Privacy Measurement for Social Network Actor Model

Yong Wang
College of Business and Information Systems
Dakota State University
Madison, SD, 57103
yong.wang@dsu.edu

Raj Kumar Nepali
College of Business and Information Systems
Dakota State University
Madison, SD, 57103
rknepali@ pluto.dsu.edu

***Abstract*—Privacy measurement is critical to evaluate privacy risks for business, public policy and legislation. However, there is a lack of effective and practical way to quantify, measure, and evaluate privacy. In this paper, we propose three privacy indexes for privacy measurement (ranking) for social network actor model, i.e., weighted privacy index (*w-PIDX*), maximum privacy index (*m-PIDX*), and composite privacy index (*c-PIDX*). We also introduce a novel virtual attribute for social network actor model to describe the combined attributes' behavior. We further evaluate and demonstrate the effectiveness of these *PIDX*es for various user groups in different testing scenarios. Our tests and analysis show that composite privacy index, *c-PIDX*, is the best to measure privacy for social network actor model. A practical approach to evaluate *w-PIDX*, *m-PIDX*, and *c-PIDX* is also presented in the paper.**

Keywords: privacy measurement, privacy index, w-PIDX, m-PIDX, c-PIDX

I. INTRODUCTION

Many concerns have been raised regarding the privacy issues in social media. The risks, as well as the security and privacy issues of social media in business, public policy, and legislation need to be evaluated and studied. However, there is lack of effective and practical way to quantify, measure, and evaluate privacy. This paper focuses on privacy measurement issues in social media. We consider all public accessible information as social media, such as online social networks, public records of law court proceedings, records of births, marriages, etc.

Recent incidents indicate that risks, security, and privacy issues of social media to business, public policy, and legislation need to be further evaluated and studied. For example, between April 17 and April 19, 2011, user account information for the Sony PlayStation Network and its Qriocity service was compromised [1]. Sony suffered a massive breach in its video game online network that led to the theft of names, addresses and possibly credit card data belonging to 77 million user accounts. It is one of the largest-ever Internet security breaches. Sony has estimated a total cost of $172 million for the PlayStation Network breach.

In such incidents, it is very important for entrepreneurs, administrators, public policy makers, and legislature to evaluate the impact of these incidents to general public, such as, how many people were affected, and how much of their privacy would be disclosed. However, how to quantify, measure, and evaluate privacy is very challenging. First, the definition of privacy is subjective. People may have different opinions about privacy. Second, privacy can be identified by certain disclosed attributes, such as, name, credit card number, SSN, biometric details, etc. One attribute may have different impact to individuals. Third, hidden information exists and could be further inferred from known attributes. This makes privacy measurement more challenging. Fourth, certain attributes could be combined and used to reveal more personal information. The impact of combined attributes' behavior needs to be further studied.

We presented a social network model, SONET, for social media privacy monitoring and ranking in [2]. SONET includes three models, actor model, community model, and social network model. We proposed to use privacy index (*PIDX*) for privacy measurement. The work in [2] focuses on the social network modeling and does not include privacy measurement functions to evaluate *PIDX*. In this paper, we propose three privacy measurement functions, i.e., weighted privacy measurement function, maximum privacy measurement function, and composite privacy measurement function for privacy measurement. Three *PIDX*es, weighted privacy index (*w-PIDX*), maximum privacy index (*m-PIDX*), and composite privacy index (*c-PIDX*) can be further built on these proposed privacy measurement functions. We also introduce a novel virtual attribute for actor model to describe combined attributes' behavior.

Our main contributions in the paper include:

- We propose three privacy measurement functions and three privacy indexes for privacy measurement.
- We introduce a novel virtual attribute concept for social network actor model to describe combined attributes' behavior.
- We further test and demonstrate the effectiveness of these three *PIDX*es.

Our tests and analysis show that the proposed composite *PIDX* (*c-PIDX*) is the best to measure privacy for social network actor model. Few works have been conducted in the literature and have limitations when used to measure privacy. There is significant difference between our proposed approach and the existing work. Our proposed *PIDX*es provides an effective and practical way to measure privacy and thus is very useful to evaluate privacy risks for business, public policy, and legislation.

The paper is organized as follows: Section II discusses the related work. Section III introduces the actor model and our proposed privacy measurement functions and privacy

978-0-7695-5137-1/13 $26.00 © 2013 IEEE
DOI 10.1109/SocialCom.2013.99

IEEE computer society

indexes, followed by testing and results in Section IV, comparison and analysis in Section V. Section VI summarizes the paper and future works.

II. Related Work

Social media privacy has raised many concerns and may affect individuals, enterprises, legislatures, and government agencies. Different types of attacks have been investigated in [3][4][5][6][7][8]. Previous works on social media privacy focus on privacy preserving [9] [10] [11] [12][13] and privacy policy conflicts [14][15]. Few works [16][17][18] [19][20][21] have been conducted on privacy measurement due to the challenges to quantify the privacy risk associated with online social network users.

Recent works attempting to quantify the privacy risks associated with the usage of online social networks can be found in [16][17][18] [19][20]. In [16], the authors propose to use the amount of information revealed in online social networks to quantify the privacy risks. However, there is no measurement functions developed in [16]. In [17], 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 [17]. The authors extend their works to another approach in [18]. Item Response Theory (IRT) is used to evaluate sensitivity and visibility of attributes when evaluating privacy scores [18]. The authors in [19] develop a tool, Privometer, to measure information leakage. The leakage is measured by a numerical value derived from combined probability of inference. The tool can suggest self-sanitization actions based on the numerical value. The work in [20] proposes to measure the privacy risk based on social networks' privacy policy and practices when handling user [20]. Privacy scores are calculated in a debatable manner. In [21], the authors use risk labeling approach to tag users based on the community members' feedback. Active learning method is used to correctly label strangers.

III. Privacy Indexes for Actor Model

Actor model is first introduced in [2]. In this section, we briefly introduce actor model. Then, we introduce virtual attributes, privacy measurement functions, and privacy indexes.

1. Actor Model

Definition *An actor is a social entity (e.g. people, organization, etc.) in a social network.*

Definition *Actor has certain characteristics that describe its features known as attributes.*

Definition *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.*

An actor has attributes like name, address, social security number (SSN), phone number, education, marital status, etc. In general, SSN has higher impact on privacy than phone number. Let A_i be an actor and $L_i = \{a_{i1}, a_{i2}, \dots, a_{in}\}$ represent its attributes. Then, $A_i(a_{i1}, a_{i2}, \dots, a_{in})$ is a representation of an actor with attributes.

We consider privacy impact factor for full privacy disclosure is 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. We use $S_i = \{s_{i1}, s_{i2}, \dots, s_{in}\}$ to represent attributes' privacy impact factors.

There are certain attributes which could be inferred from other attributes. For example, if an actor's occupation is known, the actor's salary information can be inferred. Such information is known as hidden information.

Definition *Hidden information is indirect information which is not available firsthand but could be inferred from existing data.*

We use (a_{ij}, a_{ik}, p_{jk}) to further indicate hidden relationship $a_{ij} \xrightarrow{p_{jk}} a_{ik}$ between two attributes. (a_{ij}, a_{ik}, p_{jk}) indicates that a_{ik} can be inferred from a_{ij} in probability p_{jk}.

2. Virtual Attributes

Each attribute has an impact on privacy. Certain attributes could be combined and used to further disclose privacy. For example, the work in [22] shows that 87% of Americans can be uniquely identified by five digit zip code, gender, and date of birth. However, none of them alone can significantly affect privacy [22][23]. We propose to use virtual attributes to describe attribute group behavior.

Definition *A virtual attribute describes a group of attributes behavior and their impact on privacy. A virtual attribute may have significant impact on the privacy of an actor. However, none of the attributes alone can significantly affect privacy.*

Let v_{ij} be a virtual attribute and v_{ij} is decided by a group of attributes R_{ij} ($R_{ij} \subset L$). We use (R_{ij}, v_{ij}, p_{ij}) to indicate virtual attribute relationship $R_{ij} \xrightarrow{p_{ij}} v_{ij}$. $R_{ij} \xrightarrow{p_{ij}} v_{ij}$ indicates that v_{ij} can be inferred from R_{ij} in probability p_{ij}. Assume we have k virtual attributes. Then, we have

$$\begin{cases} R_{i1} \xrightarrow{p_{i1}} v_{i1} \\ \dots \\ R_{ik} \xrightarrow{p_{ik}} v_{ik} \end{cases}$$

where $R_{ij} \subset L$ $(1 \le j \le k)$.

3. Privacy Index

Among all the attributes $L = \{a_1, a_2, \dots, a_n\}$ in an actor model, there is a subset of attributes which might be known. If certain attributes are known, privacy might be disclosed. Let $L_k = \{a'_1, a'_2, \dots, a'_m\}$ be a known attribute list ($L_k \subseteq L$) and $S_k = \{s'_1, s'_2, \dots, s'_m\}$ be the corresponding privacy impact factors of these attributes. Let f be a privacy

measurement function which returns a numeric value on L_k. Let

$$w_{L_k} = f(L_k, S_k)$$

be the total privacy impact of L_k. **Privacy Index** can be used to measure privacy exposure.

Definition *Privacy Index (PIDX) is used to describe an entity's privacy exposure factor based on the known attributes. Higher PIDX indicates higher exposure of privacy. Privacy Index PIDX is between 0 and 100.*

$$PIDX = \frac{w_{L_k}}{w_L} \times 100$$

Let T represent a threshold which is critical to an actor's privacy. We define **privacy invasion** as

$$PIDX \geq T$$

If $PIDX$ is lower than T, privacy is considered to be preserved.

a. Senstivity

Each attribute is given a privacy impact factor. The privacy impact factor reflects the sensitivity of the attribute. Large number indicates more sensitive information. Attributes' sensitivity is represented by $S_i = \{s_{i1}, s_{i2}, \dots, s_{in}\}$.

b. Visibility

The actor model also uses probability to describe hidden information and virtual attributes. The probability is a reflection of information visibility. The visibility can be represented using $V_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}$ $(0 \leq p_{ij} \leq 1)$. 0 indicates an unknown attribute and 1 indicates a known attributes. A value between 0 and 1 indicates partial disclosure of an attribute.

Privacy index provides a practical way to measure and evaluate privacy. We consider both sensitivity and visibility of attributes in the model and privacy index can be calculated accordingly.

4. Privacy Measurement Function

Let (L, S) represent an actor model's complete attribute list, and the attributes' privacy impact factors.

$$\begin{cases} L = \{a_1, a_2, \dots, a_n\} \\ S = \{s_1, s_2, \dots, s_n\} \end{cases}$$

Let (L_k, S_k, V_k) represent the known attribute list, and their sensitivities, and visibilities.

$$\begin{cases} L_k = \{a'_1, a'_2, \dots, a'_m\} \\ S_k = \{s'_1, s'_2, \dots, s'_m\} \\ V_k = \{p'_1, p'_2, \dots, p'_m\} \end{cases}$$

Without losing generality, we use $V = \{p_1, p_2, \dots, p_n\}$ $p_i = 1$ $(1 \leq i \leq n)$ to represent $L's$ visibilities.

Privacy can be evaluated based on three measurement metrics, i.e., known attribute list, attribute sensitivities, and attribute visibilities. Thus, a privacy measurement function can be evaluated based on these three inputs. We use $f(L, S, V)$ to represent a privacy measurement function.

In this paper, we propose three privacy measurement functions, i.e., weighted privacy measurement function f_w, maximum privacy measurement function f_m, and composite privacy measurement function f_c. Using these three functions, we present three privacy indexes correspondingly, i.e., weighted privacy index (***w-PIDX***), maximum privacy index (***m-PIDX***), composite privacy index (***c-PIDX***). These privacy measurement functions and privacy indexes are discussed below.

a. Weighted Privacy Measurement Function and Weighted Privacy Index (w-PIDX)

Weighted privacy measurement function is defined as

$$f_w(L_k, S_k, V_k) = p'_1 s'_1 + p'_2 s'_2 + \cdots + p'_m s'_m = \sum_{j=1}^{m} p'_j s'_j$$

w-PIDX is an index which measures an entity's privacy based on known attribute list weight. *w-PIDX* is defined as

$$\begin{aligned} w - PIDX &= \frac{w_{L_k}}{w_L} \times 100 \\ &= \frac{f_w(L_k, S_k, V_k)}{f_w(L, S, V)} \times 100 \\ &= \frac{\sum_{j=1}^{m} p'_j s'_j}{\sum_{j=1}^{n} s_j} \times 100 \end{aligned}$$

b. Maximum Privacy Measuement Function and Maximum Privacy Index (m-PIDX)

Maximum privacy measurement function is defined as

$$f_m(L_k, S_k, V_k) = max(p'_1 s'_1, p'_2 s'_2, \dots, p'_m s'_m)$$

m-PIDX is an index which measures an entity's privacy based on the maximum attribute impact factor of all the known attributes.

$$\begin{aligned} m - PIDX &= f_m(L_k, S_k, V_k) \times 100 \\ &= max\,(p'_1 s'_1, p'_2 s'_2, \dots, p'_m s'_m) \times 100 \end{aligned}$$

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

c. Composite Privacy Measurement Function and Composite Privacy Index (c-PIDX)

Composite privacy measurement function is defined as

$$f_c(L_k, S_k, V_k) = f_m(L_k, S_k, V_k) + (1 - f_m(L_k, S_k, V_k)) \times \frac{f_w(L_k, S_k, V_k)}{f_w(L, S, V)})$$

c-PIDX is an index which measures an entity's privacy based on composite privacy measurement factors. *c-PIDX* is defined as

$$c - PIDX = f_c(L_k, S_k, V_k) \times 100$$

$c - PIDX$ can be represented using *m-PIDX* and *w-PIDX*.

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

The equation above can be verified easily and the detailed proof is skipped in the paper.

5. Actor Model Privacy Measurement

We suggest a three-step procedure to calculate privacy indexes for actor model as shown in Figure 1.

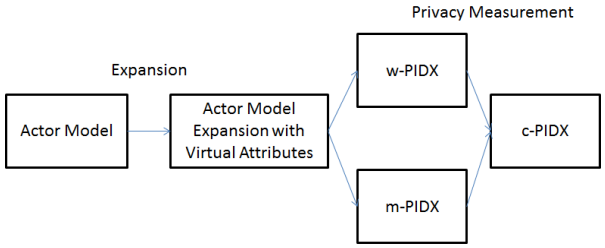

Figure 1. Actor Model Privacy Measurement

Let (L, S, V) represent an actor model's attribute list, sensitivities, and visibilities. First, we expand (L, S, V) to (L_E, S_E, V_E) to include virtual attributes. Assume we have k virtual attributes, we have

$$L_E = L \cup \{v_1, \dots, v_k\}$$

Without losing generality, we assume in L_E, a_1 to a_n are attributes, and a_{n+1} to a_{n+k} are virtual attributes. The same expansion will be applied to S and V too. A virtual attribute a_{n+t} has privacy impact factor s_{n+t} and default visibility $p_{n+t} = 0$. Its visibility is further decided by $R_t \xrightarrow{p_t} v_t$. We assign $p_{n+t} = p_t$ if all the attributes in R_t is known. For each attribute a_i $(1 \le i \le n)$, we assume $p_i = 1$ if a_i is known and $p_i = 0$ if a_i is unknown. For hidden information $a_i \xrightarrow{p_{ij}} a_j$, we will assign $p_j = p_{ij}$ only if a_i is known and p_{ij} is greater than a_j's current visibility. After expansion with virtual attributes and hidden information, we have Table 1.

	Attributes				Virtual Attributes		
L_E	a_1	a_2	...	a_n	a_{n+1}	...	a_{n+k}
S_E	s_1	s_2	...	s_n	s_{n+1}	...	s_{n+k}
V_E	p_1	p_2	...	p_n	p_{n+1}	...	p_{n+k}

Table 1. Actor Model Expansion

w-PIDX, *m-PIDX*, and *c-PIDX* can be calculated as

$$w - PIDX = \frac{\sum_{j=1}^{n+k} p_j s_j}{\sum_{j=1}^{n+k} s_j} \times 100$$

$$m - PIDX = \max(p_1 s_1, p_2 s_2, \dots, p_{n+k} a_{n+k}) \times 100$$

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

IV. PIDX Testing and Results

In this section, we describe our testing and results.

1. Attribute Selection

Our testing starts with identifying privacy attributes. We extracted preliminary attributes from social networking sites' personal profile and privacy settings. We then developed a survey to collect user's rating of the privacy impact of each attribute. After the survey, we selected 20 attributes for our testing. These 20 attributes are shown in Table 2. Biometric Details include information such as height, weight etc.

2. Attribute Privacy Impact Factor Assignation

According to the survey, we assign a privacy impact factor for each attribute. The attribute and its privacy impact factor are listed in Table 2. Each attribute is also identified by an ID number.

1	Full Name	0.05	11	Friend List	0.60
2	Education	0.15	12	Email	0.65
3	Marital Status	0.15	13	Hometown	0.65
4	Family Members	0.25	14	Places Visited	0.65
5	Gender	0.25	15	Date of Birth	0.65
6	City	0.45	16	Personal Phone Number	0.70
7	State	0.45	17	Current Location	0.80
8	Father's Name	0.45	18	Mother's Maiden Name	0.80
9	Mother's Name	0.45	19	Biometric Details	0.90
10	Photos	0.55	20	SSN	0.90

Table 2. Privacy Impact Factor

Attributes selection and privacy impact factor assignation are important for privacy measurement. However, these two issues are out of the scope of this paper. In this paper, we use the static model to demonstrate the privacy measurement functions and privacy indexes.

3. Privacy Measurement for Attribute Incremental Changes

We first evaluate the three *PIDX*es when known attributes change incrementally. We use T_i to represent a testing case where i is an integer between 1 and 20. T_i represents all the attributes between 1 and i are known. For example, T_3 indicates that attributes 1 to 3, full name, education, and marital status, are known. We evaluate the three privacy indexes for 20 testing cases from T_1 to T_{20}. In these testing cases, no hidden information and virtual attributes are considered. Figure 2 shows the testing results.

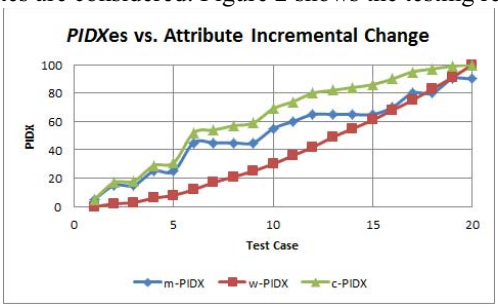

Figure 2. *w-PIDX* vs. *m-PIDX* vs. *c-PIDX*

Figure 2 shows that *w-PIDX* is good at reflecting attribute incremental changes. However, *w-PIDX* does not reflect the actual privacy exposure. *m-PIDX* can be used for privacy ranking. However, *m-PIDX* does not reflect the attribute incremental changes. For example, from T_{12} to T_{15}, *m-PIDX* is the same. *c-PIDX* has both the advantage of *w-PIDX* and *m-PIDX*.

4. User Groups

We further tested the three privacy indexes for different user groups. User groups are selected according to the work in [24]. The authors in [24] divided users into three categories according to their privacy preferences:

- 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

We further divide Privacy Fundamentalist (PF) group into two sub groups and Pragmatic Majority (PM) group into four sub groups. In each group or sub group, users have different privacy preferences and certain attributes are known. User group settings are shown in Table 3.

	Attributes	PIF	PF I	PF II	PM I	PM II	PMIII	PM IV	MC
1	Full Name	0.05	x	x	X	x	x	X	x
2	Education	0.15	x	x	X	x	x	X	x
3	Marital Status	0.15				x	x	X	x
4	Family Members	0.25				x	x	X	x
5	Gender	0.25	x	x	X	x	x	X	x
6	City	0.45		x			x	X	x
7	State	0.45		x			x	x	x
8	Father's Name	0.45						x	x
9	Mother's Name	0.45						x	x
10	Photos	0.55			X	x	x	x	x
11	Friend List	0.60			X	x	x	x	x
12	Email	0.65				x	x	x	x
13	Hometown	0.65					x	x	x
16	Places Visited	0.65							x
18	Date of Birth	0.65							x
14	Personal Phone Number	0.70							x
15	Current Location	0.80							x
17	Mother's Maiden Name	0.80							
19	Biometric Details	0.90							
20	SSN	0.90							

Table 3. User Groups

Three privacy indexes are evaluated for all these user groups and the results are shown in Figure 3.

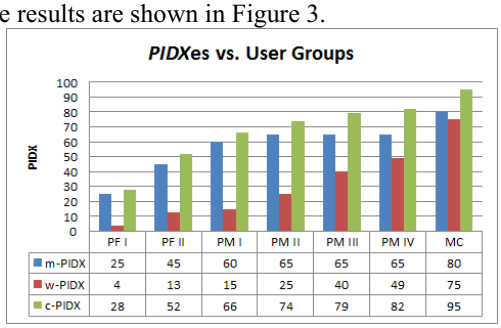

Figure 3. *PIDX*es for Different User Groups

Figure 3 shows that *c-PIDX* is good for privacy ranking for different user groups. *m-PIDX* does not differentiate privacy risks among PM II, PM III, and PM IV.

5. Hidden Information and *PIDX*es

Hidden information may affect *PIDX*es too. For example, from a person's education, it is likely to guess the person's hometown. Let

$$a_2: Education \xrightarrow{p=0.9} a_{13}: Hometown$$

In consideration of hidden information, we evaluate its impact to *PIDX*es for Privacy Fundamentalist Group I and Group II. The results are shown in Figure 4.

In Figure 4, *c-PIDX* of PF I changes from 28 to 63, a 125% change on the *c-PIDX*. It shows that privacy might be disclosed even for users who are unwilling to share data. Hidden information is desirable to measure the real privacy risks for a user.

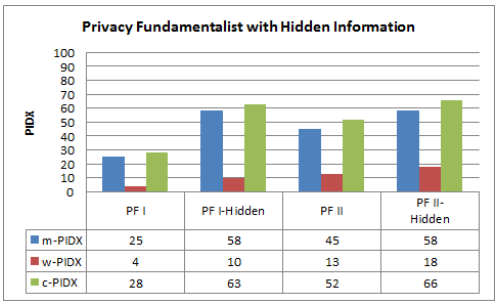

Figure 4. *PIDX*es and Hidden Information

6. Virtual Attributes

Virtual attribute is another factor which may affect *PIDX*es. Let v_1 be a virtual attribute and it can be decided by gender, city, state, and date of birth in probability p_1.

$$R_1(\text{Gender, city, state, Date of Birth}) \xrightarrow{p_1} v_1$$

v_1 is assigned a privacy impact factor 0.9 since it might be used to identify a person. In scenario R, we assume gender, city, state, date of birth is known, we evaluate *PIDX*es without considering virtual attribute. We further evaluate R with virtual attribute v_1 in two scenarios, $p_1 = 0.87$ and $p_1 = 1$. The results are shown in Figure 5.

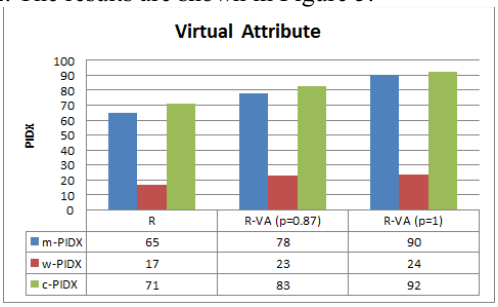

Figure 5. *PIDX*es and Virtual Attributes

Figure 5 shows that, in consideration of virtual attributes, *c-PIDX* changes from 71 to 83 ($p_1 = 0.87$), a 16% increase, and to 92 ($p_1 = 1$), a 30% increase. Thus, virtual attribute is critical to measure privacy risks.

V. Comparison and Analysis

The proposed model differs significantly from the privacy scores in [17][18]. The proposed approach in [18] is based on IRT. However, IRT is not designed for a complex behavioral network like social networks. IRT has three basic assumptions: items are independent, users are independent, and items and users are independent. However, these assumptions do not apply to social networks. Further, the work in [18] assumes attributes are independent and does not consider relationships between attributes. However, as found in [22][23], 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.

Our model does not have these limitations and it considers all the relationships between actors and attributes.

The sensitivity and visibility of attributes are further characterized by the possibilities in attribute to attribute relationships. In this paper, we propose three index for privacy measurement.

w-PIDX is a simple way to calculate privacy index. It is good to measure attribute incremental changes. However, *w-PIDX* is not good for privacy ranking. For example, let v_1, v_2 $(v_1 > v_2)$ be two *w-PIDX*es, v_1 does not indicate it has more privacy disclosure than v_2 although $v_1 > v_2$.

m-PIDX is another way to calculate privacy index. *m-PIDX* is good to measure privacy relative value. For example, let v_1, v_2 $(v_1 > v_2)$ be two *m-PIDX*es, it is safe to assume that v_1 indicates more privacy disclosure than v_2. However, *m-PIDX* is not good at measuring privacy increment change.

c-PIDX is a composite measurement for privacy. It combines the advantages of both weighted privacy index and maximum privacy index. *c-PIDX* can not only be used to measure privacy increment change, but also be used to compare privacy relative value.

In summary, *c-PIDX* is the best to measure privacy for social network actor model.

VI. Conclusion and Future Works

Many concerns have been raised regarding the security and privacy issues of social media. The risks, as well as the security and privacy issues of social media in business, public policy, and legislation need to be evaluated and studied. In this paper, we propose three privacy indexes for social network actor model, i.e., weighted privacy index (*w-PIDX*), maximum privacy index (*m-PIDX*), and composite privacy index (*c-PIDX*). We also introduce a novel virtual attribute to describe combined attributes' behavior. We further evaluate and demonstrate the effectiveness of these *PIDX*es for various user groups in different testing scenarios. Our tests and analysis show that composite privacy index, *c-PIDX*, is the best to measure privacy for social network actor model. Hidden information and virtual attributes are also critical to measure privacy. A practical approach to evaluate *w-PIDX*, *m-PIDX*, and *c-PIDX* is also presented in the paper. Our future work includes evaluation of these three *PIDX*es for social network community model.

References

[1] C. Morris, SONY: PlayStation Breach Involves 70 Million Subscribers, April 26, 2011, CNBC.com.

[2] Raj Kumar Nepali and Yong Wang, SONET: A Social Network Model for Privacy Monitoring and Ranking, The 2nd International Workshop on Network Forensics, Security and Privacy, July 08, 2013.

[3] J. Tang, T. Lou, and J. Kleinberg, "Inferring social ties across heterogeneous networks," in WSDM'12, 2012, pp. 743–752.

[4] E. Zheleva and L. Getoor, "To join or not to join?: The illusion of privacy in social networks with mixed public and private user profiles," in WWW 2009, 2009, pp. 531–540.

[5] G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, "A Practical Attack to De-anonymize Social Network Users," 2010 IEEE Symposium on Security and Privacy, pp. 223–238, 2010.

[6] A. Narayanan and V. Shmatikov, "De-anonymizing Social Networks," in 2009 30th IEEE Symposium on Security and Privacy, 2009, pp. 173–187.

[7] B. Zhou and J. Pei, "Preserving Privacy in Social Networks Against Neighborhood Attacks," in *2008 IEEE 24th International Conference on Data Engineering*, 2008, pp. 506–515.

[8] G. Brown, T. Howe, M. Ihbe, A. Prakash, K. Borders, and A. Arbor, "Social Networks and Context-Aware Spam," in *CSCW'08*, 2008, pp. 403–412.

[9] L. Sweeney, "K-anonymity: a model for protecting privacy," International Journal on uncertainity, Fuzziness and knowledge-based system, vol. 10, no. 5, pp. 557–570, 2002.

[10] A. Machanavajjhala, J. Gehrke, D. Kifler, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," in Proceedings of the 22nd IEEE International Conference on Data Engineering, 2006.

[11] F. Buccafurri, G. Lax, and V. Graziella, "Privacy-Preserving Resource Evaluation in Social Networks," in *Tenth Annual International Conference on Privacy, Security and Trust*, 2012, pp. 51–58.

[12] J. Anderson, C. Diaz, F. Stajano, K. U. Leuven, and J. Bonneau, "Privacy-Enabling Social Networking over untrusted networks," in WONS, 2009, pp. 2–7.

[13] D. Starin, R. Baden, A. Bender, N. Spring, and B. Bhattacharjee, "Persona?: An Online Social Network with User-Defined Privacy Categories and Subject Descriptors," in SIGCOMM'09, 2009, pp. 135–146.

[14] A. Yamada, T. H. Kim, and A. Perrig, "Exploiting Privacy Policy Conflicts in Online Social Exploiting privacy policy conflicts in online social networks," 2012.

[15] Y. Liu, K. P. Gummadi, and A. Mislove, "Analyzing Facebook Privacy Settings : User Expectations vs . Reality," in *IMC' 11*, 2011.

[16] J. Becker and H. Chen, "Measuring Privacy Risk in Online Social Networks," in Web 2.0 security and privacy Workshop, 2009.

[17] 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.

[18] 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.

[19] N. Talukder, M. Ouzzani, A. K. Elmagarmid, H. Elmeleegy, and M. Yakout, "Privometer: Privacy protection in social networks," 2010 IEEE 26th International Conference on Data Engineering Workshops (ICDEW 2010), pp. 266–269, 2010.

[20] 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.

[21] 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.

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

[23] Golle, Philippe. "Revisiting the uniqueness of simple demographics in the US population." *Proceedings of the 5th ACM workshop on Privacy in electronic society*. ACM, 2006.

[24] Ackerman, M. S., Cranor, L. F., & Reagle, J. (1999, November). Privacy in e-commerce: examining user scenarios and privacy preferences. In Proceedings of the 1st ACM conference on Electronic commerce (pp. 1-8). ACM.