Trend Analysis of Privacy Settings and User Classifications in Social Network Services

申 立明　　岩井原 瑞穂
早稲田大学大学院情報生産システム研究科　北九州市若松区ひびきの　2-7

E-mail:　shenlimikng@fuji.waseda.jp,　iwaihara@waseda.jp

Abstract Many SNSs are based on access control that selectively discloses privacy information to the public, which causes numerous privacy issues, while, sharing information is the purpose of using social media. Our aim is to analyze privacy setting tendencies through measuring distributions of different user categories and by evaluating their privacy scores in Facebook. We also investigate how a user's privacy setting is affected by his/her neighbors, and observe the relationship between privacy settings and social activity. These findings can be used for guiding SNSs users in designing their privacy settings.

Keyword　Social Network,　Privacy,　Facebook

1. INTRODUCTION

Social network services (SNSs) are rapidly becoming popular among internet users. The web is empowered with a new mode to adopt the real life social relationships into the digital world [2]. SNSs are individually centered, and their primal objective is to uphold user connections and communities having the same interests. SNSs make it very convenient to keep in touch with friends, relatives, classmates and colleagues in real social life. Facebook has nearly 800 million users which is more friendship-oriented and primarily used for communication, news feeds, entertainment and photo and video sharing.

However, the fast growth of SNSs has also raised issues and concerns about trust, privacy and security. It is very important to improve the security and privacy mechanisms in SNSs. With the rapidly increase of the amount of users in the SNSs all over the world, there exists an increased threat regarding trust and privacy leak as compared to the traditional web sites. Trust has been on the research agenda in several disciplines such as computer science, psychology, philosophy and sociology. Moreover, the SNS users are encouraged to share and distribute variety of personal information, including interests, cultural, religious and social attributes.

While analyzing privacy in SNSs, it is obvious that even if the individual user has high privacy awareness and a number of methods are adopted to enhance privacy, the SNSs users could still encounter privacy violations performed by the provider of the social network service. SNSs providers gain all the access to user data that are provide by users themselves and they could explore the data in many ways. The providers' success in attracting a large amount of users who are prepared to share their user profile has increased the market value for each SNS. For the providers to continue in growing and attracting more users it is of great importance to consider the privacy concerns. One of the major challenges will probably be to find a good trade-off between the level of privacy and the ease of use of modern SNSs. If a user's settings are too restrictive, communications with his/her friends become difficult and may lose opportunities to be known to a larger scope of users. On the other hand, if the settings are too public, the risk of his/her personal information being misused or unexpectedly used is raised [3].

In this paper, we conduct trend analysis of privacy settings based on a large dataset of Facebook users. We divide users into groups by their profile attributes, such as gender, number of friends and so on. To evaluate the openness level of a user's privacy settings we utilize privacy score [2]. The results show the connection between the level of openness of privacy settings and the scale of friendship on social network services. In Section 2, we survey related work. In Section 3, our method of evaluating users' privacy settings was introduced. In Section 4, we show several graphs to illustrate the results of our experiments. Section 5, as a summery to our work, also talked about works of next step.

2. RELATED WORK

A considerable number of works have been done about

the privacy setting issues of SNSs. In [2], a technique of computing privacy score on social networks by implementing Item Response Theory is introduced. This score indicates the user's potential risk caused by his or her participation in the network. They proposed that the definition of privacy score satisfies the following intuitive properties: the more sensitive information a user discloses, the higher his or her privacy risk. Also, the more visible the disclosed information becomes in the network, the higher the privacy risk. Also, mathematical models were developed to estimate both sensitivity and visibility of the information.

In [3] a framework was proposed on assisting privacy settings of SNSs by visualizing tendencies of similar users' settings and recommending based on attribute co-occurrence. Their approach was based on understanding of privacy practice of others and tried to collect other users' privacy settings as many as possible, and present overall tendencies of the privacy settings to the target user.

3. METHOD

Facebook provides very detailed profile privacy settings. Profile attributes (such as Work and Education, Living, Family, Basic Information and Contact Information including birthday, sex, Mobile Phones etc.) can be set to different openness levels, including "public", "friends", "only me" and "custom". In the "custom" option user can set attribute to "Friends of Friends", "Friends", "Specific People or Lists" or "Only me". We use a bit vector to represent a user's open privacy settings. If an attribute in a user profile is open to the public then we assign "1" to the corresponding attribute in the vector, otherwise we assign "0" to this attribute.

Profile attributes have different sensitivity; the easier an attribute could be used to identify a person the higher sensitive this attribute is. Attributes like phone numbers, address are high sensitive could be used to identify a user in real world. Exposing the profile attribute of high sensitivity to the public causes high risk in profile information security. We weigh profile attributes by incorporating public knowledge on the risk of private information leakage. JNSA published surveys on information security incidents [4]. Then we could evaluate a user's privacy score [2] to represent the level of risk undertaken by users of online social networks. A user reveals his/her much more sensitive information to the public tends to gain higher privacy score.

We compute a user's privacy score by two factors. One is the visibility of a profile attribute from the public; another is the weight that represents sensibility of this attribute.

The privacy score of individual j due to item i, denoted by Ps (i, j), can be any combination of sensitivity and visibility. That is,

$$\mathrm{Ps}(i,j) = \beta_i \otimes \mathrm{V}(i,j)$$

β_i denotes the sensitivity of item i. Operator $\otimes$ is used to represent any arbitrary combination function that respects the fact that $\mathrm{Ps}(i,j)$ is monotonically increasing with both sensitivity and visibility. For simplicity, throughout our discussion we use the product operator to combine sensitivity and visibility values. In order to evaluate the overall privacy score of user j, denoted by $\mathrm{Ps}(j)$, we can combine the privacy score of j due to different items. Again, any combination function can be employed to combine the per-item privacy scores. For simplicity, we use a summation operator here. That is, we compute the privacy score of individual j as follows:

$$\mathrm{Ps}(j) = \sum_{i=1}^{n} \mathrm{Ps}(i,j) = \sum_{i=1}^{n} \beta_i \times \mathrm{V}(i,j)$$

In the above, the privacy score can be computed using the observed visibility.

4. DATA ANALYSIS

In this paper we are focusing on the privacy settings of Facebook users. If a profile attribute can be accessed from a user ID which has no friends implies that this profile attribute is set to visible to the public. To obtain user dataset randomly, we made an ID generator to collect user profile page by trial and failure method. In this way, we randomly collected 93135 users' profile (including 42068 female users and 51067 male users) through a Facebook ID without any connection to other users.

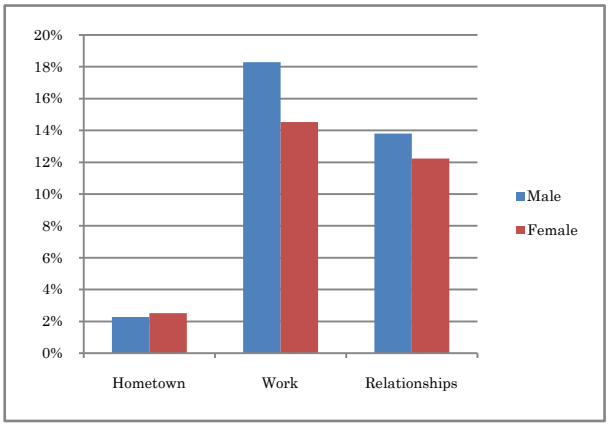

Fig.1. Disclosure rate by gender 1

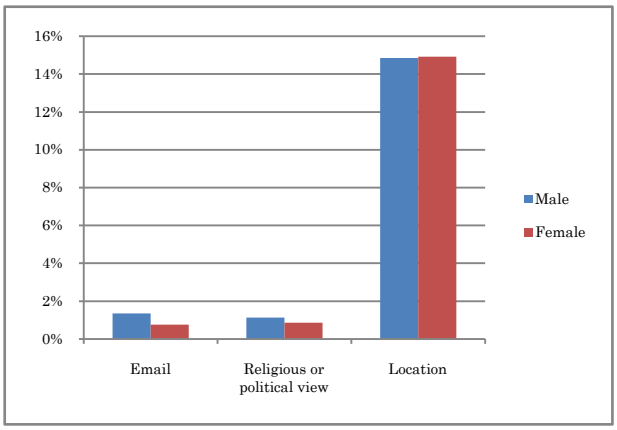

Fig.2. Disclosure rate by gender 2

Figure 1 and Figure 2 show the disclosure rates (the proportion that users who set a profile attribute to be visible from public in all the users of the same category) for profile attributes grouped by gender. Notice that female users are less likely to disclose attributes than males. Few users are willing to disclose certain attributes like Email address to the public. This kind of information could be collected for advertising or other purpose that might be annoying to users. Attributes like religious or political view also have low disclosure rate. This may because such things are unpopular so that many users did not fill these kinds of attributes.

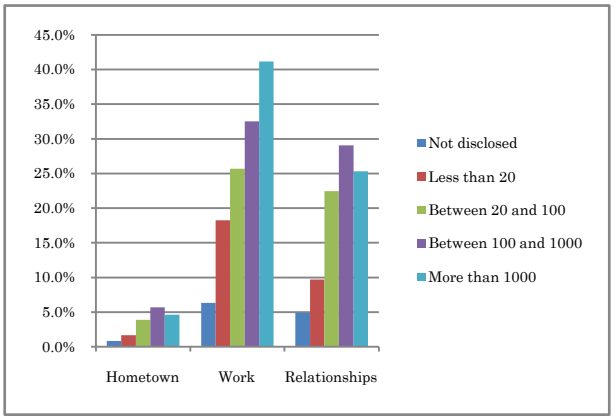

Fig.3. Disclosure rate by number of friends 1

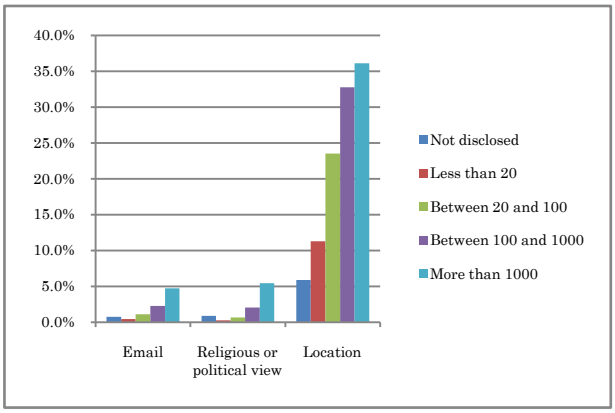

Fig.4. Disclosure rate by number of friends 2

Figure 3 and Figure 4 show the disclosure rates for some profile attributes by users grouped by their number of friends. Among 93135 users, 44694 (that is 48%) users chose not to disclose their friends list. From the chart we can see the users who don't show their friend list have very low disclosure rates of all the profile attributes. That can be explained as these users adopt a very conservative privacy policy. Generally, users have large number of friends are more likely to disclose their information. Comparing users who have less than 20 friends and who have more than 1000 friends the difference of their disclose rate is so huge. Users having friends over 1000 are much more active in the social networks. With detailed privacy information these users are more likely to be identified by their friends in real life and even for unknown people rich information will increase the chance to be accepted as a new friend.

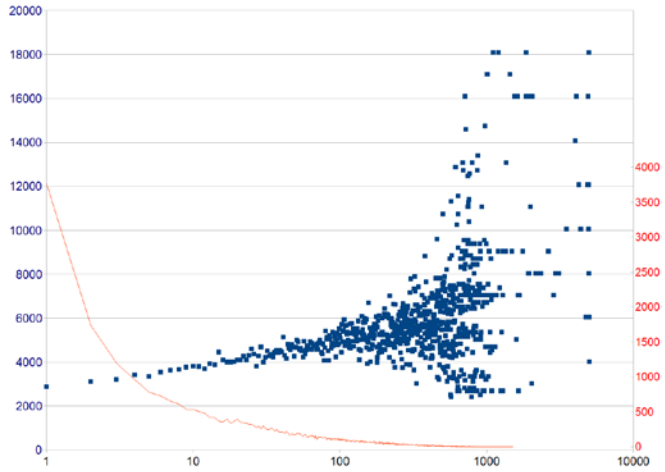

Fig.5. Privacy score distribution by Number of Friends

The blue dots in Figure 5 shows the privacy score distribution scattering over the number of friends in a logarithmic scale. The red line depicts the population

distribution by the number of friends in a logarithmic scale.

The correlation between Privacy score(Y) and Number of Friends(X) is 0.492, indicating a positive association between them. After linear regression we got the regression equation: Y = 2.308 * X + 4905.814.

From Figure 5 we can see the trend of privacy score such that while the number of friends is increasing, users' average privacy score is increasing. This confirms the connection between the level of openness of privacy settings and the scale of friendship. Users have more friends are tend to have high openness levels, which is captured by privacy score.

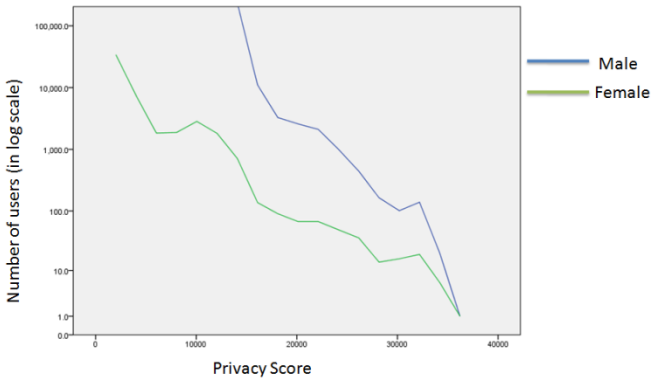

Fig.6. Privacy score distribution by gender

Figure 6 illustrates privacy score distribution by gender. Higher privacy score implies more open privacy settings and more disclosed attributes. From the privacy score, we can see the outcomes consistent with the former observation that female users are less likely to disclose their information.

Fig.7. Privacy score distribution by user language

Figure 7 shows privacy score distribution by user language. From it we can see that the trend of privacy score distribution is not very clear. The vertical axis is average privacy score, and the horizontal axis is language code of Facebook users. Facebook language codes are following the ISO Language and country codes respectively, concatenated by an underscore. The basic format is "ll_CC", where "ll" is a two-letter language code, and "CC" is a two-letter country code. For instance, 'en_US' represents US English. There are two exceptions that do not follow the ISO standard: ar_AR and es_LA. These denote Arabic and Spanish, and the latter case indicates a few more specialized localizations of Spanish. The language codes lying on the horizontal axis are sorted by their user populations.

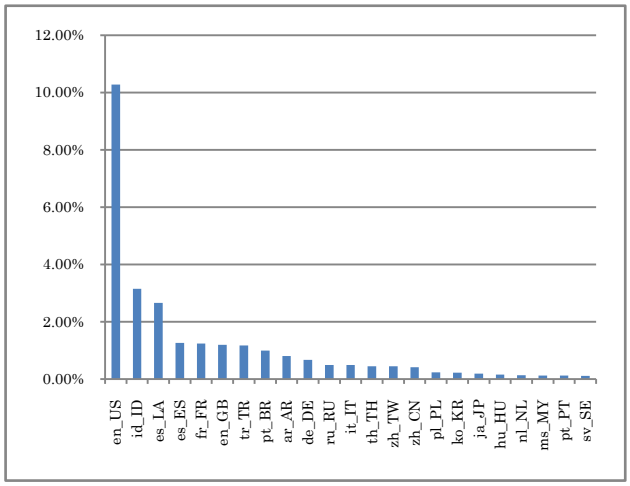

Fig.8. user language code distribution

From Figure 8 we can see the proportions of each language used on Facebook. Combining with Figure 7, we can see that the fluctuation at the right part in Figure 7 is caused by the reduction of samples. It is hardly to say that the openness level of privacy settings have a strong connection with users' languages.

5. CONCLUSION

In this paper, we conducted trend analysis of privacy settings based on a large set of Facebook users' open profiles. Our method on crawling open profile is not depending on full access to a user's profile, therefore we could collect a large amount of users' privacy settings information, where users can be categorized by available user attributes. In the future we will classify users into more categories by their location and languages. Also, we will monitor changes of privacy settings over a long period of time, and we also consider evaluating correlation between user's activities and privacy settings, and influence of privacy practice of friends.

REFERENCES

[1] Niklas Lavesson , Henric Jonhnson , “Measuring Profile Distance in Online Social Networks”, WIMS 2011.

[2] K. Liu, E. Terzi, “A framework for computing the privacy score of users in online social networks,” Proc. Int. Conf. Data Mining, 2009.

[3] Toshikazu Munemasa and Mizuho Iwaihara, “Trend Analysis and Recommendation of Users’ Privacy Settings on Social Networking Services”, Social Informatics, 2011.

[4] NPO Japan Network Security Association. 2009 Survey Report of Information Security Incident, Apr. 2010.