Social Network Privacy Measurement and Simulation

Yong Wang, Raj Kumar Nepali, and Jason Nikolai
College of Business and Information Systems
Dakota State University
Madison, SD, 57042
yong.wang@dsu.edu, {rknepali,janikolai}@pluto.dsu.edu

***Abstract* – Privacy has become an important concern in online social networks. One of the fundamental challenging issues is privacy measurement. Without a practical and effective way to quantify, measure and evaluate privacy, it is hard for social networking sites and users to make and adjust privacy settings to protect privacy. In this paper, we introduce a practical and effective approach for privacy measurement in social networks. We use Privacy Index (PIDX) to measure a user's privacy exposure in a social network. PIDX is a numerical value between 0 and 100. High PIDX value indicates high privacy risk in social networks. A privacy index function *PIDX(i, j)* is proposed to evaluate actor A_j's privacy exposure to actor A_i. Using this model, it is convenient to evaluate any users' privacy exposure to their friends, friends of friends, and public. We further develop a social network privacy simulation tool, OSNPIDX, to verify the effectiveness of our approach.**

Index Terms – social networks, privacy measurement, privacy index, simulation

I. INTRODUCTION

Online social networking sites (OSNs) such as Facebook have attracted millions of users worldwide. According to Facebook, as of December 31, 2012, there were 1.06 billion monthly active users. As more people are sharing information in social networks, it also raises many privacy concerns. One of fundamental challenging issues is privacy measurement.

A social networking site is a place where people are willing to share information. The shared information is scattered in various data sources, such as user profiles, instant messages, posts, blogs, etc. The shared information is generally available to public and can be retrieved by users around the world. However, the shared information may also include users' personal data and could be manipulated by a malicious user against a person. Online social networking sites usually provide privacy settings to help protect a person's privacy. However, many of these privacy settings are confusing and depend on users to set them correctly. There is lack of effective practice to alarm users how much their privacy will be exposed or changed if certain information is posted or certain changes are made in their privacy settings. Without a practical and effective approach to quantify, measure, and evaluate privacy, it is hard for users to decide how much information they are willing to share and how much risk they are going to take. It is also impossible for online social networks to make appropriate policies and adjustments to protect user privacy.

Privacy measurement studies how to quantify, measure, and evaluate user privacy in a social network. Privacy measurement is a challenging issue. First, the definition of privacy is subjective. People have different opinions and expectation about privacy. Second, privacy can be further identified by certain disclosed attributes, such as, name, credit card number, SSN, biometric details, etc. Each of these attributes may have different privacy impact on individuals. Third, hidden information exists and could be further inferred from known attributes and social graph. This makes privacy measurement more challenging. Fourth, certain attributes could be combined and used to reveal more personal data. Attributes also show group properties which make privacy measurement more complicated.

In this paper, we introduce a practical and effective approach for privacy measurement in online social networks based on SONET model. We use Privacy Index (PIDX) to measure a user's privacy exposure in a social network. PIDX is a numerical value between 0 and 100. High PIDX value indicates high privacy risk in social networks. A privacy index function *PIDX(i, j)* is proposed to evaluate actor A_j's privacy exposure to actor A_i. Using this model, it is convenient to evaluate any user's privacy exposure to friends, friends of friends, and public. We further develop a social network privacy measurement tool, OSNPIDX, to verify the effectiveness of our approach.

The paper is organized as follows: Section II discusses the related work. Section III introduces social network model and the proposed privacy index function *PIDX(i,j)*. Section IV demonstrates a practical and effective approach for privacy measurement in online social networks using the proposed model. Experiments and results are also given in the section to verify the effectiveness of the approach. Section V summarizes the paper and future works.

II. RELATED WORK

Few works have been conducted on privacy measurement for online social networks. In [1], the authors propose to use the amount of information that can be inferred from social networks to quantify the privacy risks. A tool, PrivAware, is designed to detect and report unintended information disclosures in online social networks. PrivAware employs inference model which is based on the fact that information about users can be inferred from their social graph. Privacy score is calculated as total number of attributes visible to the third party applications divided by total number of attributes per participant. The measured percentage is then mapped to a letter grade, e.g., *A*, *B*, *C*, *D*, *E*, or *F*, where *A* score represents very few attributes being revealed and *F* score indicates that

privacy risk to the threat of a malicious third party application is high.

In [2], the authors present an approach in which privacy score is calculated by computing sensitivity and visibility of attributes. Naïve approach for evaluating sensitivity and visibility of attributes is demonstrated in [2]. The authors further extend their works to another approach in [3]. They use Item Response Theory (IRT) to evaluate sensitivity and visibility of attributes when calculating privacy score. The authors use both synthetic and real-world data to show the effectiveness of their approach.

The authors in [4] develop a tool, Privometer, to measure information leakage based on user profiles and their social graph. The leakage is indicated by probability. Privometer is based on an augmented inference model where a potentially malicious application installed in the user's friend profiles can access substantially more information. Privometer is implemented as a Facebook application. It operates in two modes. In online mode, inference is performed based on the friend's profile where most frequently value is selected. In offline mode, it uses only immediate friends and 'network-only Bayes classifier' to measure the probability of inference.

In addition, the authors in [5] propose an approach to compute privacy score as a function of sensitivity and visibility. However, no datasets are used or shown to measure effectiveness of their proposed model. In [6], the authors use risk labelling approach to tag users based on the community members' feedback. Active learning method is used to correctly label strangers. In [7], the authors conduct a thorough study and analysis of the privacy practices and policies of various online social networking sites. 45 different sites are studied with 260 criterions to evaluate them. The work in [7] does not measure the privacy scores of an individual user. However, it presents the concept of the privacy ranking of the websites which can allow the users to make their decisions on the basis of website privacy scores.

III. SOCIAL NETWORK PRIVACY MEASUREMENT

SONET model was first proposed in [8] for privacy monitoring and ranking. In [9], Privacy Index (PIDX) was introduced for privacy measurement in actor model. The work in [9] focuses on actor model and does not discuss privacy measurement among friends in social networks. In this paper, we further extend SONET model and PIDX to support privacy measurement in social networks and introduce a privacy index function *PIDX(i,j)* to evaluate any two users' privacy exposure in a social network.

A. Social Network Model

Definition *A social network is a network of actors tied together.*

An actor is a social entity (e.g. people, organization, etc.) in a social network. An actor has certain characteristics that describe its features known as attributes. Each attribute has a different impact on privacy. This impact is referred as Attribute Privacy Impact Factor (APIF). Privacy impact factor is a numerical value. We consider privacy impact factor for full privacy disclosure as 1. An attribute's privacy impact factor is a ratio of its privacy impact to full privacy disclosure. Thus, an attribute's privacy impact factor has a value between 0 and 1. Privacy impact factor reflects sensitivity of an attribute's privacy risk.

Let A_i be an actor and $L_i = \{a_{i1}, a_{i2}, \dots, a_{in}\}$ represent $A_i's$ attributes. Then, $A_i(a_{i1}, a_{i2}, \dots, a_{in})$ is a representation of an actor with attributes. We use $S = \{s_1, s_2, \dots, s_n\}$ to represent each attribute's privacy impact factor. Actor and attribute relationship can be represented using a graph $G_i = (V_i, E_i)$. In this graph, we have $V_i = (A_i, a_{i1}, a_{i2}, \dots, a_{in})$ where A_i is an actor and a_{ij} is one of the actor's attributes. An edge e_j belongs to E_i if

$$e_j: (a_{ij}, A_i) \in E_i \text{ if } a_{ij} \text{ is one of } A_i\text{'s attributes}$$

The actor model is further extended with hidden information and virtual attributes to reveal more privacy risks [8].

A social network consists of a group of actors and the relationships among them. Relationships can be represented as undirected relationship as well as directed relationship. In undirected relationship, either the relationship exists or does not exist. In directed relationship, only one user contributes to the relationship. For example, actor *A* knows *B* while *B* does not know *A*. Most online social networks use friends to define the relationships among users. Let *A*, *B* be two actors in a social network, the friend relationship between *A* and *B* is mutual, i.e., *A* is a friend of *B* and vice versa. Thus, we can use an undirected notation *A*—*B* to represent the relationship.

Let $G_S = (V_S, E_S)$ represent a social network. It includes n actors, i.e., $A_1, A_2, \dots, A_n$. We use $G_i = (V_i, E_i)$ to represent each actor's actor model. Then, we have

$$\begin{cases} V_S = V_1 \cup V_2 \cup \dots \cup V_n \\ e \in E_S \; if \; e \in E_i \end{cases}$$

We use (A_i, A_j) to further indicate actor to actor relationship. Thus, we have

$$e_{ij}: (A_i, A_j) \in E_S if \; A_i \; and A_j are \; friends$$

A social graph G_S can thus be formed based on the social network model.

The visibility of actors A_i and A_j can be further characterized by their privacy settings. For any two actors in a social network model, they might be separated by a few steps away. The **degrees of separation** is used to describe the steps between two actors. We use degrees of separation function h

$$d_{ij} = h(A_i, A_j)$$

to represent the steps between actor A_i and A_j.

According to the value of degrees of separation, friends can be divided into different groups, e.g., friends (1 degree actors), friends of friends (2 degree actors), public (3 and above degree actors). Privacy settings can thus be defined as shown in Table 1. 'X' indicates that an attribute is visible.

	User Profile Items	Friends	Friends of Friends	Public
1	Full Name	X	X	X
2	Education	X	X	X
3	Marital Status	X		
4	Family Members	X		
5	Gender	X	X	X
6	City	X	X	

Table 1 Examples of User's Privacy Settings

For any two actors $A_i(a_{i1}, a_{i2}, \dots, a_{in})$ and $A_j(a_{j1}, a_{j2}, \dots, a_{jn})$, the visibility of A_j to A_i depends on the degrees of separation d_{ij} and A_j's privacy settings. Let g_j be A_j's attribute visibility function

$$p_{jt}(i) = g_j(a_{jt}, d_{ij}), 1 \le t \le n$$

g_j returns a value to indicate a_{jt}'s visibility to actor A_i. The visibility $p_{jt}(i)$ is a numeric value between 0 and 1. 0 indicates an invisible attribute and 1 indicates a visible attribute. A value between 0 and 1 indicates a partial visible attribute. We further assume $p_{jt}(i) = 1$ if $i = j$ since a user has visibility of itself.

$A_j's$ privacy settings to A_i can be represented by

$$P_j(i) = \{p_{j1}(i), p_{j2}(i), \cdots, p_{jn}(i)\}$$

where $p_{jt}(i)$ is decided by $A_j's$ attribute visibility function g_j.

B. Privacy Index (PIDX)

Privacy Index (PIDX) is used to describe an entity's privacy exposure factor based on known attributes in actor model. We extend PIDX to social network model to measure an actor's privacy exposure to another.

Let A_i and A_j be two actors in the social network model G_S. $L_i = (a_{i1}, a_{i2}, \dots, a_{in})$ and $L_j = (a_{j1}, a_{j2}, \dots, a_{jn})$ represent their attributes. Let $S = \{s_1, s_2, \dots, s_n\}$ represent the corresponding privacy impact factors of these attributes and $P_j(i) = \{p_{j1}(i), p_{j2}(i), \cdots, p_{jn}(i)\}$ be the privacy settings of A_j to A_i. We also assume that the degrees of separation function $h(A_i, A_j)$ and attribute visibility function $p_{jt}(i) = g_j(a_{jt}, d_{ij})$ are available to calculate actor A_j's privacy settings against A_i.

Let f be a privacy measurement function which returns a numeric value on L_j, S, $P_j(i)$. We use $w(i,j)$ to represent the privacy weight of $A_j's$ visible attributes to A_i. We have

$$w(i,j) = f(L_j, S, P_j(i))$$

We use $w(j)$ to represent the maximum privacy weight of $A_j's$ attributes. According to our definition, we have $p_{jt}(j) = 1$ and $w(j,j)$ returns the maximum privacy weight of actor A_j. Thus,

$$w(j) = w(j,j) = f(L_j, S, P_j(j))$$

Definition *Privacy Index PIDX(i,j) is used to describe actor A_j's privacy exposure to A_i based on $A_j's$ visible attributes to A_i. High PIDX value indicates high exposure of privacy. Privacy Index PIDX is between 0 and 100.*

$$PIDX(i,j) = \frac{w(i,j)}{w(j)} \times 100$$

C. Privacy Measurement for Social Network Model

In this paper, we introduce three privacy measurement functions to measure A_j's privacy exposure to A_i, i.e., weighted privacy measurement function, maximum privacy measurement function, and composite privacy measurement function. Three privacy indexes are defined accordingly.

1) Weighted Privacy Measurement Function and Weighted Privacy Index (w-PIDX(i,j))

Weighted privacy measurement function is defined as

$$f_w\left(L_j, S, P_j(i)\right) = s_1 p_{j1}(i) + s_2 p_{j2}(i) + \cdots + s_n p_{jn}(i) = \sum_{t=1}^{n} s_t p_{jt}(i)$$

w-PIDX(i,j) is an index which measures actor $A_j's$ privacy exposure to A_i.

$$w - PIDX(i,j) = \frac{w(i,j)}{w(j)} \times 100 = \frac{f\left(L_j, S, P_j(i)\right)}{f\left(L_j, S, P_j(j)\right)} \times 100 = \frac{\sum_{t=1}^{n} s_t g_j(a_{jt}, d_{ij})}{\sum_{t=1}^{n} s_j} \times 100$$

2) Maximum Privacy Measurement Function and Maximum Privacy Index (m-PIDX(i,j))

Maximum privacy measurement function is defined as

$$f_m\left(L_j, S, P_j(i)\right) = \max(s_1 p_{j1}(i), s_2 p_{j2}(i), \cdots, s_n p_{jn}(i))$$

where *max* is a function returning the maximum value in the list.

m-PIDX(i,j) is an index which measures actor $A_j's$ maximum privacy exposure to A_i

$$m - PIDX(i,j) = f\left(L_j, S, P_j(i)\right) \times 100 = max(s_1 g_j(a_{j1}, d_{ij}), s_2 g_j(a_{j2}, d_{ij}) \dots, s_n g_j(a_{jn}, d_{ij})) \times 100$$

w-PIDX(i,j) is good at reflecting attribute incremental changes. However, *w-PIDX(i,j)* does not reflect the actual privacy exposure. *m-PIDX(i,j)* can be used for privacy ranking. However, *m-PIDX(i,j)* does not reflect the attribute incremental changes. Thus, we develop a new privacy measurement function, composite privacy measurement function, for privacy measurement [9].

3) Composite Privacy Measurement Function and Composite Privacy Index (c-PIDX(i,j))

Composite privacy measurement function is defined as

$$f_c\left(L_j, S, P_j(i)\right) = f_m\left(L_j, S, P_j(i)\right) + \left(1 - f_m\left(L_j, S, P_j(i)\right)\right) \times \frac{f_w\left(L_j, S, P_j(i)\right)}{f_w\left(L_j, S, P_j(j)\right)}$$

c-PIDX(i,j) is an index which measures actor $A_j's$ privacy exposure to A_i based on $A_j's$ composite privacy measurement function. *c-PIDX(i,j)* is defined as

$$c - PIDX(i,j) = f_c\left(L_j, S_j, P_j(i)\right) \times 100$$

c-PIDX(i,j) can be represented using *w-PIDX* and *m-PIDX* as below:

$$c - PIDX(i,j) = m - PIDX(i,j) + (100 - m - PIDX(i,j)) * \frac{w - PIDX(i,j)}{100}$$

The detailed proof is skipped due to page limits. *c-PIDX(i, j)* is a good indication of actor $A_j's$ privacy exposure to A_i. Note that *c-PIDX(i, j)* might not equal *c-PIDX(j,i)* because A_i and A_j may have different privacy settings.

In the remaining of the paper, we use *c-PIDX(i, j)* for privacy measurement in social networks and assign *c-PIDX(i,j)* as the default value of *PIDX(i,j)*.

IV. PRIVACY MEASUREMENT FOR SOCIAL NETWORKS USING SONET MODEL

An online social network may attract millions of users. These users are connected together through ties with friend. Each user can be described by user profile, privacy settings, and a friend list. A profile consists of personal data of a user. Privacy settings describe how users want to distribute their personal data. A friend list includes a group of people who are connected together. The friend list can be further categorized as different groups, such as friends, friends of friends, or public, etc. Privacy settings can thus be defined according to the groups.

A. *SONET Model for Social Networks*

SONET model provides an effective and practical way to model privacy in social networks. Figure 1 shows a mapping between a social network and the SONET model. Only two users are demonstrated in the figure.

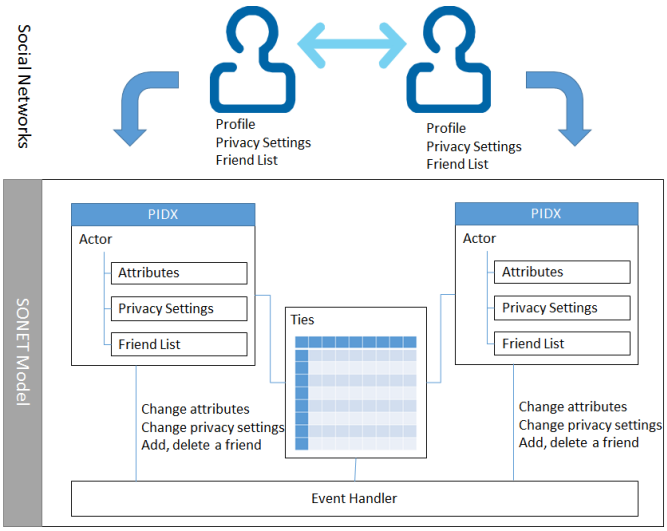

Figure 1 SONET Model for Social Networks

In SONET model, users are represented by actors. Profile is described by attribute list. The attributes are further extended with hidden information and virtual attributes. Friend list is described by a degrees of separation function $h(A_i, A_j)$. Privacy settings are described by an attribute visibility function $g_j(a_{jt}, d)$ which returns a numeric value to indicate if a particular attribute is visible to another actor. PIDX can thus be evaluated to reflect a user's privacy exposure. Further, SONET model also supports events such as attribute value changes, privacy settings update, and friend addition/deletion. SONET model can be used to simulate privacy changes in a social network and assess privacy impact in case user accounts are compromised.

B. *OSNPIDX: Social Network Privacy Simulation Tool*

OSNPIDX is an implementation of the SONET model for social networks. The tool is developed using Visual C++ 2010 in Windows 7. Each actor in OSNPIDX tool is described by 20 attributes and each attribute is assigned a privacy impact factor as shown in Table 2. Default attributes and values come from a survey we did in the past. Attributes and their privacy impact factors can be customized and changed in a configuration file.

Actors are divided into three groups according to their privacy preferences. User groups are selected according to the work in [10]:

- Privacy Fundamentalist (PF): extremely concerned users unwilling to share data.
- Pragmatic Majority (PM): concerned but willing to share information with privacy control.
- Marginally Concerned (MC): willing to provide any data.

Each user group is also associated with some default privacy settings as shown in Table 2.

The simulation starts with creating a social graph which consists of n actors. The tool provides two options to create a social graph, based on random graph or based on power-law graph. Studies have confirmed that social networks have the properties of power-law distribution [11]. The default setting is thus power-law distribution. Actors are then created and randomly assigned to a user group, i.e., PF, PM, or MC. After actors are created, the tool will also characterize each actor's friend list according to their ties in the social graph. During simulation, users have choices to select an actor and change its profiles, privacy settings, and friend list. *PIDX(i, j)* can be calculated based on an actor's attribute list, sensitivity and visibility as shown in Section III.C.

	PIF	Privacy Fundamentalist			Pragmatic Majority			Marginally Concerned		
	.05	F	FF	P	F	FF	P	F	FF	P
Full Name	.15	1	1	1	1	1	1	1	1	1
Education	.15	1	1		1	1	1	1	1	1
Marital Status	.25				1			1	1	1
Family Members	.25				1			1	1	1
Gender	.25	1	1	1	1	1	1	1	1	1
City	.45	1	1		1	1	1	1	1	1
State	.45	1	1		1	1	1	1	1	1
Father's Name	.45				1			1	1	
Mother's Name	.45				1			1	1	
Photos	.45				1	1		1	1	1
Friend List	.60	1			1	1		1	1	1
Email	.65				1	1		1	1	1
Hometown	.65				1	1		1	1	1
Places Visited	.65							1	1	1
Date of Birth	.65							1	1	1
Personal Phone Number	.70							1	1	1
Current Location	.80							1	1	1
Mother's Maiden Name	.80							1		
Biometric Details	.90									
SSN	.90									

Table 2 Group Privacy Settings

C. *Experiments and Evaluation*

Let A be an actor in the social network. We use the default attributes and privacy impact factors to describe A's profile (Table 2). To simplify our discussion, we only consider actors with degrees less than 3, i.e., friends (F), friends of friends (FF), and public (P), and we use the default privacy settings for each group as shown in Table 2. We assume A has all the information available in the profile except biometric details and SSN. We evaluate A's privacy exposure to A's friends, friends of friends, and public.

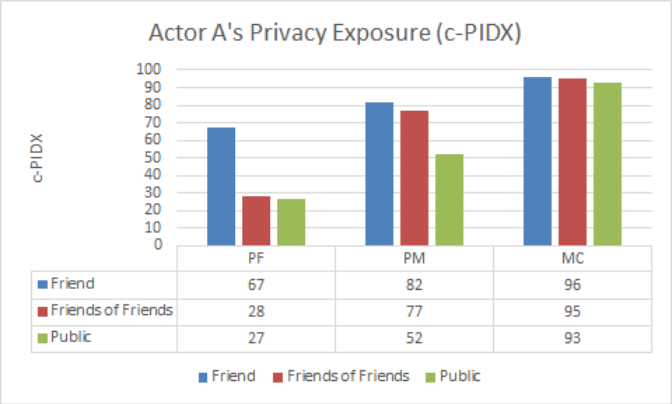

Figure 2 Privacy Exposure in Social Networks

Figure 2 shows *A*'s privacy exposure to friends, friends of friends, and public when *A* selects different privacy settings. The data in the figure matches with the privacy settings defined in the user groups. Privacy fundamentalist users are unwilling to share data. Their privacy exposure to friends of friends drops from 67 to 28, a 56.6% decrease than the friends. On the contrary, marginally concerned users do not have any privacy settings and thus their privacy exposure to friends of friends (*c-PIDX*=95), and public (*c-PIDX* =93) are very high. Many online social networks' privacy settings are very confusing and lack of numerical value to alarm users their privacy exposure to others. PIDX provides a practical and effective way to tell the difference.

Further, in case user *A*'s account is compromised, a malicious user could steal *A*'s profile. The privacy exposure of user *A* to a hacker is as high as 96. Users must be cautious when putting their personal data in social networking sites.

D. Comparison with Other Privacy Measurement Approaches

The proposed model differs significantly from the privacy scores in [2][3]. Authors in [2] have shown a simple approach to compute privacy score as a function of sensitivity and visibility. However, no datasets are used or shown to measure effectiveness of their proposed model. The proposed approach in [3] is based on IRT. However, IRT is not designed for a complex behavioural network like social networks. IRT has three basic assumptions: items are independent, users are independent, and items and users are independent. These assumptions do not apply to social networks. Further, the work in [2][3] assumes attributes are independent and does not consider relationships between attributes. However, as found in [2][12], revelation of a combination of a few attributes can jeopardize the privacy of a user since it can lead to easy access of other attributes.

PrivAware [1] and Privometer [4] are both based on inference model. PrivAware calculated privacy score based on the total number of attributes visible to the third party applications to the total number of attributes of a user. It does not consider sensitivity of an attribute. SONET model considers both visibility and sensitivity of attributes. It also introduces virtual groups to consider attribute group properties. SONET model provides better modeling for social networks through hidden attributes, virtual attributes, and privacy settings.

V. Conclusions and Future Works

Privacy measurement is a challenging issue in social networks. In this paper, we extend SONET model to support privacy measurement in social networks. We propose to use *PIDX(i, j)* to measure actor A_j's privacy exposure to A_i. We consider both sensitivity and visibility of attributes for privacy measurement. *PIDX(i, j)* can be used to evaluate privacy exposure between any two actors in a social network. SONET model provides a practical and effective way for online users and social networking sites to measure privacy.

SONET model currently uses a static model to assign privacy impact factor. The value of the privacy impact factors comes from pilot survey data we did in the past. Privacy impact factor is important for privacy measurement. We will consider a dynamic model next to better characterize privacy impact factors. Further, we are also considering extending OSNPIDX tool to use real data from online social networking sites.

References

[1] J. Becker, "Measuring Privacy Risk in Online Social Networks," *Design*, vol. 2, p. 8, 2009.

[2] E. M. Maximilien, T. Grandison, T. Sun, D. Richardson, S. Guo, and K. Liu, "Privacy-as-a-Service : Models , algorithms , and results on the facebook platform," in *Web 2.0 Security and privacy workshop*, 2009.

[3] K. U. N. Liu, "A Framework for Computing the Privacy Scores of Users in Online Social Networks," *Knowl. Discov. Data*, vol. 5, no. 1, pp. 1–30, 2010.

[4] N. Talukder, M. Ouzzani, A. K. Elmagarmid, H. Elmeleegy, and M. Yakout, *Privometer: Privacy protection in social networks*, vol. 1, no. 2. VLDB Endowment, 2010, pp. 141–150.

[5] E. M. Maximilien, T. Grandison, T. Sun, D. Richardson, S. Guo, and K. Liu, "Privacy-as-a-service: Models, algorithms, and results on the facebook platform," in *Proceedings of Web*, 2009, vol. 2.

[6] C. Akcora, B. Carminati, and E. Ferrari, "Privacy in Social Networks: How Risky is Your Social Graph?," in *2012 IEEE 28th International Conference on Data Engineering*, 2012, pp. 9–19.

[7] J. Bonneau and S. Priebusch, "The Privacy Jungle : On the Market for Data Protection in Social Networks," in *The Eighth Workshop on the Economics of Information Security*, 2009, pp. 1–45.

[8] R. N. Kumar and Y. Wang, "SONET: A SOcial NETwork Model for Privacy Monitoring and Ranking," in *The 2nd International Workshop on Network Forensics, Security and Privacy*, 2013.

[9] Y. Wang and R. N. Kumar, "Privacy Measurement for Social Network Actor Model," in *The 5th ASE/IEEE International Conference on Information Privacy, Security, Risk and Trust*, 2013.

[10] M. S. Ackerman, L. F. Cranor, and J. Reagle, "Privacy in e-commerce: examining user scenarios and privacy preferences," in *Proceedings of the 1st ACM conference on Electronic commerce*, 1999, vol. 99, no. 1998, pp. 1–8.

[11] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," *Proc. 7th ACM SIGCOMM Conf. Internet Meas. IMC 07*, vol. 40, no. 6, p. 29, 2007.

[12] L. Sweeney, "Uniqueness of simple demographics in the U. S. population," in *Data privacy Lab white paper series LIDAP-WP4*, 2000.