified and heterogeneous and users cannot be similar to all their friends in all dimensions, thus social recommendations that assume the homogeneity of social relationships may even diminish the recommendation quality as a whole.

***Recall and precision*.** The results of recall and precision of all methods on travel package recommendation and mobile app recommendation are shown in Figure 3 and Figure 4, respectively. It shows that *ICTM* outperforms all the benchmarks on both datasets. Besides, ItemKNN performs worse on travel package recommendation than on mobile app recommendation as the size of recommendation lists increases, because the average purchases of travel packages are only about 6, making it difficult to accurately compute item similarity based on co-selections.

Dataset	ICTM	SBPR	LinkLDA	LDA	ItemKNN
Travel package dataset	0.5491	0.4781	0.4209	0.4949	0.4407
Mobile app dataset	0.4582	0.4401	0.4365	0.4464	0.3778

Table 2. NDCG comparison of all methods on two datasets

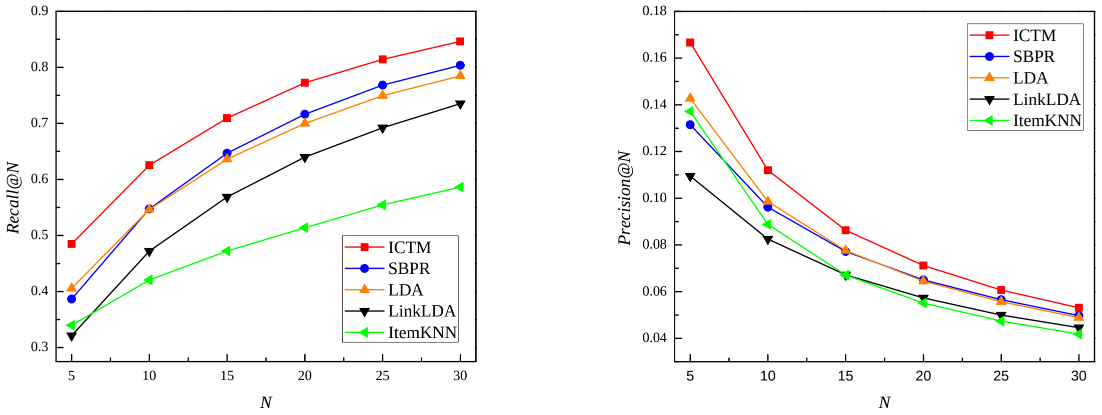

Figure 3. Recall and precision comparison of all methods on travel package recommendation

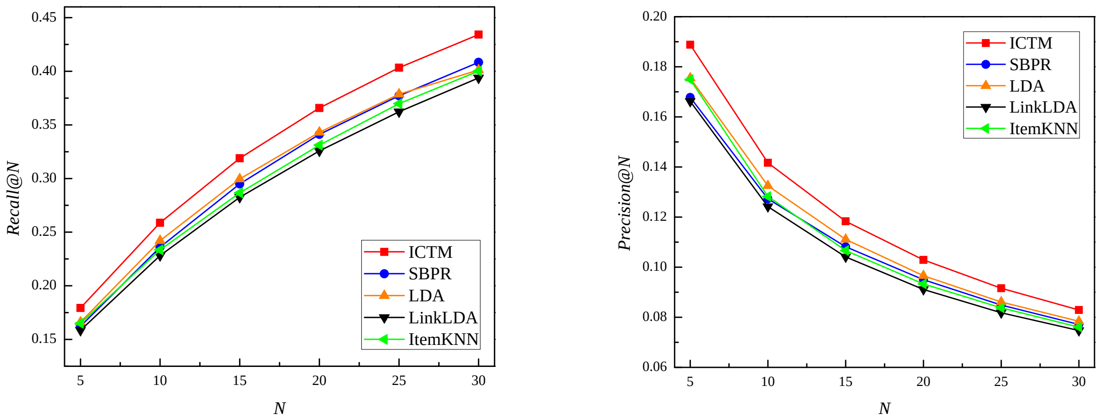

Figure 4. Recall and precision comparison of all methods on mobile app recommendation

4.3 Case study of topic and community detection

The *ICTM* model is able to cluster items into topics through the topic-item distribution φ_z^i, which indicates how well a single item *i* represents a topic *z*. Thus, to capture the basic semantics of a topic, a simple but effective way is to list top representative items of the topic combined with some descriptive information of the representative items. In this sense, we list three topics detected in the travel package dataset, each described with titles of Top three travel packages in Table 3, and the topic labels are summarized based on the descriptive information of top representative packages such as the topic label "Mountain Climbing" for topic 41.

Topic 41 **"Mountain Climbing"**	**Topic 17** **"Tour Conductor Activities"**	**Topic 26** **"Surburb Night-walk"**
Western Mountain traversing on Saturday of autumn	A photography ceremony with famous models	A night-walk around Baoshi Mountain around West Lake
Hui-hang ancient path traversing on National Day	Outdoor development: A competition between professional and part-time leaders	A night-walk around Linyin Mountain
Mount Jiuhua climbing on National Day	Treasure hunt and social gatherings between tour conductors and models	A night-walk around lakes through food street

Table 3. Semantics of topics described by titles of top 3 representative travel packages

Another merit of the proposed model lies in its ability of automatically detecting interest-based communities based on structural and behavioral information. To demonstrate this, we randomly select a user in the travel package dataset and plot its ego-network with an open-source network visualization tool Gephi[5] in *Figure 5*. We first select *main communities* the user belongs to according to the user-community distribution σ_m^c obtained by *ICTM* model such that the sum of the user-community probabilities belonging to the main communities is higher than a threshold, for example 0.95. Then, each of the user's friends is assigned to one of the main communities with the highest community-user probability according to the estimated community-level user distribution π_c^f.

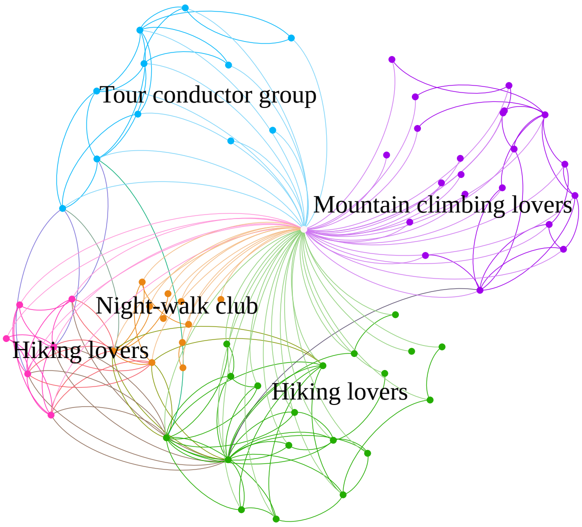

Figure 5. Detected communities of one real user

As shown in *Figure 5*, the ego-network of the user mainly consists of five social groups, namely communities 2, 20, 15, 24 and 29, denoted with colors: purple, green, blue, orange and pink, separately, and the user-community probabilities are 0.314, 0.279, 0.128, 0.118, and 0.117, respectively. Each social group developed distinct travel interests illustrated from the estimated

[5] https://gephi.org

community-level topic distribution ω_c^z, as listed in Table 4. Based on the modeling results, the proposed model successfully recommends 3 travel packages to the user, including an autumn hiking package according to community 20, a treasure hunt activity preferred by community 15, and a night-walk activity of community 24. In this sense, the proposed model is advantageous in combining social group detection and interests mining for personalized item recommendation.

Index	Community interests	Community label
2	Prefer to climb mountains (topic 41)	Mountain climbing lovers
20, 29	Prefer leisure hiking in spring and autumn	Hiking lovers
15	Prefer photography and professional outdoor development (topic 17)	Tour conductor group
24	Prefer regularly walking on every Saturday night (topic 26)	Night-walk club

Table 4. Semantics of community interests and labels

4.4 Factors influencing the level of conformity

The learned parameter λ by *ICTM* represents the conformity tendency of each user. Different users exhibit different levels of conformity tendency. To explore factors influencing the level of individual conformity tendency, we use multiple linear regression to analyze correlations between conformity and some factors including the number of selected items, entropy of user-topic distribution and entropy of user-community distribution, as shown in Table 5. It shows that conformity is positively correlated with the number of selected items and entropy of user-topic distribution, which can be explained that as users select more items and develop more diversified individual interests (higher entropy of user-topic distribution), they are more likely to be open and influenced by social environments, and thus exhibit higher level of conformity. Besides, conformity is negatively correlated with the entropy of user-community distribution. The rational behind is that, as the heterogeneity of social groups increases (higher entropy of user-community distribution), users are exposed to more conflicts of different social norms and thus become less likely to conform to a group.

	Travel package dataset		Mobile app dataset	
	Estimate	P-value	Estimate	P-value
Number of selected items N_m	1.02E-02	<2.00E-16	3.46E-03	<2.00E-16
Entropy of user-topic distribution θ_m	6.81E-02	<2.00E-16	1.69E-01	<2.00E-16
Entropy of user-community distribution σ_m	-2.01E-02	<2.00E-16	-5.81E-03	<2.00E-16
Adjusted R-squared	0.329		0.152	

Table 5. Regression analysis of factors influencing personal conformity on two datasets

5 RELATED WORK

In this section, we will give a brief review of works mainly related with our research, including social recommendation and topic models.

Social recommendation has attracted increasing attention both in industry and academia recently. The mainstream of social recommendation aims to leverage trust relationships to enhance rating prediction based on the assumption that users' tastes are similar to their trusted neighbors'. Such social recommenders widely exploit matrix factorization techniques, such as STE (Ma et al. 2009), SocialMF (Jamali and Ester 2010), SoReg (Ma et al. 2011), and so on. For example, Jamali and Ester (2010) assumed a user' preference to be closer to the average preference of the user' socially connected neighbors. And Ma et al. (2011) proposed a pair-wise social regularization by formulating the closeness of two connected users as their similarity based on previous ratings. Recently, some researchers started to pay attention to the diversity of social relationships. For instance, Yang et al. (2012) argued that users may trust different subsets of friends for different categories of items and proposed a circle based recommender system focusing on inferring category-specific trust circles to predict ratings for user-item pairs. Besides, Kabassi (2010) developed a personal and social latent factor model by considering the heterogeneity of people's social roles in effecting the formation of social relationships. However, to our best knowledge, none of existing works focused on simultaneously discovering heterogeneous interest-based communities and interest topics, and incorporating the community-level social influence to improve recommendation quality, especially for recommendation scenarios with implicit feedback, an open field worthy of further research.

Topic models are notable for their ability of discovering topics in documents. Following the classical topic model LDA (Blei et al. 2003) which clusters items into topics based on the co-occurrences of words, a large variations of topic models are proposed considering different factors, such as authorships (Rosen-Zvi et al. 2004), citation relationships (Erosheva et al. 2004; Nallapati et al. 2008), and so on. Recently, topic models have been applied to the problem of recommendation with implicit feedback, benefiting from the fact that topic models are able to summarize interests profiles of users and only positive feedbacks are needed in topic models. For example, a tourist-area-season topic model was proposed for travel package recommendation (Liu et al. 2011; Tan et al. 2014), and a collaborative topic regression model was developed for recommendation scientific articles (Wang and Blei 2011). However, a gap exists in combining topic models with community detection into a unified model for item recommendation, which is addressed in this paper.

6 CONCLUSION

In this paper, we focus on discriminating social influence of distinct social groups to enhance item recommendation. First, we develop a novel recommendation model combining topic model with community detection to model selection behaviors influenced by both individuality and conformity, which is experimentally demonstrated to be superior to the state-of-art recommendation methods in terms of recommendation accuracy. Second, the proposed model demonstrates good explainability and interpretability in terms of topic semantics and community interests. Finally, we find out that different users exhibit different levels of conformity tendency and the conformity is positively correlated with the number of selected items and the diversity of individual interests, and negatively correlated with the heterogeneity of social groups. In our future research, we will further study how to utilize the findings about conformity to applications such as recommendation and find relevant theory behind it. In addition, it is worthy of exploring the factor of social dynamics, namely whether the evolution of friend groups is correlated with the dynamics of user interests and how the correlations can be leveraged to enhance recommendation quality.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China under grant No. 71272029, 71432004 and 71490724, the 863 program under grant No. 2014AA015204, and the Beijing Municipal Natural Science Foundation under grant No. 4152026.

References

Bernheim, B. D. (1994). A theory of conformity. Journal of Political Economy, 102(5), 841-877.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation. the Journal of Machine Learning Research, 3, 993-1022.

Cialdini, R. B. and Goldstein, N. J. (2004). Social influence: Compliance and conformity. Annu. Rev. Psychol., 55, 591-621.

Erosheva, E., Fienberg, S. and Lafferty, J. (2004). Mixed-membership models of scientific publications. Proceedings of the National Academy of Sciences of the United States of America, 101 Suppl 1(Suppl 1), 2004.

Fortunato, S. (2009). Community detection in graphs. Physics Reports, 486(3-5), 75-174.

Hu, Y., Koren, Y. and Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In Proceedings of the Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on.

Jamali, M. and Ester, M. (2010). A matrix factorization technique with trust propagation for recommendation in social networks. In Proceedings of the Fourth ACM Conference on Recommender Systems.

Kabassi, K. (2010). Personalizing recommendations for tourists. Telematics and Informatics, 27(1), 51-66.

King, I., Lyu, M. R. and Ma, H. (2010). Introduction to social recommendation. In Proceedings of the 19th International Conference on World Wide Web.

Liu, Q., Chen, E., Xiong, H., Ge, Y., Li, Z. and Wu, X. (2014). A cocktail approach for travel package recommendation. Knowledge and Data Engineering, IEEE Transactions on, 26(2), 278-293.

Liu, Q., Ge, Y., Li, Z., Chen, E. and Xiong, H. (2011). Personalized travel package recommendation. In Proceedings of the Data Mining (ICDM), 2011 IEEE 11th International Conference on.

Ma, H., King, I. and Lyu, M. R. (2009). Learning to recommend with social trust ensemble. In Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval.

Ma, H., Zhou, D., Liu, C., Lyu, M. R. and King, I. (2011). Recommender systems with social regularization. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining.

Marsden, P. V. and Friedkin, N. E. (1993). Network Studies of Social Influence. Sociological Methods & Research, 22(1), 127-151.

Mcpherson, M., Smithlovin, L. and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. Annual Review of Sociology, 15(4), 344–349.

Nallapati, R. M., Ahmed, A., Xing, E. P. and Cohen, W. W. (2008). Joint latent topic models for text and citations. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge

Discovery and Data Mining.
Porteous, I., Newman, D., Ihler, A., Asuncion, A., Smyth, P. and Welling, M. (2008). Fast collapsed Gibbs sampling for latent Dirichlet allocation. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
Rendle, S., Freudenthaler, C., Gantner, Z. and Schmidt-Thieme, L. (2012). BPR: Bayesian personalized ranking from implicit feedback. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence.
Rosen-Zvi, M., Griffiths, T., Steyvers, M. and Smyth, P. (2004). The Author-Topic model for authors and documents. In Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence.
Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th International Conference on World Wide Web.
Tan, C., Liu, Q., Chen, E., Xiong, H. and Wu, X. (2014). Object-oriented Travel Package Recommendation. ACM Transactions on Intelligent Systems and Technology (TIST), 5(3), 43.
Tang, J., Hu, X. and Liu, H. (2013). Social recommendation: a review. Social Network Analysis & Mining, 3(4), 1113-1133.
Wang, C. and Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
Yang, X., Steck, H. and Liu, Y. (2012). Circle-based recommendation in online social networks. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
Zhang, H., Qiu, B., Giles, C. L. and Foley, H. C. (2007). An LDA-based community structure discovery approach for large-scale social networks. In Proceedings of the Intelligence and Security Informatics, 2007 IEEE.
Zhang, J., Tang, J., Zhuang, H., Leung, W. K. and Li, J. (2014). Role-aware conformity influence modeling and analysis in social networks. In Proceedings of the AAAI Conference on Artificial Intelligence.
Zhao, T., McAuley, J. and King, I. (2014). 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.
Zhou, D., Manavoglu, E., Li, J., Giles, C. L. and Zha, H. (2006). Probabilistic models for discovering e-communities. In Proceedings of the 15th International Conference on World Wide Web.