Demographics of Social Network Users - A Case Study on Plurk

Chun-Chao YEH, Yan-Shao LIN, Ting-Hsiang LIN
Department of Computer Science, National Taiwan Ocean University, Taiwan
ccyeh@mail.ntou.edu.tw, Fax: +886 2 24623249, Tel: +886 2 24622192 ext. 6604

Abstract— **In this research study, we present an empirical study on Plurk social network (http://www.plurk.com). Plurk's unique features, such as public available user information, provide us a good opportunity to explored different user behaviors corresponding to different Plurk user groups based on their genders, ages, and nationalities. We crawled Plurk website and collected over hundred thousands of unique Plurk user pages in early 2011. These up-to-date sampled user data enable us to observe possible user behaviour on a social network. Basing on the user pages we collected, we characterize behaviours of Plurk user groups. In this paper we share what we found from this empirical case study. As far as we know, this study is first Plurk user study based on large Plurk user data randomly sampled from Plurk web site. While this study is to some extent primitive, we hope that the results of this study provide some insights and/or some references for comparison to researchers working on the topics about Internet social network user behaviours and related issues at different aspects.**

Keywords— **Social Network, Plurk, Demographics**

I. INTRODUCTION

Remarkable Advances of communication and network technologies enable Internet users to access varieties of Internet services with higher bandwidth then ever, anytime and anywhere. Such a pervasive internet access infrastructure tightly connects not only internet devices but also internet users all over the world. Meanwhile, the concept of Web 2.0 advocates new paradigms of Internet web services, in which Internet users not only being an Internet content consumer but Internet content providers as well. Today, Internet has become a perfect platform for Internet users to share/express their personal stuff, visible (such as pictures, files, works) or invisible (such as knowledge, feelings, thoughts, opinions, responses), to a huge number of Internet users instantly, anytime anywhere. Such a new Internet service paradigm provides a new mechanism of social communication and interaction, which is greatly different from the way we have before in human history, in terms of scale, distance, and time. A great example is social network services.

Most of social network services provide users with a personal web space, in which users can post their messages of any types (e.g. personal opinions, comments, and feelings) in any forms (e.g. text, pictures, or files). The messages could be a well organized article shared to the public or just gossip between friends. Social networks, such as Facebook[1], Twitter[2], and Plurk[3], get great popularity in recent years. By July 2011, there are 750 million active users hook on to Facebook, according to an official report [4]. The number is over 800 million few months later (October 2011), and more than 50% of Facebook users log on to Facebook in any given day [5].

In this research study, we present an empirical study on Plurk social network (**http://www.plurk.com**). Three main reasons drive us to choose Plurk for the case study. First, only few works have been done for Plurk user behaviour study, comparing with the works on top social networks such as Facebook or Twitter. From internet search we found a similar work conducted by Premraj Narkhede *etc.* [6] , which analyses Plurk user behaviour based on about forty thousands (44,457) user profiles collected as of May 2010. Since the collection is based on friends links initial from a chosen seed profile. The data set is bias to a specific group of users (e.g. bias to the users using same languages, or living in same regions/countries). The results from our study and other reference sources justify our concerns. In contrast, we took a random sample approach to collect the Plurk user profiles. Near one hundred thousands (98,766) of user profiles have been collected. Second, the population size of Plurk users is moderate to conduct random sampling for a feasible size of random sample. Our intention is to capture user behaviors of a social network without too much bias, which we think is important to drive any conclusions for the population from the sample date. Consequently, we need enough random samples comparing to the population size. Third, Plurk's unique features, such as public available user information, provide us a good opportunity to explored different user behaviours corresponding to different Plurk user groups based on their genders, ages, and nationalities.

We crawled Plurk website and collected over hundred thousands of unique Plurk user pages in early 2011. These up-to-date sampled user data enable us to observe possible user behaviour on a social network. Basing on the user pages we collected, we characterize behaviours of Plurk user groups. Results showed that the gender distribution between female and male of the sampled Plurk users is around 59% female) vs. 41% (male). Most of Plurk users are at the ages between 16 to 30 years old (70%), of which the user group with ages 19 to 23 (roughly at same age of undergraduate college students) contributes about 35% of the total population of the sample data. Meanwhile we found the majority of the sampled users are from Asia countries (contribute more that 80%), of which

users identifying themselves as from Taiwan is dominated. Approximately 46% of the sampled users claims coming from Taiwan.

The rest of the paper is organized as follows. Section 2 provides a brief introduction of Plurk social network and related works. Section 3 presents Plurk user data collection and the data features we retrieved for analysis. Section 4 presents the results of the case study. Finally, remarks in Section 5 conclude the paper.

II. BACKGROUND AND RELATED WORKS

In this section, we present some background knowledge about Plurk social network and the related works on the topics.

A. Plurk Social Network

Plurk, launched on May 12, 2008, is a free social networking [7]. Plurk provides micro-blog services. Plurk users can post/update/response with short messages or links up to 140 words, similar to twitter. Plurk provides a timeline presentation format to updates user messages received, either posted by the users or by others to respond to the message, in chronological order. The unique timeline presentation format provides a clear perception to viewers about the timing of the events the message associated with, and provides a quick summary of possible activities/status of the Plurk webpage owner during a fixed time window, such as in recent minutes, hours, days, or weeks.

Most of Plurk users are from Asia, especially Taiwan and South-East Asia, according to our observation. The Plurk traffic data observed by Alexa support our observation. Table 1 shows the Internet traffic statistics summary for Plurk.com, observed by Alexa [8], queried on Oct. 15, 2011. The statistics data count those internet traffics seen by Alexa (usually based on tracking information of users of its Alexa Toolbar), and thus exclude those traffics not seen. Nonetheless, the Alexa data provides a good reference of what the real traffics could be.

TABLE 1. ALEXA WEB SITE DEMOGRAPHICS OF PLURK.COM (DATA QUERIED ON OCT. 15, 2011)[7]

	Country	Percent of Visitors	Regional traffic rank
1	Taiwan	40.20%	33
2	India	14.20%	718
3	United States	9.20%	3,974
4	Philippines	6.90%	126
5	Indonesia	6.00%	388
6	Pakistan	2.80%	554
7	Russia	1.90%	5,104
8	Hong Kong	1.90%	398
9	Canada	1.60%	2,409
10	Brazil	1.50%	3,729

B. Related Works

Social networks have received great popularity in recent years. The hype of social networks also attracts researchers working in different disciplines to investigate various issues and consequences related to the subject. Among many others, Twitter is one of hot targets for case studies. Research issues are such as news recommendation [14], detection of real-time events [15], dynamics of information spread [16], and exploring user knowledge profile for e-learning [17]. Besides, similar to our works, Mislove et at.[13] presented demographics of Twitter users, based on over 54 millions of Twitter users collected between March 2006 and August 2009. Due to the limited user information (self-reported) available in Twitte, only three features (geographic location, gender, and race/ethnicity (referred by name)) are profiled.

In contrast, few works have been done for Plurk user behaviour study. From internet search we found only one work similar to what we done in this paper. Premraj Narkhede and his colleagues posted their work on Plurk user behaviour study in 2010[6]. They analysed Plurk user behaviour based on about forty thousands (44,457) user profiles collected as of May 2010. They collected the data starting with one seed profile and gathered information for all friends of seed profile followed by all friends of friends of seed profile and so on. Based on the data they analysed the collect data from different aspects such as nationality, gender, age, and Karma score (the indicator defined by Plurk to represent the degree of activity a user engagement). The work is similar to the context we present in this paper.

However, the data collection method, the scale of sample data and the timing of data collection are different. We collect more than twice of user profiles than the authors in [6] did. Our data collect is as from early 2011, server months later than the authors did. And, thus our data collection is fresher. Besides, the most critical point is the collection scheme. Our scheme is truly a random sampling. In contrast, the authors' scheme is bias to a specific group of users (e.g. bias to the users using same languages, or living in same regions/countries), since the sampling is not truly random but bias to the friendship relation to those user profiles being collected earlier. For example, their results showed that users with nationality of Philippines is dominated (16,346/44,457, 36.77%) in their collecting data, followed by those coming from Taiwan (14410/44457, 32.41%), and then by United States (2,378/44,457, 5.35%). This is much different from what we found. Our data show 46.4% (45,837/98,766) of the collected users claim coming from Taiwan, followed by Indonesia (17.4%) and then by Philippines (12.0%). According to the Alexa internet traffic survey (Table 1), our data are seemingly more consistent with the Alexa data.

III. DATA COLLECTION AND DATA FEATURES

In this section, we depict the sample data and the features we used for analysis.

A. Data Collection

Plurk assigns each user with an unique id. Basing on the fact, we randomly selected a set of ids, and tried to get the user profile if the guess id belongs to a valid Plurk user. By

the random sampling method mentioned above, we successfully collected 98,767 unique user profiles as from March 2011.

We retrieved following user profile data from each Plurk user we collected.

- User Personal data: including Gender, Nationality, and Date of Birth
- System maintained data: including registration date (member_since), friend counts, fan counts, Plurk counts, profile views, Plurk response counts, and Karma value.

Like most of social networks, the user personal data are maintained by user themselves. No validation procedure is imposed to verify the correctness of the user personal data. While possible cheating would be happened, we believe most of the data could be trusted or at least "reasonable". We have randomly checked about tens of user profiles manually by comparing their personal data (age, gender, and nationality) with the context of messages posted by the user or by other Plurk users. Nonetheless, like the cases in most of online social networks, since there is no ground truth for us to check with the reliability of the personal data provided by individual users, we just put aside the argument of data reliability to treat them as possible noise to the true data. On the other hand, system maintain data are maintained by Plurk servers automatically, and thus with more confidence to trust with.

B. User Profiles

Explanations for each data features belonging to the system maintain data are presented as follows:

- Member since: the date the user first registers himself/herself to the Plurk social network.
- Friend counts: the number of Plurk users adding the user as their friends. Plurk users can make a request to others as a friend. Once the requested user grands the friend request, the two become friends. The functionality of "friend" is similar to same term in Facebook social networks.
- Fan counts: the number of Plurk users adding the user to become the user's fans. If user X is user Y's friend or fan, Plurk servers will forward Y's plurks to X's Plurk timeline once Y post his message on his timeline. The functionality of "fan" is similar to "follower" in Twitter social networks.
- Plurk counts: the number of messages the user post to his timeline so far.
- Profile views: the number of times the user's Plurk webpage being browsed.
- Plurk response counts: the number of times the user post messages to response other user posts.
- Karma value: the indicator defined by Plurk to represent the degree of Plurk activities the user engaged. Karma value is floating over time. If the user engages in positive activities (e.g. post messages, response to messages posted by others, or invite new friends), Karma value will increase; on the other hand, if the user engages in negative activities (e.g. post spam messages, reject by others, or not posting message for a long time), Karma value will decrease. Users receiving high Karma value gain privileges to use advance features of system functionalities.

IV. RESULTS

In this section, we present results of the case study.

A. Demograhics of the Sample Data

First we give an overview of the sample data we collected, including gender distribution, nationality distribution, age structure, and member-since distribution. We compare the results with Premraj Narkhede's study [6] mentioned in Section II.B, and the traffic statistics provided by Alexa [8], whenever possible. We should note readers again that Premraj Narkhede's study is bias to some specific user groups, due to their data sampling scheme.

1) *Gender Distribution:* Table 2 shows the gender distribution of the sample data. Results show females are over-represented at *plurk.com,* with a ratio 58.8% versus 41.2%. The gender distribution is similar to the result of Premraj's study [6], in which the authors pointed out the gender distribution in their collection data is 56.86% versus 43.14%, based on 44,457 user profiles. Meanwhile, our results are consistent with Alexa's observation [8] as well, which indicates females are over-represented at plurk.com, relative to the general internet population.

TABLE 2. GENDER DISTRIBUTION OF THE SAMPLE DATA

Sample	users	ratio
Female	58,107	58.8%
Male	40,659	41.2%
Total	98,766	100%

2) *Nationality Distribution:* Table 3 shows the nationality distribution of the sample data. Our data show that users with nationality as Taiwan dominate the population (46.40%), followed by Indonesia (17.40%) and Philippines (12.00%). Users from Asia represent over 70% of the population, and the population of top 10 nationalities counts as high as 83.9% of the population. Our results are different from the survey reported in [6], which shows users with nationality of Philippines is dominated (36.77%) and then followed by Taiwan (32.41%) and United States (5.35%). Comparing the results with Alexa Internet Traffic survey (Table 1), our data are seemingly more consistent with the Alexa data. Possible differences have been discussed in Section II.B.

3) *Age Structure*: Table 4 shows the age structure of the sample data. The age group is partition according to school age for ages below 30, and then with a step of 10 for each age group. It is worthy of noting that we only have the age information, but no information to indicate where the user is a student or not. The results show Plurk users are dominated by young people with ages between 16 to 30 (72%), which are

around at the ages of senior high school or college (including graduate school) students or at the ages just few years after they leave schools. Besides, we found a non-negligible portion (0.6%) of users claimed to be with ages over 80 in the sample data. We've checked with some of the users in the age group, and found they most of them either not represent a real single person (for example, a media/company, or a special interest/supporting group) or represent a real person but like to be age cheating.

Our results are similar to the study conducted in [6], which shows near 80% of the sampled users are with ages below 30. Meanwhile, according to the analysis of Alexa, Plurk's visitors tend to be young people [8].

TABLE 3. NATIONALITY DISTRIBUTION OF THE SAMPLE DATA

(a) Taiwan versus non-Taiwan			
	nationality	users	ratio
	Taiwan	45,837	46.40%
	non-Taiwan	41,536	42.10%
	error	1	0.00%
	not shown	11,392	11.50%
	total	98,766	100.00%
(b) top 10 nationality			
	nationality	users	ratio
1	Taiwan (tw)	45,837	46.40%
2	Indonesia (id)	17,224	17.40%
3	Philippines (ph)	11,847	12.00%
4	United States (us)	2,783	2.80%
5	Singapore (sg)	1,759	1.80%
6	Malaysia (my)	1,036	1.00%
7	Brazil (br)	658	0.70%
8	Japan (jp)	646	0.70%
9	Hong Kong (hk)	639	0.60%
10	Canada (ca)	458	0.50%
	summation	82,887	83.90%

TABLE 4. AGE STRUCTURE OF THE SAMPLE DATA

Age group	users	ratio
00—06 (preschool)	1	0.00%
07—12 (elementary school)	1	0.00%
13—15 (junior high sch.)	4,860	4.90%
16—18 (senior high sch.)	18,759	19.00%
19—23 (college)	33,999	34.40%
24—30 (graduate school)	18,496	18.70%
31—40	9,014	9.10%
41—50	1,839	1.90%
51—60	402	0.40%
61—70	183	0.20%
71—80	161	0.20%
81—100	632	0.60%
not shown	10,419	10.50%
Total	98,766	100.00%

B. Overview of User Behaviour Data

In this subsection, we focus on system maintain data, including registration date (*member_since*), friend counts, fan counts, Plurk counts, profile views, and Plurk response counts. Table 5 shows statistics of the data. The *use_month* data is derived from member_since, which indicates how long (in month) the account exists since the account is registered by the user.

1) Friend_counts, Fan_counsts, and Profile_views: These three features are in general presenting the degree of interactions contributed by other users to the specific user. To some extend, the three features represent how popular the user is. Our sample data show each Plurk user, on average, has 34.3 friends and 15.7 fans. The number is far less than the number reported in [6], which shows on average Plurk user has 183 friends and 93 fans. And we note Facebook user on average has 130 friends, according to Facebook statistics [5] on Oct. 2011. We have browed the friend counts of the user profiles, and found a great portion of the users have only few friends. Half of the users have friend number less than or equal to 10. The 25 percentile and 75 percentile of the friend count are 3 and 31 respectively, as shown in Table 5. The statistics of *fan_count* is similar, with half to one-third smaller than the scale of *friend_count*. It seems that Plurk users prefer adding others as "friends" instead of "fans". Meanwhile, the statistics of *profile_views* is about ten times larger than the number of *friend_counts*.

TABLE 5. USER BEHAVIOUR STATISTICS OF THE SAMPLE DATA

(A) friend-counts, fan-counts, and profile-views			
	friend counts	fan counts	profile views
average	34.32	15.68	471.67
std deviation	733.94	135.10	1409.82
minimum	1.00	0.00	3.00
25 percentile	3.00	1.00	43.00
50 percentile	10.00	4.00	104.00
75 percentile	31.00	11.00	383.00
maximum	193,294.00	28,033.00	168,181.00
(B) Plurks, Responses, and use month			
	Plurks	Responses	use month
average	957,018.13	1,713.69	23.03
std deviation	64,094,573.17	33,095.58	7.46
minimum	0.00	0.00	0.00
25 percentile	4.00	2.00	17.00
50 percentile	36.00	54.00	24.00
75 percentile	275.00	766.00	28.00
maximum	4,294,967,295.00	7,204,172.00	42.00
Total user counts: 98,766			

2) Plurk_counts and Responses_counts: The values for *plurk_counts* and responses counts represent the degree of engagement of the user to the Plurk. The value *plurk_count*, counts the number of message the user post (call "plurks"), while the value, *responses_count*, counts the number of message the user response to other user's plurks. Our data

show the distribution of *plurk_count* is quit unbalanced. Most of plurks are posted by top few percent of users. The mean value of *plurk_count* is over 957,000, while 75 percentile of the users post less than 300 messages. Again, we browsed the user profile to find possible reasons for this unique phenomenon. We found some users post lots of messages everyday. Most of them are promoters to promote some products or something else, for example a news media, a company, or a fan club. The distribution of *responses_count* is unbalanced as well, but not as large as the distribution of *plurk_count*. The mean value of *responses_count* is 1713, two times more than the value of 75 percentile (766). Again, we found some users response to users plurks arbitrarily. Many of the responses are related to advertisement, and though as spam.

TABLE 6. TOP TEN FRIEND USERS OF THE SAMPLE DATA

(a) Friends, Gender, Description

	Friends	Gender	Account/ Description
1	193,294	Male	Song_eliver/ Music service
2	111,675	Female	Gugigi/ fortuneteller service
3	25,456	Male	Plurkfu/ plurk recommandation serv.
4	23,916	Female	yam_hercafe/female subject discussion
5	22,104	Female	nciku_tw/ English learning service
6	17,139	Male	Bodhisattva_/ fortuneteller service
7	15,879	Female	ShengKuan/ Taiwan politician
8	10,174	Male	Blessu/ bless offer service
9	7,792	Male	pct_astro/ constel. fortuneteller serv.
10	5,420	Male	Bikhim/ Taiwan politician

(b) Age, Fans, Views, Plurks, Responses, Use_month

	Age	Fans	Views	Plurks	Resp.	Use_ month
1	19	2,723	68,591	891	3,672,780	25
2	31	1,827	54,632	87	6,285,044	25
3	26	894	26,887	168	7,204,172	18
4	30	3,164	62,127	2,777	381	30
5	--	15,062	168,181	1,408	676	35
6	--	501	20,223	47	473,174	21
7	--	659	47,259	724	7,569	33
8	81	1,259	33,330	12	329,158	28
9	29	1,227	30,745	8,142	1	31
10	--	1,818	43,646	690	1,765	28

--: age not shown

3) *Heavy Users:* According to Robin Dunbar, a British anthropologist and evolutionary psychologist, the number of individuals with whom a person can maintain stable relations is around 150 [9]. The number is referred as Dunbar's number, which is a theoretical cognitive limit to the number of people with whom one can maintain stable social relationships [10]. However, in Table 5 we found some users with friend number as large as over one hundred thousand, which is 6,000+ times larger than 75 percentile of the distribution (31). While, Plurk's friend is different from real friends in our real life, we use this number as a threshold to check with the friend count in the user profiles. We found about 97.5% (96258/98766) of the users are with friend count less then or equal to 150; 2508 users (2.5%) in the sample data are with friend count larger than 150. In Table 6, we list the top 10 users with most friend counts. We found that all the top 10 users are with nationality of Taiwan. Most of them provide some special services. Only two of them (ranked 7 and 10) are exception. Both of them are politicians and own many fans in real life, due to their personal charisma.

V. CONCLUSSIONS

In this paper we share what we found from this empirical case study. As far as we know, this study is first Plurk user study based on large Plurk user data randomly sampled from Plurk web site. While this study is to some extent primitive, we hope that the results of this study provide some insights and/or some references for comparison to researchers working on the topics about Internet social network user behaviours and related issues at different aspects.

REFERENCES

[1] The Facebook website. [online]. Available: http://www.facebook.com.
[2] The Twitter .website. [online]. Available: http://www.twitter.com.
[3] The Plurk website. [online]. Available: http://www.plurk.com.
[4] Facebook Press Room. [online]. Available: http://www.facebook.com/press.php)
[5] Facebook statistics. [online] . Available: http://www.facebook.com/press/info.php?statistics.
[6] PremRaj Narkhede, P. Venkata Rajesh, Sumit Kumar, "Analysis on the General Profile of Users on Plurk.com" [online]. Available: http://www.buzzom.com/2010/05/demographic-analysis-of-users-on-plurk-com/.
[7] Plurk, wikipedia [online]. Available: http://en.wikipedia.org/wiki/Plurk.
[8] Alexa internet traffic statistics for Plurk.com [online]. Available: http://www.alexa.com/siteinfo/plurk.com.
[9] Robin Dunbar, "How Many Friends Dose one Person Need?: Dunbar's Number and Other Evolutionary Quirks", Harvard University Press, 2010.
[10] Dunbar's number, wikipedia [online]. Available: http://en.wikipedia.org/wiki/Dunbar%27s_number
[11] Barry Wellman, "Computer Networks as Social Networks", SCIENCE, vol 293:2031-2034, 2001
[12] Chun-Yao Huang, Yung-Cheng Shen, I-Ping Chiang, Chen-Shun Lin, "Characterizing Web user's online information behaviour", Journal of the America Society of Information Science and Technology, vol 58(13):1988-1997, 2007.
[13] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, J. Niels Rosenquist, "Understanding the Demographics of Twitter Users", in Proceedings of ICWSM 2011.
[14] O. Phelan, K. McCarthy, B. Smyth, "Using Twitter to Recommend Real-time topical news", in Proceedings of RecSys 2009, pp. 385-388.
[15] T. Sakaki, M. Okazaki, Y. Matsuo, "Earthquake shkes Twitter Users: Real-time Event Detection by Social Sensors", in Proceedings of WWW 2010, pp. 851-860.
[16] K. Lerman, R. Ghosh, "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks", in Proceeding of ICWSM 210, pp. 90-97.
[17] Claudia Hauff, Geert-Jan Houben, "Deriving Knowledge Profiles form Twitter", in Proceedings of EC-TEL 2011 (LNCS 6964), pp. 139-152.