World Wide Web (2020) 23:853–871
https://doi.org/10.1007/s11280-019-00724-7

Matching of social events and users: a two-way selection perspective

Zikai Yin[1] · Tong Xu[1] · Hengshu Zhu[2] · Chen Zhu[2] · Enhong Chen[1] · Hui Xiong[3]

Received: 29 May 2019 / Revised: 31 July 2019 / Accepted: 22 August 2019 /
Published online: 27 November 2019

Abstract
Recent years have witnessed a growing trend that offline social events are organized via online platforms. Along this line, large efforts have been devoted to recommending appropriate social events for users. However, most prior arts only pay attention to the selections of users, while the selections of events (organizers), which lead to the "*two-way selection*" process, are usually ignored. Intuitively, distinguishing the two-way selections in historical attendances can help us better understand the social event participation and decision making process in a holistic manner. To that end, in this paper, we propose a novel two-stage framework for social event participation analysis. To be specific, by adapting the classic *Gale-Shapley* algorithm for stable matching, we design utility functions for both users and event organizers, and then solve two layers of optimization tasks to estimate parameters, i.e., capturing user profiling for event selection, as well as event rules for attender selection. Experimental results on real-world data set validate that our method can effectively predict the event invitation and acceptance, compared with the combinations of one-way-selection baselines. This phenomenon clearly demonstrates the hypothesis that two-way selection process could better reflect the decision making of social event participation.

Keywords Social event · Two-way selection · Stable matching

1 Introduction

Nowadays, with the deep integration of social factors and new business modes [4], it is commonly seen that an offline social event is organized via online social network services, in this way cyber strangers can be connected in physical world. This phenomenon raises the significant challenge for event-oriented services to help their users overcome the information explosion, especially for users who face the *conflict of invitation* [42], which means

This article belongs to the Topical Collection: *Special Issue on Graph Data Management in Online Social Networks*
Guest Editors: Kai Zheng, Guanfeng Liu, Mehmet A. Orgun, and Junping Du

✉ Tong Xu
tongxu@ustc.edu.cn

Extended author information available on the last page of the article.

they have to make choices among multiple options. Thus, effective techniques for event recommendation are urgently required.

Traditionally, large efforts have been made to recommend proper events for users based on estimating their profiles. For instance, [18] and [36] focused on event recommendation with user profiling and geographical information, and [33, 34] further explored the impact of social influence. However, they may fail to consider one crucial fact that, when users select preferred event to attend, the events (or their organizers) will also select appropriate attenders according to certain rules, which results in the so-called "**two-way selection**" process. For instance, as shown in Figure 1, the event organizers require attenders to submit resumes, and only those who are selected will receive the future invitations. Under this situation, some users may quit since they are not selected by the organizers, but not due to their own preferences. Obviously, the historical attendances here may not fully reflect users' first choice, but only the balance between user preferences and event selections. Thus, traditional recommender techniques could be severely misled in these cases.

Indeed, these cases are not occasional. In traditional scenario of recommender system, like movie recommendation in Netflix or item recommendation in Amazon. Usually, there are no constraint on the quota of items, in other words, they could accept as many users as they wish, thus no further selection for users are required. However, for offline events, due to the limitation of venue or event scale, the quota for attenders is usually limited. Moreover, in traditional scenarios like location recommendation [16], routes recommendation [14, 44] and task assignment problem [43], only one side, e.g., the customers, will make decisions. However, in our problem, both two sides, namely organizers and attenders, will make decisions based on their unique preferences. For instance, as mentioned above that organizers are intuitively urged to select proper attenders to ensure the quality of events. This phenomenon results in the "*two-way selection*" that could not be fully described by traditional methods, thus more comprehensive solution is still required.

At the same time, some researchers noticed these limitations and proposed some algorithms based on reciprocal recommendation. For instance, [24] implemented a reciprocal recommender for online dating with higher success rate and partly solved the cold start problem. Also, [1] used some content-based collaborative filtering algorithms in reciprocal recommendation to further enhance the performance. However, the core idea of these methods is to combine the scores of both sides through some aggregation functions(e.g. algebraic mean, harmonic mean) and cannot model the decision making processes for both sides effectively in social event participation analysis.

While reviewing the process of "two-way selection", we realize that this process could be approximately described by the so-called "**stable matching**" task, which has been widely studied in the field of economics and management science, e.g., the classic *Gale-Shapley* (G-S) algorithm [5] and its variants [26]. In a stable matching task, each entity in one side of selection (like recruiter or job seeker) is required to rate *all* the entities in opposite side, and then G-S algorithm will provide a matching plan without "unstable" pairs, i.e., no pairs will break up due to the attraction of a better choice. In that way, "two-way selections" among events and attenders seem to be well formulated. Unfortunately, *G-S* algorithm and variants rely on the complete ranking list, which are usually unavailable in real world. For instance, in the example in Figure 1, there is no explicit rule to explain how to select resumes. Thus, these heuristic algorithms could not be directly deployed for our task.

To deal with that problem, in this paper, we propose a novel two-stage framework to reflect the two-way selection process between users and events. Specifically, we intuitively

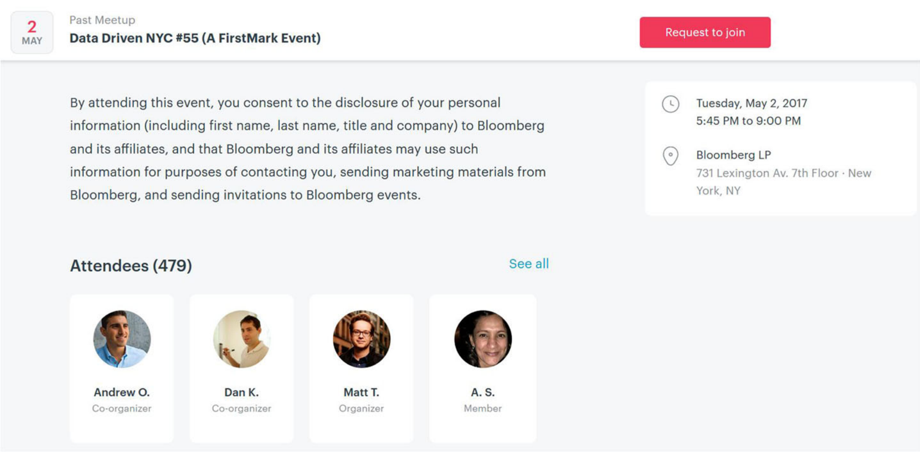

Figure 1 Example of selection for potential attenders by event organizers

assume that the historical attendances reflect the final balance of mutual selection. Thus, if we could accurately reveal the user/event profiles, we will optimally "*reproduce*" the two-way selection process which best fits the historical attendances. To that end, we first design utility functions for both users and events (organizers), and then conduct two layers of optimization tasks, where the inner layer of optimization targets at estimating parameters (e.g., user profiling) based on current matching results, and the outer layer targets at simulating the two-way selection where we adapt the classic G-S algorithm for stable matching. As parameters estimated, the future attendance will be easily predicted only via the outer layer matching task. The contribution of our paper could be summarized as follows:

1. To the best of our knowledge, we are the first to re-consider the task of social event analysis in the perspective of "*two-way selection*".
2. With adapting and enhancing the classic G-S algorithm, we propose a novel framework to formulate the two-way selection process, and reveal accurate user/event profiles simultaneously.
3. Experiments based on real-world data sets validate the effectiveness of our framework and reveal some interesting rules for selection process.

Overview The rest of this paper is organized as follows. Prior arts are summarized in Section 2. In Section 3, we re-formulate the event organization process from the perspective of two-way selection, and then propose our technical framework. The technical details for optimization tasks are explained in Section 4. The experimental results are illustrated in Section 5. Finally, this paper is concluded in Section 6.

2 Related work

In this section, we will summarize prior arts on social event analysis, as well as stable matching problem.

2.1 Social event analysis

As mentioned above, during the event participation analysis, both users and events (organizers) should be considered. Thus, related prior arts will be separately introduced.

In the viewpoint of *users* (attenders), prior arts mainly focus on the event recommendation tasks. For example, [18] integrated several contextual signals from event-based social networks to recommend proper events, and [23] targeted at learning personalized weights for criterion to improve the performance, while [6] mainly focused on the influence of social group, and then proposed a new Bayesian latent factor model. At the same time, some other researches did not directly focus on recommendation social events, but attempted to recommend the event-driven social groups. For example, [41] integrated comprehensive features into a unified model to recommend groups for users, and [10] studied the event team formation by combining the ideas of team formation and influence maximization. Besides, some other practical issues have been discussed. For example, [15] presented a hybrid collaborative filtering model to predict users' social influences on upcoming events, and [40] discussed the cold-start problem with comprehensively utilizing content and context information for event representation.

Meanwhile, in the viewpoint of *events* (organizers), the basic task is recommending participants for events. For example, [8] proposed the feature-based matrix factorization model to optimize the pairwise rankings of participants, and [38] investigated how to select potential participants in EBSNs by measuring user preferences and concerning social influence which attracts the largest group of attenders. Along this line, some researchers targeted at revealing the rules to design an event. For example, [12] attempted to assign set of events for a group of users to attend via a greedy-based algorithm, and [7] considered the roles of attenders to maximize the "harmony" of the event, while [39] conducted modeling analysis of several contextual factors for developing a group-based social influence propagation network. Finally, some other works targeted at forming the proper group for maximal participation of event, like [29] and [11].

At the same time, some researchers considered the recommender system for both sides, which is named reciprocal recommendation. For instance, [24] calculated the harmonic mean of the score of both sides as a reciprocal score for recommendation in online dating platform and improved the success rate. This method could also solve the cold-start problem partly. Also, [1] took more features into consider, namely user profiles and user interactions, to better estimate the score of both sides and achieved better performance. Besides, some others prior arts, e.g., [37] and [22] applied reciprocal recommendation to other application scenarios such as recruitment and both had good experimental results. Although these works outperform some traditional ones, the core idea of them is to combine the two scores of both sides for recommendation, which leads to a result that these methods cannot model well the decision-making processes separately for both sides.

Different from those prior arts, we consider both user and event separately and then propose the comprehensive framework to reproduce the two-way selection process, which could be novel and effective in practice.

2.2 Stable matching

Then, we turn to summarize the related works of stable matching, which was first studied in [5]. In this paper, two typical two-way selection scenarios were discussed, i.e., the college admission and stable marriage problem, and then the famous Gale-Shapley (G-S) Algorithm was proposed to deal with these tasks. Afterwards, [26] formally defined the concept of two-sided matching. In 2012, authors of the former two papers won the Nobel Prize of economics. Along this line, large efforts in the fields of mathematics, economics and computer science have been made on this problem. For example, [19] extended the classic one-to-one matching to the many-to-one matching problem, and [28] even discussed the more general many-to-many matching problem. Besides, some variants have been studied, e.g., [20] focused on the stable matching for unequal sets, and [21] proposed the so-called *break-marriage*, and then put forward a method to find all the solutions for the stable marriage problem.

Based on these theoretic achievement, some practical problems were also analyzed. For example, [30] considered dynamic controller assignment so as to minimize the average response time of the control plane, and [2] studied how to incorporate social effects into partner selection and matching scenarios. Moreover, [31] considered the stability of matching between drivers and riders to solve the assignment problem in the dynamic ride-sharing system, [46] presented a group pattern discovery method in trajectory streams to model various group incidents. Along this line, [45] put forward a novel reference-based Spatio-temporal trajectory compression framework, and [17] proposed a multithreading heuristic algorithm to solve the multi-constrained graph pattern matching problem. Besides, [13] proposed a novel end-to-end framework with multi-level attention mechanism for matching process with quantitative interpretation.

Different from those prior arts which mainly rely on the complete ranking list of opposite side, in our study, we conduct two-layers optimization task to estimate the parameters, thus ranking list, as well as profiles for both sides could be accurately estimated.

3 Technical framework for two-way selection

In this section, we will propose our framework to reproduce the two-way selection in detail, including problem definition, utility functions and introduction of interation process.

3.1 Problem definition with preliminaries

First, we will formulate our event participation analysis task with preliminaries. As mentioned above, prior arts only recommend events to users, without considering reversed selection from events (organizers). However, in real world applications, events may select potential attenders, or even actively **invite** preferred users, which results in the *two-way selection process* as follows:

1. The event organizers (blue nodes in Figure 2) will rank all the potential attenders (red nodes), and then send invitations to those who are in the top list, which corresponds to all the lines in Figure 2.
2. Every user will select the most interested invitation to accept, which result in the *full line* in Figure 2, while the rest ones are rejected invitations, which are labeled as the *dashed lines*. Sometimes, users may even reject all the invitations, just like u_2 in Figure 2.
3. As all the invitations are responded, events with remaining quota will send invitations to the rest users in the ranking list, until all the quotas are filled, or there are no more users to invite.

Along this line, the G-S algorithm could be utilized to formulate the two-way selection process in our event participation analysis. Indeed, in this process, there are actually

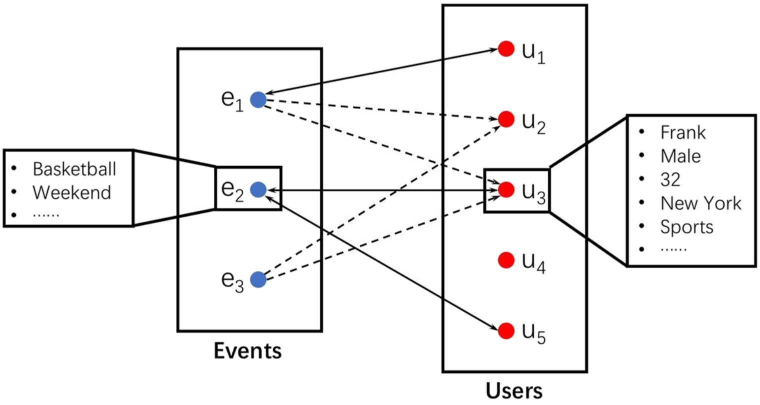

Figure 2 An toy example of two-way decision making process

two problems to be solved. First, event organizers will select users to send invitations, which leads to the "**Invitation Task**". In this case, the results reflect *event rules* to select users. Second, users will select the most interested invitation to accept, which leads to the "**Acceptance Task**" and reflects users' *preferences*. Thus, different from traditional recommendation task that only considers one target, we attempt to predict two values, namely the event invitation and acceptance results. At the same time, their reasons for selections, i.e., *event rules* or *user preferences*, are summarized as **utility functions** for both sides. To be specific, the event participation analysis problem could be defined as follow:

Definition 3.1 (Problem Definition) Given the target user set **U** and organizers set **G**, as well as their events **E**, in which each $e_{jk} \in \mathbf{E_j}$ organized by $g_j \in \mathbf{G}$ contains a certain limited quota, for each $u_i \in \mathbf{U}$, we target at predicting ϕ^i_{jk} for "*Invitation Task*", as well as ψ^i_{jk} for "*Acceptance Task*".

Here $\phi^i_{jk} = 1$ indicates that u_i is invited by e_{jk}, and $\phi^i_{jk} = 0$ indicates the disregard. Similarly, $\psi^i_{jk} = 1$ means that u_i accepts the invitation from e_{jk}, and $\psi^i_{jk} = 0$ means the rejection. In the following section, we will introduce the details about utility functions.

3.2 Utility functions

We design utility functions for both sides as reasons of selection. Similar with prior arts, first, we have a normalized vector $\boldsymbol{p_i}$ to describe preference of u_i, in which each dimension indicates u_i's interest in one certain topic. Correspondingly, we have another normalized vector $\boldsymbol{a_{jk}}$ to describe the attributes of event e_{jk}, which is organized by group (organizer) g_j. Intuitively, the similarity between $\boldsymbol{p_i}$ and $\boldsymbol{a_{jk}}$ could roughly estimate the acceptance of u_i to e_{jk}. Meanwhile, organizers also prefer those attenders who are interested in the event topics. Thus, we select the widely used *Cosine* value, denoted as $\cos < \boldsymbol{p_i}, \boldsymbol{a_{jk}} >$ to measure the similarity, which will be treated as basic part of utility functions for both sides.

Second, we further consider the social factors due to the "word-of-mouth" effect. When a user compares the invitations, usually, they may listen to the suggestions from his/her friends. Along this line, we introduce the adjacent matrix $W = \{w_{ir}\}$, where each element w_{ir} measures the social influence strength from u_r to u_i. Then, following the basic idea of Independent Cascade model [9], we define the social influence for u_i of e_{jk} as $1 - \prod_{r \in \boldsymbol{N}^i_{jk}}(1 - w_{ir})$, where $\boldsymbol{N}_{jk}$ indicates the friends who decided to accept invitation of e_{jk}.

At the same time, for event organizers, they wish to invite those influential users who could attract more attenders in the future, which could be usually captured by influence models. Intuitively, as mentioned in prior arts like [35] and [32] that a linear social influence model is essentially PageRank with prior, following this idea, we measure the overall influence of each user u_i based on the PageRank value on adjacent matrix W, denoted as $pagerank(W)[i]$.

Third, we agree that users' decisions could be affected by some implicit factors, e.g., loyalty to one certain group. Obviously, higher loyalty leads to higher probability to accept the invitation. Since this factor could be hardly estimated by explicit statistical index, we formulate it as h_{ij}, which indicates the loyalty of u_i to the group g_j. Intuitively, if for user u_i, all the h_{ij} are relatively high, it means that u_i could be active to attend any event.

Based on all the definitions above, finally, we could summarize the utility functions for both sides. For the event group (organizer) g_j, we denote X^i_{jk} to measure the **utility** score of u_i for event e_{jk}, and two factors will be considered, namely social and interest factor as follow:

$$X^i_{jk} = pagerank(W)[i] \cdot \cos < \boldsymbol{p_i}, \boldsymbol{a_{jk}} > \tag{1}$$

Correspondingly, for user u_i, denote by Y^i_{jk}, the **utility** score of event e_{jk} organized by g_j, by considering loyalty, interest and social factor as follow:

$$Y^i_{jk} = h_{ij} \cdot \cos < \boldsymbol{p_i}, \boldsymbol{a_{jk}} > \cdot \left(1 - \prod_{r \in \boldsymbol{N}^i_{jk}} (1 - w_{ir})\right) \tag{2}$$

Given the utility functions, we could "*reproduce*" two-way selection process as stable matching with estimated ranking lists. In the future, we will consider more comprehensive factors with additional data sources.

What should be noted is that, some users may **reject all the invitations** (like u_2 in Figure 2), which means that the scores for all the events may not surpass their threshold. Thus, to describe the threshold, we design a "**virtual**" group of organizers as g_{m+1}, which keeps inviting all the users, i.e., $X^i_{m+1,k} \equiv 1, \forall u_i \in \boldsymbol{U}$. Correspondingly, for each user u_i, he/she accepts the invitation from g_{m+1} with a fixed probability as gr_{ik}, which will be estimated based on the acceptance probability of u_i in recent several months. Some related notations are summarized in Table 1.

3.3 Two layers of iterations

So far, given the utility functions for both sides, we rank the events and users for the two-way selection. However, some parameters are initially unknown in real world as mentioned above, to achieve the complete ranking list, we design two iterations separately for matching and parameter estimation, which is intuitively shown in Figure 3, and listed step-by-step as follows:

1. First of all, we initialize all the parameters, and then the utility scores X^i_{jk} and Y^i_{jk} between u_i and e_{jk} could be roughly initialized, which is shown in Step **1** in Figure 3.
2. **Outer Iteration for Matching**. Secondly, we adapt the G-S algorithm to achieve the stable matching, and then ϕ^i_{jk} and ψ^i_{jk} will be predicted. Afterwards, we compare them with the ground truth $\hat{\phi}^i_{jk}$ and $\hat{\psi}^i_{jk}$ to calculate the loss, which is shown in Step **2** and **3**.

Table 1 Mathematical notations

Notation	Description
$\boldsymbol{U} = \{u_i\}$	the set of users
$\boldsymbol{G} = \{g_j\}$	the set of groups
$\boldsymbol{E} = \{e_{jk}\}$	the set of events
$\boldsymbol{p_i}$	the preference vector of user u_i
$\boldsymbol{a_{jk}}$	the attribute vector of event e_{jk}
w_{ir}	the social impact of u_r on u_i
h_{ij}	the tolerance of u_i to g_j
gr_{ik}	the trends for u_i to refuse all the invitations
ϕ^i_{jk}	invitation for e_{jk} to u_i
ψ^i_{jk}	attendacce for u_i to e_{jk}
X^i_{jk}	the utility score of e_{jk} on u_i
Y^i_{jk}	the utility score of u_i on e_{jk}

3. **Inner Iteration for Parameters**. Thirdly, we update all the parameters based on the loss function, and then the utility score X^i_{jk} and Y^i_{jk} are re-calculated, which is shown in Step **4** and **5**.
4. We repeat Step **2-5** until parameters and predictions are converged.

Technical details of these two iterations will be shown in the following section. Based on all the definitions and formulations above, finally, we now have our **two-stage framework** to reproduce the two-way selection for event participation analysis. Specifically, in training stage, both iterations for matching and parameter estimation will be conducted. However,

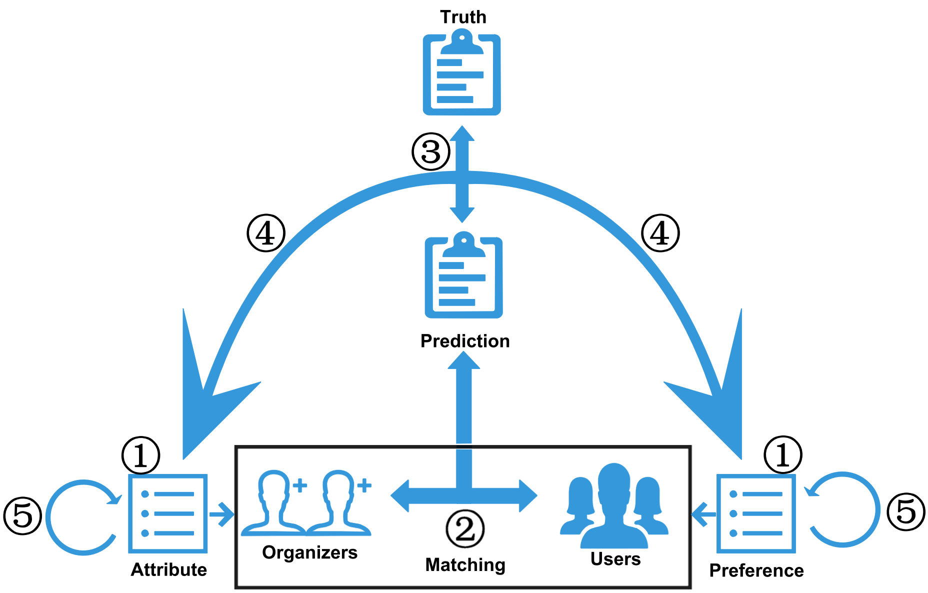

Figure 3 The iteration process for parameter estimation of two-way selection

in test stage, as parameters have been estimated, only iteration for matching works. Details for iterative process in training stage are summarized in Algorithm 1.

Algorithm 1 Iterative process for training stage.

Input: user set $\mathbf{U} = \{u_i\}$, community set $\mathbf{G} = \{g_j\}$, event set $\mathbf{E} = \{e_{jk}\}$, invitation records $\hat{\phi}^i_{jk}$, attendance records $\hat{\psi}^i_{jk}$;
Output: user profile $< \boldsymbol{p_i}, h_{ij} >$, social impact w_{ir};
Initialize $\boldsymbol{p_i}, \boldsymbol{a_{jk}}, h_{ij}, gr_{ik}, w_{ir}$;
$Iteration\_State = 1$;
while $Iteration\_State \neq 0$ **do**
    $Iteration\_State = 0$;
    Update X^i_{jk} and Y^i_{jk};
    ϕ^i_{jk}, ψ^i_{jk} = Matching(X^i_{jk}, Y^i_{jk});
    if I_{ijk} or A_{ijk} changed **then**
        $Iteration\_State = 1$;
        Update $\boldsymbol{p_i}, h_{ij}, gr_{ik}, w_{ir}$ with Gradient Descent method;
    end if
end while
return $\{\boldsymbol{p_i}, h_{ij}, gr_{ik}, w_{ir}\}$;

4 Technical details for two iterations

In this section, we will introduce the technical details for two layers of optimization tasks in two iterations, separately for matching and parameter estimation. Also, some related properties will be proved.

4.1 Outer iteration for matching

As mentioned in the preceding section, the *Outer Iteration* for stable matching process will be solved by adapted *G-S* algorithm according to the utility scores X and Y. Specifically, since each event e_{jk} is limited by the quota η_{jk}, we constrain that e_{jk} could invite at most η_{jk} attenders in total. At the same time, for each user u_r who was not invited by the event e_{jk}, related Y^r_{jk} will be set as 0, despite his/her preference. Along this line, matching results for both *Invitation* and *Acceptance* tasks will be simultaneously achieved. Details for the matching process are summarized in Algorithm 2.

Though matching results could be easily achieved with *G-S* algorithm, we would like to ensure the optimality and uniqueness of our solution. Indeed, related properties of basic *G-S* algorithm have been well proved. However, there exist two major differences between basic *G-S* algorithm and our solution. First, in our task, the amount of users and event quotas are usually unequal. In this case, some users/events will not be selected during the two-way selection process. Thus, the order of selection may affect the final matching result, as intuitively, those who are selected at last may be probably ignored. Luckily, according to the prior art [20], the optimality and uniqueness of stable matching are ensured even for the unequal sets, which means that the order of selection will not affect the matching results.

Second, basic *G-S* algorithm was designed for the one-to-one matching problem, e.g., the classic "stable marriage" problem. But in our task, each event will accept at most η_{jk} attenders. Luckily, this difference could be easily solved, if one event could be treated as η_{jk} individual "events", where new "events" share the same parameters but each "event" could only accept one attender. In this case, the many-to-one matching problem will be transferred as a basic one-to-one matching problem. As mentioned above that the optimality of classic one-to-one matching has been proved, the rationality of our solution could be realiable.

Algorithm 2 Outer iteration for matching.

Input: set of users $\boldsymbol{U} = \{u_i\}$, set of events $\boldsymbol{E}_k = \{e_{jk}\}$, utility score X^i_{jk}, Y^i_{jk}, invitation quotas η_{jk}, number of participants θ_{jk};
Output: the matching result ϕ^i_{jk}, ψ^i_{jk};
for $e_{jk} \in \boldsymbol{E}_k$ **do**
  for $u_i \in \boldsymbol{U}$ **do**
    if X^i_{jk} in top-η_{jk} of X_{jk} **then**
      $\phi^i_{jk} = 1$;
    else
      $\phi^i_{jk} = 0$, $X^i_{jk} = -1$;
    end if
  end for
end for
while not every user has matched an event **do**
  for u_i who has not matched any event **do**
    e_{jk} is the event with the highest Y^i_{jk};
    if the number of users matching with $e_{jk} < \theta_{jk}$ **then**
      $\psi^i_{jk} = 1$;
    else
      u_r is the user with the lowest X^r_{jk} satisfies $\psi^r_{jk} = 1$;
      if $X^r_{jk} > X^i_{jk}$ **then**
        $\psi^i_{jk} = 0$, $Y^i_{jk} = -1$;
      else
        $\psi^i_{jk} = 1$, $\psi^r_{jk} = 0$, $Y^r_{jk} = -1$;
      end if
    end if
  end for
end while
return $\{\phi^i_{jk}, \psi^i_{jk}\}$

4.2 Inner iteration for parameter estimation

Afterwards, we turn to introduce the optimization task in the inner iteration for parameter estimation. Specifically, we define the loss function based on prediction of both *Invitation* and *Acceptance* tasks as follow:

$$loss = \sum_{i=1}^{n} \sum_{j=1}^{l_1} \sum_{k=1}^{m+1} \left[\alpha^i_{jk} \cdot \left(X^i_{jk} - \widetilde{X}_{jk} \right) + \beta^i_{jk} \cdot \left(Y^i_{jk} - \widetilde{Y}^i_j \right) \right] \tag{3}$$

where α and β act as two signals for wrong prediction. For instance, $\alpha^i_{jk} = 1$ if $\phi^i_{jk} = 1$ and ground truth $\hat{\phi}^i_{jk} = 0$, which indicates that u_i is wrongly invited. On the contrary, $\alpha^i_{jk} = -1$ if $\phi^i_{jk} = 0$ and $\hat{\phi}^i_{jk} = 1$, which indicates that u_i is wrongly missed. Otherwise, $\alpha^i_{jk} = 0$. Similarly, we could define β^i_{jk} based on ψ^i_{jk} and $\hat{\psi}^i_{jk}$. Also, $\widetilde{X}_{jk} = \min_{\phi^r_{jk}=1} X^r_{jk}$ to presents the boundary of invitation, i.e., $X^i_{jk} > \widetilde{X}_{jk}$ indicates that u_i should be invited by e_{jk}. Similarly, we could define $\widetilde{Y}^i_j$ as the boundary of acceptance.

Convergence analysis Considering that all the parameters here could be normalized vectors, or within the value range as [0,1], which means that they are bounded according to the definitions. Also, the loss function is L-smooth. Thus, all the derivatives for parameters have the property of Lipschitz Continuity [27] based on the boundedness of the parameters and the L-smooth property of the loss function. As a result, the parameter iteration process will be convergent due to the Lipschitz Continuity property of the derivatives.

Complexity analysis The **Algorithm** 1 leads to the time complexity for one round of the Inner Iteration for Parameter Estimation is $\Theta\left(\mathcal{P}+\mathcal{H}+\mathcal{G}r+\mathcal{W}\right)$, which results in the overall time complexity is $\Theta\left(\left(\left(\mathcal{P}+\mathcal{H}+\mathcal{G}r+\mathcal{W}\right)R_1+\mathcal{M}\right)R_2\right)$ for the training stage. Here, $\mathcal{P}$, $\mathcal{H}$, $\mathcal{G}r$, $\mathcal{W}$ represent for the training time required for the parameters $\boldsymbol{p_i}$, h_{ij}, gr_{ik} and w_{ir} each inner round; $\mathcal{M}$ represents for the time required for matching each outer round; R_1 and R_2 are the numbers of rounds for inner and outer iterations.

According to (1), (2), (3) and **Algorithm** 2, the overall time complexity for the training stage is $\Theta\left(n^3 lmR_1R_2\right)$, where $n = |U|$, $m = |G|$ and l is the number of events for each group used in the training stage.

5 Experiments

To validate the performance, in this section, we conduct a series of experiments on a real-world data set with empirical case study and discussions.

5.1 Experimental settings

5.1.1 Data profiling

We perform our experiments on the data set extracted from Meetup,[1] one of the most popular websites that facilitates offline events. In Meetup, each event is organized by a group, corresponds to the "*organizer*" in our framework and users can join as many groups as they want. In this case, those overlapping users may face to multiple invitations of events from different groups simultaneously, and they have to select the most preferred invitation to accept, as discussed in the prior art [42], or they could reject all of them (i.e., selecting the "virtual" event as mentioned at the end of Section 3.2).

Considering that each event is organized by a group, and all the group members could receive the notice of event, thus, we approximated the invitation process in the perspective of group. In detail, we first intuitively assumed that the events organized by the same group share the same rules to select users. Second, we could approximately treat that only current members are **invited**, while the non-members, or those who quit the group are **rejected**. As mentioned in prior art [33] that most attenders will quit the group after attending only a few events, which means that they could not survive in the group. In this case, our settings could be reasonable.

5.1.2 Data preprocessing

For each group, to ensure the quality of data set, we first picked up those users who have participated in at least one event, who are treated as "historically active users". Along this

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

line, we compared the overlapping rate of historical active users for each pair of groups. Specifically, if the overlapping rate is larger than a pre-set threshold, then this pair of groups are selected as "*competitive group pair*", which indicates that, when these two groups hold events almost simultaneously, their overlapping users could probably face to the multiple invitations, which will definitely lead to the "two-way selection process".

To that end, for each competitive group pair, we traversed all the events organized by these two groups on the timeline, and then screened for the pairs of events that are held at the same time (within a short period). Intuitively, these pair of events are named as "competitive event pairs". Finally, we have 108 pairs of groups with more than 10% overlapping members. In average, each group contains 25 events and 116 users. Also, we have the brief textual introduction for different groups, events and users, as well as the invitation and responses records.

To represent the textual information, we took the descriptions of all the groups, events and users as a corpus, and then conducted the Latent Dirichlet Allocation (LDA) model. As a result, 2,856 key words were exploited and all the $\boldsymbol{p_i}$ of u_i and $\boldsymbol{a_{jk}}$ of e_{jk} were represented as a 30-dimensional vector. Here 30 topics are similar to the 34 categories of groups defined by Meetup.

Besides, h_{ij} and gr_i are initialized as 1, and the social impact w_{ir} is initialized by the ratio of events that u_i and u_r have attended together, which is a heuristic solution widely used in location-based social network. Clearly, the social impact is asymmetrical for user pairs.

5.1.3 Evaluation metrics

Since both *Invitation* task and *Acceptance* task could be treated as a typical classification problem. Thus, we evaluated the prediction performance with the widely-used **Precision** and **Recall** for both two tasks.

5.1.4 Baseline selection

Since most prior arts focused on one-way selection, we combined several widely-used methods to simulate the two-way selection process. Specifically, four kinds of methods were considered:

– Logistic Regression (**LR**), basic pointwise solution for classification task.
– Binary classification. Here we chose Discrete Choice Model (**DCM**) [3] to achieve the exclusive selection, which well fits user selection process. Besides, we chose Factorization Machine (**FM**) [25] as one of the novel method in recent years.
– Learning to Rank (**LTR**), which transfers the classification task as a ranking problem to pick the top ones. Here we select SVMRank as an example.
– Social Model (**SM**). With considering social factors, we could treat the invitation process as diffusion of social influence. Here we select Independent Cascade (**IC**) [9] as an example.
– Reciprocal Recommendation (**RR**). These recommender techniques could be quite similar with our framework, in which both sides will be partially satisfied based on their preference. However, these techniques rely on explicit preference and pre-defined aggregation functions.

Along this line, three kinds of combinations were generated. First, we selected *LR* for *Invitation* task, and classification mehod for *Acceptance* task, thus we have **LRDCM** and **LRFM**. Second, we selected *LTR* for *Invitation* task, and *LTR* or classification method for *Acceptance* task, which leads to **LTR2**, **LTRDCM** and **LTRFM**. Finally, we selected social diffusion model for *Invitation* task, and classification method for *Acceptance* task, thus we have **ICDCM** and **ICFM**.

At the same time, we also considered two reciprocal recommender methods with different aggregation functions, i.e. algebraic mean and harmonic mean namely **AMRR** and **HMRR**.

In addition, for evaluating the power of G-S algorithm, we also design a heuristic method of stable matching, where users/events are ranked based on the topical similarity based on their description, named as **SBSM**. Totally, 10 baselines were generated.

5.2 Experimental results

5.2.1 Overall performance.

First, we will introduce the experimental results for overall performance. Specifically, 75% of events were treated as training samples, and the rest 25% were test samples. The performance is shown in Tables 2 and 3, where our solution is named as *Two-Way Selection Method* (**TWSM**). What should be noted is that, as each e_{jk} invited as many as η_{jk} attenders, i.e, the the quota of e_{jk}, the amount of users in prediction is exactly the same with the ground truth. Thus, metric for precision and recall are the same for our **TWSM** and **SBSM**, so do ranking-based baselines and reciprocal recommender baselines, e.g., LTR2, LTRDCM, LTRFM, AMRR, and HMRR.

According to the results, **TWSM** outperforms all the baselines in both two tasks, especially in event acceptance prediction task, where the improvement achieves 5-50%. At the same time, SBSM also performs better than all the one-way baselines, which further indicates the power of two-way selection process. Besides, we realize that ranking algorithms may better reflect the selection process, as they perform better than classification or social algorithms.

Table 2 Performance comparison for invitation task

Model	Invitation Task					
	Precision			Recall		
	Val	Imp(%)	Var	Val	Imp(%)	Var
TWSM	**87.39**	–	0.013	**87.39**	–	0.013
LRDCM	76.91	13.63	0.0243	76.9	13.64	0.0259
LTR2	86.4	1.15	0.0167	86.4	1.15	0.0167
LTRDCM	86.4	1.15	0.0167	86.4	1.15	0.0167
ICDCM	73.14	19.48	0.0242	80.11	9.09	0.0207
LRFM	76.91	13.63	0.0243	76.9	13.64	0.0259
LTRFM	86.4	1.15	0.0167	86.4	1.15	0.0167
ICFM	73.93	18.21	0.024	80.27	8.87	0.0222
AMRR	82.58	5.82	0.0332	82.58	5.82	0.0332
HMRR	83.03	5.25	0.0397	83.03	5.25	0.0397
SBSM	85.86	1.78	0.015	85.86	1.78	0.015

Table 3 Performance comparison for acceptance task

Model	Acceptance Task					
	Precision			Recall		
	Val	Imp(%)	Var	Val	Imp(%)	Var
TWSM	**76.93**	–	0.075	**76.93**	–	0.075
LRDCM	51.95	48.08	0.0829	53.35	44.2	0.0636
LTR2	62.73	22.64	0.0823	70.65	8.89	0.0698
LTRDCM	58.8	30.83	0.0735	58.2	32.18	0.0662
ICDCM	50.33	52.85	0.0725	48.19	59.64	0.0567
LRFM	51.71	48.77	0.066	55.34	39.01	0.0608
LTRFM	56.08	37.18	0.0771	56.51	36.14	0.0718
ICFM	52.05	47.8	0.058	50.64	51.92	0.0483
AMRR	59.36	29.6	0.0934	59.36	29.6	0.0934
HMRR	61.29	25.52	0.0872	61.29	25.52	0.0872
SBSM	72.91	5.51	0.0966	72.91	5.51	0.0966

5.2.2 Parametric sensitivity analysis

At the same time, we measured the sensitiveness of the parameter User Overlap Rate (Ur), which is defined by the proportion of overlapping users, i.e., how many users face multiple choices. As shown in Figure 4, for the task of event invitation prediction, we found that when the value of Ur is relatively small (≤ 0.100), both **Precision** rate and **Recall** rate show a large randomness, which is due to the impact of accidental errors. However, the randomness gradually disappears when Ur becomes larger (≥ 0.125). Generally, for our two tasks, the impact of Ur could be limited.

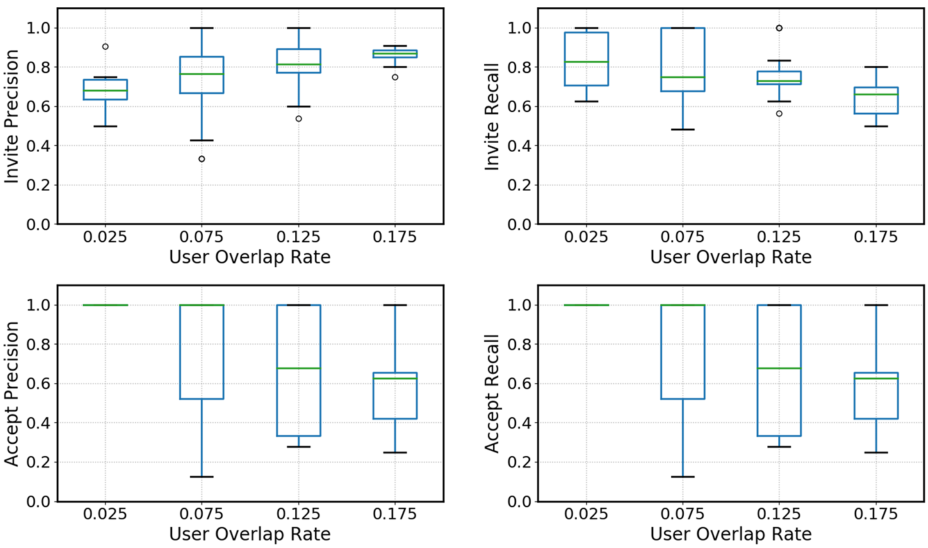

Figure 4 Parametric sensitivity analysis for user overlap rate

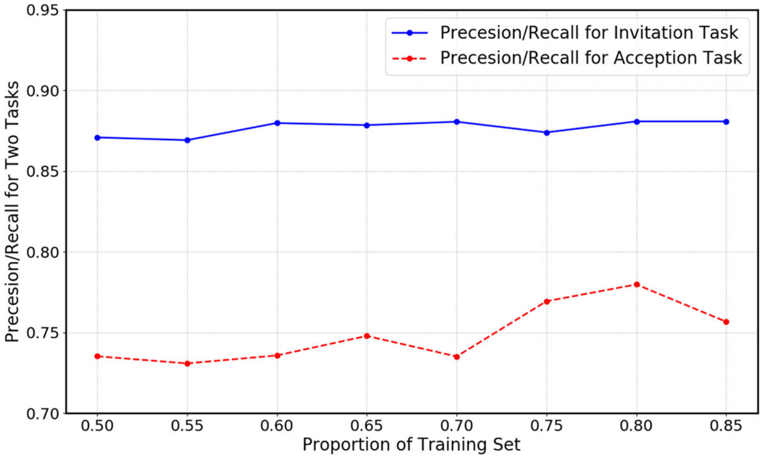

Figure 5 Parametric sensitivity analysis for the partition of the training and test sets

We also measured the influence of the partition for the training and test sets in Figure 5. According to the curves, we can find that different partitions for the training and test sets indeed have little effect on the performance of event invitation prediction, but great interference on the task of event acceptance prediction. It is obvious that when the proportion of training set is too small($< 70\%$), the Precision and Recall rates are less than 0.75. But if when the proportion too big($> 80\%$), the rates will be also degraded, which might be due to the randomness caused by the small sample size in the test set.

5.3 Discussion with case study

To better understand the performance, we further conducted discussion with three case studies as follows.

◇ **What kind of groups could be easily predicted?**

For this issue, obviously, an important factor for each pair of groups is the proportion of new users. Based on the statistical result in Table 4, where five pairs of groups were randomly selected as examples for quantitative discussions, we could find out from the data that higher proportion of new users leads to poor performance. For example, p_1 and p_2 who contain more than 20% new users, achieved a precision rate lower than 0.5. On the contrary, the rest three pairs achieved precision more than 0.6 with about 10% new users. With deep looking into the data, we find that the precision in the test of social event participation of new users is lower than 30%, which severely impaired the overall performance.

Table 4 Examples of pairs of groups for case study

Pairs	p_1	p_2	p_3	p_4	p_5
Invite_Precision (%)	85.29	89.39	87.10	83.47	93.10
Accept_Precision (%)	34.92	49.67	67.86	73.33	84.72
New User Rate (%)	21.79	25.10	9.63	9.46	10.34
Clustering Coefficient	0.63	0.71	0.75	0.76	0.82

Table 5 Examples of users for case study

User	u_1	u_2	u_3	u_4	u_5
Recent Invited Rate (%)	10.00	21.43	79.16	61.11	8.33
Participation	0	14	53	69	89
Recent Ac tivity (%)	NaN	28.57	41.67	40.90	15.38

At the same time, we found that though the proportion of new users in p_2 is higher than p_1, it also performed better in prediction. Then, we notice another important factor – the clustering coefficient of users' network, which is usually used to describe the coincidence degree of friends between users in a social network. We find that although the proportion of new users in p_2 is about 25%, 63% of them have mutual connections before, which leads to a fact that some of the new users may be familiar with the community through his friends' introduction.

In summary, more fresh users lead to worse performance. If you wish to improve the effectiveness, you'd better try to encourage more connections.

◇ **What kind of users could be easily invited?**

For each user, an important factor is the total number of events he has attended. As shown in Table 5, where five users were randomly selected as examples, we could find that for u_1 and u_2, they have attended less than 20 events, with invitation rates less than 25%. Correspondingly, u_3 and u_4 have attended more than 50 events, which leads to higher invitation rate achieve as more than 60%. Clearly, more attendances lead to more chances to be invited.

Another important factor is the proportion for each user to accept the invitation in the previous six months, which can measure user's recent loyalty. We can find that in general, with more attendance, the user loyalty will also improve. However, the last user, namely u_5, is a special case. We noticed that u_5 continuously rejected more than 10 invitations by the target groups in the previous six months. Therefore, he received only 1 invitation in the 12 newly organized events and eventually withdrew from the target groups.

Indeed, these two factors imply that event organizers may select users voluntarily, instead of just being selected by users. Also, as a conclusion, user who has attended more events and maintains a high level of acceptance rate will become more likely to be invited.

◇ **What kind of invitations may be accepted?**

Finally, we discuss about the invitations from different groups. The results are shown in Table 6. Obviously, one crucial factor is the number of organized events. For the former four groups, generally more events result in higher rate of accepted invitations.

At the same time, g_5 could be a special case. With deep analysis, we find that the *proportion of new users* could be also important. With the development of groups, organizers need to kick some inactive users and invite new users, so that their events could be attractive to the mass. However, too many freshmen may also cause some problem. As g_5 holds

Table 6 Examples of Groups for Case Study

Group	g_1	g_2	g_3	g_4	g_5
Accepted Rate (%)	37.98	52.10	61.76	73.33	29.63
Past Events	17	23	57	95	25
New User Rate (%)	5.73	16.28	27.02	22.65	41.95

an average new user rate over 40%, obviously, they could not keep enough loyal members, which leads to the low acceptance rate as lower than 30%, far below the average.

In summary, if a group wants to make its event more attractive with a higher rate of acceptance, it needs to organize more events to enrich its experience, and then seek the balance between absorbing new users, and keeping the loyal members simultaneously.

6 Conclusion

In this paper, we proposed a novel two-stage framework to simulate the two-way selection process between users and events for social event participation analysis. To be specific, with adapting the classic *Gale-Shapley Algorithm* for stable matching, we first designed utility functions for both sides, and then solved two layers of optimization tasks for stable matching and parameter estimation, i.e., capturing user profiling and event rules. Experimental results on real-world data set validated that our method can effectively predict the event invitation and acceptance.

Acknowledgments This research was partially supported by grants from the National Natural Science Foundation of China (Grant No. 61703386, 91746301, U1605251), the Anhui Provincial Natural Science Foundation (Grant No. 1708085QF140), and the Baidu Songguo Project.

References

1. Akehurst, J., Koprinska, I., Yacef, K., Pizzato, L., Kay, J., Rej, T.: Ccra content-collaborative reciprocal recommender for online dating. In: 22nd international joint conference on artificial intelligence (2011)
2. Anshelevich, E., Bhardwaj, O., Hoefer, M.: Friendship and stable matching. In: European symposium on algorithms, pp. 49–60. Springer (2013)
3. Berry, S.T.: Estimating discrete-choice models of product differentiation. The RAND Journal of Economics pp. 242–262 (1994)
4. Chang, B., Xu, T., Liu, Q., Chen, E.H.: Study on information diffusion analysis in social networks and its applications. Int. J. Autom. Comput. **15**(4), 377–401 (2018)
5. Gale, D., Shapley, L.S.: College admissions and the stability of marriage. Am. Math. Mon. **69**(1), 9–15 (1962)
6. Gao, L., Wu, J., Qiao, Z., Zhou, C., Yang, H., Hu, Y.: Collaborative social group influence for event recommendation. In: Proceedings of the 25th ACM international on conference on information and knowledge management, pp. 1941–1944. ACM (2016)
7. Gao, S., Zhang, Z., Su, S., Zia, M.A.: Multi-role event organization in social networks. Inform. Sci. **447**, 229–243 (2018)
8. Jiang, J.Y., Li, C.T.: Who should i invite: predicting event participants for a host user. Knowledge and Information Systems pp. 1–22 (2018)
9. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 137–146. ACM (2003)
10. Li, C.T., Huang, M.Y., Yan, R.: Team formation with influence maximization for influential event organization on social networks. World Wide Web pp. 1–21 (2017)
11. Li, C.T., Shan, M.K.: Composing activity groups in social networks. In: Proceedings of the 21st ACM international conference on Information and knowledge management, pp. 2375–2378. ACM (2012)
12. Li, K., Lu, W., Bhagat, S., Lakshmanan, L.V., Yu, C.: On social event organization. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1206–1215. ACM (2014)
13. Qin, C., Zhu, H., Xu, T., Zhu, C., Jiang, L., Chen, E., Xiong, H.: Enhancing person-job fit for talent recruitment: an ability-aware neural network approach (2018)

14. Li, L., Zheng, K., Wang, S., Hua, W., Zhou, X.: Go slow to go fast: minimal on-road time route scheduling with parking facilities using historical trajectory. The VLDB Journal—The International Journal on Very Large Data Bases **27**(3), 321–345 (2018)
15. Li, X., Cheng, X., Su, S., Li, S., Yang, J.: A hybrid collaborative filtering model for social influence prediction in event-based social networks. Neurocomputing **230**, 197–209 (2017)
16. Lian, D., Zheng, K., Ge, Y., Cao, L., Chen, E., Xie, X.: Geomf++: Scalable location recommendation via joint geographical modeling and matrix factorization. ACM Trans. Inf. Syst. (TOIS) **36**(3), 33 (2018)
17. Liu, G., Liu, Y., Zheng, K., Liu, A., Li, Z., Wang, Y., Zhou, X.: Mcs-gpm: multi-constrained simulation based graph pattern matching in contextual social graphs. IEEE Trans. Knowl. Data Eng. **30**(6), 1050–1064 (2017)
18. Macedo, A.Q., Marinho, L.B., Santos, R.L.: Context-aware event recommendation in event-based social networks. In: Proceedings of the 9th ACM conference on recommender systems, pp. 123–130. ACM (2015)
19. Martínez, R., Massó, J., Neme, A., Oviedo, J.: The blocking lemma for a many-to-one matching model. J. Math. Econ. **46**(5), 937–949 (2010)
20. McVitie, D.G., Wilson, L.B.: Stable marriage assignment for unequal sets. BIT Numer. Math. **10**(3), 295–309 (1970)
21. McVitie, D.G., Wilson, L.B.: The stable marriage problem. Commun. ACM **14**(7), 486–490 (1971)
22. Mine, T., Kakuta, T., Ono, A.: Reciprocal recommendation for job matching with bidirectional feedback. In: 2013 2nd IIAI international conference on advanced applied informatics, pp. 39–44. IEEE (2013)
23. Ogundele, T.J., Chow, C.Y., Zhang, J.D.: Socast*: Personalized event recommendations for event-based social networks: A multi-criteria decision making approach. IEEE Access (2018)
24. Pizzato, L., Rej, T., Chung, T., Koprinska, I., Kay, J.: Recon: a reciprocal recommender for online dating. In: Proceedings of the 4th ACM conference on recommender systems, pp. 207–214. ACM (2010)
25. Rendle, S.: Factorization machines. In: 2010 IEEE 10th international conference on data mining (ICDM), pp. 995–1000. IEEE (2010)
26. Roth, A.E.: On the allocation of residents to rural hospitals: a general property of two-sided matching markets. Econometrica: Journal of the Econometric Society pp. 425–427 (1986)
27. Schwartz, J.T.: Nonlinear functional analysis, vol. 4. CRC Press (1969)
28. Sotomayor, M.: Three remarks on the many-to-many stable matching problem. Math. Soc. Sci. **38**(1), 55–70 (1999)
29. Sun, T., Chen, W., Liu, Z., Wang, Y., Sun, X., Zhang, M., Lin, C.Y.: Participation maximization based on social influence in online discussion forums. ICWSM **11**, 361–368 (2011)
30. Wang, T., Liu, F., Guo, J., Xu, H.: Dynamic sdn controller assignment in data center networks: Stable matching with transfers. In: INFOCOM 2016-The 35th annual IEEE international conference on computer communications, IEEE, pp. 1–9. IEEE (2016)
31. Wang, X., Agatz, N., Erera, A.: Stable matching for dynamic ride-sharing systems. Transportation Science (2017)
32. Xiang, B., Liu, Q., Chen, E., Xiong, H., Zheng, Y., Yang, Y.: Pagerank with priors: An influence propagation perspective. In: 23rd international joint conference on artificial intelligence (2013)
33. Xu, T., Zhong, H., Zhu, H., Xiong, H., Chen, E., Liu, G.: Exploring the impact of dynamic mutual influence on social event participation. In: Proceedings of the 2015 SIAM international conference on data mining, pp. 262–270. SIAM (2015)
34. Xu, T., Zhu, H., Zhong, H., Liu, G., Xiong, H., Chen, E.: Exploiting the dynamic mutual influence for predicting social event participation. IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/TKDE.2018.2851222 (2018)
35. Yang, Y., Chen, E., Liu, Q., Xiang, B., Xu, T., Shad, S.A.: On approximation of real-world influence spread. In: Joint European conference on machine learning and knowledge discovery in databases, pp. 548–564. Springer (2012)
36. Yin, H., Zou, L., Nguyen, Q.V.H., Huang, Z., Zhou, X.: Joint eventpartner recommendation in event-based social networks. ICDE (2018)
37. Yu, H., Liu, C., Zhang, F.: Reciprocal recommendation algorithm for the field of recruitment. J. Inf. comput. Sci. **8**(16), 4061–4068 (2011)
38. Yu, Z., Du, R., Guo, B., Xu, H., Gu, T., Wang, Z., Zhang, D.: Who should i invite for my party?: Combining user preference and influence maximization for social events. In: Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing, pp. 879–883. ACM (2015)
39. Zhang, S., Lv, Q.: Event organization 101: Understanding latent factors of event popularity. arXiv:1709.02024 (2017)

40. Zhang, W., Wang, J.: A collective bayesian poisson factorization model for cold-start local event recommendation. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1455–1464. ACM (2015)
41. Zhang, W., Wang, J., Feng, W.: Combining latent factor model with location features for event-based group recommendation. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 910–918. ACM (2013)
42. Zhao, X., Xu, T., Liu, Q., Guo, H.: Exploring the choice under conflict for social event participation. In: International conference on database systems for advanced applications, pp. 396–411. Springer (2016)
43. Zhao, Y., Zheng, K., Li, Y., Su, H., Liu, J., Zhou, X.: Destination-aware task assignment in spatial crowdsourcing: A worker decomposition approach. IEEE Transactions on Knowledge and Data Engineering (2019)
44. Zheng, B., Su, H., Hua, W., Zheng, K., Zhou, X., Li, G.: Efficient clue-based route search on road networks. IEEE Trans. Knowl. Data Eng. **29**(9), 1846–1859 (2017)
45. Zheng, K., Zhao, Y., Lian, D., Zheng, B., Liu, G., Zhou, X.: Reference-based framework for spatio-temporal trajectory compression and query processing. IEEE Transactions on Knowledge and Data Engineering (2019)
46. Zheng, K., Zheng, Y., Yuan, N.J., Shang, S., Zhou, X.: Online discovery of gathering patterns over trajectories. IEEE Trans. Knowl. Data Eng. **26**(8), 1974–1988 (2013)

Affiliations

Zikai Yin[1] · Tong Xu[1] · Hengshu Zhu[2] · Chen Zhu[2] · Enhong Chen[1] · Hui Xiong[3]

Zikai Yin
yinzikai@mail.ustc.edu.cn

Hengshu Zhu
zhuhengshu@baidu.com

Chen Zhu
zhuchen02@baidu.com

Enhong Chen
cheneh@ustc.edu.cn

Hui Xiong
hxiong@rutgers.edu

[1] Anhui Province Key Lab of Big Data Analysis and Application, University of Science and Technology of China, Hefei, China

[2] Baidu Talent Intelligence Center, Beijing, China

[3] Rutgers University, Newark, NJ, USA