# Exploring Personal Impact for Group Recommendation

Xingjie Liu Yuan Tian Mao Ye Wang-Chien Lee
The Pennsylvania State University
State College, PA 16801, USA
{xzl106, yxt144, mxy177, wlee}@cse.psu.edu

## ABSTRACT

Group activities are essential ingredients of people's social life. The rapid growth of online social networking services has greatly boosted group activities by providing convenient platform for users to organize and participate in such activities. Therefore, recommender systems, as a critical component in social networking services, now face new challenges in supporting group activities. In this paper, we study the *group recommendation* problem, i.e., making recommendations to a group of people in social networking services. We analyze the decision making process in a group to propose a *personal impact topic* (PIT) model for group recommendations. The PIT model effectively identifies the group preference profile for a given group by considering the personal preferences and personal impacts of group members. Moreover, we further enhance the discovery of personal impact with social network information to obtain an *extended personal impact topic* (E-PIT) model. We have conducted comprehensive data analysis and evaluations on three real datasets. The results show that our proposed group recommendation techniques outperform baseline approaches.

## Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information Filtering; J.4 [**Computer Applications**]: Social and Behavior Sciences

## General Terms

Algorithms, Experimentation.

## Keywords

Recommender Systems, Probabilistic Generative Model, Group Recommendation.

## 1. INTRODUCTION

Group activities are essential ingredients of people's social life. With recent development of social networking services, such as Facebook[1] and Meetup[2], organizing and participating in group activities (e.g., dining out, movie watching, vacations and parties) become very easy. As more and more people are willing to share their social relationships and activities online, new web services today need to target not only on individual users but also user groups. As a key technology in modern web services, recommender systems are now facing vigorous new demands and challenges in making recommendations for a group of people.

While making significant advances in the recommender system technology, most of the prior research studies in this topic area have focused on providing recommendations for individuals, which unfortunately can not be effectively applied for *group recommendations*, i.e., making recommendations for a group of people. Notice that a common idea behind recommender systems that make personalized recommendations for individuals is to discover users' preference profiles (either from user ratings or item text descriptions) in order to identify items that best match the profiles of targeted users. However, for group recommendations, we argue that a good recommender system not only needs to model users' individual preferences but also understands how a decision among group members is reached.

The pioneering researches on group recommendation have been dated back to MusicFX [14], a music recommender system built for gym users in 1998. Over these years, important findings and solutions have been reported in various studies on web/news pages [18], tourism [15], music [8], TV programs [25] and movies [17]. In this paper, we aim to study the problem of group recommendation, under the context of group activities in social events. Our research is unique from previous research in both of the problem and solution aspects.

- Problem aspect - Prior research studies on group recommendations [18, 17], mostly assume static groups which consist of stable membership (e.g., a movie interest group). In this study, we are targeting on "ad hoc" groups (e.g., people gathering for a dinner). In addition, different from [14, 17], where individual user's ratings or profiles are made available explicitly, we avoid such a strong assumption and consider only the group activity history.
- Solution aspect - Prior studies on group recommendation systems usually explore heuristics to apply some interesting strategies in recommendations, for example, PloyLens [17] adopts *least misery* (i.e., the least

[1]www.facebook.com
[2]www.meetup.com


*CIKM'12,* October 29–November 2, 2012, Maui, HI, USA.

satisfied user's preference) as the main strategy for its group recommendations. Aiming to better understand the latent factors behind group activities, in our study, we propose a probabilistic model to capture both of a user's own preference and her influence (termed as *personal impact*) to a group.

In this paper, we propose a probabilistic model, namely, the personal impact topic (PIT) model, for group recommendations. This model naturally captures the process in which an item is selected collectively by an ad hoc group. In this model, user preferences over items are abstracted into a number of latent topics. As such, each user's personal preference is modeled as a distinct distribution over all these topics, while each topic is expressed as a distinct distribution over all the items. Now a key issue is how an item is selected by a group of users, each of whom has her own preferences. While each group member may contribute to the group decision, we model the decision making process for an ad-hoc group with two key points. First, group decision on item selection follows the personal preferences of group members. Second, for a given item selection event (i.e., a decision on a group activity), a group member is probabilistically chosen to be a *group decision representative* (or representative for short)[3]. In this model, we introduce the personal impact parameter to control this process such that users with higher personal impact values shall have higher probabilities to serve as a group representative.

Our contributions can be summarized as follows:

- Group recommendation is essential in today's web services, especially in social networking systems that encourage various group activities. In this paper, we tackle a general group recommendation problem, where a group is formed in an ad hoc manner. In our study, we require only groups' item selection history (i.e., we do not need individuals' item selection history).
- We capture a group item selection process in a probabilistic model, called *personal impact topic* (PIT) model, that introduces the notion of personal impact to differentiate contributions of group members to a group decision. We perform statistical tests on real datasets to show that people do have different personal impacts in group decisions.
- Based on PIT model, we develop learning algorithms to learn personal impacts. Accordingly, we are able to use these parameters to construct group profiles from individual's user profile to make group recommendations. Additionally, we utilize social information as additional features to improve the personal impact estimations and thus further alleviate the potential problem of over-fitting.
- Finally, we evaluate our proposed group recommendation methods using three real dataset collected from whrrl.com and meetup.com. We consider three scenarios of group recommendations, i.e., Whrrl POI visit recommendation , Meetup tag selection recommendation and Meetup event venue selection recommendation. Experimental results show that our proposals significantly outperform the compared baselines.

[3]Please note that this representative is not designated as the decision maker but to denote the "probable source" of group decision made.

The rest of the paper is organized as follows. In Section 2, we review related works for both recommender systems designed for individual users as well as those for a group of users. In Section 3, we define the tackled group recommendation problem and provide some background for the proposed probabilistic models. The personal impact parameters are introduced in Section 4. Accordingly, we propose the personal impact topic (PIT) model to address the proposed recommender problem and extend the PIT model by exploring additional social features. In Section 5, we evaluate our proposed models using three datasets and conclude the paper in Section 6.

## 2. RELATED WORKS

Here, we review some related works, including general recommender systems, recommender systems for groups, and latent semantic models applied to recommender system.

**Item recommender system:** Item recommendation is an essential service for e-commerce and web services (e.g. netflix.com and amazon.com). The goal is to recommend a list of items that a targeted user may be interested in. Collaborative filtering and content-based techniques are two widely adopted approaches for recommender systems [1, 22]. Both of them discover users' personal profiles and utilize these profiles to find relevant items. Collaborative filtering techniques [7, 9, 21] automatically predict relevant items for a given user by referencing item rating information from other similar users. Content-based techniques [16] make recommendations by matching a user's personal profile to descriptive information of items. Although both techniques are complimentary to each other, collaborative filtering gains huge popularity among modern web services because it is easier to obtain item ratings explicitly (e.g. from movie rating by netflix.com) or implicitly (e.g. click information from amazon.com) from users than to collect interest descriptive information from users. Therefore, many techniques adopt the ideas of collaborative filtering. The classic *memory-based* collaborative filtering technique [19, 21] identifies a group of $k$ users that have the most closed rating results with the targeted user, and merge ratings from these $k$ users to find the top-$N$ recommendations. However, there are several limitations for the memory-based CF techniques, such as when data are sparse and the common items are few, it is problematic to find similar users. To achieve better prediction performance and overcome shortcomings of memory-based CF algorithms, *model-based* CF approaches, including matrix factorization [6], clustering CF models [23], and latent semantic models [10], have been investigated.

**Group recommender system:** Group recommender systems make item recommendations for a user group. Existing group recommender systems have been applied in web/news pages [18], tourism [15], music [14, 8], and TV programs and movies [17, 25]. Generally speaking, two main approaches have been proposed for group recommendation [11]. The first one creates an aggregated profile for a group based on individual profiles of its group members and makes recommendations based on the aggregated group profile [14, 25]. The second approach *aggregates* the recommendation results for individual members into a single group recommendation list. In other words, recommendations (i.e., ranked item lists) for individual members are created independently and then aggregated into a joint group recommendation list [4], where the aggregation functions could be based on average or least misery strategies [13, 17]. However, the group recommendation system can go beyond simple strategies such as averaging or least misery. In fact, the group member's personality and social status may determine how much weight his preference would determine group selections. Intrigue [3] take roles of tourist group members into account to recommend places to visit, (such that children would follow their parents' plans). The group members' disagreement to candidate items are considered in the recommendation score function in [2]. In this paper, we introduce the notion of personal impact which differentiates the contribution of group members to a decision in a different way. Personal impacts are useful for constructing group profiles to make good recommendations for groups.

**Probabilistic Latent Topic Models**. Latent topic models analyze relationships between a set of documents and the words, and identify latent topics that are represented as distributions over words in a corpus. In [10], a probability topic model is proposed to recommend item for individual users. In [24], a latent topic model incorporating social influence has been proposed to make better item recommendations. Latent Dirichlet Allocation (LDA) [5], a well known latent semantic model for text summarization, adopts Dirichlet priors for document topic distributions and word distributions. The principle of LDA is applicable under the context of item recommendation by mapping document/word entities to user/item entities. As an extension to LDA, author topic (AT) model [20] analyze co-authored document text, and therefore discover each author's expertise and contributions to each co-authored documents. Our approach focus on group recommendation problem, and is related to LDA [5] and author topic model [20].

# 3. PRELIMINARY

In this section, we first formally define the group recommendation problem studied in this paper. Then we adopt a probabilistic modeling approach for document analysis, called author topic model, to serve as the basis for addressing the group recommendation problem.

## 3.1 Problem Definition

Let $U = \{u_1, u_2, \cdots, u_V\}$ be a set of $V$ users, and $W = \{w_1, w_2, \cdots, w_M\}$ denote a set of $M$ items. A history log of group-item selection can be written in the form of $\{\langle G_1, s_1\rangle, \langle G_2, s_2\rangle, \cdots \langle G_N, s_N\rangle\}$ where $G_n = \{u_i\}$ and $s_n = w_j$ denote an ad hoc group and the item selected by the group, respectively[4]. Let $\vec{G}$ collectively denote the $N$ ad hoc groups and $\vec{s}$ be the collection of $N$ selections made by them. $H = \langle \vec{G}, \vec{s}\rangle$ represents the history of all the $N$ group-item pairs. Given $H$, the group recommendation problem is defined as to construct a ranked recommended list of items for a targeted group $G_x$, in accordance with the likelihood of the group to select those items.

We aim to adopt the collaborative filtering paradigm to address the group recommendation problem without requiring additional context information for users, groups or items. Also, we assume an input dataset of implicit ratings among groups and items in forms of clicks, tags, likes, checkins, etc, which are generally available in various web services, especially social networking systems.

The group recommendation problem defined above brings us several challenges. First, the groups are formed in ad hoc manner and usually ephemeral. For example, three friends may have dinner together and checkin the restaurant in a social network service. The user groups for such one-off social activities are usually not persistent, as different groups may be formed on another day. Thus, we do not treat these groups as pseudo users to apply single user recommendation techniques. There are very few item selections logged for each group, and some targeted group even has no previous history at all, leading to severe cold-start problem. Second, many services are group-centric which do not keep track of individual users' item selections/ratings. Thus, it requires a technique that would construct user preference profiles purely from group activities. Finally, how groups reach a decision to select an item is not well answered in previous studies. In [3, 2, 17, 25], group recommendations are made based on various strategies. In this study, we aim to model the impacts from different group members on group item selections, and then exploit such information in making recommendation for groups.

## 3.2 Extended Author Topic Model

To solve the group recommendation problem defined above, we need to address two issues: 1) how to profile preferences of individual users, and 2) how to model the group selection decision making process, given the preferences of its members. To address these issues, we propose to use the latent topic modeling approach which is known for achieving good recommendations even when dataset is sparse [10, 24]. Moreover, the model provides an interpretation of the item selection process.

Without loss of generality, we assume there are $K$ latent topics $T = \{t_1, t_2, \cdots, t_K\}$, each of them represents a different multinomial distributions over items. Every topic can be considered as a soft-cluster of items, where items with higher probabilities in the same topic tend to be picked by similar groups. A user's preference is modeled as a mixture of topics, which is also a multinomial distribution. To model the group preference profile, we assume the preference profile is a mixture of the group members' preferences. Thus, we use an election process to describe this mixture. More specifically, for each item selection observation (logged as a historical transaction) , we assume a representative is probabilistically selected from the group. Then, the item is probabilistically picked based on this representative's personal preferences. Because every member has a chance to be elected as the representative, the overall probability of an item selected by the group can reflect the mixture of the group members' personal preferences.

While there is no prior latent topic model proposed for the group recommendation problem, we observe that the Author Topic (AT) model [20] proposed for document-authorship analysis may provide some hints as the authors of a paper naturally form a group. In the AT model, the creation of a document is modeled as follows. For each word in the document, an author is drawn from the document's author group. Next, a topic is selected following the author's topic distributions, and from the topic's word distributions, the word is selected. Each author's topic distribution and each topic's

[4]Note that it is possible that a group is duplicated (e.g. $G_x = G_y, (x \neq y)$ ). However, this situation is rare since there is quite a few item selection history for a distinct group.

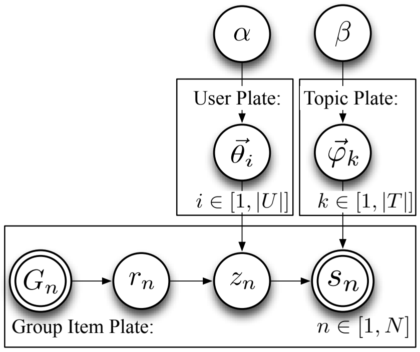



**Figure 1: Author topic model under the group item selection context. $G_n$ and $s_n$ (in double circled nodes) represent the observed group and the selected item, respectively. $r_n$ is the corresponding representative, and $z_n$ is the generated topic.**

word distribution are modeled as multinomial distribution with Dirichlet prior.

Through a careful analysis, we find that the above-described document generation process shares some similar properties with our group item selection that we envisage. More specifically, if we consider the authors of a document a group and items selected by this group collected as words of the document, we may potentially apply the AT model to the group item selection process (as depicted in the plate representation of an author topic model in Fig. 1). As shown, the $n$th group $G_n$ generates the $n$th representative $r_n \in G_n$ following the uniform multinomial distribution. The latent topic $z_n$ is then generated following the $r_n$'s topic distribution, which is a multinomial distribution $\vec{\theta}_i$ following the Dirichlet prior $\mathcal{D}(\vec{\alpha})$. Finally, the item $s_n$ is selected following the topic $z_n$'s word distribution, which is also a multinomial distribution $\vec{\varphi}_k$ following the Dirichlet prior $\mathcal{D}(\vec{\beta})$.

Furthermore, in AT model, the latent variables $\vec{r}$ and $\vec{z}$ are learned from the observations of both $\vec{G}$ and $\vec{s}$. In group recommendation, we need to not only learn the latent variables, but also make prediction of $s$ when a targeted group $G$ is given. Here, we extend the AT model with a prediction function. Specifically, after the latent parameters $\vec{\theta}$ and $\vec{\varphi}$ are learned, the probability that a user $u_i$ selects an item $w_j$ can be derived as:

$$\Pr^{\text{AT}}(w_j|u_i) = \sum_{k=1}^{|T|} \Pr(w_j|t_k) \cdot \Pr(t_k|u_i) = \sum_{k=1}^{|T|} \varphi_{j,k} \cdot \theta_{k,i} \quad (1)$$

Following the generative process in Fig. 1, each group representative is generated with equal probabilities. Thus an item $s_n$ is accessed by a given group $G_n$ with the following probability:

$$\begin{aligned}\Pr^{\text{E-AT}}(s_n = w_j|G_n) &= \sum_{\forall u_i \in G_n} \Pr(r_n = u_i|G_n) \cdot \Pr^{\text{AT}}(w_j|u_i) \\ &= \sum_{\forall u_i \in G_n} 1/|G_n| \cdot \Pr^{\text{AT}}(w_j|u_i)\end{aligned} \quad (2)$$

Eqn. (2) directly uses the AT model parameters to estimate the future group item selections. We call this recommendation method as Extended-AT (E-AT) model and use it as a baseline in this study.

# 4. PERSONAL IMPACT TOPIC MODEL

The E-AT model assumes that each group member has an equal probability to be a representative (see Eqn. (2)). In this section, we argue that group members have different impacts in group item selection decisions. Hence, we introduce the notion of *personal impact* parameter to model this difference. Accordingly, we propose a personal impact topic (PIT) model by incorporating the personal impacts to control the generation of group representatives. We also provide the model learning algorithm to infer personal impacts, topic distribution of users and item distributions of topics. Finally, to over-come the over-fitting problem when input dataset is sparse, we extend the PIT model by exploring additional social features.

## 4.1 Personal Impact

From our daily life experience, we may observe that many group activities are usually proposed and organized by some active people, while other participants mostly follow the agenda. Similarly, in decisions such as item selections, some group members may out-speak others in expressing their preferences (due to expertise, authority, or other personality factors) and thus have more influence on the decision. Therefore, the group selection decisions may reflect their opinions more than other indiscriminate group members. To quantify such influence, we introduce a latent variable of *personal impact* in our model to better capture how likely a user would influence a decision made within a group.

In our model, the personal impact of each individual user $u_i$ can be abstracted as a positive numerical value $\gamma_i$, where a higher value indicates greater impact on a group's decisions. For $|U|$ users, we use $\vec{\gamma}$ to represent all the personal impacts. More specifically, $\gamma_i$ reflects the *relative* probability to influence a group decision. In other words, the probability that a user is the representative of the group is:

$$\Pr(r_n = u_i|G_n) = \begin{cases} \frac{\gamma_i}{\sum_{\forall u_j \in G_n} \gamma_j} & \text{if } u_i \in G_n \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

In Eqn.(3), the probability for a group member $u_i \in G_n$ to be selected as a representative is proportional to his personal impact. Otherwise, a user has zero probability to be a representative if he does not belong to the group. Following this definition, Eqn. (2) in E-AT model assumes that group members have the same personal impact, i.e., $Pr(r_n = u_i|G_n) \equiv 1/|G_n|$, which is a strong assumption. Next, we introduce the personal impact topic (PIT) model and develop techniques to learn personal impacts from group item selection dataset.

## 4.2 Personal Impact Topic Model

Fig. 2 illustrates the personal impact topic (PIT) model in plate graph. In the figure, personal impact parameters are introduced in the user plate, one for each user. Thus, the group item selection process is composed as three steps: 1) representative $r_n = u_i$ is drawn from $G_n$ according to group members' personal impacts in Eqn. (3), 2) representative $u_i$ probabilistically selects a topic $z_n = t_k$ following his topic distribution $\vec{\theta}_i$, and then 3) the item $s_n$ is probabilistically selected following $t_k$'s item distribution $\vec{\varphi}_k$. Since users may have different personal impacts, the group representative selection is biased towards those with higher personal impacts. That is, people with higher personal impacts have higher

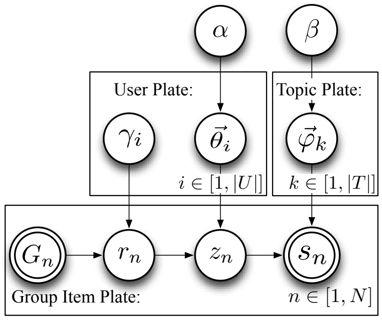



**Figure 2: Personal Impact Topic (PIT) Model**

probabilities to influence the group to select items following their preferences. For simplicity, we do not assume a Bayes prior for personal impacts in the PIT model.

## 4.3 PIT Model Parameter Inference

We adopt the Gibbs sampling method to infer model parameters as it provides fast and effective solutions to our model. Gibbs sampling is a MCMC method that samples latent variables based on the model's posterior distribution. We use $\vec{s}$ to represent all the items previously selected by groups $\vec{G}$, and the latent representatives and topics for those group-item pairs can be denoted as $\vec{r}$ and $\vec{z}$, respectively. Following the PIT model in Fig. 2, the posterior is:

$$\begin{aligned}
&\Pr(\vec{r}, \vec{z} | \vec{G}, \vec{s}, \vec{\gamma}, \alpha, \beta) \\
&\propto \Pr(\vec{r} | \vec{G}, \vec{\gamma}) \cdot \Pr(\vec{z} | \vec{r}, \alpha) \cdot \Pr(\vec{s} | \vec{z}, \beta) \\
&= \prod_{n=1}^{N} \frac{\gamma_{r_n}}{\sum_{u_j \in G_n} \gamma_j} \cdot \prod_{i=1}^{|U|} \frac{B(\vec{C}^{UT}_{i,*} + \vec{\alpha})}{B(\vec{\alpha})} \cdot \prod_{k=1}^{|T|} \frac{B(\vec{C}^{TW}_{k,*} + \vec{\beta})}{B(\vec{\beta})}
\end{aligned} \tag{4}$$

In Eqn. (4), $\vec{C}^{UT}$ represents a counting matrix where $\vec{C}^{UT}_{i,k}$ is the number of times that the $k$th topic is sampled for the $i$th user. Hence, $\vec{C}^{UT}_{i,*}$ denotes a vector of the $i$th user's sample topic counts. Similarly, $\vec{C}^{TW}_{k,*}$ is a vector of the $k$th topic's item counts. In Eqn. (4) the first term is the product of all the representative selection probabilities under the personal impacts. The second term is the joint probability of the respective users' selected topics. Because the Dirichlet prior is conjugate to the multinomial distribution of each users' topic selection, the joint probability can be simplified as $\frac{B(\vec{C}^{UT}_{i,*} + \vec{\alpha})}{B(\vec{\alpha})}$, where $B(\cdot)$ represents extended beta function. Similarly, the last term is the joint probability for each topic's item selections.

In order to sample the posterior probability, a Gibbs sampling method computes the full conditional probability for one single representative user and topic assignment $\langle r_n, z_n \rangle$ conditioned on all the other assignments $\vec{r}_{\neg n}, \vec{z}_{\neg n}$ as:

$$\begin{aligned}
&\Pr(r_n = u_i, z_n = t_k | \vec{G}, \vec{s}, \vec{r}_{\neg n}, \vec{z}_{\neg n}, \vec{\gamma}, \alpha, \beta) \\
&\propto \frac{\gamma_i}{\sum_{u_j \in G_n} \gamma_j} \cdot \frac{\vec{C}^{UT}_{i,k,\neg n} + \alpha}{\sum_{k'} (\vec{C}^{UT}_{i,k',\neg n} + \alpha)} \cdot \frac{\vec{C}^{TW}_{k,j,\neg n} + \beta}{\sum_{j'} (\vec{C}^{TW}_{k,j',\neg n} + \beta)}
\end{aligned} \tag{5}$$

The subscript $\neg n$ indicates that the counting matrix or the sample vector *excludes* the $n$th sample of $\langle r_n, z_n \rangle$. The derivation of Eqn. (5) is similar to the AT model's posterior in [20], except that the first personal impact term. Eqn. (5) provides a method to sample $\langle r_n, z_n \rangle$ in a batch. Note that user profiles $\vec{\theta}$ and topic profiles $\vec{\varphi}$ are not presented in Eqn. (5). These two parameter spaces are integrated out to produce a simplified probability computation. The actual $\vec{\theta}$ and $\vec{\varphi}$ parameters are inferred from sampled $\vec{r}$ and $\vec{z}$ after the Gibbs sampling converges. Hence, this is called a *collapsed* Gibbs sampling method.

Although by repeatedly sampling $\vec{r}$ and $\vec{z}$, we can produce fairly good posterior samples, Eqn. (5) assumes the personal impacts $\vec{\gamma}$ are known and fixed. However, in our study, we also need to estimate personal impact values as well. To do this, we follow the maximum likelihood paradigm. The rationale is that after each round of Gibbs sampling, $\vec{r}$ is updated to reflect the new representative distributions. Thus personal impact parameters $\vec{\gamma}$ are also updated so that the likelihood of generating samples $\vec{r}$ is maximized. In particular, the likelihood of observing $\vec{r}$ given the groups $\vec{G}$ and personal impacts $\vec{\gamma}$ is:

$$\mathcal{L}(\vec{r} | \vec{\gamma}, \vec{G}) = \log \left( \prod_{n=1}^{N} \Pr(r_n | G_n) \right) = \sum_{n=1}^{N} \log \frac{\gamma_{r_n}}{\sum_{u_j \in G_n} \gamma_j} \tag{6}$$

To find the maximum likelihood estimation of $\vec{\gamma}$ from Eqn.(6), we adopt the gradient descent method. Specifically, by taking a derivative regarding to each $\gamma_i$, we have:

$$\frac{\partial \mathcal{L}(\vec{r} | \vec{\gamma}, \vec{G})}{\partial \gamma_i} = \sum_{\forall r_n = u_i} \frac{1}{\gamma_i} - \sum_{\forall G_n, u_i \in G_n} \frac{1}{\sum_{\forall u_j \in G_n} \gamma_j} \tag{7}$$

First we initialize all the personal impacts $\vec{\gamma}$ to a default value 1, which indicates that all users are assumed to have equal impacts *when there are no observations*. By iteratively take steps proportional to the personal impact gradients (see Eqn. (7)), we are able to find a set of local optimal personal impact parameters $\hat{\vec{\gamma}}$ that fits the observations $\vec{r}$. Note that, by definition, the personal impacts $\vec{\gamma}$ should be non-negative. Thus, we introduce a positive number $\delta$ as the lower bound of personal impacts. That is, whenever a personal impact is made lower than $\delta$, its value is set to $\delta$ (fixed to 0.1 in our experiments). As such, we avoid some overfitting effects by ensuring that every user has a positive probability to be selected as a representative.

To put everything together, the Gibbs sampling algorithm for PIT model inference can be presented in Algorithm 1. This algorithm does two things: 1) output a Gibbs sampling $\vec{r}$ and $\vec{z}$ that converge to the posterior distribution of Eqn. (4) and 2) output a maximum likelihood estimation of personal impacts $\hat{\vec{\gamma}}$ so that the representative sampling of $\vec{r}$ can be explained. The first part of Algorithm 1 from line 1 to 5 initializes the latent variable samples $\vec{r}$ and $\vec{z}$ by randomly drawing the users from the corresponding groups $G_n$ and topics $1 \sim |T|$. During the initialization, the algorithm maintains two counting matrices $\vec{C}^{UT}$ and $\vec{C}^{TW}$. These two counting matrices are used to efficiently compute the sampling posterior as in Eqn. (5). From line 6, The transitions between successive states of the Markov chain result from repeatedly drawing $\langle r_n, z_n \rangle$ from their distributions conditioned on all other variables as in Eqn. (5). Also, after each iteration of $\vec{r}$, $\vec{z}$ sampling, we use the gradient descend method to fit $\vec{\gamma}$ in line 11.

After Algorithm 1 finishes, we obtain the samples of $\vec{r}$, $\vec{z}$ and the maximum likelihood estimation of personal impact

**Algorithm 1:** PIT model Gibbs sampling algorithm.

**Input**: Group Item Access History $\vec{G}, \vec{s}$, super-parameters: $\alpha$, $\beta$, max_iter.
**Output**: Sampled $\vec{r}, \vec{z}$, personal impact $\vec{\gamma}$.

$\vec{\gamma} \leftarrow \vec{1}$;
**for** $n \leftarrow 1, 2, 3, \cdots, N$ **do**
  Sample user $r_n \leftarrow x \sim \text{Mult}(1/|G_n|)$;
  Sample topic $z_n \leftarrow y \sim \text{Mult}(1/|T|)$;
  Update counters $\vec{C}^{UT}$ and $\vec{C}^{TW}$;
**for** *iter* $\leftarrow 1, 2, 3 \cdots$ *max_iter* **do**
  **for** $n \leftarrow 1, 2, 3, \cdots, N$ **do**
    Remove $r_n$, $z_n$ from counters $\vec{C}^{UT}$ and $\vec{C}^{TW}$;
    Sample representative $r_n$ topic $z_n$ block $r_n, z_n \sim \Pr(r_n, z_n|\vec{G}, \vec{s}, \vec{r}_{\neg n}, \vec{z}_{\neg n}, \vec{\gamma}, \alpha, \beta)$ following Eqn. (5);
    Update counters $\vec{C}^{UT}$ and $\vec{C}^{TW}$;
  Gradient descend fit $\hat{\vec{\gamma}}$ given $\vec{G}, \vec{r}$;
**return** $\vec{r}, \vec{z}$ and $\hat{\vec{\gamma}}$;

$\vec{\gamma}$. To infer the user and item profiles $\vec{\theta}$ and $\vec{\varphi}$, we have them following the Dirichlet distributions:

$$\begin{aligned} \Pr(\vec{\theta}_i|\vec{r}, \vec{z}, \alpha) &\sim \texttt{Dir}(\vec{C}^{UT}_{i,*} + \vec{\alpha}) \\ \Pr(\vec{\varphi}_k|\vec{r}, \vec{z}, \beta) &\sim \texttt{Dir}(\vec{C}^{TW}_{k,*} + \vec{\beta}) \end{aligned} \tag{8}$$

Usually, we only need to use the maximized probability estimation of these parameters, which can be written as

$$\begin{aligned} \hat{\theta}_{i,k} &= \hat{\Pr}(t_k|u_i) = \frac{\vec{C}^{UT}_{i,k} + \alpha}{\sum_{k'}(\vec{C}^{UT}_{i,k'}) + \alpha)} \\ \hat{\varphi}_{k,j} &= \hat{\Pr}(w_j|t_k) = \frac{\vec{C}^{TW}_{k,j} + \beta}{(\sum_{j'}(\vec{C}^{TW}_{k,j'}) + \beta)} \end{aligned} \tag{9}$$

Therefore, the probability for a targeted group $G_x$ to select an item $w_j$ is :

$$\begin{aligned} \hat{\Pr}^{\text{PIT}}(w_j|G_x, \hat{\vec{\theta}}, \hat{\vec{\varphi}}, \hat{\vec{\gamma}}) &\propto \sum_{\forall u_i \in G_x} \hat{\gamma}_i \cdot \left( \sum_{k=1}^{|T|} \hat{\theta}_{i,k} \cdot \hat{\varphi}_{k,j} \right) \\ &\propto \sum_{\forall u_i \in G_x} \hat{\gamma}_i \cdot \hat{\vec{\theta}}_i \cdot \hat{\vec{\varphi}}_j^{\,T} \end{aligned} \tag{10}$$

Eqn. (10) is then used to make item recommendations.

## 4.4 PIT Model Extension

In the previous section, we discussed the PIT model to exploit the differences of personal impacts in group item selections. Yet, since the PIT model introduces an additional parameter $\gamma$ for each user, if the group-item dataset is sparse, learning personal impact parameters suffers from overfitting problems. More specifically, if a user $u_i$ has only participated in very few group activities, then estimating personal impact $\gamma_i$ using PIT model may not be practical. In this section, we explore the relationship between the personal impacts and other features (especially social network features). By including these related features to PIT model, we expect these additional data shall further alleviate the over-fitting problem and thus achieve more reliable group recommendations.

As shown in Fig. 3, we introduce another feature $e_i$ as the evidence of the personal impact parameter. Consider $e_i$ as a

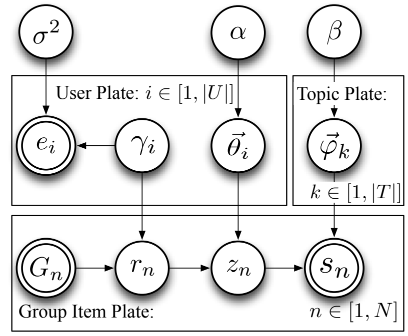



**Figure 3: Extended Personal Impact Topic Model**

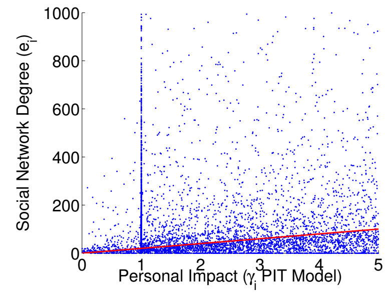



**Figure 4: Personal Impacts vs. Degree distributions**

numerical observation feature for the user $u_i$, and personal impact $\gamma_i$ is correlated with $e_i$. Candidates for this feature can be network degree, clustering coefficient, betweenness, page rank of the social network etc. Here, we studied the correlation between personal impact and social network degree as an example. In Fig. 4, we plot the personal impact parameters learned from PIT model against corresponding user social network degrees on the Whrrl dataset (dataset details can be found in Section 5.1). From this figure, although the personal impacts learned form PIT model may have overfitted values, users with higher personal impacts still are more likely to have larger social network degrees. Actually, the computed correlation between $\vec{\gamma}$ and $\vec{e}$ in this dataset is 0.596, and the slope of the red fitted line is 6.13.

To model the correlation between $e_i$ and $\gamma_i$, we assume a linear model between them with Gaussian error, leaving other possible non-linear models to future works. Under our linear model assumption, the personal impact log likelihood can be expressed as:

$$\mathcal{L}(\vec{r}, \vec{e}|\vec{G}, \vec{\gamma}) = \sum_{n=1}^{N} \log \frac{\gamma_i}{\sum_{u_j \in G_n} \gamma_j} - \frac{1}{2} \sum_{i}^{|U|} (\frac{a + b\gamma_i - e_i}{\sigma})^2 \tag{11}$$

Under this log-likelihood, we not only need to estimate $\vec{\gamma}$ but also need to find the linear relation's parameter $a$ and $b$. Because the correlation is modeled as linear, parameter $a$ and $b$ can be found the same as linear regression, but the gradient for each $\gamma_i$ should be updated as:

$$\begin{aligned} \frac{\partial \mathcal{L}(\vec{r}, \vec{e}|\vec{G}, \vec{\gamma})}{\partial \gamma_i} &= \sum_{\forall r_n = u_i} \frac{1}{\gamma_i} - \sum_{\forall G_n, u_i \in G_n} \frac{1}{\sum_{\forall u_j \in G_n} \gamma_j} \\ &\quad - \sum_{i}^{|U|} \frac{b(a + b\gamma_i - e_i)}{\sigma^2} \end{aligned} \tag{12}$$

By replacing Eqn. (7) with Eqn. (12), we incorperate the personal impact related evidence into the PIT model. Intuitively, the evidence value for each user gives a hint to his personal impact estimation. The inclusion of the personal impact evidence penalizes those personal impacts that deviate $e_i$ more than $\sigma$ away. Therefore, a smaller $\sigma$ penalizes more the error in Eqn. (12), and therefore forces the learned personal impacts to be closer to the observed evidence. In this paper, we use empirical methods to determine a good selection of $\sigma$.

# 5. EVALUATION

In this section, we evaluate the proposed models with three real datasets, one collected from a location-based social network service whrrl.com and the other two from an event-based social network service meetup.com. We evaluate our proposed group recommendation models with three recommendation problems: 1) Whrrl group POI checkin recommendation, 2) Meetup group tag recommendation, and 3) Meetup event venue recommendation. These three recommendation problems have different data space and group-item selection behaviors, which reflect comprehensive performances of proposed approaches. For each dataset, we evaluate the proposed PIT model against the baseline methods. Since the Meetup datasets do not contain additional social network information, E-PIT model is only tested in Whrrl dataset.

## 5.1 Dataset Description

Whrrl.com is an online location-based social network that helps people to explore points-of-interests (POIs) (e.g. restaurants, movie theaters, parks etc). A group of people can checkin POIs and share the experiences via mobile devices. Given the group-POI checkin history, a recommendation task is to identify POIs that a target group has potential to visit. On the other hand, meetup.com is an online event-based social network that provides a platform for users to organize face-to-face social events. From the website, we collected two datasets that support two types of group recommendation problems. First, each meetup online group can specify some tags to represent common interests of group members, so we may recommend additional tags for groups based on existing ones. Second, to help people to organize social events at venues that fit participants' preferences on meetup.com, we may recommend event venues to users depending on event participations and past venue selections.

The datasets collected for these three group recommendation problems are summarized as in Table 1[5]. The first three rows show the total number of distinct users and groups, as well as the recommendation items (POIs, tags and venues). The fourth and fifth rows summarize average sizes of groups and how many groups each user participated in. Generally, smaller group sizes and more user-group participations lead to easier group recommendations. Finally, the sixth and seventh rows show how many items each *unique* group has selected, and how many selections each item received, respectively. The fewer items a group selects or the fewer group selections an item receives, the more sparse the dataset is, which results in more difficult recommendations. With different statistics on the three datasets, we are able to comprehensively test the proposed approaches in various application scenarios.

[5]Meetup dataset [12] can be downloaded from http://largenetwork.org/ebsn

## 5.2 Evaluated Recommendation Methods

| | |
|---|---|
| CF | Neighborhood collaborative filtering methods. Treat each group members as equal. |
| LM-CF | Neighborhood collaborative filtering methods. Use least miserable strategy to merge individual users' recommendations. |
| E-AT | Extended author topic model, introduced in Section 3.2 |
| PIT | Personal impact topic model, introduced in Section 4.3. |
| E-PIT | Extended personal impact topic model, introduced in Section 4.4. |

**Table 2: Evaluated Recommendation Methods**

In Table 2, we summarize the recommendation methods evaluated in this paper. In addition to the E-AT, PIT, E-PIT models, we also introduce two memory-based collaborative filtering methods as the baselines, namely, the classic memory-based CF method [19] (CF) and the least miserable CF [17] (LM-CF). Since our datasets provide group-item pairs instead of user-item pairs, we first project a group-item pair $\langle G_n, s_n\rangle$ into user-item selection pairs $\{\langle u_i, s_n\rangle | \forall u_i \in G_n\}$. After applying memory-based CF to the projected dataset, we make recommendations for a target group by averaging group members' individual preference scores. LM-CF was studied in [17] as an important strategy to make group recommendations. LM-CF is similar to CF method, but when making recommendation to a target group, each item's lowest score from group members is used to make recommendations.

For each method, we randomly sample 20% group-item pairs from the entire dataset as validation dataset. The other 80% data are used as training data. We use recall as the evaluation metric. Therefore, for each top-$n$ recommended items, we calculate

$$\texttt{recall} = \frac{\#\text{ of items that are included in top-}n\text{ recommendations}}{\#\text{ group-item pairs validated}} \qquad (13)$$

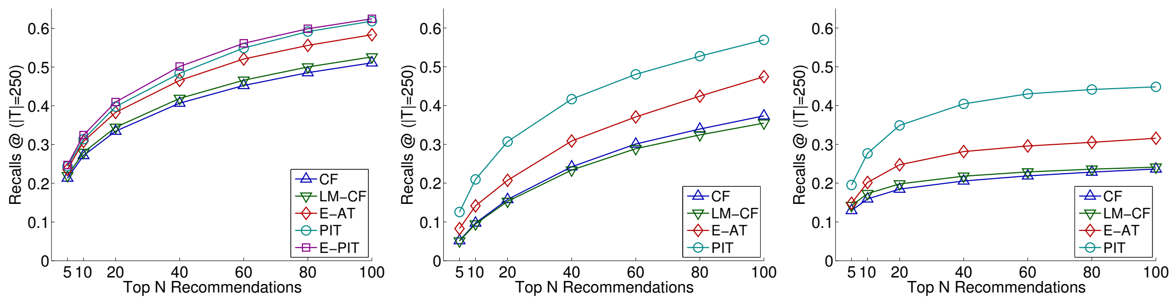
Therefore, the recall of each recommendation method is the percentage of targeted group's selected items that are included within recommendations.

## 5.3 Evaluation Results

The first set of results on recall vs. number of recommendations are reported in Fig. 5. We fix the total number of topics $K$ at 250, and vary the number of recommendations. As shown in Fig. 5, while the recall grows as the number of recommendations increases, PIT models achieve the best performance, followed by E-AT, and the CF approaches perform the worst. The results validate our proposed methods, in which the personal impacts of group members effectively improved recommendation results. CF and LM-CF fall behind others because these two methods not only ignore the difference among users, but also treat each group activity as a trivial union of individual behaviors per group member. By breaking up the group selection transactions into personal activities, the CF methods overlook the social indication in group activities. The extended author topic (E-AT) model considers each group as a whole, which effectively preserves the group's social information. It also

| Whrrl POI | | Meetup Group Tag | | Meetup Event Venue | |
|---|---|---|---|---|---|
| Parameters | Value | Parameters | Value | Parameters | Value |
| # Users ($\|U\|$) | $10,147$ | # Users ($\|U\|$) | $124,047$ | # Users ($\|U\|$) | $329,590$ |
| # Checkin Groups | $34,338$ | # Online Groups | $1,613$ | # Event Groups | $115,634$ |
| # POIs ($\|W\|$) | $15,940$ | # Tags ($\|W\|$) | $2,118$ | # Venues ($\|W\|$) | $47,606$ |
| Avg. Checkin Group Size | 3.98 | Avg. Group Size | 97.64 | Avg. Event Participants | 10.75 |
| Avg. Checkins for a User | 13.48 | Avg. Groups joined by a User | 1.205 | Avg. Events for a user | 3.77 |
| Avg. POI for a Group | 1.21 | Avg. Tags for a Group | 4.92 | Avg. Venues for an Event Group | 1.03 |
| Avg. Checkins for a POI | 2.15 | Avg. Groups for a Tag | 3.74 | Avg. Events for an Venues | 2.43 |

**Table 1: Data Set Summary**




(a) Whrrl POI dataset (b) Meetup group tag dataset (c) Meetup event venue dataset

**Figure 5: Recalls under # recommendations made ($n$)**

utilizes Dirichlet prior to prevent over-fitting, thus obtains better results. However, since the E-AT model is not aware of the differences of personal impacts, every group member is treated the same while making prediction for a target group, which degrades the recommendation performance. By mining the personal impacts from group-item pair histories, our proposed PIT model improves the recommendation performance significantly. To further alleviate the over-fitting of personal impact parameters, our E-PIT model exploits social network degrees as an additional feature $\vec{e}$ to guide the personal impact estimation, and achieves the best result on the Whrrl dataset.

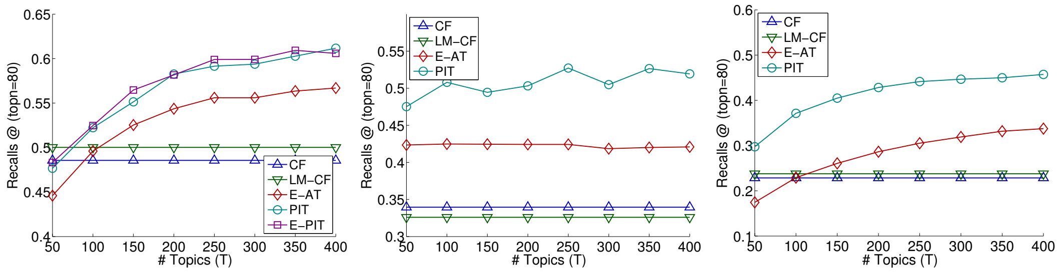
Fig. 6 illustrates the recall results when the number of recommendations is fixed at 80, and the number of topics varies. As CF and LM-CF do not involve topics, they have constant recalls. In Fig. 6 ($a$) and Fig. 6(c), when the number of topics increases, the recalls first increase and then flat out. This is because items in these two datasets have large data space (see row 3 of Table 1), a large number of topics can help to precisely differentiate the items. In Fig. 6 ($b$), because the number of tags (item space) is much smaller, the Figure shows that 100 topics are enough to model tags. In all three recommendation test cases, we can see that the proposed PIT and E-PIT models improve over the E-AT model by about 10%, and outperform the CF approaches by 15%-25%. These significant improvements demonstrated the effectiveness of our proposals.

## 5.4 Personal Impact Distribution and Significance

By learning the PIT and E-PIT models, we obtain a set of estimated personal impacts. One fundamental question that we need to answer is whether people do exhibit different personal impacts. Because only if most of people's personal impacts are significantly different from 1, the proposed PIT and E-PIT model can improve over E-AT model which assumes equal personal impacts.

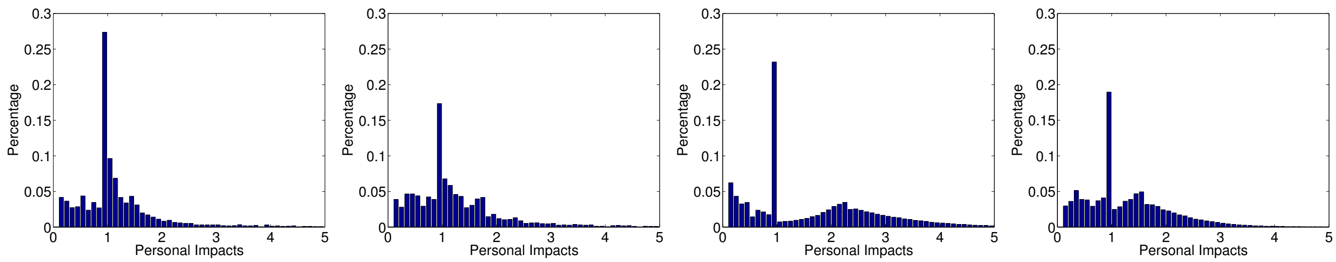
In Fig. 7, the histograms of the learned personal impacts are plotted. Fig. 7(a) and Fig. 7(b) demonstrate learned personal impacts from PIT model and E-PIT model for the Whrrl POI checkin dataset, while Fig. 7(c) and Fig. 7(d) show personal impact distributions for the Meetup group tag and event venue dataset obtained from PIT model, respectively. In all four results, we can see that there are also a great portion of users with personal impacts close to 1. These users are either because 1) they have limited historical data to derive personal impacts, or 2) as their group representative probabilities are balanced, their personal impacts are indeed 1. Through careful examination of the data, we found that most of these users have extremely sparse historical information in the dataset (e.g., they have only participated in one group and involved in only one item selection, and the other group members are inactive as well). Hence, we are not able to differentiate the personal impact of such users from other group members. This result reflects the case that these users have equal impact on group decisions. By comparing Fig. 7(a) and Fig. 7(b), we can find that by introducing additional social network degree features, E-PIT model improves over the PIT model by reducing the number of users with personal impacts closed to 1 from 28% to 17%. Thus, we believe about 10% of ambiguous personal impacts can be refined after introducing the social network degree.

On the other hand, Fig. 7 demonstrate the rest of users' personal impacts distributed relatively evenly above and below the value of 1, meaning that some of them influence their group's decision more, while others influence much less. Interestingly, Fig. 7(c) shows a drop of the distribution on $[0.5, 1)$ and $(1, 2]$. The reason for this is that each of Meetup group selects an average of 4.92 tags. Based on our analysis, tags selected by one group are usually related. Thus, one user with the interest matching the group's tags may have multiple chances to be sampled as representatives. The personal impacts of these users are therefore boosted multiple times, leading to higher impact values. Meanwhile, users




(a) Whrrl POI dataset (b) Meetup group tag dataset (c) Meetup event venue dataset

**Figure 6: Recalls under # topics (T)**




(a) Whrrl POI dataset (PIT Model) (b) Whrrl POI dataset (E-PIT Model) (c) Meetup group tag dataset (PIT Model) (d) Meetup event venue dataset (PIT Model)

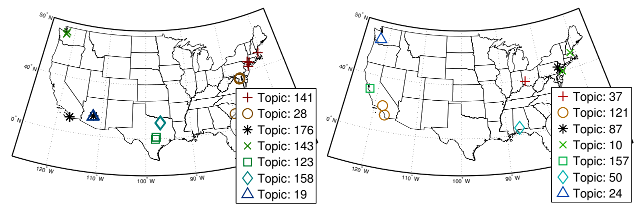
**Figure 7: Personal Impacts Learned from Dataset**

who do not match the group preference profile tend to gain even lower personal impact values. The learned personal impacts thus deviate more from 1. Fig. 7(d) has similar trends as in Fig.7(a), which does not have extreme personal impacts as Fig. 7(c) does, because each social event group usually only selects one venue.

To statistically confirm that most users have their personal impact significantly deviate from 1, we applied a statistic test on a non-hypothesis that the personal impact values equal to 1 for all the users. Accordingly, for an arbitrary user $u_i$, the expected number of times that $u_i$ is selected as representatives under this non-hypothesis is $E_i = \sum_{\forall G_n|u_i \in G_n} 1/|G_n|$. The observed times that $u_i$ is selected as representatives is $O_i = \sum_{\forall r_n|u_i=r_n} 1$. Therefore, for each user, the $p$-value for his personal impact is the probability that a more extreme statistic value to $O_i$ is observed. By taking this statistic test for each individual user's personal impact, we are able to find the percentage of users whose personal impacts are significantly not equal to 1 under the Type I error $\alpha = 0.05$. For these four results, the percentages are 56.4%, 68.0%, 50.7% and 61.0%, respectively. Thus, due to the limitation of our dataset size and user activities, we cannot show that all the users' personal impacts are significant other than 1. However, these results demonstrate that lots of users' personal impacts are indeed statistically significant, which proves our theory on utilizing personal impacts.

## 5.5 Topic Analysis

In this subsection, we conduct semantic investigations on the topics established by the proposed PIT model. In general, if the topics generated by the PIT model are semantically meaningful, we believe that the model is effective. Here, we show the analysis results with both location-centric data and text tag data.




(a) Whrrl POI dataset (b) Meetup event venue dataset

**Figure 8: Geographic distribution of sampled topics**

Fig 8 depicts item geographic location distributions of 7 selected topics on the Whrrl dataset and the Meetup event venue dataset. Each topic is represented by locations of 5 items with the highest generating probabilities. In both datasets, we can observe that the items of the same topic locate close to each other. The reason is that location-related group activities such as checkins and offline events are usually constrained by spatial distance, and people tend to move within small geographic areas most of the time. Thus, the items selected in these group activities fall into such small geographic areas. These results demonstrate that the PIT model has good ability to extract topics from pure group-item selections.

In Table 3, we present a list of topics composed of text tags. Given a set of topics $T = t_k$ identified by the PIT model, we associate each user $u_i$ with a topic that he/she is most interested in (i.e., $t_k$ with the largest $Pr(t_k|u_i)$). Then we sort the topics by number of users they are associated with. Intuitively, the more users associated with the topic, the more popular the topic is. As shown in Table 3, each row represents a topic and we list the top three tags for each topic ($w_j$ with highest $Pr(w_j|t_k)$). The most popular topic

| # Users | Top 3 Tags |
|---|---|
| 14,126 | hiking, outdoors, outdoor recreation |
| 6,850 | business networking, business referral networking, business connection |
| 6,279 | dating and relationships, singles, newly single |
| 6,248 | active dogs, dog play groups, socializing dogs |
| 6,186 | photo, digital camera, photography workshops |
| ...... | |
| 126 | meeting new people, social network, women |
| 119 | personal growth, self-improvement, law of attraction |
| 45 | social network, women, Christian parent |

**Table 3: Meetup group tag dataset PIT model topics**

is about hiking and outdoor activities, which matches our daily life experience that such activities are usually held on a group basis. It also indicates that people love to use online social network services to organize such outdoor events. Further, we find that more elaborated and specific tags attract more users, such as business networking, photography and dogs, and this indicates that people are willing to participate in group activities with focused topics. On the opposite, generic and vague tags like "meeting new people" degrades the model's ability of separating topics, and they are not as attractive to users. Therefore, we can conclude that our PIT model works well on identifying interesting topics.

# 6. CONCLUSION

As social group oriented services are growing rapidly in recent years, recommender systems need to adapt to this new trend to offer better service. This paper studies the group recommendation problem that aims at recommending items to ad hocly formed target groups. We proposed a probablistic *personal impact topic* (PIT) model to effectively address this problem. In the PIT model, we introduced the personal impact parameter to describe different influence of users on the group decision making process, and identify the recommended items by collectively consider users personal preferences and their personal influences. Therefore, we naturally capture the fact that influencial users have higher chances to affect group selection decisions. Under the general assumption that individual user preference profiles are not available, we designed a learning algorithm to inference the personal impacts and user preferences purely from group selection history. Furthermore, we exploit other social network features such as degree distributions to help refine the parameter inference in an extended PIT (E-PIT) model. Via comprehensive evaluations, we demonstrate that our proposed models bring great recommendation improvements over baseline approaches.

# 7. REFERENCES

[1] G. Adomavicius, G. Adomavicius, and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE TKDE*, 17(6):734–749, 2005.

[2] S. Amer-Yahia, S. B. Roy, A. Chawlat, G. Das, and C. Yu. Group recommendation: semantics and efficiency. *PVLDB*, 2:754–765, 2009.

[3] L. Ardissono, A. Goy, G. Petrone, M. Segnan, and P. Torasso. Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. *Applied Artificial Intelligence*, 17(8):687–714, 2003.

[4] L. Baltrunas, T. Makcinskas, and F. Ricci. Group recommendations with rank aggregation and collaborative filtering. In *RecSys*, pages 119–126, 2010.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[6] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *UAI*, pages 43–52, 1998.

[7] J. S. Breese, D. Heckerman, and C. M. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *UAI*, pages 43–52, 1998.

[8] A. Crossen, J. Budzik, and K. J. Hammond. Flytrap: intelligent group music recommendation. In *IUI*, pages 184–185, 2002.

[9] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *CACM*, 35:61–70, 1992.

[10] T. Hofmann. Latent semantic models for collaborative filtering. *TOIS*, 22(1):89–115, 2004.

[11] A. Jameson and B. Smyth. The adaptive web. chapter Recommendation to groups, pages 596–627. 2007.

[12] X. Liu, Q. He, Y. Tian, W.-C. Lee, J. McPherson, and J. Han. Event-based social networks: Linking the online and offline social worlds. In *KDD*, 2012.

[13] J. Masthoff. Group modeling: Selecting a sequence of television items to suit a group of viewers. *UMUAI*, 14:37–85, 2004.

[14] J. F. McCarthy and T. D. Anagnost. Musicfx: an arbiter of group preferences for computer supported collaborative workouts. In *CSCW*, pages 363–372, 1998.

[15] K. McCarthy, M. Salamó, L. Coyle, L. McGinty, B. Smyth, and P. Nixon. Cats: A synchronous approach to collaborative group recommendation. In *FLAIRS*, pages 86–91, 2006.

[16] R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In *DL*, pages 195–204, 2000.

[17] M. O'Connor, D. Cosley, J. A. Konstan, and J. Riedl. Polylens: a recommender system for groups of users. In *ECSCW*, pages 199–218, 2001.

[18] S. Pizzutilo, B. De Carolis, G. Cozzolongo, and F. Ambruoso. Group modeling in a public space: methods, techniques, experiences. In *WSEAS*, pages 175–180, 2005.

[19] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *CSCW*, pages 175–186, 1994.

[20] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. In *UAI*, pages 487–494, 2004.

[21] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, pages 285–295, 2001.

[22] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advance in Artificial Intelligence*, pages 4:2–4:2, 2009.

[23] L. Ungar, L. Ungar, D. Foster, E. Andre, S. Wars, F. S. Wars, D. S. Wars, and J. H. Whispers. Clustering methods for collaborative filtering. In *AAAI Workshop on Recommendation Systems*, 1998.

[24] M. Ye, X. Liu, and W.-C. Lee. Exploring social influence for recommendation - a generative model approach. In *SIGIR*, 2012.

[25] Z. Yu, X. Zhou, Y. Hao, and J. Gu. Tv program recommendation for multiple viewers based on user profile merging. *UMUAI*, 16:63–82, 2006.