Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

Analyzing Social Event Participants for a Single Organizer

Jyun-Yu Jiang[1,2] and Cheng-Te Li[1]
[2]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
[1]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan
jyunyu.jiang@gmail.com, ctli@citi.sinica.edu.tw

Abstract

Online social networking services allow people to initialize various kinds of offline social events (e.g., cocktail parties, group buying, and study groups), and invite friends or strangers to participate the events in either manual or collaborative manners. However, such invitation manners are tediously long, and irrelevant, uninterested and even spammers can unexpectedly be added into the event. In this paper, we aim at investigating the characteristics of social events participants for a specific organizer. Specifically, we are wondering how social network, user profiles and geo-locations affect user participation when the social event is held by a single organizer. An extensive analysis has been conducted on the real-world event-based social network Meetup dataset. The results of data analysis also demonstrate that these factors actually influence users' event participation.

1 Introduction

While geo-social networking services are ubiquitous due to the persuasiveness and accessibility of mobile devices, users are allowed to interact with each other through both social connections in the online virtual world and geographical activities in the offline physical world. Among geo-social platforms that encourage and record both online social interactions and offline geographical footprints of users, *Event-based Social Networking* (EBSN) services, such as Meetup[1], Plancast[2], and Facebook Events are the most representative. The general function of EBSNs is enabling users to initiate social events or activities and invite their friends or strangers to participate, and those feel interested or available can accept the invitations. The *organizer* of a social event is the user who specifies its objective, tags, and geo-location, and send RSVP (i.e., a request for a response from the invited person) to either his/her friends or strangers. People can accept to participate the event by responding to the RSVP. In some EBSNs, the invited persons can further invite others by sending RSVP while in other services, only the host user has the authority to invite people to participate. However, such invitation procedure can be tediously long and inefficient, and irrelevant users and even spammers could be included into the event group due to human subjectivity and personal bias. In addition, both manual and collaborative invitation manners have some potential to neglect those individuals who are perfectly matched or can significantly contribute to the event. Without a certain intelligent mechanism, the quality of the event participants can be lowered down.

A possible solution to discover potential participants is to exploit users' social network, demographic and semantic information. Given a set of early participants of an event, *Event-Centric Diffusion Analysis* (ECA) (Liu et al. 2012) analyzes which users in the EBSN will accept to attend the event in the future by measuring user relationship. Many previous work also attempt to recommend events to users by their locations (Georgiev, Noulas, and Mascolo 2014)(Zhang, Zhao, and Cao 2015) and social network (Qiao et al. 2014)(Tu et al. 2015). Despite such methods are able to recommend users proper events and find participants for an existing event, they cannot well predict the potential participants for a social event with only one organizer. Especially, analyzing the participants for a single organizer is a cold-start version of ECA: *only the organizer* of the event is given. Therefore, this task is apparently challenging.

As the pioneer of the study about social events with a single organizer, in this paper, we conduct several in-depth data analysis on social event participants with only a single organizer based on the event-based social network Meetup dataset. First, we adopt tags in user profiles as the semantic information, which can represent their interests. The results show that participants generally have more common tags to the organizer than non-participants. It is reasonable because users may be more likely to attend the events held by organizers with more similar interests. Second, the locations of users may affect their participation. Users may not participate in events organized in far places. The data analysis also demonstrates that users tend to participate in closer events. Finally, the relationship between users and the event organizer must be an important factor of user participation. To measure the relationship between users, the social network is one of the most representative approaches. To verify it, we observe the distance and the number of co-friends on the social network. The results indicate that event participants actually have closer relationship to the organizers on the social network. In sum, we analyze three factors which may affect users' participation when only the event organizer is

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

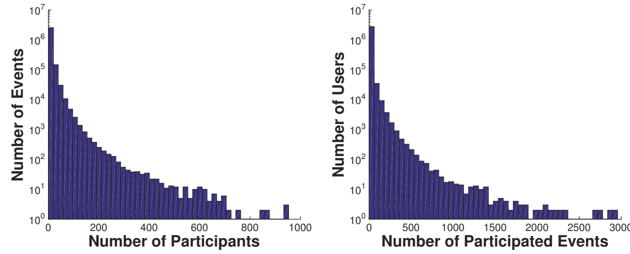
[2]http://plancast.com/

given. These factors may also be important to predict participants for the event organizer.

2 Related Work

The relevant studies about event participant prediction can be divided into two parts: *Event-Centric Diffusion Analysis* and *Event Group Recommendation*. (Liu et al. 2012) is the first attempt to analyze event-based social networks, and first propose to identify event participants by diffusion, termed *Event-Centric Diffusion Analysis* (ECA). Their general setting is that assume an event is created at time t_c and ends at time t_s, which is the time that the event takes place in the real world. Given the event e at time t, where $t_c < t < t_s$, the task is to identify which users will accept the RSVP to event e between t and t_s. All the users who accept to participate event e between t_c and t can be regarded as positive training instances to build the predictive model. It is apparent that our setting is more realistic and general than ECA (Liu et al. 2012): we propose to analyze participants for a single organizer who creates the event at time t_c. It is believed that our setting is more useful to help the host organize the event efficiently and effectively in real-world applications. Note that although ECA (Liu et al. 2012) also examine the cold-start case (i.e., only the organizer is used), the performance is not satisfying. Although a recent study (Yu et al. 2015) also aims to find the potential participants of for an organizer of an event, they formulate the problem as a preference-based influence maximization problem to select the participants that can lead to higher influence spread.

Event Recommendation in event-based social networks (EBSN) aims at recommending either online social groups (Zhang, Wang, and Feng 2013) or offline geographical events (Georgiev, Noulas, and Mascolo 2014)(Qiao et al. 2014)(Zhang, Zhao, and Cao 2015)(Tu et al. 2015) to users and predict whether each user will be willing to join the groups and/or participate the events (held by other members in the group). Note the organizer's information is not considered in EBSN event recommendation. but have the User spatio-temporal mobility patterns, place semantics, and social factors are jointly analyzed and used for event recommendation (Georgiev, Noulas, and Mascolo 2014). A heterogeneous graph-based recommender (Pham et al. 2015) is developed to jointly recommend groups and events to users. Some studies (Du et al. 2014)(Macedo, Marinho, and Santos 2015) further combine both content (event description) and context information (social, spatial, and temporal) to recommend events for a target user. A dynamic social influence approach (Xu et al. 2015), which models how user mutually influence their willingness on event participation, is devised to recommend events to users. Another recent study (Hu, Farnham, and Talamadupula 2015) quantifies degree of user engagement on events in Twitter for event recommendation for users. While studies in this task focus on treating "events" as "items" in the context of traditional recommendation, considering different spatial, temporal, and social factors, our study alternatively provides an extensive analysis on investigating the underlying factors that derive users to participate in the event organized by a certain organizer.



(a) Event Distribution (b) User Distribution

Figure 1: The distributions of events and users.

3 Data Analysis

In this section, we first conduct several data analysis to learn what factors may affect event participant prediction.

3.1 Data Settings

For the data analysis, we adopt the real-world dataset of Meetup.com, which is an event-based social service. Users can express their interests to offline events by sending RSVPs. In the dataset, each RSVP represents that a user participated in a certain event. Each user has a registered location with its latitude and longitude. Events are also associated with the geographical information of event locations. To represent the personal interests, users can apply some tags such as "travel-photography" to describe themselves. Moreover, users can join online social groups and share comments and photos with other members in the same groups.

We show the distribution of number of events for different event sizes in Figure 1(a). It can be observed that most events have very limited number of participants (e.g. less than 50), and only few events are large-scale (i.e., the participant number is higher than 400). In addition, the distribution of number of users for different numbers of event participations is shown in Figure 1(b). Similarly, most users participant in fewer events (e.g. ≤ 50) while rare users attend events very frequently (e.g. ≥ 1000). These two distributions exhibit the severe data sparsity problem, along with that our setting is based on *only one* organizer, which leads to a kind of cold-start analysis, so it is a very challenging task.

To remove inactive users, we do not consider those users who participate in less than 10 events and those events with less than 10 participants. Since the data does not provide the time that a user participates in an event, in the experiments, we randomly select a user as the organizer for each event. Besides, to eliminate the location bias, we separate users and events into the subsets of eight cities, including New York City (NYC), Log Angeles (LA), Chicago (CHI), San Diego (SD), San Jose (SJ), Phoenix (PHX), London (LDN) and Paris (PA), by the user locations.

3.2 Semantic Information

The tags can be treated as the semantic information for determining users' interests. In other words, these information is able to be a hint to find the relationships between users and the organizer. Users may be more likely to attend the events

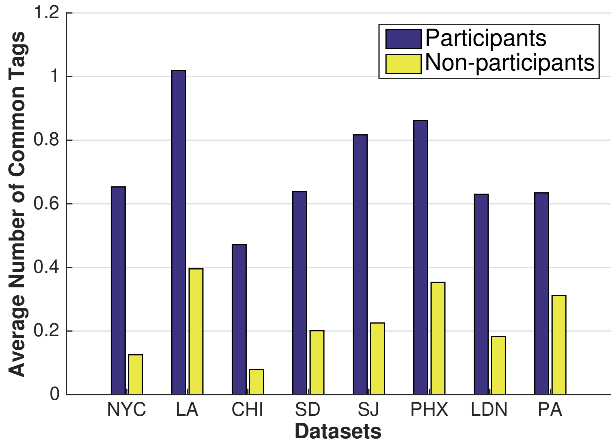


Figure 2: The average numbers of common tags to the organizer in eight datasets for all users whose profiles include tags.

of the organizers with more common interests. For example, people who like outdoor activities may join the events created by the organizers who also enjoy going outside. In contrast, if the organizer has no any common interest, it will be so absurd for users to participate in the event created by that organizer.

For users with tags in their profiles, Figure 2 shows the average number of common tags to the organizer in each dataset. It is so obvious that participants generally have more common tags to the organizer than non-participants in every dataset. In the NYC dataset, the participants averagely have more than five times as the number of common tags to the event organizer than the non-participants. The results also demonstrate that the tag similarity between users and the organizer may be very important to finding potential participants. If the tag information of users can be well incorporated, the system will be able to understand more semantic knowledge.

3.3 Geographical Information

The geographical limitation may be an important deciding whether a user participates in the event. If the event location is far from users' homes, they may be unwilling to pay the enormous effort of transportation for attending the event. On the contrary, people may be more willing to participate in the events organized in closer places so that they can save the transportation costs in both of time and money.

Table 1 shows the median distance from users' homes to event locations in each dataset. We adopt the great-circle distance (Moritz 1980) to calculate the distance on the geographic coordinate system. To avoid the bias from extremely long or short distance, we compute the median distance instead of the average distance. Except the datasets of two European cities, participants are closer to the event locations than non-participants in almost of all datasets. The two European datasets contradict the others because users in the two datasets registered their locations so close to each other. The results show that people actually tend to join events held in closer places. Hence, the users who live in closer locations should have higher priority to be considered as participants. We think a method can more precisely learn users' preference if considering their location information.

Table 1: The median distances (miles) to the event locations for participants and non-participants in eight datasets.

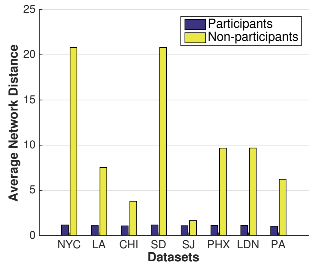
Dataset	NYC	LA	CHI	SD
Participants	**10.8973**	**14.9825**	**6.2850**	**19.7326**
Non-participants	11.0904	19.8551	7.6397	20.5993
Dataset	SJ	PHX	LDN	PA
Participants	**29.9376**	**20.7983**	40.2062	36.5179
Non-participants	30.4025	22.8540	**40.2024**	36.5179

3.4 Social Network Information

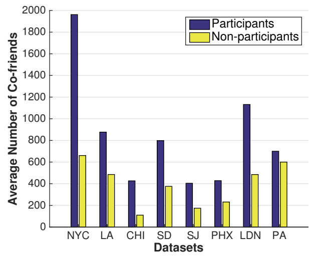
The online social groups in Meetup.com can well form an online social network. The interaction among users in that social network may affect users' participation. Users may be more likely to participate in the events organized by users in the same online social groups. On the contrary, people may not like to join an offline event organized by an unknown person.

To construct the online social network, we first treat each user as a node in a social network. For each pair of users, an edge between two users will be created if they are the members of the same online social group. A direct way to measure the relationship between two nodes in a social network is the distance on the graph. If a user has closer relationship to the organizer, the network distance between them may be also shorter. Figure 3(a) shows the average network distance between users and the event organizers in each dataset. Note that we define the network distance between two nodes as the number of nodes in the shortest path in the network. The results show that the network distances of participants are significantly shorter than the distances of non-participants. We also measure the relationship between users by the number of co-friends on the social network. Here we define the friends are the users who are the members in the same online social group. Figure 3(b) represents the average number of co-friends between users and the event organizers in each dataset. The participants generally have more co-friends on the online social networks of all datasets. In sum, both of the measures show that the online social network may be helpful to discover the users who will join the offline events.

In addition to online social networks, the event participation history of users may also form an offline social network. It is intuitive that users would like to participate in the events held by people who had ever attended the same events. To conduct the data analysis, we use 50% events in each dataset to construct the offline social network, and present some statistics with the remaining events. For each pair of users, an edge is created if they had ever participated in the same offline events. Figure 4(a) and Figure 4(b) represent the average network distance and the number of co-friends between the event organizer and users in the offline social network. The results are consistent with the analysis of online social networks. Participants of events have also

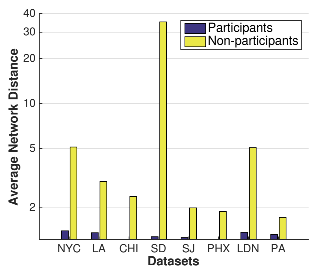


(a) Network Distance

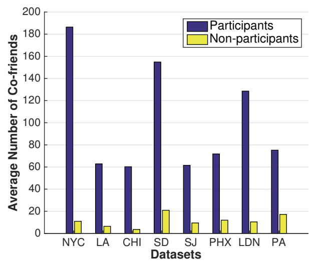


(b) Number of co-Friends

Figure 3: The average network distances and numbers of co-friends between the event organizer and users in the online social networks (groups) of eight datasets.



(a) Network Distance



(b) Number of co-Friends

Figure 4: The average network distances and numbers of co-friends between the event organizer and users in the offline social networks (events) of eight datasets. Note that here we build the offline social network with 50% events, and do statistics with the remaining events.

shorter distances and more co-friends to the organizers than non-participants. Moreover, the difference on the offline social networks between participants and non-participants is much more significant than the difference on the online social networks. Hence, offline social networks may be more effective than online social networks for determining the event participants.

To summarize the results of data analysis, all three kinds of information may be so helpful to find event participants. If we can incorporate these useful information into our approach, the performance of predicting event participants may be much boosted.

4 Conclusions

In this paper, we propose to analyze event participants for a single organizer. As the first study of this task, we analyze three factors relevant to users' participation, including semantic information, geographical information and social network information, with the event-based social network Meetup dataset. The results show that the participants actually have closer relationship to the event organizer in three factors. As the future work, we would like to apply these information to actually predict the participants for a single organizer.

5 Acknowledgments

This work was sponsored by Ministry of Science and Technology (MOST) of Taiwan under grant 104-2221-E-001-027-MY2. This work is also supported by Multidisciplinary Health Cloud Research Program: Technology Development and Application of Big Health Data, Academia Sinica, Taipei, Taiwan under grant MP10212-0318.

References

Du, R.; Yu, Z.; Mei, T.; Wang, Z.; Wang, Z.; and Guo, B. 2014. Predicting activity attendance in event-based social networks: Content, context and social influence. In *Proceedings of ACM UbiComp 2014*, 425–434.

Georgiev, P.; Noulas, A.; and Mascolo, C. 2014. The call of the crowd: Event participation in location-based social services. In *Proceedings of AAAI ICWSM 2014*, 341–350.

Hu, Y.; Farnham, S.; and Talamadupula, K. 2015. Predicting user engagement on twitter with real-world events. In *Proceedings of AAAI ICWSM 2015*, 168–177.

Liu, X.; He, Q.; Tian, Y.; Lee, W.-C.; McPherson, J.; and Han, J. 2012. Event-based social networks: linking the online and offline social worlds. In *Proceedings of ACM KDD 2012*, 1032–1040.

Macedo, A. Q.; Marinho, L. B.; and Santos, R. L. 2015. Context-aware event recommendation in event-based social networks. In *Proceedings of ACM Recommender Systems*, 123–130.

Moritz, H. 1980. Geodetic reference system 1980. *Journal of Geodesy* 54(3):395–405.

Pham, T.-A. N.; Li, X.; Cong, G.; and Zhang, Z. 2015. A general graph-based model for recommendation in event-based social networks. In *Proceedings of IEEE ICDE 2015*, 567–578.

Qiao, Z.; Zhang, P.; Zhou, C.; Cao, Y.; Guo, L.; and Zhang, Y. 2014. Event recommendation in event-based social networks. 567–578.

Tu, W.; Cheung, D. W.-L.; Mamoulis, N.; Yang, M.; and Lu, Z. 2015. Activity-partner recommendation. In *Proceedings of PAKDD 2015*, 591–604.

Xu, T.; Zhong, H.; Zhu, H.; Xiong, H.; Chen, E.; and Liu, G. 2015. Exploring the impact of dynamic mutual influence on social event participation. In *Proceedings of SIAM SDM*, 262–270.

Yu, Z.; Du, R.; Guo, B.; Xu, H.; Gu, T.; Wang, Z.; and Zhang, D. 2015. Who should i invite for my party?: Combining user preference and influence maximization for social events. In *Proceedings of ACM UbiComp 2015*, 879–883.

Zhang, W.; Wang, J.; and Feng, W. 2013. Combining latent factor model with location features for event-based group recommendation. In *Proceedings of ACM KDD 2013*, 910–918.

Zhang, X.; Zhao, J.; and Cao, G. 2015. Who will attend? - predicting event attendance in event-based social network. In *Proceedings of IEEE MDM*, 74–83.