Exploring the Impact of Dynamic Mutual Influence on Social Event Participation *

Tong Xu Hao Zhong Hengshu Zhu Hui Xiong Enhong Chen Guannan Liu

Abstract

Nowadays, it is commonly seen that an offline social event is organized through online social network services (SNS), in this way cyber strangers can be connected in physical world. While there are some preliminary studies on social event participation through SNS, they usually have more focus on the mining of event profiles and have less focus on the social relationships among target users. In particular, the importance of **dynamic mutual influence** among potential event participants has been largely ignored. In this paper, we develop a novel discriminant framework, which allows to integrate the dynamic mutual dependence of potential event participants into the discrimination process. Specifically, we formulate the group-oriented event participation problem as a variant two-stage discriminant framework to capture the users' preferences as well as their latent social connections. The experimental results on real-world data show that our method can effectively predict the event participation with a significant margin compared with several state-of-the-art baselines, which validates the hypothesis that dynamic mutual influence could play an important role in the decision-making process of social event participation.

1 Introduction

The newly emerged *event-driven* Social Network Services (SNS) target at providing the opportunities for online people to gather together in offline events, which has become popular and attractive for millions of users all around the world. For instance, at Meetup.com, more than 10,000 events are organized every day, and invitations may even exceed 100 times per minute. This new business model imposes new challenges on social event analysis with considering *social factors*, and raises the difficulties for the event organizers to draw the event plan and predict the attendance.

Indeed, the "word-of-mouth" effect can strongly affect the decision-making process of social event participation, especially for the users without strong opinions. It is common to see that users simply follow the advices from their friends when they are hesitating. For instance, prior study has revealed that 10%-30% of human movement could be explained by social factors, even more evident on long-ranged travel [3]. Since face-to-face communication is inevitable for offline social gatherings, people usually tend to stay with the familiars, which leads to more cohesive communities for event-driven social networks than the ordinary ones [11], and definitely, more explicit social effects. Therefore, there is a critical need to investigate the social effects on the social event decision-making.

For the past several years, some researchers like [23] and [21] have considered the social factors as features or constraints in their studies, which can effectively improve the performance. However, these techniques cannot capture the dynamic network evolution and time-variant mutual influence within potential attendants of event series. Some other works like [10] attempt for social group decision-making, where personal impact, social relations and game equilibrium are integrated together to provide a unified decision. However, social factors here are simplified as game-playing or delegate-voting, and personalized analysis could not be provided. In fact, the importance of dynamic mutual influence has not been fully exploited in the above studies.

When we describe the effect of social factors in event participation, we realize that mutual influence should be considered as dynamic. For example, when making a decision within a group, people may listen to and can be influenced by some friends, and they will further influence the others. Also, if two friends hold the same idea, their tendency will be mutually strengthened; on the contrary, opposite ideas lead to weaken confidence. As a result, the *chain reaction* based on **D**ynamic **S**ocial **I**nfluence (**DSI**) will be formed. As shown in Figure 1, potential participants share their ideas in the social network, where mutual influence is digested to form new decisions and further spread. The iterative process will

*Author affiliations and emails:
Tong Xu and Enhong Chen (Corresponding Author), *University of Science and Technology of China*, tongxu@mail.ustc.edu.cn, cheneh@ustc.edu.cn,
Hao Zhong and Hui Xiong, *Rutgers, the State University of New Jersey*, {h.zhong31, hxiong}@rutgers.edu,
Hengshu Zhu, *Baidu Research - Big Data Lab*, zhuhengshu@baidu.com,
Guannan Liu, *Tsinghua University*, guannliu@gmail.com

Figure 1: An illustration of dynamic mutual influence within event participation.

repeat until the final decision achieved stably.

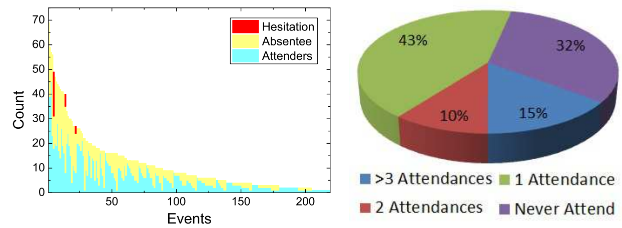
To that end, in this paper, we aim at exploring dynamic mutual influence in the decision-making process of social event participation. To be specific, we propose a novel two-stage discriminant framework, which allows integrating the dynamic mutual dependence of potential participants into the learning process. Based on the framework, we can model the group-oriented decision-making process to capture the users' preferences as well as their latent social connections. To the best of our knowledge, we are the first to investigate the impact of dynamic mutual influence on social event participation.

Finally, experimental results on real-world social network data indicate our framework can effectively improve the event participation prediction. This validates the hypothesis that mutual social influence indeed plays an important role in the decision making process of potential event participants.

Overview. The rest of this paper is organized as follows. Section 2 further illustrates the motivation of this study with related statistics. In Section 3, we define the participation prediction task and formulate our discriminant framework, then technical details are explained in Section 4. In Section 5, we evaluate the performances and discuss some interesting findings with a case study. Section 6 presents the related works. Finally, in Section 7, we conclude the paper.

2 Investigation on Social Effects: Are Participation Affected by the "Social"?

In this section, we will deeply discuss the social effects on event participation as some questions still remain. First, social effects on event series have not been studied before. Second, homophily is not distinguished, i.e., attendance might be due to similar preference but not mutual influence. To study on these questions, we provide some discussion based on related statistics.



(a) (b)

Figure 2: User distribution on attendance. (a) Distribution of responses for each event (b) Percentages for different amount of attendances.

2.1 Data Set Description We conduct our study on a real-world data set collected from Meetup.com [1], which is one of the most popular websites that facilitates offline activities. Specifically, we extracted event logs and user profiles via the official APIs of Meetup, which totally consists of 422 user groups, 9,605 social events and 24,107 related users. Summary of event participation is shown in Figure 2(a).

What should be noted is that Meetup executes the group-based scheme, thus users have to join the groups in order to receive invitations, which results in the long-term group-oriented event series. Also, Meetup highlights group structure but ignore point-to-point connection, thus we build connection based on co-occurrence of pairwise users, and the frequency is treated as link strength, which is commonly used in related work like [11] and [9].

To describe the attributes of social events, we exploited key words in group descriptions, and users' preferences are summarized by the events they attended. Similar with other relevant studies in social user profiling like [25], LDA model [2] was introduced to learn the topics. Finally, all profiles are presented as attribute vectors.

2.2 Social Effects on User Engagement We first discuss on the long-term social effects in event series. Active users, i.e., those who have attended at least 3 events are labeled as valuable to groups. Totally, only 14.74% of attendants are labeled as active ones, who attended 11.08 events in average, much more than 3.24 for overall users. The details are shown in Figure 2(b).

Two sets of statistics are conducted. First, we count the degree and weights of users at their first attendance, to explore whether their initial status may influence their long-term activity. Second, we would like to know whether small communities formed by active ones are indeed denser than the ordinary ones, so we

[1] http://www.meetup.com/

Table 1: Comparison for social factors in event series.

	Average for All Events		First Attendance	
	Density	Ave. Weight	Degree	Ave. Weight
Active	0.7849	0.2343	0.1249	0.0109
Overall	0.4694	0.1305	0.0498	0.0062
P-Value	0.000	0.000	0.001	0.004

Table 2: Comparison for user preference to events.

Single Event		All Events	
Attendant	0.108	Active Users	0.106
Absentee	0.094	Overall Users	0.105
P-Value	0.016	P-Value	0.334

count the average density and weights during the event series. The results are shown in Table 1, in which *P-Value* presents the T-test result for the assumption that measures of active users are higher than ordinary ones. Unsurprisingly, differences on all the measures are significant, however, the initial connection on the first attendance is extremely low. It is quite similar to the related work [18] that usually attendants are not connected to the majority, but only a few initiators, which results in the typical "star network" structure.

Though social factors might be insufficient initially, they will soon be enhanced. As shown in Figure 3, when users attend more events, both the degree and weight grow rapidly. Interestingly, the degree turns stable soon, then decreases slowly, while the weight still keeps increasing. Clearly, some friends leave, but retained connection become more strong due to more co-occurrences. In summary, long-term active users hold denser communities than ordinary ones, which definitely means more significant social effects.

2.3 Preference in Social Event Participation

Then, we study the preference factors in event participation. As preferences are presented in vectors, *Cosine similarity* is introduced to estimate participation without considering the social factors. Two pairs are compared: attendants vs. absentees of single events, and active users vs. overall users for all the events. The results are shown in Table 2 with corresponding T-test result (assuming the former one is higher than the latter). Interestingly, though the attendants hold clear interest than the absentees, we found that the active users do not express explicit preference than ordinary ones.

Considering that most people, who hold similar preference with the active ones, quit after attending only one or two events, we could do fair reasoning. For inactive users, they may be attracted by the topics at first; however, they quit soon as they don't like the group or event hardly fit in the group. At the same time, for active users, though sometimes they don't like the events, they attend due to invitation from their friends.

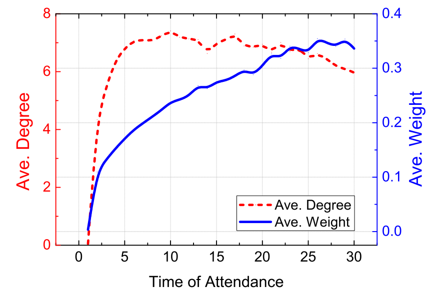


Figure 3: Average degree and weight for attendants on different times of occurrence.

This phenomenon validates our motivation that the social factors indeed affect the decision-making process of event participation, especially for the well-connected group. Also, the social factors might not be reflected by similar preferences, but direct effect on decisions.

3 Social Effects Formulation and Framework

As our motivation has been intuitively supported, in this section, we first formally define the problem and introduce some preliminaries. Then, our novel discriminant approach with social-influence-based threshold will be formulated. And finally, we demonstrate our two-stage framework for social event participation prediction.

3.1 Problem Statement In this paper, we focus on the individual participation. However, to consider the social effects within users, it is necessary to put individuals into a group. Therefore, here we use the definition *target user group* to represent the group of users to be predicted instead of individual users. Formally, the problem can be defined as follows.

Definition 3.1. (Problem Statement) *Given the target user group* $\mathbf{U}$ *with weighted connections, the problem is to predict the individual participation* $s_{i,k}$ *of target event* e_k *for each user* $u_i \in \mathbf{U}$*, here* $s_{i,k} = 1$ *indicates the attendance, while* $s_{i,k} = 0$ *means the absence.*

To model user profiles, we first exploit a vector $\mathbf{p}_i$ to present the preferences of user u_i, in which each element denotes the preference level on a specific aspect. Correspondingly, we also have a vector $\mathbf{a}_k$ for each event e_k to indicate the attributes, which has the same dimensions with $\mathbf{p}_i$. Indeed, the similarities between $\mathbf{p}_i$ and $\mathbf{a}_k$ may roughly indicate the probability of u_i attending e_k without considering the social factors.

For the weighted connection within target user group $\mathbf{U}$, we use $W = \{w_{ij}\}$ to indicate the set of connection strengths (weights) that are not achieved

Table 3: Mathematical Notations.

SYMBOL	DESCRIPTION
$\mathbf{U} = \{u_i\}$	the set of users
$\mathbf{E} = \{e_k\}$	the set of events
$\mathbf{p_i}$	preference vector for u_i
$\mathbf{a_k}$	attributes vector for e_k
w_{ij}	social connection strength from u_i to u_j
$f_{i,k}$	intention for u_i to attend e_k
$h_{i,k}$	threshold for u_i to attend e_k
$s_{i,k}$	attendance for u_i to e_k

initially, and w_{ij} corresponds to the social influence strength from u_i to u_j. The mathematical notations used throughout this paper are summarized in Table 3.

3.2 Discrimination with Social-based Threshold When we treat the event participation as a discriminant problem, specifically, we have similarity function $f(u_i, e_k)$ and a threshold $h(u_i, e_k)$ for user u_i and event e_k, the individual participation $s_{i,k} = 1$, iff. $f(u_i, e_k) \geq h(u_i, e_k)$ and $s_{i,k} = 0$, iff. $f(u_i, e_k) < h(u_i, e_k)$.

To formulate the social effects in decision-making process, we have two choices, i.e., merging the social factors with $f(u_i, e_k)$ or $h(u_i, e_k)$. Traditionally, prior works integrate the social factors into $f(u_i, e_k)$ following the assumption that the social connections leads similar preferences, thus usually the social factors are formulated as constraint or features. However, as discussed above, we realize that social influence may directly affect the decision-making, but not the preference indirectly. Therefore, here we choose to merge the social effects with calculation of threshold $h(u_i, e_k)$.

Along this line, we assume that $h(u_i, e_k)$ depends on the participation of friends. In our approach, the dependence is reflected by the variance of threshold $h(u_i, e_k)$. To formulate the dependence, we borrow and adapt the classic Independent Cascade (IC) model [6] for simulating the dynamic mutual influence within users. What should be noted is that IC model here could be replaced by other social influence model if needed, we choose IC here since it is widely used as one of the basic models, and its effectiveness has been well proved. Particularly, if we denote $f(u_i, e_k)$ as $f_{i,k}$ and $h(u_i, e_k)$ as $h_{i,k}$, the threshold can be formulated as follows.

$$h(u_i, e_k) = h_{i,0} \cdot \prod_{j \in N_i} [1 - \mathcal{I}(f_{j,k} - h_{j,k}) \cdot w_{ji}], \tag{3.1}$$

where $h_{i,0}$ denotes the parameter for personal participating activity, i.e., active users will hold lower $h_{i,0}$. Also, N_i means friends of u_i in the target group $\mathbf{U}$, and w_{ji} indicates the strength of social influence from u_j to u_i. Interestingly, higher w_{ji} may not only indicate the strong connection from u_j to u_i, but also indicate that u_i could be easily influenced, especially when all the w_{li} for

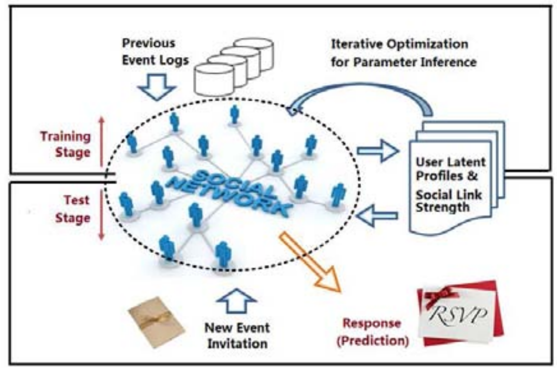


Figure 4: Overview of our framework for event participation prediction.

$u_l \in N_i$ are relatively high. Furthermore, $w_{ij} < 0$ is eligible to present the situation that u_i and u_j usually hold opposite opinions, which is different from the setting of traditional social influence problems. Besides, $\mathcal{I}(x)$ presents the discriminant function to indicate friend's option. To smooth the variation and ease the optimization, we introduce the *sigmoid function* here to approximate the sign function as follows.

$$\mathcal{I}(x) = \frac{1}{1 + e^{-\alpha x}}, \tag{3.2}$$

where α presents the parameter to regulate the slope. Definitely, since the sigmoid function is smooth and derivable on the $\mathbb{R}$ set, the related optimization task will be much easier to solve.

For the preference function $f(u_i, e_k)$, which depends on the characteristics of data set, the details will be introduced in experimental part.

With the above formulation, we can integrate the users' profiles and mutual influence into the unified discriminant framework. Indeed, this framework can reflect the intuition that users usually make their own decision for event participation, then they are influenced by friends to change their mind, this process repeats until they finally achieve the final equilibrium.

3.3 Two-stage Framework Based on the definitions above, now we can formally present our two-stages framework for event participation prediction. Figure 4 demonstrates the overview of s framework.

Training Stage. Given a target user group $\mathbf{U} = \{u_i\}$ and a set of historical events $\mathbf{E} = \{e_k\}$, in which corresponding attendance record $\mathbf{S} = \{s^0_{i,k}\}$ for each pair of u_i and e_k are pre-known. Also, we have the event attributes $\mathbf{a}_k$ for each e_k. In this stage, we aim at inferring the latent profile $\mathbf{p}_i$ and activity measure $h_{i,0}$ for each u_i, as well as learning the connection strength $\{w_{ij}\}$ for pairwise friends in the social network.

Algorithm 1 Iterative Solution for Training Stage.

Input: target user group $\mathbf{U} = \{u_i\}$, event set $\mathbf{E} = \{e_k\}$ and attendance records $\{s_{i,k}^0\}$;
Store: event attributes $\mathbf{a}_k$ for each $e_k \in \mathbf{E}$;
Output: users' profile $\langle \mathbf{p}_i, h_{i,0} \rangle$ and social strength w_{ij}

Iteration = *True*;
while (Iteration)
  Iteration = *False*;
  for $u_i \in \mathbf{U}$, $e_k \in \mathbf{E}$
    update $\langle \mathbf{p}_i, h_{i,0} \rangle$ and $\{w_{ij}\}$ until convergency;
    update $f_{i,k}$, $h_{i,k}$ based on Equation 3.1;
    update $s_{i,k}$ as $\mathcal{I}(f_{i,k} - h_{i,k})$;
    if $s_{i,k}$ changed **then** Iteration = *True*;
    end if
  end for
end while
return $\{\langle \mathbf{p}_i, h_{i,0} \rangle\}$, $\{w_{ij}\}$;

Test Stage. After obtaining the users' profiles $\langle \mathbf{p}_i, h_{i,0} \rangle$ and mutual affection strength $\{w_{ij}\}$ in the training stage, in the test stage, given a certain event e_k with attributes $\mathbf{a}_k$ and the corresponding target user group $\mathbf{U}_k = \{u_i\}$, we aim at predicting event participation $s_{i,k}$ for each $u_i \in \mathbf{U}_k$.

4 Technical Details for the Prediction of Social Event Participation

In this section, we introduce the technical solutions for both training and test stage of our framework.

4.1 Iterative Optimization for Training Stage Indeed, the task in training stage can be regarded as a supervised learning problem, which targets at minimizing the cost of discrimination errors on training data. Therefore, we can formulate the objective function, i.e., the cost function, for training stage as follows.

$$\arg\min_{\mathbf{p},h_0,w} \sum_{u_i \in U} \sum_{e_k \in E} [s_{i,k}^0 - \mathcal{I}(f_{i,k} - h_{i,k})]^2, \tag{4.3}$$

Here $s_{i,k}^0$ presents the ground truth of attendances. Intuitively, discrimination errors lead to higher cost, and minimizing the cost function may result in the optimized inference of users' profiles and social strength. However, since the calculation of $h_{i,k}$ of users depends on the $f_{j,k}$ and $h_{j,k}$ of their friends, to optimize the mutual dependence is extremely difficult.

To address this challenge, we propose a step-by-step iterative approach. To be specific, we treat the dynamic social influence as an iterative generation process, where the decision made in current round will only affect friends' thresholds in next round. Iteration of our objective function is formulated as follows.

$$F^t(U,E) = \sum_{u_i \in U} \sum_{e_k \in E} [s_{i,k}^0 - \mathcal{I}(f_{i,k}^t - h_{i,k}^t)]^2, \tag{4.4}$$

where we have

$$h_{i,k}^t = h_{i,0}^t \cdot \prod_{j \in N_i} [1 - \mathcal{I}(f_{j,k}^{t-1} - h_{j,k}^{t-1}) \cdot w_{ji}^t]. \tag{4.5}$$

After each round, all parameters will be updated and the mutual influence will be digested to achieve the new discrimination results. During this process, some ones may change their mind, e.g., they may be activated to join the event, or quit due to negative influence. Besides, discriminant errors will also affect the iteration. This process will repeat until the cost is stable, which indicates that no one will further change their minds. To optimize the new objective function (i.e., Equation 4.5), we exploit the gradient descent method.

4.2 Prediction for Test Stage With the users' profiles and social connection strength inferred in the training stage, in the test stage, we aim at predicting the participation for the target user group to a certain event. Since all the parameters are pre-learned, here we even don't need an objective function, but directly achieve the prediction with iteration.

To be specific, we first calculate raw intention for each user. Then, in each step, threshold will be updated based on Equation 3.1, and attendance will be re-predicted afterwards. This process will repeat until the prediction results are stable. What should be noted is that as negative influence may exist, which is different from traditional social influence problem. Thus, active users may not keep rising, which increase the steps to iterate. The details are illustrated in Algorithm 2.

5 Experiments and Discussions

To validate our hypothesis that dynamic mutual influence may affect the decision making of social event participation, in this section, we conduct series of experiments on a real-world data set. Also, some empirical case studies and discussion will be presented.

5.1 Experimental Setup In this section, we summarize the data set pre-processing and selected baseline algorithms for the experiments.

5.1.1 Data Set Pre-processing As introduced in Section 2, we conduct our experiments on the real-world data set collected from Meetup. To describe the users' profiles as well as the events' attributes, 30 topics, similar with 34 categories defined by Meetup, were learned by leveraging the LDA model.

Algorithm 2 Iterative Solution for Test Stage.

Input: target user $\mathbf{U}$ and event e_k with attributes $\mathbf{a}_k$;
Store: users' profiles $\langle \mathbf{p}_i, h_{i,0} \rangle$ and connection weights w_{ij};
Output: $s_{i,k}$ for each $u_i \in \mathbf{U}$

for $u_i \in \mathbf{U}$
  calculate $f_{i,k}$ for each u_i;
end for
Iteration = *True*;
while (Iteration)
  Iteration = *False*;
  for $u_i \in \mathbf{U}$
    update $h_{i,k}$ based on Equation 3.1;
    update $s_{i,k}$ based on $f_{i,k}$ and $h_{i,k}$;
    if $s_{i,k}$ changed **then** Iteration = *True*;
    end if
  end for
end while
return $\mathbf{S} = \{s_{i,k}\}$;

For the preference function $f(u_i, e_k)$ mentioned in Section 3.2, as description could be easily normalized and presented in vectors, we could intuitively select the **Cosine similarity**. However, some cost factors, e.g., distance, time spending or financial cost may also affect the decision. Since these factors could hardly be normalized, we further multiply the Cosine similarity with **Gaussian probability** for each cost factor, where means are learned during the training stage, and variances are set based on statistics. To be specific, $f(u_i, e_k)$ will be estimated as follows,

$$f(u_i, e_k) = Cosine(\mathbf{a_k^T}, \mathbf{p_i^T}) \cdot \prod_c \mathcal{N}(p_i^c | a_k^c, \sigma_c^2) \tag{5.6}$$

where $\mathbf{a_k^T}$ and $\mathbf{p_i^T}$ present the vector corresponds to topics, and $\{c\}$ presents the cost factors. σ_c presents the variance which is determined by statistics of samples.

Besides, since we introduce the IC model to describe the dynamic mutual influence within potential attendants, we treat the event organizers as *"seed users"* to start the influence process. Those organizers will be treated as input for certain event (although we still attempt to learn their preferences, since they may appear in other event as ordinary attendants).

5.1.2 Evaluation Baselines

As we integrate the mutual social influence into the event participation prediction analysis, we choose three state-of-the-art baselines which correspond to both the traditional recommendation methods and social influence simulation for more comprehensive comparison.

1) Cost-aware PMF (GcPMF) [5]. Probabilistic matrix factorization (PMF) is one of the basic tools for recommender system. To be specific, we utilize the GcPMF [5] as baseline, in which the cost factors mentioned in above subsection are also integrated.

Table 4: Overall performance of each approach.

	DSI	SoRec	GcPMF	PSS
Precision (%)	**75.88**	60.23	47.47	46.15
Improvement (%)	-	+25.98	+59.85	+64.42
Variance	**0.022**	0.102	0.134	0.059
P-Value	-	0.000	0.000	0.000
Recall (%)	**75.34**	75.21	21.73	41.82
Improvement (%)	-	+0.17	+246.71	+80.16
Variance	**0.030**	0.112	0.234	0.180
P-Value	-	0.063	0.000	0.000

2) Social-based PMF (SoRec) [13]. Following the intuition that a user's social network will affect personal behaviors on the Web, the SoRec model introduces another matrix which indicates the social network into the PMF framework for representing the social constraint. Indeed, it is a enhanced PMF with merging the **static social factors**. As no explicit social network could be achieved in Meetup data set, we constructed the connection based on the rules described in Section 2.1.

3) Preference-sensitive Social Spread (PSS) [20]. To analyze the event participation in the perspective of social spread, we also introduce the preference-sensitive social spread (PSS) method as baseline, which is a typical two-stages framework with the basic assumption that the social spread is sensitive to users' common preference. Note that as social spread is actually a series of random events, thus, we repeat experiments for 500 times for each prediction to reduce the uncertainty.

5.2 Experiment Results

Due to the group-based scheme of Meetup, we treat *group* as the unit of our experiments. In other words, each group leads to a set of independent experiments, and the average results for 422 groups are presented as the finals.

For the evaluation metric, as a typical discrimination problem, we select the common used **Precision** and **Recall** rates. In our framework, both discriminant function and threshold are learned, while for the former two PMF-based baselines, we choose the best threshold based on the ROC curve.

5.2.1 Comparison of Overall Performance

First of all, we validate our prediction performance of our novel discriminant approach comparing with different baselines. Since we face to the severe sparse data issue that only less than 20% users attended at least 3 events in a group, we assign 90% events within one group as training samples to ensure the quality of training, while the rest 10% are test samples.

The experimental results of overall precision performance are shown in Table 4. Our DSI approach achieves the performance above 70% and outperforms the baselines with significant margin, and the results are quite stable, which indicates that the dynamic social influence indeed affects the event participation. At the same time, based on the comparison between GcPMF and SoRec, as social constraints on preference lead to significant improvement, the effects of social factors has been further proven, regardless they may function in different perspectives (static or dynamic).

Finally, we find that the baseline with preference-sensitive social spread (PSS) achieves the worst performance. Indeed, though preferences are integrated between pairwise users, PSS assumes that all the attenders interact with each other, and the strength depends on their common interests. However, as discussed in Section 2, social factors mainly function within active users, and the connections could hardly be estimated only by similar preferences. Besides, considering the cold-start problem, which leads to sparse social network and interaction records, the performance may be further hurt.

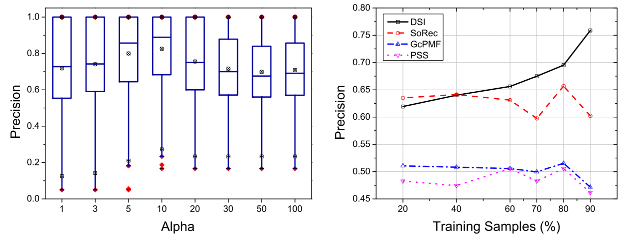
5.2.2 Parameter Sensitiveness Then, we evaluate the sensitiveness of the slope parameter α in Sigmoid function. Results are shown in Figure 5(a). In our experiments, the default α is set as 10.

For the slope α, as mentioned in Section 3.2, we utilize the Sigmoid function to approximate the sign function jumping from 0 to 1, thus a higher α might be better for approximation. However, smooth variation is still needed especially for those who don't hold clear preference, thus their hesitation shall not lead to dramatic change in social influence. That may explain why performance achieves the peak when α is around 10, but not monotonously increasing as we expected.

5.2.3 Sample Allocation Sensitiveness Also, we discuss about the sensitiveness of the training sample proportion, which is summarized in Figure 5(b). We unsurprisingly find that our performance degenerates with less training samples, which indicates that our framework is sensitive social network structure. However, it still works better than almost all the baselines.

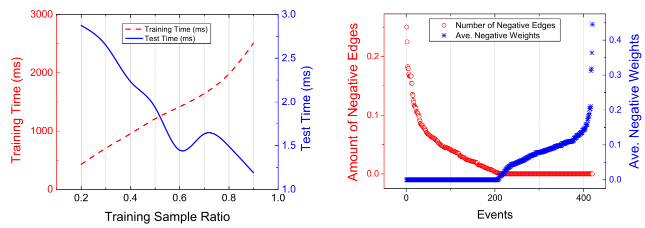
On the contrary, SoRec keeps relatively stable during the change, since it requires only some social-related statistics, but not the network structure. Besides, the high ratio of freshmen leads to the severe "cold-start" problem, which impacts the result severely.

5.2.4 Discussion on Complexity Finally, we discuss about the complexity. Though social influence simulation is integrated, we still believe that it could apply



(a) (b)

Figure 5: Verification on robustness: (a) Performance with different α. (b) Performance with different percentage of training samples.



(a) (b)

Figure 6: Statistics for complexity: (a) Execution times of our framework for each group in average with different ratio of training samples. (b) Amount and average weights of negative edges for each group.

to large-scale network. On one hand, for the test stage, if there is no negative edges, we realize that the computational complexity is the same with *Linear Threshold* [6] model, which is fast enough for large-scale computation. As shown in Figure 6(b), usually we have less than 10% negative edges and relatively low weights (lower than 0.1), thus they won't interfere a lot. On the other hand, for the training stage, which could be conducted offline and updated infrequently, the overall computational cost will be limited.

We summarize the execution time for different proportion of training samples in Figure 6(a). Indeed, we find for each group, which contains more than 20 events in average, it costs only a few milliseconds to predict potential attendants for all the events. Clearly, it demonstrates the potential of our framework to deal with large-scale social network.

5.3 Case Study To better understand the performance, i.e., how the dynamic social influence affects the prediction, we randomly select four groups as examples to illustrate some interesting discoveries concerning about the social factors. Details about these four groups are listed in Table 5, in which the precision is also listed for clear comparison. Considering about the social factors, two crucial issues should be studied: 1) who influences the others and who are influenced; 2) how the mutual social influence functions.

5.3.1 Participants of Social Influence Process

For the first issue, two types of potential attendants should be carefully observed, namely the event organizers (also “seed users” in social influence), and the new comers who are fresh to the group and causing the “cold-start” problem. For the former two groups, namely A and B, we realize that at least 5 members have been organizers in each group, and for every event, usually there are at least 2 hosts, even up to 5. On the contrary, for the latter two groups who suffer relatively poor precision, we found that they have stable hosts, i.e., only one or two users act as hosts for all the events. As organizers are usually authoritative in the group, for group C and D, though they are huge groups with hundreds of members, the “authority” nodes are quite limited, which result in the limited social influence and definitely interfere the prediction.

Also, we find that almost all the groups suffer severe “cold-start” problem, i.e., former members quitting and new ones coming. Usually, higher ratio of freshmen causes problems in prediction. Interestingly, we find that though the ratio of freshmen for Group B reaches more than 50%, it still gains more than 94% precision. With deep looking into the data, we realized that actually there is only one event in the test samples, and for this event, nearly half of the participators are active users, while the rest are all strongly connected to them. There is even a freshman who acts as the organizer directly, which is rarely found in other groups. This phenomenon implies that the strong social influence may help to overcome the “cold-start” problem, which also supports our hypothesis of social effects.

5.3.2 Affection of Negative Social Influence

Then, we discuss about the function of mutual influence. As mentioned in Section 3.2 that the social influence strength could be negative to indicate that the two users are usually conflicting. For the four groups here, we find that the former two groups with better performance contain almost 100% positive edges, while the Group C suffers 7% negative edges, and nearly 4% for D, which definitely increases the level of uncertainty.

It might indicate that strongly connected community with common goal will lead to better prediction, while an intricate group, in which members who are conflicting with each other will be in confusion. Interestingly, it might also answer the high ratio of freshman and low level of activity of these groups, e.g., most of members attend only 1-2 events and only around 10 attendants for each event. It seems that a smaller size leads to more intensive connection and more sufficient interaction, which is accordance with the idea in [19].

Table 5: Examples for Case Study

Group	A	B	C	D
Precision	96.15%	94.64%	48.20%	47.01%
Members	129	160	1088	273
Ave. Freshmen	20%	50%	35%	35%
Negative Edges	$< 1\%$	$< 1\%$	7%	4%

6 Related Work

Generally, two types of social event analysis have been intensively studied in recent years, namely social event recommendation, and decision-making analysis. Specifically, some researchers focused on the conformity between users' profiles and event attributes. For example, [7] proposed a hybrid approach that is enriched with social influential features and user diversity model on decision making, and [9] studied the offline ephemeral social networks to infer the latent preferences and social relations for ranking the recommended social events. Furthermore, there are some related works focused on other practical problems. For example, [14] built the connections between events at different times by borrowing the feedback from past events to deal with the deficiency of explicit feedback, and [15] attempted to solve the cold-start problem of mobility via discovering the rule of popular events among the residents of an certain area. Finally, some researches focus on the event-driven social groups. For example, [23] considered the geographical features, social features, and implicit patterns simultaneously in an unified model to achieve the recommendation of event-based groups.

Another related topic of this paper is the group-based recommendation, i.e., to recommend events to a social group but not individuals. Usually, previous works mainly follow two different schemes. The first is to select a representative from the group, and then the representative will draw the overall conclusion. For example, [10] proposed a personal impact topic (PIT) model to enhance the group preference profile. Another direction is to achieve the agreement within group based on a certain consensus function, like, [4] captures the social, expertise, and interest dissimilarity among multiple group members. Indeed, some other factors might also be considered during the agreement process, e.g., [16] analyzed that how the personality of cooperation and trust could influence the group recommendation results, and [1] discussed the monotonicity and efficiency for group recommendation. Finally, some prior works, e.g. targeted at combining social recommendation with traditional personalized recommendation [12].

In addition, although some works do not directly focus on the event participation problem, they still concentrate on some related topics. For instance, [11] concerned about the comparison of social structure between online and offline social network, and discussed about the information flow within event-based social network. Also, [18] studied the spatial and temporal characteristics of event participation, and revealed the group evolution rule for event organization. Some other works target at forming the proper group for social event maximal participation, such as [17] and [8]. Finally, some related work study the offline user behaviors in the perspective of *ephemeral social networks*, like [24] developed a factor graph model based framework to infer the likelihood of future encounter, and [22] recommended offline geo-friends based on pattern-based heterogeneous information network analysis.

7 Conclusion

In this paper, we investigated how to exploit the impact of dynamic mutual influence on the decision-making process for social event participation. A unique characteristic of our method is that the social influence is integrated into the threshold calculation for the discriminant function, which reflects the dynamic mutual dependence within friends for event participation. Then, we formulated the group-oriented event participation problem as a variant two-stage discriminant framework to capture both users' preferences and their latent social connections. Finally, experimental results on the real-world data showed that our method could effectively predict the participation with a significant margin compared with several state-of-the-art baselines. This shows that social connections may not only affect the user preferences, but also directly affect the decision-making process of event participation.

As we discussed that dynamic social influence mainly functions on active users, further study will be conducted on how to integrate our framework with existing techniques to improve the applicability. Also, since the proposed techniques could be applied for some other application problems, such as social group formation and group target design, in the future, we would like to exploit more applications of the proposed method and develop the techniques to integrate more types of social constraints into the learning framework.

8 Acknowledgement

This research was partially supported by grants from the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010), and the National High Technology Research and Development Program of China (Grant No. 2014AA015203). Also, it was supported in part by Natural Science Foundation of China (Grant No. 71329201). Finally, thanks for Qi Liu's help during the paper writing.

References

[1] S. Amer-Yahia, S.B. Roy, A. Chawlat, G. Das, and C. Yu, Group recommendation: Semantics and efficiency, in *VLDB'09*, 2009, pp. 754–765.
[2] D.M. Blei. A.Y.Ng, and M.I. Jordan, Latent dirichlet allocation, in *JMLR*, 2003, pp. 993–1022.
[3] E. Cho, S.A. Myers, and J. Leskovec, Friendship and mobility: user movement in location-based social networks, in *KDD'11*, 2011, pp. 1082-1090.
[4] M. Gartrell, X. Xing, Q. Lv, A. Beach, R. Han, S. Mishra, and K. Seada, Enhancing group recommendation by incorporating social relationship interactions, in *Group'10*, 2010, pp. 97–106.
[5] Y. Ge, Q. Liu, H. Xiong, A. Tuzhilin, and J. Chen, Cost-aware travel tour recommendation, in *KDD'11*, 2011, pp. 983–991.
[6] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in *KDD'03*, 2003, pp. 137–146.
[7] H. Khrouf, and R. Troncy, Hybrid event recommendation using linked data and user diversity, in *RecSys'13*, 2013, pp. 185–192.
[8] C.T. Li, and M.K. Shan, Composing activity groups in social networks, in *CIKM'12*, 2013, pp. 2375–2378.
[9] G. Liao, Y. Zhao, S. Xie, and P.S. Yu, An effective latent networks fusion based model for event recommendation in offline ephemeral social networks, in *CIKM'13*, 2013, pp. 1655–1660.
[10] X. Liu, Y. Tian, M. Ye, and W.C. Lee, Exploring personal impact for group recommendation, in *CIKM'12*, 2012, pp. 674–683.
[11] 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'12*, 2012, pp. 1032–1040.
[12] Q. Liu, Z. Dong, C. Liu, X. Xie, E. Chen, and H. Xiong, Social Marketing Meets Targeted Customers: A Typical User Selection and Coverage Perspective, in *ICDM'14*, 2014, pp. 350–359.
[13] H. Ma, H. Yang, M. Lyu, and I. King, Sorec: social recommendation using probabilistic matrix factorization, in *CIKM'08*, 2008, pp. 931–940.
[14] E. Minkov, B. Charrow, J. Ledlie, S. Teller, and T. Jaakkola, Collaborative future event recommendation, in *CIKM'10*, 2010, pp. 819–828.
[15] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft, Recommending social events from mobile phone location data, in *ICDM'10*, 2010, pp. 971–976.
[16] L. Quijano-Sanchez, J.A. Recio-Garcia, and B. Diaz-Agudo, Social factors in group recommender systems, in *ACM TIST*, 4(1): 8, 2013.
[17] T. Sun, W. Chen, Z. Liu, Y. Wang, X. Sun, M. Zhang, and C.Y. Lin, Participation Maximization Based on Social Influence in Online Discussion Forums, in *ICWSM'11*, 2011, pp. 361–368.
[18] C. Wang, M. Ye, and W. Lee, From face-to-face gathering to social structure, in *CIKM'12*, 2012, pp. 465–474.
[19] B. Xu, A. Chin, and D. Cosley, On how event size and interactivity affect social networks, in *CHI'13*, 2013, pp. 865–870.
[20] T. Xu, D. Liu, E. Chen, H. Cao, and J. Tian, Towards annotating media contents through social diffusion analysis, in *ICDM'12*, 2012, pp. 1158–1163.
[21] T. Xu, H. Zhu, E. Chen, B. Huai, H. Xiong and J. Tian, Learning to Annotate via Social Interaction Analytics, in *Knowledge and Information System*, 2014(41):251-276.
[22] X. Yu, A. Pan, L.A. Tang, Z. Li, and J. Han, Geo-friends recommendation in gps-based cyber-physical social network, in *ASONAM'11*, 2011, pp. 361–368.
[23] W. Zhang, J. Wang, and W. Feng, Combining latent factor model with location features for event-based group recommendation, in *KDD'13*, 2013, pp. 910–918.
[24] H. Zhuang, A. Chin, S. Wu, W. Wang, X. Wang, and J. Tang, Inferring geographic coincidence in ephemeral social networks, in *ECML-PKDD'12*, 2012, pp. 613–628.
[25] H. Zhu, E. Chen, H. Xiong, H. Cao, and J. Tian, Ranking user authority with relevant knowledge categories for expert finding, in *World Wide Web*, 2014(17):1081-1107.