Perceived and Realized Risk of Online Social Networks

December 7, 2011

Abstract

With the fast development of Online Social Networks (OSN), privacy issue has become a great concern. Privacy settings of OSN may involve hundreds of items, indicating how the user presents himself/herself to other users. Hiding or disclosing private information to other users is basically a personal choice. However, disclosing private information also comes with certain risks such as information spam, insurance discrimination, and financial fraud. Meanwhile, people disclose private information to gain some social benefits, such as getting reconnected with old friends and open to job opportunities. OSN users may implicitly make decision on privacy disclosure based on a personal tradeoff between the risks and the benefits. Is there a global pattern representing most users concern about the implicit privacy risk? How to model the privacy risk in terms of social network structure? How can users quantitatively evaluate the privacy risks and social benefits by disclosing or hiding a private item? In this paper, we mine real data from online social network to investigate peoples feeling about the tradeoff between privacy risk and social benefits of disclosing certain personal information item. Either privacy risk or social benefit is realized only when a certain profile item is visited by another OSN user. We also model the probability of visit based on the users personal social network structure and derive the expectation of realized privacy risk. The social benefits can be modeled in a similar way. These models can help users assess their privacy settings of profile items according to potential benefits and risks.

1 Introduction

In recent years, online social networking is becoming a popular life style for many users. [rewrite benefits of social networking, it provides people a virtual world for social relations. In order to enjoy these benefits, users have to disclose their personal information to allow other users search, identify and link them.

Disclosing private information also raises certain risks such as information spam [], insurance discrimination [], and finacial fraud [1]. **[some examples show how malicious parties can use the private information to damage users' interests.] To be added later** These examples may educate people how risk to disclose personal information. People are more and more aware of the risk of disclosing private information. As a result, hiding or disclosing a profile item is a personal tradeoff between privacy risks and social benefits.

Intuitively, different personal information items may be associated with different levels of risk. To make this tradeoff between risks and benefits, people may want to know what the level of risk is if they disclose a personal information item. Although people understand more and more private-information based attacks, it is difficult to quantitively evaluate the level of risk and rank the profile items by the damages caused by disclosing them. A comprehensive approach is to understand all the possible attacks that utilize the exposed personal information and then evaluate the overall damages of disclosing a profile item. We refer the privacy risk defined by this approach as *real privacy risk*, which is our long-term goal.

Alternatively, it is also useful to learn what people feel about privacy risks. The hypothesis is that most people have taken the risk factors into consideration when they adjust their privacy settings. The aggregated user opinions about privacy setting of a profile item could provide some clues about the level of risk of this item. We name the aggregated user judgement as the *perceived privacy risk*.

The privacy risk is not *realized*, if the profile item is not seen by another user. We argue that the probability of a profile item is seen is determined by the user's privacy setting and the structure of his/her social network. Therefore, we also study how network structure affects the realized privacy risk.

In this paper, we study the perceived privacy risk with real facebook data and build a model for estimating the expectation of realized privacy risk. This paper has three unique contributions.

- We believe that privacy setting of social network profile items is not only a personal preference, but also related to the risks associated with exposing the specific profile items. A better understanding of the privacy risk can be built around the concepts such as real privacy risks, perceived privacy risk, and realized privacy risk.
- We propose an empirical method to study users' implicit judgements on the privacy risk, with which we can guide new users of social networks or existing users unaware of privacy risks to evaluate their privacy settings.
- We propose a network based model fro estimating the expected realized privacy risk. With this model and the privacy risk measure (either real or perceived), a user can estimate the potential privacy risk by disclosing one specific item to different levels of audience.

2 Definitions and Notations

We define social network as a directed graph $\mathcal{G} = (U, F)$, where U is the set of users in the social network, and we use N to denote the number of users of the social network, and F is a set of edges that represent friendship in social networks. We use $u_i, 1 \leq i \leq N$ to represent a user in social network, $f_{ij} : u_i \to u_j, i, j \in \{1, 2, \ldots, N\}$ to represent the friendship between user u_i and user u_j.

In a social network, user u_i has N_i friends, and these friends form a micro-community denoted as $\mathcal{C}_F = \{u_j | \exists u_i \to u_j, 1 \leq j \leq N_i\}$, and friends of these friends form a even larger micro-community denoted as $\mathcal{C}_{FOF} = \{u_k | \exists u_i \to u_j \to u_k, 1 \leq j \leq N_i \text{ and } 1 \leq k \leq \sum_{m=1}^{N_i} N_m\}$, and the extreme larger community consists of all users in the given social network, and we denote it as $\mathcal{C}_{ALL} = \{u_i | u_i \in \mathcal{G}\}$. And for comparison, we also denote the smallest micro-community, the user him/her-self, for user u_i as $\mathcal{C}_{USER} = u_i$. The community is illustrated in Figure 1.

Every node u_i in social network $\mathcal{G}$ has P_i number of features $F_i = \{f_{ij} | 1 \leq j \leq P_i\}$. These features, in social network, corresponds to the profile items of a user, e.g. name, date of birth, address, education etc., which are the identities of user u_i in the social network. Profile items are basic information for social interaction and can also be used by social network providers to provide personalized services such as friends recommendation, search and targetted advertising.

Usually, social networks require a minimum number of items upon registration, for example, sex, birthday etc.. And other optional items can be updated later. Because profile items are carriers of personal private information, social network sites provide configuration options for user to change their targetted audience for each item. For a user u_i, we denote the setting of social network profile item f_{ij} as s_{ij}. For example, Facebook provides settings that include only me, friends only, friends of friends and everybody from the most conservative to the most open. So for the case of facebook, we have $s_{ij} \in \{S_{me}, S_f, S_{fof}, S_{all}\}$. And we will use these levels of settings in our privacy model. Although Facebook also provide finer grained privacy setting by specifying specific users to view a certain profile item, in our model we ignore this situation and will leave it for future work. The meaning and corresponding audiences of each privacy settings are listed in Table 1.

3 The Item Response Theory Model

In this section, we will have a brief review of the item response theory(IRT). The original IRT model is only used to model dichotomous tests, which include only correct or incorrect outcomes. The polytomous model was proposed to handle cases with multiple outcomes.

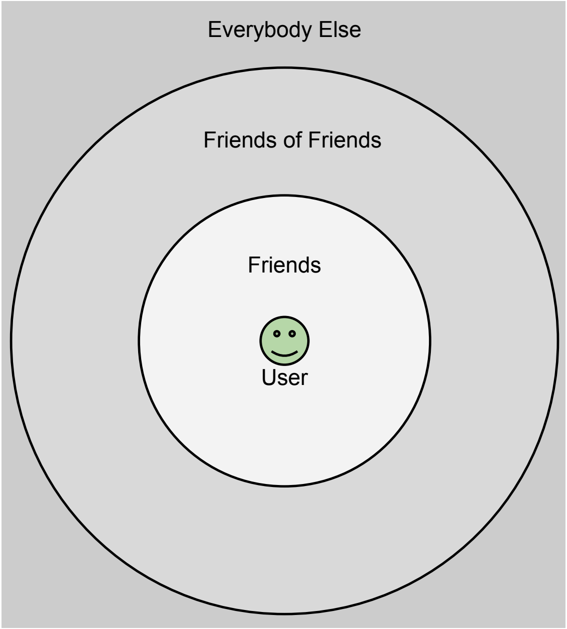

Figure 1: User Centric Social Network Communities.

Access Setting	**Audiences**
Only Me(S_{me})	$\mathcal{C}_{USER}$
Friends (S_f)	$\mathcal{C}_F$
Friends of Friends (S_{fof})	$\mathcal{C}_{FOF}$
Everybody (S_{all})	$\mathcal{C}_{ALL}$

Table 1: Profile Item Access Setting and Corresponding Audiences.

Item Response Theory is a paradigm for the design, analysis, scoring of tests and questionnaires measuring abilities, attitudes and other variables. It is based on the idea that the probability of a correct response to an item is mathematically determined by some parameters of persons and items. The person parameter is usually called latent trait, ability or the strength of an attitude.

The common approach to test the ability of an examinee is to develop a test consisting a number of items, each item is used to test a certain ability of interest, the items are dichotomously scored such that correct answers with get score of ones and zeros for incorrect answers. The probability that an examinees can correctly answers an question is determined on the one hand by the ability of the examinee, and by the item difficulty on the other hand. The ability is denoted as θ and the probability of answering an question item correct with ability θ is $P(\theta)$. And for a specific test item $P(\theta)$ will be larger for high ability examinees and lower for low ability examinees.

3.1 Application of test analysis of IRT model

Two general processes are involved for IRT to be used in test analysis.

1. Test calibration. The test calibration stage is used to determine the item characteristic curve(ICC) for a specific test item. Alan Birnbaum proposed a two stage iterative method with maximum likelihood estimation. In one stage the parameters of N items in a test are estimated and in the second stage, the ability parameters of the M examinees are estimated. These two stages are performed iteratively until a stable set of parameter sets are obtained. At that point, a test has been calibrated and a ability scale metric defined.

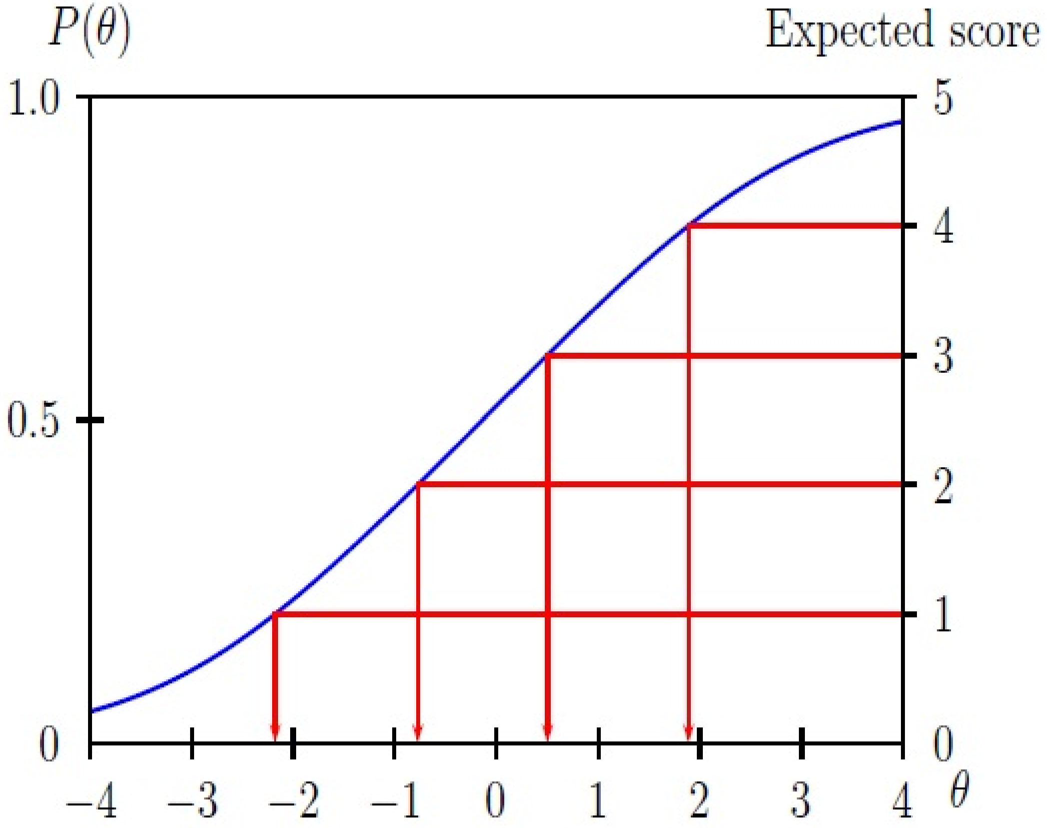

Figure 2: Ability estimation using a test with five items.

2. Examinee ability prediction with the IRT model. The task of this step is to estimate the ability of an examinee with a test. Let's suppose a test with N item is given a an examinee, and the examinee responses to each of these questions resulting in a dichotomous score vector for this test. By using the test score as a prior ability of the examinee and along with the parameters of each item, we can iteratively estimate the ability of an examinee with maximum likelihood estimation with the following formula.

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^{N} -a_i[u_i - P_i(\hat{\theta}_s)]}{\sum_{i=1}^{N} a_i^2 P_i(\hat{\theta}_s) Q_i(\hat{\theta}_s)}$$

where $\hat{\theta}_s$ is the estimated ability of the examinee within iteration s; a_i is the discrimination parameter of item $i \in \{1, 2, \ldots, N\}$; u_i is the response made by the examinee to item i; $P_i(\hat{\theta}_s)$ is the probability of item i under the given item characteristic curve model, at ability level $\hat{\theta}$ within iteration s; and $Q_i(\hat{\theta}_s) = 1 - P_i(\hat{\theta}_s)$ is the probability of incorrect response to item i, under the given item characteristic model, at ability level $\hat{\theta}$ within iteration s. An example is shown shown in figure 2.

The One parameter IRT model

This IRT model has only one parameter b, and can be shown as in Eq 1.

$$P(\theta|b) = \frac{1}{1 + e^{-(\theta - b)}} \tag{1}$$

The Two parameter IRT model
This model has an ability parameter and a discrimination parameter.

$$P(\theta|a,b) = \frac{1}{1+e^{-a(\theta-b)}} \tag{2}$$

The Three Parametric Model

$$P(\theta|a,b,c) = c + (1-c)\frac{1}{1+e^{-a(\theta-b)}} \tag{3}$$

3.2 Concrete example

1	1	1	1	1
0	1	1	1	1
0	0	1	1	1
0	0	0	1	1
0	0	0	0	1

4 Exploring Perceived Privacy Risk

Risks and attacks of online social network reflect the potential damage to social network users. And users are learning, they start to make careful decisions about what information to put online under which circumstances, at the same time, more and more people are changing to a more restrictive privacy setting [2]. Usually, social network users will have risk considerations in mind when disclosing their private information on the social network. And it has been shown [3] that stronger privacy concerns resulted in more restrictive profile settings.

Definition 1. *Perceived risk is the direct reflection of user's opinion on how sensitive a specific profile item is to him/her-self. And this opinion will guide user make privacy settings on social network.*

5 Modeling Realized Privacy Risk

5.1 Realized Social Network Privacy

Definition 2. *The social benefits and/or social risks of user u_i are realized when user $u_j, j \neq i$ visits the profile item f_{ij} of u_i, and we call the the social network privacy of u_i is realized upon event $e_{u_j \to u_i}$.*

We define the probability that the profile item f_{im} of user u_i is visited by user u_j as p_{jim}.
Use b_{jim} to denote benefit of u_j access profile item f_{im} of u_i.
The realized benefit of user u_i is defined as:

$$\mathcal{B}_i = \alpha_i \sum_{m\in\{1,2,\ldots,N_i\}} \sum_{j\in\mathcal{C}} b_{jim} \times p_{jim}$$

where α_i is parameter specific to each user, N is the number of items that user u_i has, and $\mathcal{C} \in \{\mathcal{C}_{USER}, \mathcal{C}_F, \mathcal{C}_{FOF}$ and $\mathcal{C}_{ALL}\}$ is the community of friendship.
The realized risk of user u_i is defined as:

$$\mathcal{R}_i = \beta_i \sum_{m\in\{1,2,\ldots,N_i\}} \sum_{j\in\mathcal{C}} r_{jim} \times p_{jim}$$

where β_i is the risk factor for user u_i, r_{jim} is the risk when u_j visits profile item f_{im} of user u_i, and p_{jim} is the probability of user u_j visits profile item f_{im} of user u_i.

For every friend $u_j, j \in \{1, \ldots, N_i\}$ of user u_i, the probability that u_j visits u_i is determined by the number of friends u_j has. And by assuming the simplest equal probability model we have $p_{ji} = \frac{1}{N_j}$, where N_j is the number of friends that u_j has.

For every friend of friend $u_j, j \in \{1, \ldots, N_i\}$ of user u_i, the probability that u_j visits u_i is determined by the number of friends of friends u_j has. And by assuming the simplest equal probability model we have $p_{ji} = \frac{1}{\sum_{k=1}^{N_j} N_k}$, where N_k is the number of friends of u_j's friend u_k has.

5.2 Deriving Benefit b_{jim} From Perceived Benefit

Social network benefit should be a perception of u_i about the benefit he can get by setting a certain privacy level. So, it is user specific. So, $b_{jim} = o_{jim}$, where o_{jim} is the observed benefit when u_j visits profile item f_{im} of u_i, which is actually the specific setting of privacy level for profile item f_{im} of user u_i.

5.3 Deriving Risk r_{jim} From Perceived Risk

Social network risk r_{jim}, is a natural property of a profile item f_m, it is the nature of the profile item, for example, some profile items are more risky than others in nature. Example, *phone number* is more risky than *education.* So, by this defnition, we can rewrite r_{jim} as r_m. We say that profile item f_m is more *risky* when more people set this item to be a more restrictive privacy level.

By using linear model, assign integer values to each privacy level $\mathcal{S} \in \{\mathcal{S}_{me} = 1, \mathcal{S}_f = 2, \mathcal{S}_{fof} = 3, \mathcal{S}_{all} = 4\}$, we can define r_m as mean value of aggregated privacy level settings.

$$r_m = \frac{1}{N} \sum_{i=1}^{N} o_{im}$$

where o_{im} is observed privacy setting of profile item f_{im} of u_i.

5.4 Item Response Theory (IRT) model

IRT We use the two parametric logistic model to fit our privacy framework.

$$P_{ij}(\theta) = \frac{e^{\alpha_i(\theta - \beta_j)}}{1 + e^{\alpha_i(\theta - \beta_j)}}$$

$P_{ij}(\theta)$ is the probability that the privacy setting for profile item f_{ij} of user u_i can satisfy his/her expectation of using social network, α_i is the social network benefit expectation factor, which can be seen as a factor of openness of user u_i and/or benefit expectation of using the social network; β_j is the risk level for profile item f_j. And θ is the observed privacy level setting for f_j.

We expect that, when fixed risk levels β_j, the higher α_i, the harder to get satisfied; and when fixed α_i, less risky profile items can be easily satisfied. Figure 3 and Figure 4 have better illustration of these expectations.

And we can use optimization techniques to estimate parameters of benefit expectation factor α_i and risk β_j.

6 Experiments

Facebook is one of the largest social networking sites, with more than 800 million users[4] and it provides a very comprehensive privacy configuration page that let users make their privacy choices. We derive the privacy settings by crawling user's facebook pages using two accounts, one is my account which has about 100 friends and the other is nobody's account with no friends to represent a random user on facebook.

6.1 Deriving User's Privacy Settings

7 Related Work

In recent years, research community has been trying hard to deal with privacy issues of online social networks, related topics include spamming and phishing [5, 6, 7, 8, 9], attack analysis [10, 11, 12, 13, 1, 14, 15] etc., which are also directly related to traditional web privacy and security. Fang et al. [16] proposes a classification model to help social network users automate the privacy related settings. Social network privacy control is also considered an access control problem. Carminati et al. [17, 18] propose client-based semi-decentralized access control model, access is granted based on the attestation of access authorization by the access requestor. Mohd et al. [19] proposes a reflective policy assessmenst method based on visualization to help user understand the implications of access control policies. Another research branch related to social network privacy is related to the social network platform, which targets privacy risks by third party application and social network providers. Adrienne et al. [20] addresses the privacy risks associated with social network APIs through proxy. Singh et al. [21] propose an information flow model to control what untrusted applications can do with the information they receive.

Spamming, phishing and attack of social network sites is another topic that is related with privacy. Kyumin Lee et al. [22], Stringhini et al. [23], Gao, Hongyu et al. [24] and Huber, Markus et al. [6]. It is found that the identified spam data contains contents that are strongly correlated with observable profile features. Tom Jagatic et al. [25] studied phishing attacks by using the publicly available personal information from social networks. They find that the phishing attack was easy and effective with a success rate of 72%. Bilge, Leyla et al.[1] studied identity theft attacks, and find that existing users of OSN can be compromised, and their identity can be used to request friendship with other cloned victims. Lars Backstrom et al. [13] and Gilbert Wondracek et al. [12] deanonymization of private data can cause severe privacy breach in the context of anonymized data publication. Besides, social network friendship can also bring privacy threats to social network users. [26] and [10] studied the risks that social interaction and frienship can bring.

With the increasing concern of users and various social network incidents, social network sites are pressed to provide more and more finer grained privacy control configurations. For example, facebook has been refining their privacy setting over time. But on the other hand, although user awareness of social network users are increasing, and more and more users change their privacy settings. It is hard for them to fully understand and configure these settings. Fang Lujun et al. [16] propose a machine learning based method to help users automatically make the settings.

References

[1] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, and Engin Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 551–560, New York, NY, USA, 2009. ACM.

[2] Amanda Lenhart. Adults and social network websites, January 2009.

[3] Nicole Krmer Sonja Utz. The privacy paradox on social network sites revisited: The role of individual characteristics and group norms, 2009.

[4] Wikipedia. Facebook wiki, 2011.

[5] Chris Grier, Kurt Thomas, Vern Paxson, and Michael Zhang. @spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*, CCS '10, pages 27–37, New York, NY, USA, 2010. ACM.

[6] Markus Huber, Martin Mulazzani, Edgar Weippl, Gerhard Kitzler, and Sigrun Goluch. Exploiting social networking sites for spam. In *Proceedings of the 17th ACM conference on*

Computer and communications security, CCS '10, pages 693–695, New York, NY, USA, 2010. ACM.

[7] Markus Huber, Martin Mulazzani, Sebastian Schrittwieser, and Edgar Weippl. Cheap and automated socio-technical attacks based on social networking sites. In *Proceedings of the 3rd ACM workshop on Artificial intelligence and security*, AISec '10, pages 61–64, New York, NY, USA, 2010. ACM.

[8] Benjamin Markines, Ciro Cattuto, and Filippo Menczer. Social spam detection. In *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*, AIRWeb '09, pages 41–48, New York, NY, USA, 2009. ACM.

[9] Fabricio Benevenuto, Tiago Rodrigues, Virgilio Almeida, Jussara Almeida, Chao Zhang, and Keith Ross. Identifying video spammers in online social networks. In *Proceedings of the 4th international workshop on Adversarial information retrieval on the web*, AIRWeb '08, pages 45–52, New York, NY, USA, 2008. ACM.

[10] Bin Zhou and Jian Pei. Preserving privacy in social networks against neighborhood attacks. In *ICDE '08: Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, pages 506–515, Washington, DC, USA, 2008. IEEE Computer Society.

[11] Krishna P.N. Puttaswamy, Alessandra Sala, and Ben Y. Zhao. Starclique: guaranteeing user privacy in social networks against intersection attacks. In *CoNEXT '09: Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pages 157–168, New York, NY, USA, 2009. ACM.

[12] Gilbert Wondracek, Thorsten Holz, Engin Kirda, and Christopher Kruegel. A practical attack to de-anonymize social network users. *Security and Privacy, IEEE Symposium on*, 0:223–238, 2010.

[13] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 181–190, New York, NY, USA, 2007. ACM.

[14] John R. Douceur. The sybil attack. In *Revised Papers from the First International Workshop on Peer-to-Peer Systems*, IPTPS '01, pages 251–260, London, UK, 2002. Springer-Verlag.

[15] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *SP '09: Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, pages 173–187, Washington, DC, USA, 2009. IEEE Computer Society.

[16] Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. In *WWW '10: Proceedings of the 19th international conference on World wide web*, pages 351–360, New York, NY, USA, 2010. ACM.

[17] Barbara Carminati and Elena Ferrari. Privacy-aware collaborative access control in web-based social networks. In Vijay Atluri, editor, *Data and Applications Security XXII*, volume 5094 of *Lecture Notes in Computer Science*, pages 81–96. Springer Berlin / Heidelberg, 2008.

[18] Barbara Carminati, Elena Ferrari, and Andrea Perego. Rule-based access control for social networks. In Robert Meersman, Zahir Tari, and Pilar Herrero, editors, *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, volume 4278 of *Lecture Notes in Computer Science*, pages 1734–1744. Springer Berlin / Heidelberg, 2006.

[19] Mohd Anwar, Philip W. L. Fong, Xue dong Yang, and Howard Hamilton. Visualizing privacy implications of access control policies in social network systems.

[20] Adrienne Felt and David Evans. Privacy protection for social networking apis, 2008.

[21] Kapil Singh, Sumeer Bhola, and Wenke Lee. xbook: redesigning privacy control in social networking platforms. In *Proceedings of the 18th conference on USENIX security symposium*, SSYM'09, pages 249–266, Berkeley, CA, USA, 2009. USENIX Association.

[22] Kyumin Lee, James Caverlee, and Steve Webb. Uncovering social spammers: social honeypots + machine learning. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 435–442, New York, NY, USA, 2010. ACM.

[23] Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. Detecting spammers on social networks. In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, pages 1–9, New York, NY, USA, 2010. ACM.

[24] Hongyu Gao, Jun Hu, Christo Wilson, Zhichun Li, Yan Chen, and Ben Y. Zhao. Detecting and characterizing social spam campaigns. In *Proceedings of the 10th annual conference on Internet measurement*, IMC '10, pages 35–47, New York, NY, USA, 2010. ACM.

[25] Tom N. Jagatic, Nathaniel A. Johnson, Markus Jakobsson, and Filippo Menczer. Social phishing. *Commun. ACM*, 50(10):94–100, 2007.

[26] Christo Wilson, Bryce Boe, Alessandra Sala, Krishna P.N. Puttaswamy, and Ben Y. Zhao. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, EuroSys '09, pages 205–218, New York, NY, USA, 2009. ACM.

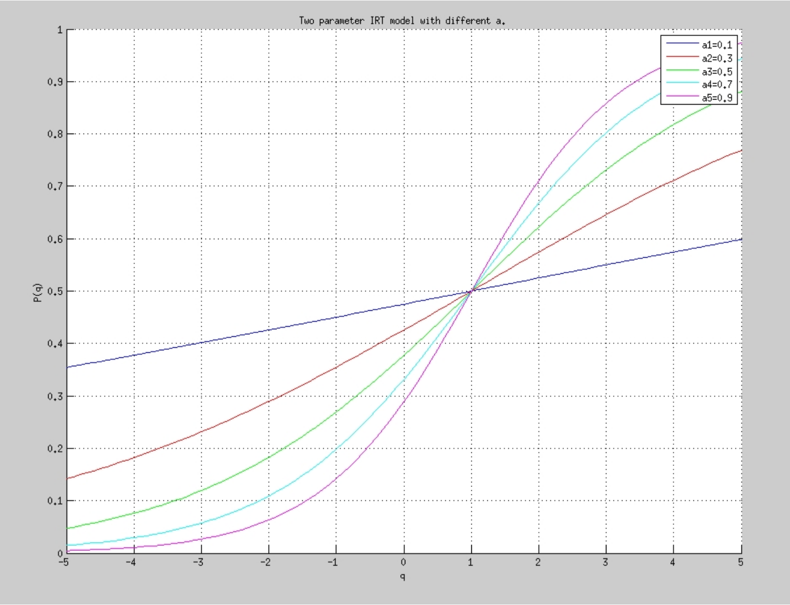

Figure 3: Satisfaction With different perception level.

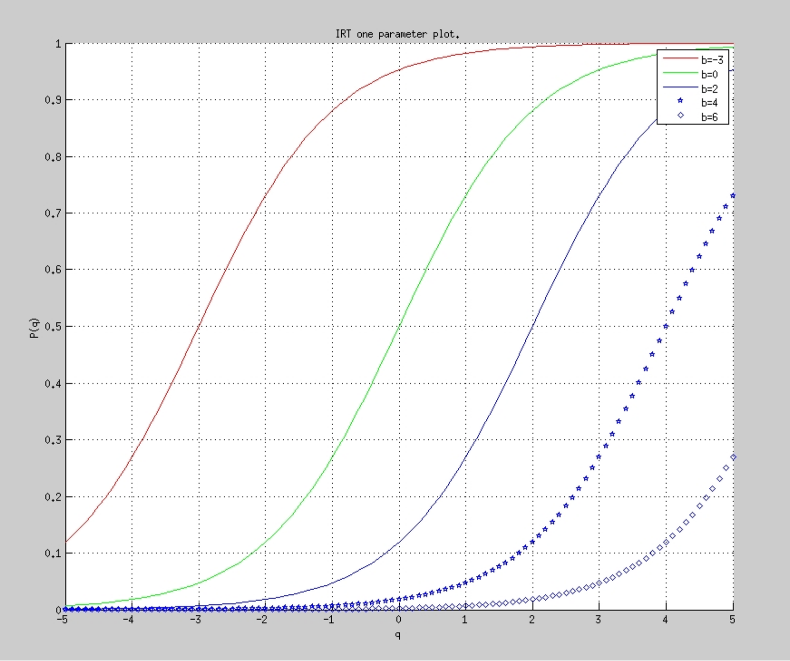

Figure 4: Satisfaction With different profile risk level.