How Online Social Network Affects Offline Events: A Case Study on Douban

Junwei Han[1,2], Alvin Chin[1], Jianwei Niu[2*], Wei Wang[1], Chao Tong[2], Xia Wang[1]

[1]Nokia Research Center
100176 Beijing, China

[2]Department of Computer Science and Technology
Beihang University 100191 Beijing, Chin

{ext-junwei.han, alvin.chin, wei.41.wang, xia.s.wang}@nokia.com,
{niujianwei, tongchao}@buaa.edu.cn

***Abstract*—Social networking sites (SNS) such as Facebook and Twitter are becoming popular forms for finding, promoting and attending offline events and activities. Much work has looked into characterizing these social networks and their user behavior. With the emergence of event and activity-based applications such as Facebook Events, LinkedIn Events, Xinghui, Zaizher, and Douban, it is easier to connect with other people online and meet them offline. However, few have looked into analyzing these offline events and activities that are shared online. This paper seeks to gain insights into the user behavior around people attending offline events which are promoted online. By studying the events in Douban, we present results around the event properties, user behavior of participants and wishers to an event, and social influence to an event. We show that event distribution by participants and wishers follow the typical power-law distribution; most users attend or like short events that last several days or regular events that last less than 3 months; participants attend an event within one day after the publish time; and that there is an exponential relationship between follow probability and number of common events attended between two users and a linear relationship for common events interested in. These findings provide a better understanding on how SNS could affect user behavior in attending events, and provide guidelines on how to improve the design of event-based applications.**

Keywords: ***user behavior, social network, offline, online, event***

I. INTRODUCTION

Recently, social networking sites (SNS) such as Facebook, Twitter, Google Plus and Flickr, have become mainstream web services and are changing the way we contact with each other. With the help of these sites, people can contact friends, share information such as photos and status, and track the latest and hottest news and so on. Researchers have paid a lot of attention to study topological properties [1, 2], community structure [3, 4, 5, 6], user behaviors [7, 8], evolution [9, 10], social influence [11], and information diffusion [2, 12, 13].

All this work has focused on online environments and address questions such as what are the topological properties of Twitter? How does information flow within an online social network? What is the user behavior of following other people? However, we live offline, and most of the time we communicate with others in real life like attending an event. Nowadays, a lot of SNSs are trying to guide their users to attend offline events and activities such as the events features on Facebook and LinkedIn, and specific applications such as Xinghui and Zaizher. Applications are now beginning to exploit the offline to online business model to link the online connections and transactions to offline consumption and activity, such as GroupOn [14] who mainly guides its users to buy discounted deals online that can be used at physical stores and restaurants. However, few research has investigated the relationship between offline and online and user behavior especially within the context of an event. In this paper, we focus on offline events which are published and shared in an online social network. In this paper, we aim to address the following research questions:

- What is the relationship between social influence and event participation?
- What is the user behavior of those that are interested but cannot attend an event (wishers) versus those who want to attend an event (participants) and their relationship?
- Do events affect user behavior?

To answer these questions, we use data from Douban which is a famous online social network in China for finding and attending offline events. We describe the properties of Douban events from the dataset that we obtain in terms of event distribution by category, duration, and user type (participant or wisher). Then, we explain the user behavior of participants and wishers to the events, and the social influence of followers on the user behaviors of those attending an offline event. These are fundamental issues as their answers would provide a better understanding of how to promote an event online, and allow social networking service providers to improve their systems to attract more users, and aid researchers to have an initial understanding of user behavior about attending events and construct an accurate model to characterize user behavior.

Our results show that three out of ten event categories dominate the Douban events (salon, music and party), and the event distribution by participants and wishers follow a power-law distribution while event duration deviates from a power-law. Most of the participants and wishers join a particular event within one day after an event is published and participants and wishers have a strong correlation in terms of numbers to the event. For social influence, the probability to follow a person is correlated with the number of common events both have attended or are interested in, with an exponential distribution for the former and linear distribution for the latter.

Our major contributions in this paper are as follows. First, we characterize the properties of offline events that are

*Corresponding author: niujianwei@buaa.edu.cn

published online using Douban and the relationship between offline events and online connections, of which few have been done. Second, we discuss the implications of our work for improving social networking for offline events and for integrating offline and online in a mobile application.

The outline of this paper is as follows. Related works are discussed in Section II and we give a detailed description of Douban and the associated dataset in Section III. Then we analyze the event properties of Douban in Section IV. In Section V and Section VI we present our analysis of event evolution, relationship between participants and wishers as well as social influence respectively. We discuss the implications of our work in Section VII and Section VIII concludes the paper and discusses areas for future work.

II. Related Work

There is much work on studying online activities in online social networks [2, 15], however few have studied offline activities and events. Now, researchers are beginning to study the relationship between offline activities and online activities and interactions. For example, Cranshaw et al [16] explored connections between an online social network and the location traces of its users to infer online social relationship based on offline encounters and Noulas et al [17] analyzed geo-temporal rhythms of user check-ins and user similarity based on places they visit, in order to make user recommendations. Wang et al [18] analyzed the effect of offline encounters on online friendship, and discovered that offline encounters do have an effect on adding a person as a friend. However, these studies only disclose user behavior from offline to online direction. In this paper, we use data from an events network extracted from Douban to study the relationship from online to offline.

A lot of phenomena in social networks show that they exhibit a power-law distribution, for example, follower and friendship degree distributions [2, 8]. However, previous studies tend to find the entire network's statistical properties, for example, Zhao et al [8] provide a detailed analysis about Douban's follower network including its degree distribution, structural network, and information flow. Others study the relationship of users' properties, such as Huberman et al [11] where they provide a deep and systematic description of the relationship between the number of followers and number of posts, number of friends and number of posts, number of followers and number of friends. All of these researches are only concerned about complete online properties. A lot of researchers study user behavior in online social networks, such as Benevenut et al [12] where they presented a thorough characterization of social network workloads, based on detailed clickstream data of the number of friends that users interacted with, and Java et al [13] where they detailed why users tweet and identified different types of user intentions.

However, the above works are concerned about the complete online social network and user behavior around online activities, but they ignore studying the offline events and activities that form the basis for the online interactions. In this paper, we address this gap by focusing on offline events which are promoted online and study its characteristics and user behavior arising from these offline events.

III. Douban and Dataset Description

A. Douban Description

Douban is a Chinese Web 2.0 web site providing user reviews and recommendation services for books, movies, music and events. Douban launched on March 6, 2005. It is the largest online database for Chinese language books, movies and music. The users of Douban are mainly white collar workers and students [8]. People use Douban to find books, movies, music and online and offline events based on recommendations of other users and from Douban's recommendation system.

Figure 1 shows a screenshot of a user's homepage in Douban and its various features which are numbered from 1 to 8 below:

- (1) What offline events she attended or will likely attend.
- (2) What online events she attended or will likely attend.
- (3) What music she listened or wants to listen.
- (4) What movies she watched or wants to watch.
- (5) What books she read or wants to read.
- (6) Who she follows.
- (7) Who follows her.
- (8) What interest groups she has joined.

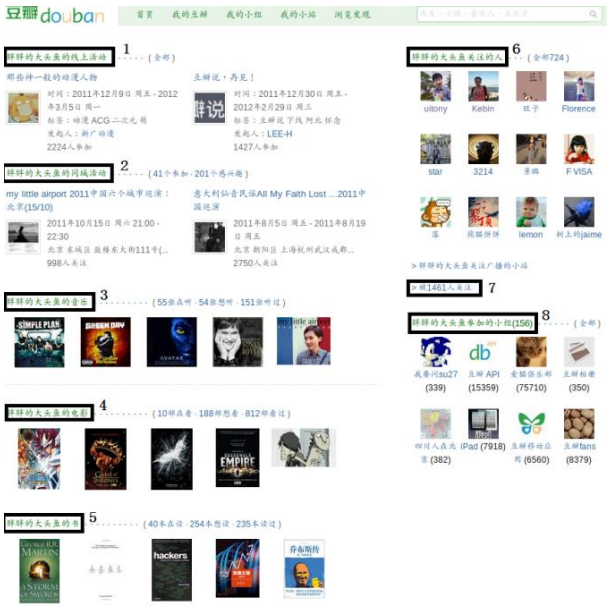

Figure 1. Screenshot of a Douban user's homepage

In this paper, we focus on only the offline events (1) and the people that the user follows (6) or the people that follow the user (7) from Figure 1. The reason why is because our objective is to study the online social connections and its impact on offline events and vice versa.

Douban Event is a platform in Douban for event organizers to promote their event by posting the related event

description on the Douban Event page. When a user selects the event page from the event list in (1) of Figure 1, she can select either the "attend" button or "like" button. If she selects the "attend" button, this indicates that she is likely to attend the event and we call her a *participant*, or if she selects the "like" button, then this indicates that she is interested in the event but cannot attend so we call her a *wisher*. Since we do not know whether or not a participant really attends an event because we do not have a user's location data, we only make an assumption that participants will attend while wishers will not. For an event, users can browse the event's participants and wishers from the event page which is public. The social relationship in Douban is similar to that in Twitter which can be viewed as a directed graph.

B. Dataset Description

We used the Douban API to crawl all valid events for each day within the specified time interval for the crawl, and obtain the event's participants and wishers as well as a particular user's followers. We define a valid event as an event whose published time is before or on the specific day of the instance of the crawl. Detailed statistics of the data that we collected are shown in Table 1.

Table 1. Basic Information Of the Douban Dataset

Time Interval	2012-03-21 ~ 2012-04-20
Number of Events	4467
Number of Participants	61654
Number of Wishers	101985
Min number of followers	1
Max number of followers	1144
Avg number of followers	16.9
Standard deviation for followers	35.04

From our dataset, we discover that more than a hundred new events are posted each day. The number of wishers is almost 2 times the number of participants, which means users like marking events that they are interested in more so than attending events. We also observe that the number of users that a particular user follows ranges from a few to more than a thousand. Since Douban has more than 50 million users [19] we can see that only a very small number of users (almost 0.1%) have expressed their interest in attending at least one event in Douban.

IV. Event Properties Analysis

In this section, we seek to understand the offline events and user behavior around events in Douban by describing its properties based from the Douban dataset introduced in the previous section. In particular, we are interested in the event distribution across different attributes in order to understand the user behavior around these events. We examine all the events within the time period of our dataset according to category, district, participants and wishers, and duration. In addition, we also examine event duration by category and the participants and wishers by category.

A. Event distribution by event category

All events in Douban are categorized into music, film, salon, sport, commonwealth, party, travel, exhibition, drama and others. However, it is natural that some event categories will be more popular among users than others. What category has the most events and why? We make a statistical analysis with our data and have the following conclusions. The top 3 categories are salon, music, and party which cover 28.2%, 17.9%, 14.6% of all events respectively. This result is reasonable since white collar workers and students like learning something from a salon and also listening to music such as their favorite singer's concert and attending parties either for contacting with old friends or just making new friends. We can also see that salon is significantly the most popular out of all the top 3 categories, indicating that users mostly publish salon events in Douban.

B. Event distribution by district

Events in Douban are also localized according to district. What is the distribution of events by district? Beijing has 15 districts: Dongcheng, Xicheng, Chaoyang, Haidian, Shunyi, Fangshan, Xuanwu, Chongwen, Fengtai, Mentougou, Huairou, Tongzhou, Pinggu, Shijingshan and Changping district. We discover that the top 3 districts that have the most events are Chaoyang, Haidian and Dongcheng districts which cover roughly 90% of all Beijing offline events on Douban. We can also explain this phenomenon based on the properties of each location. Chaoyang district has the most concert halls, gyms, and a lot of companies, and many white collar workers are in this district, therefore salon and music events are held here. However, Haidian district is where the universities are located and there are many entertainment places, high tech companies and research labs. With a large number of students and high-tech workers, this district is a better place to hold salon or party events. Dongcheng district on the other hand has a lot of markets and shopping malls, so this location is very suited for party events.

C. Event distribution by participants and wishers

For every event, Douban users can indicate their interest in the event (as a wisher) or attending the event (as a participant). How many participants and wishers do Douban events usually cover? We plot the distribution of events by number of participants and number of wishers as shown in Figure 2. The distribution of both figures resemble a power-law or long-tail distribution which demonstrates that most of the events are either attended by or are considered interesting by a small number of participants, while few events cover a large number of participants. These results are as expected since they follow typical social behavior which is similar to other social phenomena such as follower and friend distributions in social networks [2, 8].

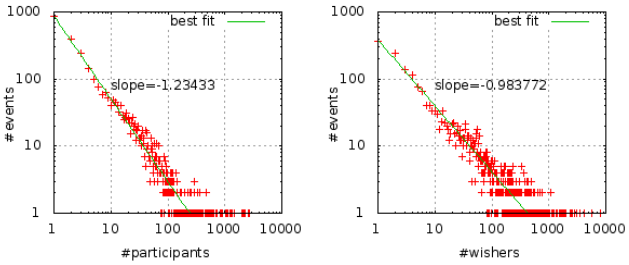

Figure 2. Event distribution by participant (left) and by wisher (right)

D. *Event distribution by duration*

Another phenomenon is how long events will typically be? When publishing an event, users can specify the event start and end time, as well as the frequency. In our Douban dataset, each event has a duration that ranges from one day to one year. Regular repeating events tend to last for a long time, for example, one event is about running every morning in a park whose duration is from 8th April to 1st July. For repeating events, it would be tedious to create the same event for every frequency. Single events, on the other hand, last for a short time. Our dataset shows that event distribution by duration does not quite satisfy a power-law distribution. More than 50% of the events last for two days, while 37% of the events last from 10 to 100 days. The average event duration is 25.3 days, and the standard deviation is 44.9 which demonstrate that the data is very scattered. At first, the results may seem surprising that it does not follow a power-law. However, if we take into account the nature of the events being single and regular ones, this becomes not surprising and is expected.

E. *Event duration by category*

Another curious question to ask is whether event category influences the event duration. As explained from our previous results, regular events tend to last longer than other events, but we also want to know which category that events tend to be regular events. From our data we find that party, sports, commonwealth and exhibition events tend to be regular (with a high duration), while film, music and salon events are shorter. This makes sense because party, sports, commonwealth and exhibition events are often held regularly. For example, a reading party may often be held once a week usually on a Sunday (when people are off work) or a photography exhibition which may last more than a month. On the other hand, film or music events are not regular events because they usually last for a single time. For example, a music event like a concert may usually last just for one to three days, and a film such as a world animation film may usually last for a week.

F. *Event participants and wishers by category*

Which categories of events are attended more and are considered interesting by Douban users? We examine the participants and wishers distribution based on event category by calculating the average number of participants and wishers for each event category. We find that more people are interested in the events, compared to wanting to attend the events, which is not unexpected. Film events have the largest on average number of attended participants, while sports events are the least attended. Exhibition and music events are also relatively popular attended events, while party, travel, salon and drama events are not as popularly attended. The same conclusion can be deduced from the wishers of each category, except that exhibition events are more interesting by wishers.

V. Event Participants and Wishers

The previous section analyzed the event properties on Douban and explained reasons why they showed these properties. In this section, we examine the user behavior of participants and wishers to the Douban events.

A. *Participants and wishers distribution by event duration*

We look at whether event duration affects users to attend an event or be interested in the event. As a result, we plot both the average number of participants and wishers distribution by duration in Figure 3.

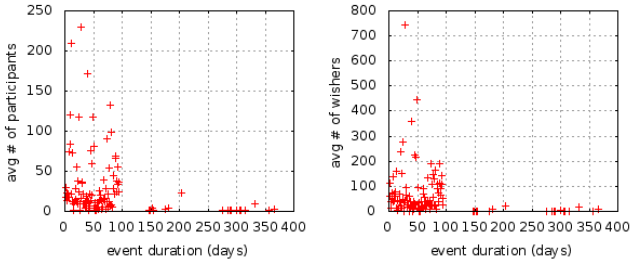

Figure 3. Average number of participants (left) / wishers (right) distribution by event duration

From Figure 3, we can see there are almost 3 peaks, a week, a month, and a quarter of a year. Almost 16% of the events have a duration of around a week, while 13% and 14% of the events for a month and a quarter respectively. The right graph of Figure 3 shows more clearly that almost 33% of all total wishers are interested in events that have a duration of around a month. These results demonstrate that people seem to like short events and attend short events more than longer ones, and regular events less than a quarter are more attractive than half or a whole year.

B. *Time effect on participating and wishing to participate in events*

After an event is published on Douban, when do users indicate their interest in the event or to attend the event? We plot the average number of participants and wishers increment (i.e. the change from the previous day) according to the number of days after the event is published in Figure 4. Almost most users indicated their interest in the event or their desire to attend the event on the day that the event is published. At first, this may seem surprising and not intuitive because one would expect that the highest number of participants/wishers would occur towards the beginning of the event when the event becomes widely publicized, rather

than after one day that the event is published, because users would browse or search for the events. However, we can explain this result by the way that Douban notifies people about events. When a new event is published, Douban sends a user an invitation through both email and its own network based on that user's profile. A user profile is built based on the particular events that a user likes and attends, and therefore Douban can know the type of events that would be suited for that user. When a user receives an invitation like an email from Douban, she will go to the website to see the details of that event and then choose whether she will attend or like the event or just ignore this invitation. Most of the users will receive email on the same day as the published event, so we can observe that almost all participants and wishers are involved on the first day. This also suggests that users normally do not browse and select what events that they are interested themselves, but rather rely on Douban's recommendation system which is Douban's expertise.

C. *Relationship between participants and wishers in events*

What is the relationship between participants and wishers in an event? We hypothesize that if an event has a large number of wishers, then it will also have a large number of participants which means they are positively correlated. Figure 4 shows the fitting result. From the scatter plot, most of the points are located around the line $y = 2.39227x$. The Pearson Correlation Coefficient is 0.6367 which also implies a strong correlation between number of participants and number of wishers to an event. It is intuitive that an event's popularity is not only reflected by its number of participants but also by its number of wishers.

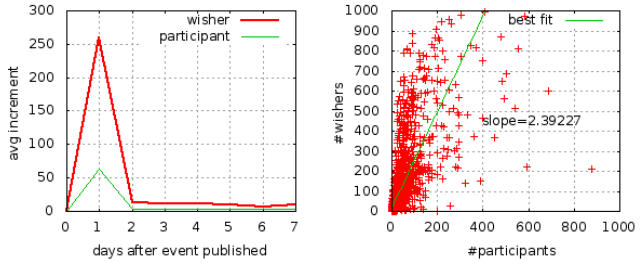

Figure 4. Average increment of participants and wishers by the day after the event is published (left) ant its participant and wisher relationship

VI. SOCIAL INFLUENCE

In this section, we examine the influence of social connections on events. Social influence is important in making social connections. Previous research shows that users generally connect with others who are similar to them through homophily [20]. We use followship in Douban for social connections. The objective is to determine if events affect followship. However, in order to do that, we first perform a user distribution around events, that is, how many events does a user attend? The user distribution by number of events satisfies a power-law distribution which means many users attend a few events, but only a small fraction of users attend a lot of events (more than ten). Therefore, we can conclude that users on Douban are not active in attending events. Part of the reason for that may be due to Douban's characteristics that its users are connected by common interests instead of by events and acquaintances.

Due to social influence, users choose to attend an event partly because their friends will attend this event. In Douban, users are aware of their friends' activities through their updated statuses or broadcast messages. As a result, we want to know whether a user who follows another will attend a lot of common events, or if two users have attended a lot of common events, what is the probability that they have a following relationship between each other. Figure 5 shows the follow probability between users who have attended common events and who have both liked common events, respectively.

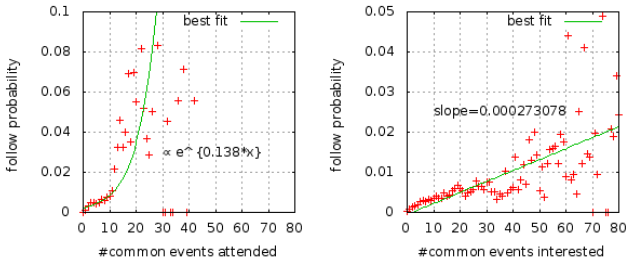

Figure 5. Follow probability based on common events attended (left) and common events interested in (right)

From the above figures, we can see that the follow probability based on common events attended and common events interested in, seem to resemble an exponential relationship and a linear relationship, respectively. We can see that the more common events two users both attend the larger probability that they have a following relationship, which seems intuitive because it is likely that the two users have similar interests that cause them to attend the same event. The limitation here is that we do not know whether two users form a following relationship before, during or after an event because we do not know the reasons why a user will follow another user.

VII. DISCUSSION

In this paper, we present a systematic and in-depth analysis on Douban's offline events, and disclose users' behavior in attending offline events. However, the results that we have presented are speculative because Douban does not record the reasons as to why a user expresses her desire to attend an event, nor does Douban record if she actually attended the event.

All these limitations arise from the lack of collecting the user's positioning data which makes it hard to study a user's offline behavior in an event. Fortunately, researchers have already noticed this problem and did some study about online and offline interactions such as [18], where they developed an application to record a user's position indoors in a conference and the location is used to spur social connections online such as follow or friendship, as well as studying the relationship between offline encounters and

online follow. In order to study the user behavior of online and offline, we believe that an application needs to be built centered around events or activities that integrate offline and online activities, through the recording of user's position and other activity context such as accelerometer and microphone. In doing so, a complete and accurate study of user behavior within offline and online can be conducted, which can aid in improving recommendations of events and people to follow or add as contact. To effectively promote an offline event online, the service provider should make full use of users' social relationship around the web and make it easy to share relative information to their friends anywhere in the web.

VIII. CONCLUSION

In this paper, we study offline events based on Douban. We analyze the event properties and reveal some interesting results, including that three event categories (salon, music, and party) dominate Douban's offline events, and event participants and wishers both satisfy power-law distribution while event duration does not. We also find that events which are public and need less interaction between participants are more popular than events which are private and have a lot of interaction or communication. Users like events with short duration (around several days), and events with duration of a week, a month and a quarter of a year tend to attract more users. Most of the users like or attend the event on the first day of its publishing on Douban which is due to Douban's notification of recommended events. We also study the social influence and show that the follow probability between two users based on the number of common events they attended is exponential, while the follow probability based on the number of common events they are interested in is linear.

There are also some limitations to our work. When we studied the follow probability based on the number of common events they attended or are interested in, we did not consider when (before, during or after an event) one follows the other user. In addition, we did not provide a detailed analysis of the follower network and its evolution such as in [8] within the context of events. Our future work will consider performing a longitudinal study of Douban Events for a longer period of time and studying different category of events to determine if category affects user behavior. For social influence, we also plan to study the friends in addition to followers. These findings provide a better understanding on how SNS could affect user behavior in events, and provide guidelines on how to improve the design of event-based applications.

ACKNOWLEDGMENTS

This work was supported by the Research Fund of the State Key Laboratory of Software Development Environment under Grant No. BUAA SKLSDE-2010ZX-13, the National Natural Science Foundation of China under Grant No. 60873241 & 61170296, the Fund of Aeronautics Science granted No. 20091951020, the Program for New Century Excellent Talents in University under Grant No. 291184, Natural Science Foundation of Beijing, China under Grant No. 4123101.

REFERENCES

[1] Y. Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," Proc. of 16th ACM WWW, 2007, pp. 835–844.

[2] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" Proc. of 19th ACM WWW, 2010, pp. 591–600.

[3] A. Clauset, M. Newman, and C. Moore, "Finding community structure in very large networks," Physical Review E, vol. 70, pp. 066111, 2004.

[4] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney, "Statistical properties of community structure in large social and information networks", Proc. of 17th ACM WWW, 2008, pp. 695–704.

[5] M. Newman, "Modularity and community structure in networks," Proceedings of the National Academy of Sciences, vol. 103, pp. 8577–8582 , 2006.

[6] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," Physical Review E, vol. 69, pp. 026113, 2004.

[7] Y. Jiang, Scott Counts, "Predicting the Speed, Scale, and Range of Information Diffusion in Twitter", in AAAI 2005.

[8] J. Zhao, J. Lui, Don Towsley, X. Guan, Y. Zhou, "Empirical Analysis of the Evolution of Follower Network: A Case Study on Douban," IEEE INFOCOM WKSHPS, 2011, pp. 924–929.

[9] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," Proc of the 7th ACM SIGKDD, 2005, pp. 177–187..

[10] J. Leskvec, L.Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks," Proc of the 14th ACM SIGKDD, 2008, pp. 462–470.

[11] B. A. Huberman, D. M. Romero, and F. Wu, "Social networks that matter: Twitter under the microscope," arXiv:0812.1045v1, Dec 2008.

[12] F. Benevenut, T. Rodrigues, M. Cha, and V. Almeida, "Charactering user behavior in online social networks," Proc, of the 9th ACM SIGCOMM Internet Measurement Conference, 2009, pp. 49–62.

[13] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," Proc. of the 9th WebKDD and 1st SNA-KDD workshop on Web mining and social network analysis, 2007, pp. 56–65.

[14] GroupOn. http://www.groupon.com

[15] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Growth of the flickr social network," WOSN'2008, 2008, pp. 25-30.

[16] J. Cranshaw, E. Toch, J. Hong, A. Kittur and N. Sadeh, "Bridging the gap between physical location and online social networks," Proceedings of Ubicomp'10, 2010, pp. 119–128.

[17] A. Noulas, S. Scellato, C. Mascolo and M. Pontil, "An empirical study of geographic user activity patterns in foursquare," ICWSM'11, 2011.

[18] L. Chang, A. Chin, H. Wang, L. Zhu, K. Zhang, F. Yin, H. Wang, L. Zhang, "Enhancing the Experience and Efficiency at a Conference with Mobile Social Networking: Case Study with Find & Connect," Proceedings of the International Conference on Human-centric Computing 2011 and Embedded and Multimedia Computing 2011, pp. 281.

[19] Wikipedia. http://en.wikipedia.org/wiki/Douban

[20] M. McPherson, L. Smith, J. Cook, "Birds of a feather: homophily in social networks," Ann Rev Sociol 2001, pp. 415–444.