Proceedings of the 51^{st} Hawaii International Conference on System Sciences | 2018

Your Privacy Is Your Friend's Privacy: Examining Interdependent Information Disclosure on Online Social Networks

Yaqoub Alsarkal
The George Washington University
alsarkal@gwmail.gwu.edu

Nan Zhang
Penn State University
nan@ist.psu.edu

Heng Xu
Penn State University
hxu@ist.psu.edu

Abstract

The highly interactive nature of interpersonal communication on online social networks (OSNs) impels us to think about privacy as a communal matter, with users' private information being revealed by not only their own voluntary disclosures, but also the activities of their social ties. The current privacy literature has identified two types of information disclosures in OSNs: self-disclosure, i.e., the disclosure of an OSN user's private information by him/herself; and co-disclosure, i.e., the disclosure of the user's private information by other users. Although co-disclosure has been increasingly identified as a new source of privacy threat inherent to the OSN context, few systematic attempts have been made to provide a framework for understanding the commonalities and distinctions between self- vs. co-disclosure, especially pertaining to different types of private information. To address this gap, this paper presents a data-driven study that builds upon an innovative measurement for quantifying the extent to which others' co-disclosure could lead to actual privacy harm. The results demonstrate the significant harm caused by co-disclosure and illustrate the differences between the identity elements revealed through self- and co-disclosure.

1. Introduction

Online Social Networks (OSNs) have increasingly facilitated users' voluntary information disclosures that may not only reveal their own identities but also their social ties' identities (e.g., tagging a friend in a status update or in a place checked-in). Such highly interactive nature of interpersonal communication and data exchange impels us to think about privacy as a communal matter. As a result, there is a growing recognition of grounding the investigation of privacy as an *interdependent* phenomenon (Biczók and Chia 2013; Jia and Xu 2015; 2016): While individuals are free to decide what personal information they disclose, they often cannot control what others disclose about them, or how others may use the private information that they disclose. Likewise, people may share information that involves others in ways that violate others' privacy preferences. In recent privacy literature, such *interdependent* nature of information disclosure has been implicitly assumed as a new source of privacy threat inherent to the OSN context.

However, there is almost no research that has empirically quantified the extent to which others' disclosure of information about an individual could lead to actual privacy harm. As a first step to address this gap in existing literature, we develop an innovative measurement to capture an individual's actual privacy loss caused by self-disclosure *vs.* by co-disclosure (i.e., disclosure of one's private information by other users of the OSN). To achieve this research objective, we considered Twitter as a case study. Specifically, we selected a sample of Twitter users located in the US, retrieved all their tweets along with the tweets of their followers, identified the identity elements of a user that can be inferred from his/her own tweets (self-disclosure) or the tweets of his/her followers (co-disclosure), and then proposed an information-theoretic measure that captures the additional amount of privacy loss caused by co-disclosure (on top of information that is already self-disclosed).

The current study contributes to existing privacy research in several important ways. First, prior privacy research has heavily relied on collecting self-reported data to measure privacy related intentions or beliefs as a proxy to study outcome variables. Given that users often behave not necessarily in rational ways when it concerns their privacy behaviors, Bélanger & Xu (2015) suggested IS researchers to "focus on actual behaviors as opposed to merely individual intentions to behave when studying information privacy." The current research addressed this methodological challenge by quantifying possible negative consequences that could result from users' real information disclosure behaviors. Second, positivist approaches have dominated in the IS field where theory-based work has defined, predicted and explained privacy-related constructs in various nomological models (Smith et al. 2011). Although the theory-based work has helped us build stronger theoretical foundation and methodological rigor, "some of us are increasingly raising doubts about

URI: http://hdl.handle.net/10125/49998
ISBN: 978-0-9981331-1-9

whether we can relate what we have found in our research to what practitioners or policymakers truly experience in reality" (Bélanger & Xu 2015). This research aims to extend the literature by pursuing a data-driven approach to measure real information privacy harms with observational Twitter data.

The rest of the paper is organized as follows. The next section presents a review of background and related literature. Following that, we define the research questions studied in this paper, and describe the two key components for measuring privacy loss: the identity linkage model used by an adversary and the information-theoretic metric of privacy disclosure. Equipped with the privacy-loss measurement, we present our research methodology, including the data collection process and the experimental design over Twitter. We then describe our research results and explain how these results address the research questions. We conclude the paper with discussions of limitations and potential extensions of our work.

2. Literature Review

2.1. Interpersonal Privacy Concerns on OSNs

Data privacy online and offline are of prime importance to individuals. By revealing their personal information, individuals make observers of their personal information co-owners of it. This introduces the problem known as interdependent privacy: users are no longer in control of their privacy as other users' actions and interactions can affect their privacy (Biczók and Chia 2013). Several research papers have enriched our understanding of interdependent privacy (Biczók and Chia 2013; Choi et al 2015; Pu and Grossklags 2015). However, very few papers focused on interdependent privacy in the OSNs domain. Shi et al. (2013) identified users' interpersonal privacy concerns arising from changes in the visibility of their social interactions in Facebook and the accessibility to it, aggregated, in a single time-lined page. Even though no new information has been revealed, but aggregating interpersonal information can reveal more details about individuals in unexpected ways (Shi et al. 2013). Choi et al (2015) offered insights on the effects of interdependent privacy by evaluating embarrassing exposures and found that even though users might feel embarrassed, they still value the social rewards of such disclosures. Although this stream of research has conceptually identified users' concerns towards interdependent privacy, there is a lack of research examining the extent to which others' disclosure of information about an individual could lead to actual privacy harm. We aim to address this gap in this research by developing a framework for understanding the commonalities and distinctions between self- vs. co-disclosure, especially pertaining to different types of private information.

2.2. Re-Identification Risks

An individual's identity could be derived from data generated from both online and offline worlds (Creese et al. 2013; Ananthula et al 2015; Li and Wang 2015). In particular, one's offline identity includes data elements such as name, date of birth, gender, nationality, and relationships; online identity includes data elements such as username, browsing history, and email. Attribute disclosure occurs when disclosed information help reveal a sensitive attribute about an individual and associate it with the individual (Li and Li 2009). No matter how mundane the information, its disclosure can have significant repercussions (Stutzman et al. 2013). That is because seemingly innocuous information such as demographic information (Sweeney 2000), movie ratings (Narayanan and Shmatikov 2008) and search queries (Barbaro and Zeller 2006) could be used to identify individuals and reveal sensitive details about them.

In today's digital world, the rising threat to privacy is further increased by the information found in various data sources. More and more personal data are collected and shared, including sensitive data such as health and financial records that can be publicly accessible online and remain available indefinitely. Thus the increasing availability of auxiliary records related to individuals has allowed records to be linked and personal identities to be revealed (Sweeney 2001). In addition, today's OSNs could afford new capabilities to combine social identity elements with personal identity elements, which will improve the accuracy of the identity linkage techniques (Li and Wang 2011; Li and Wang 2015; Park et al. 2012). This increases the threats to individuals' privacy and makes OSNs a gateway to access individuals' personal information (Acquisti 2004; Madden et al. 2007).

3. Research Questions

As discussed in the literature review section, researchers have recently recognized the existence of both self- and co-disclosure of private information in social interactions, including those occurring on OSNs (Jia and Xu 2015). Although there have been several co-disclosure-centric studies focusing on how various settings of OSNs can affect the arising of a user's privacy concerns from other users' activities (Shi et al. 2013, Choi et al. 2015, Jia and Xu 2015), these studies do not further drill into the distinctions between different types of private information (e.g., demographics, location, occupation). As a result, we do not yet sufficiently understand what types of private information are more likely to be co-disclosed by other users than self-disclosed by oneself; what types of social interactions are more likely to lead to co-disclosures of private information; and whether such co-disclosure can indeed reveal a user's identity that would otherwise be hidden from the public. Therefore, we aim to study the following two research questions in this paper:

RQ1. *What types of personal information are more likely to be disclosed through self-disclosure than co-disclosure, and vice versa?*

RQ2. *Does co-disclosure significantly increase the privacy risk for an individual? Specifically, does co-disclosure significantly increase the probability for an individual to be uniquely identified based on his/her activities on an OSN?*

To address these two questions, we need to first build two conceptual foundations. One, we need a proper understanding of the state-of-the-practice *identity linkage model*. Here we use the term "identity linkage model" to refer to the method with which a third party, i.e., someone who is neither the OSN user of interest nor his/her friend who incurs the co-disclosure, can infer the identity of the OSN user from a combination of self- and co-disclosed information about the user on the OSN. From this point onwards, we refer to such a third party as an *adversary* to the OSN user. Note that the identity linkage model includes not only the techniques used by the adversary to harvest data from the OSN, but also the potential external knowledge sources it might use to augment the self- and co-disclosed information - in order to unveil the identity of the OSN user.

Second, we need a quantitative measure for the degree of privacy leakage caused by co-disclosure *on top of* self-disclosed information. To understand why this is essential, consider two scenarios as follows: Case A where an OSN user self-discloses her real name and the city she lives in, while her friend co-discloses her home location; and Case B where the OSN user self-discloses only her first name, while her friend co-discloses her home location. While the co-disclosure activity stays the same in the two scenarios, the implication of this activity, i.e., the *additional* information it discloses beyond the user's self-disclosures, varies significantly. As such, we need a proper metric for the effect of co-disclosure activities **conditioned upon** the self-disclosed information.

In the following section, we describe the identity linkage model we shall use in the paper and an information-theoretic measure of conditional information disclosure.

4. Measurement Development

4.1. Identity Linkage Model

Recall from previous discussions the identity linkage model defines the threat faced by an OSN user on the privacy front – i.e., how an adversary can violate an OSN user's privacy based on self- and co-disclosed information, along with external knowledge sources.

Given the complexity of privacy construct, both conceptually and operationally, and how it varies from one individual to another (e.g., one user might only care about the disclosure of sexual orientation, while another might care more about the disclosure of his/her relationship with another user of the OSN), to properly define the identity linkage model, we must first define the scope of privacy concerns we focus on. For the purpose of this paper, we focus on a specific type of privacy concern: the ability for an adversary to associate a user of an OSN with the user's real-world identity. Given this focus, our goal is to define an identity linkage model with which an adversary unveils the identity of an OSN user from the user's publicly accessible activities on the OSN.

The identity linkage model has two key elements: The first is the world in which the adversary operates (Jin et al 2010) - i.e., the population from which the adversary attempts to identify the OSN user. For the purpose of this paper, we consider the world to be the entire population of US.

The second is the OSN activities used by the adversary. A well understood phenomenon of OSNs is that many OSN users self-disclose their real-world identities or identity elements (i.e., personal attributes such as age or gender) through their public profiles at the OSN (Malhotra et al. 2012). The underlying reason is straightforward - revealing identities in the user profile is a simple way to acquire OSN connections through real-world acquaintances, as the OSN user's real-world circles of family and friends can identify him/her from the profile and build OSN connections accordingly.

Additionally, identities or identity elements may also be disclosed through (publicly accessible) social interactions between an OSN user and his/her OSN connections. Here the disclosure may happen in the form of self- or co-disclosure. As an example of the latter type, consider a privacy-conscious user who chooses to never post a geo-embedded post. Even so, one of her OSN friends might tag her in a geo-embedded post, essentially revealing her location through co-disclosure.

There are two important observations associated with both self- and co-disclosures: One, privacy disclosure could happen from contents or metadata of OSN activities. An example of the former is when the location is described as text in a post, while an example of the latter is when location is revealed through a geotag associated with the post. Second, privacy disclosure could happen through explicit exposure or implicit inferences. Disclosing the city name in a post called "Hometown" is an example of explicit exposure, while implicit inference occurs when a user has numerous posts geotagged in the same city. In the latter case, even though the user has never directly disclosed the city he/she lives in, an adversary can easily identify such information through common-sense inference.

Table 1. Identity Elements and their Information Content

Identity Elements	Gender	First name	Last name	Age	Month & day of birth	State	Zip code	Home address
Privacy Loss (bits)	1†	10.7†	13†	6.3‡	8.5‡	5*	13.8*	26.9*
	10.9†							
	22.9†			14.7‡				

*Based on 2010 Census data.

†Based on 43 million 2015-2016 Voters Registration records from 9 States in the USA.

‡ Based on 24 million 2015-2016 Voters Registration records from 5 States in the USA (a subset of the previous dataset, only these 5 states report date of birth information).

4.2. Information-Theoretic Measurement of Privacy Disclosure

To quantify the loss of privacy in an objective manner, we introduce an information-theoretic metric that can capture not only the overall risk for an OSN user to be identified, but also the effect of individual identity elements and combinations of them on the identification of the OSN user. The root of the information-theoretic measurement is the concept of *information entropy* (measured in bits), which reflects the degree of uncertainty for a random variable with a probability distribution (Cover and Thomas 2006). In general, a bit of entropy represents the uncertainty associated with a single toss of a fair coin. The higher the uncertainty is, the higher the entropy will be, and vice versa. This semantic definition of information entropy (Brickel and Shmatikov 2008) has been used in the literature to measure many security and privacy related constructs, including genome and genetics privacy (Humbert et al. 2013; Erlich and Narayanan 2014).

In the context of our privacy study, consider the uncertainty an adversary faces on identifying an OSN user when it has no information about the user whatsoever. Roughly speaking, the adversary has no way of distinguishing between the 316 million people in the US population (per US Census 2013). This uncertainty – as reflected by the uniform distribution over 316 million possible values - translates to $\log_2$(316 million) = 28.2 bits of entropy. Now consider the case where the OSN user's activities reveal the user's gender. Equipped with this identity element, the adversary now has a reduced uncertainty: from one in 316 million to approximately half of it (i.e., about 158 million with the same gender). Correspondingly, the entropy is reduced by 1 bit to $\log_2$(158 million) = 27.2 bits. In this case, we say that the *privacy loss* caused by gender is 1 bit, i.e., the amount of entropy by which the adversary's uncertainty is reduced.

More formally, we define the *privacy loss* caused by a set of identity elements E having value V to be

$$L(E=V) = H(x) - H(x|E=V)$$
$$= \sum_{i=1}^{n} p(x_i|E=V)\log_2 p(x_i|E=V)$$
$$- \sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

where x_1, …, x_n form the population under consideration (the US population in our case), $p(x_i)$ is the probability for x_i to be the OSN user of interest, and $p(x_i|E = V)$ is the probability for x_i to be the OSN user of interest given knowledge that the identity elements E of the user have value V. In our previous example, there are $p(x_i)$ = 1/316 million for all $i \in$ [1, 316000000], $p(x_i|\text{Gender} = \text{Female})$ = 1/158 million for all $i \in$ [1, 158000000][1], and $p(x_i|\text{Gender} = \text{Female}) = 0$ for all $i \in$ [158000001, 316000000], leading to L(Gender = Female) = $-\log_2 158000000 + \log_2 316000000$ = 1 bit. Based on this definition, we can further define the *privacy loss* associated with a set of identity elements E to be the expected value of $L(E = V)$ for all possible values V of E. In the above example, we have L(Gender) = 1/2 L(Gender = Female) + 1/2 L(Gender = Male) = 1 bit.

It is important to note that the privacy loss function $L(\bullet)$ is not homomorphic to the set union operation, but is convex instead (Cover and Thomas 2006). That is, $L(E1) + L(E_2) \geq L(E_1 \cup E_2)$. For example, if L(First name) = 10.7 bits while L(Gender) = 1 bit, it does not mean L({First name, Gender}) = 10.7 + 1 = 11.7 bit. Instead, the total privacy loss from (an adversary learning) both First

[1] Without loss of generality, here we assume x_1, …, $x_{158000000}$ to be female and the rest of x_i to be male.

name and Gender is smaller, because some part of the information revealed by Gender is already revealed by First name – e.g., if an adversary knows that the OSN user has first name Alice, then it can already infer the user's Gender with high confidence, making the privacy loss from further disclosing Gender much less than 1 bit.

As we shall show in latter part of the paper, this convexity of privacy loss measure plays a rather important role in our understanding of the relationship between self- and co-disclosures. Table 1 summarizes the privacy loss from various identity elements and certain combinations of them. Note that, to generate the table, we used multiple datasets from the US Census to state voters' registration records, as different datasets feature different identity elements included in the table. One can make some interesting observations from the table: For example, gender and first name turn out to be strongly correlated elements, as gender only increases the privacy loss by 0.2 bits once first name is revealed. On the other hand, the correlation between first name or gender and last name is significantly less – the privacy loss caused by revealing last name (on top of first name and gender) is about 12 bits, only 1 bit less than the privacy loss caused by last name alone (13bits).

5. Research Methodology

Equipped with the identity linkage model and the information-theoretic measurement of privacy disclosure, we studied the effect of self- and co-disclosures on the privacy loss of Twitter users. We chose Twitter as the OSN platform for our study due to two main reasons: 1. Twitter users often post tweets that represent direct interactions between each other; yet such interactions can be openly accessed by third parties – constituting co-disclosure of private information. 2. Twitter provides free APIs that facilitate the extraction and analysis of large amounts of user interaction data. In the following discussions, we describe the data collection process we followed and the design of the experiments, respectively.

5.1. Data Collection

There are two key elements of the data collection process: (1) the set of users we selected for data collection; and (2) the data we collected for each selected user. We discuss these two elements respectively as follows.

Selected Users: Since Twitter uses numeric user IDs with an easily identifiable range, we started by sampling uniformly at random 50,000 numeric IDs, and then used the Twitter REST API users/show to validate the existence of the user ID and to retrieve profile information of the user. This validation process produced 44,124 valid Twitter users, on which we then applied the following filtering process. Since we decided to focus on the US population in the identity linkage model, we first filtered out all users who are not located in the US. To do so, we considered not only the location and time-zone attributes of the user profile, but also the geo-locations embedded in the user tweets. Those users with a majority of geo-tagged tweets posted from outside the US were removed from consideration. Finally, we removed those "inactive" users, i.e., those who (1) do not have any public tweets, or (2) do not have any followers or followees.

After applying these filters, we were left with 1,520 Twitter users. We then manually examined each of these 1,520 accounts and further excluded 163 of them that are obviously not personal accounts – i.e., they are self-declared accounts for businesses and organizations. We used the remaining 1,357 Twitter users to perform our subsequent analysis.

Data Collected for Each User: For each selected user, we collected its OSN activities in three categories: 1) the user's profile, 2) all tweets posted by the user, and 3) all tweets posted by the user's *followers* that mentioned the user (i.e., tagged using "@" followed by the user handle). Note that we considered only followers but not followees here because, per our experimental results, it is extremely unlikely for a user to be mentioned by a followee who is not at the same time a follower of the user. To collect the data, we used the Twitter Search API to access all information publicly available about users and their followers, and the Google Maps Geocoding API to convert the Geo-locations extracted from tweets to full addresses.

5.2. Experimental Design

The objective of experimental design is to identify and measure the privacy disclosure incurred by the collected (publicly available) data. After examining the collected data, we identified six categories of identity elements that are often disclosed through self- or co-disclosure on Twitter: name, location, gender, birthday, age, and family relationships (e.g., siblings, spouse, parents). Each category further consists of identity elements at different granularities. For example, the disclosure of name might be first name only, last name only, or full name. Similarly, the disclosure of location information might be at the metropolitan area level, at the ZIP code level, or revealing the exact address.

Table 2 summarizes how self- and co-disclosures often occur for each of these six categories of identity elements. It also depicts how we identify disclosures in our experimental study. Specifically, we followed a three-step, manual-automated-manual process. We

Table 2. Self- and Co-Disclosure of Identity Elements

Identity Elements	Self-Disclosures	Co-Disclosures
Name	The name could be extracted from the name attribute or inferred from the screen name of the user profile.	The name could be extracted from tweets that mention the user.
Location Information	The location could be extracted from the location attribute of the user profile or inferred from geo-enabled tweets.	The location could be inferred from geo-enabled tweets that tag the user.
Gender	The gender of the user could be inferred from the bio attribute of the user profile as some users describe themselves as a father or mother or wife or husband.	The gender could be inferred from tweets mentioning the user that include relational or gender specific data such as sister or bro.
Birthday	Disclosed by user tweets that mention birthday.	Birthday information could be inferred from co-owners' tweets of birthday wishes to the user, e.g., "Happy Birthday"
Age	Disclosed by user tweets that mention age, year of birth, etc.	Age information could also be inferred from birthday wishes, e.g., "Happy 43rd Birthday!"
Family Relationships	Relationships could be inferred from user tweets mentioning relatives such as siblings, spouses, and parents.	Relationships could be inferred from tweets coming from relatives such as "Miss you mom" or "Happy Birthday Dad", etc.

started with manually examining a small sample of tweets to identify the disclosure patterns summarized in Table 2. Then, we translate each pattern to a filtering condition that is automatically applied to all downloaded tweets to identify those candidates that could potentially disclose private information. For example, candidates for birthday co-disclosure were identified with a conjunctive condition of (1) "@" another user, and (2) contains keyword "birthday". In the final step, we manually examine each candidate to confirm self- or co-disclosures. This final step removes false positives that meet the candidacy criteria but do not actually reveal precise information, e.g., "your birthday is more than 6 months away".

One can make several observations from the table: First, the way identity elements are disclosed on Twitter is often ad-hoc, especially in the case of co-disclosures. For example, user A's birthday may be disclosed by a follower B's tweet "Met @C in @A's birthday party yesterday". There are two implications of such ad-hoc disclosures: First, the wide variety of ways for an identity element to be disclosed required us to manually examine the collected data instead of relying on an automated process. Second, it also calls into question the comprehensiveness of disclosures identified in our experimental design, as it is possible for a subtle disclosure to be missed in our manual examination process. We discuss the issue of comprehensiveness and the implications in the discussion section.

Second, self- and co-disclosures often take different forms. Specifically, while self-disclosures usually happen through direct statements (e.g., describing oneself as a proud mom directly discloses the gender of the user), co-disclosures tend to be subtler, and often occur through inference – e.g., being tagged in a tweet that is marked with the location of a restaurant. This difference makes co-disclosures harder to identify than self-disclosures. Again, we discuss the implication of this difference in the discussion section.

6. Research Results: RQ1 and RQ2

Following the data collection procedure and the experimental design outlined in the research methodology section, we examined the two research questions RQ1 and RQ2 respectively. Specifically, we first identified the self- and co-disclosures of the six identity elements listed in Table 2. Our study of RQ1 focused on comparing the two types of disclosures. Then, we used the information-theoretic measurement of privacy disclosure to further quantify the average privacy loss from self- and co-disclosures. The results were used in our study of RQ2, i.e., whether co-disclosures significantly increase privacy loss on top of self-disclosures.

6.1. RQ1: Self-Disclosure vs Co-Disclosure

Figure 1 depicts the percentage of the 1,357 selected Twitter users who have their identity elements self- or co-disclosed through their Twitter activities. We marked each identity element with either self- or co-disclosure, leading to 12 categories. One can see from the figure that, for each identity element, the percentage of users having the element disclosed through either self- or co-disclosure ranges from 1.7% (co-disclosure for ZIP code) to 23.7% (co-disclosure for gender). The comparison between self- and co-disclosure, however, varies significantly for different identity elements.

To further study the comparison, we performed the paired-sample *t*-test for each identity element to determine whether there is a significant statistical difference between the probability of the self- and co-disclosure of the element. Table 3 depicts the results. One can see that all but one identity element exhibit significant difference between the levels of self- and co-disclosure. Specifically, for first name, age and ZIP code, self-disclosure is significantly more frequent than co-disclosure. For birthday and gender, co-disclosure is more frequent. For all these five categories, we have *p*-value < 0.001. The disclosure of relatives (i.e., personal relationships), on the other hand, does not show a significant difference between self- and co-disclosures, with *p* = 0.2439.

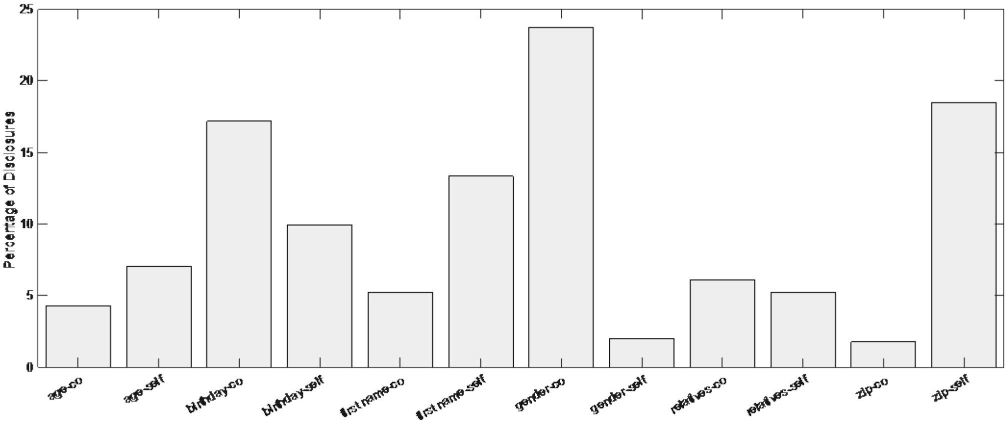

Figure 1. Percentage of Self- and Co-Disclosure for Identity Elements

Table 3. Statistical Differences between Self- and Co-Disclosures (Paired-Sample t-test)

Category	Identity Element	Self-Disclosed	Co-Disclosed	Disclosed by both	p-value
Self > Co	first name	181	71	66	1.4495e-24
	age	96	58	21	3.2160e-04
	zip	251	24	12	2.7818e-50
Co > Self	birthday	135	233	79	9.3014e-12
	gender	27	322	15	3.9476e-68
Insignificant	relatives	71	83	24	0.2439

Another interesting observation from Table 3 is that self- and co-disclosure are positively correlated with each other. We performed the χ^2-test over each identity element, and the null hypothesis of independence is rejected for every identity element with p < 0.0001.

6.2. RQ2: Does co-disclosure significantly increase the privacy risk?

To address RQ2, we measured the amount of privacy loss resulted from the disclosure of identity elements. Specifically, we first measured the amount of privacy loss from self-disclosures, and then computed the amount of *additional* privacy loss caused by co-disclosures. For example, if a user self-discloses his/her first name but not gender, and a follower of the user further co-discloses the gender of the user, then the amount of additional privacy disclosure is not 1 bit (as disclosing gender would incur), but only 0.2 bit because, per Table 1, the privacy loss from self-disclosure, i.e., first name, is 10.7 bits while the total loss from both self- and co-disclosures is 10.9 bits. As such, the amount of additional privacy loss caused by co-disclosure is 10.9 – 10.7 = 0.2 bit.

Figure 2 depicts the results from our study on users with both self- and co-disclosures. Note from the figure that we grouped all these users into 10 buckets according to the amount of privacy loss from self-disclosure alone. The first bucket, for example, consists of the users with self-disclosure privacy loss at 0 – 10 percentile (i.e., the lowest 10% of all users). Then, for each bucket, we measured the average amount of additional privacy loss caused by co-disclosures.

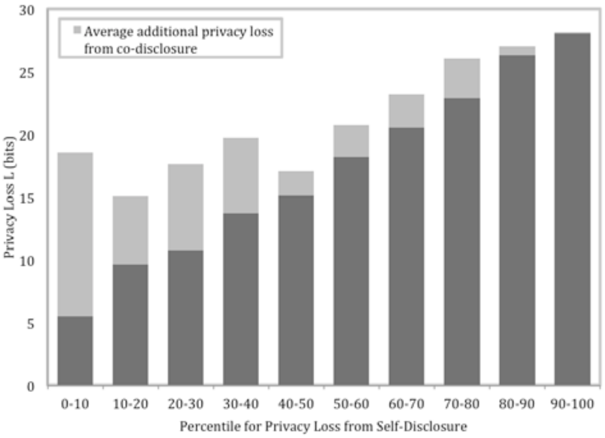

Figure 2. Additional Privacy Loss from Co-Disclosure

Somewhat surprisingly, the figure appears to suggest that, the less a user self-discloses, the more his/her followers are likely to co-disclose about him/her. While this result might appear counter-intuitive - what it indeed entails is that, for the same co-disclosure behavior, the less information a user self-discloses, the more damage (to privacy) the co-disclosure will incur. Although not shown in Figure 2, we studied the correlation between the amount of self-disclosure and co-disclosures alone, and the Pearson correlation coefficient between them is 0.0598 – demonstrating a positive, albeit weak, correlation. On the other hand, if we measure the correlation between the privacy loss from self-disclosure and the additional privacy loss from co-disclosure, the coefficient becomes -0.6689, confirming our observation of the negative correlation from Figure 2.

7. Discussions

Comprehensiveness of Self- and Co-Disclosure Identification: As mentioned in the experimental design section, we identified in our studies numerous possible ways for identity elements to be disclosed through inferences - e.g., the timestamp of a “happy birthday” message discloses the birthday of the recipient, while the tagging of a user in a “bachelor’s party” or “girls’ night out” reveals the gender.

Given the ad hoc nature of such inference channel, it is important for us to admit that the disclosures identified in our study might not be comprehensive. For example, if a user is tagged in a message describing a movie watching party for “The Sisterhood of The Traveling Pants”, then the gender and age of the user may be inferred with a high confidence, given that the movie is also perceived as targeting a female, younger, audience. It is extremely difficult, if not impossible, to identify all such disclosure channels and measure them in a systematic manner. For example, we could not find accurate statistics for the viewer distribution of the movie. Thus, in this paper, we focused on only those disclosures that *deterministically* reveal an identity element of the user, and did not further study the vast varieties of other, probabilistic, inference channels.

Despite the lack of comprehensiveness on the identification of self- and co-disclosure channels, we would like to emphasize our belief that the results presented in this paper remain valid even after considering other disclosure channels. The main reason is the observation discussed in the experimental design section: while self-disclosure often takes the form of direct statements, it is co-disclosure that usually occurs over the subtle, complex and hard-to-enumerate inference channels. Thus, the importance of co-disclosures could only be amplified if we were to include the probabilistic inference channels into consideration. Given that the results in this paper already demonstrate the additional privacy loss incurred by co-disclosures, we leave the investigations of probabilistic inference channels to future studies.

Effects of Privacy Settings: Currently, Twitter allows users to adjust their privacy settings to hide their twitters and the list of their followers / followees (but not their profile information, like self-descriptions). Interestingly, we found that making one’s Twitter account private is *not* an effective way to thwart the co-disclosures discussed in this paper. Specifically, the information that constitutes co-disclosures *remains* publicly accessible, just more difficult to find for an adversary. For example, even if User A sets her Twitter account to be private, one can still find A’s name in the follower list of user B, if B sets her account to be public. Similarly, the tagging of user A remains publicly visible in the tweets posed by User B. These policies make it impossible for a user to block co-disclosures through adjusting his/her privacy settings at Twitter. Nevertheless, setting one’s account to private does make it harder for an adversary to discover the co-disclosures, as it would have to (somehow) find the followers of the user without access to a comprehensive list. This is still feasible, however, especially when the adversary has the technical capacity to store and index all tweets (e.g., from the Twitter Firehose) that might contain information about the user of interest.

Limitations: We acknowledge several limitations in our studies. First, the scale of the experiments was constrained by the query access limitation enforced by Twitter, i.e., its APIs limit the number of requests to 15 or 180 requests per 15-minute window. This, in turn, limited the number of users that could be included in the experiments, given that we would have to query every follower of each selected user. Second, Twitter limits the maximum number of tweets from one user that can be retrieved through the Search API to 3,200. In our experiments, 168 out of 1,357 users triggered this limit - i.e., for these users, only the most recent 3,200 tweets were retrieved through the API and considered in our experiments. While these 168 users tend to be the most active ones who already have high levels of privacy disclosure (their average self-disclosure based privacy loss is 26.46 bits, compared an average of 5.00 bits for all 1,357 users), the limitation of 3,200 tweets makes it possible for our study to still underestimate the privacy loss for these users. To this end, we plan to conduct a future study on privacy loss from a longitudinal perspective, to understand the impact of such limit on the disclosure of privacy information.

Another limitation of our approach stems from the manual efforts in our experimental design. While such efforts are essential to eliminating false positives and establishing lower-bound estimates of disclosures, they prevent the study from scaling to a larger sample of Twitter users. To this end, we plan to investigate in future studies the usage of text mining and natural language processing (NLP) techniques to automate disclosure identifications.

8. Conclusions

In this paper, we presented an empirical, data-driven study quantifying the extent to which others' co-disclosure of information about an individual could lead to actual privacy harm. We developed an innovative measurement to capture an individual's actual privacy loss caused by co- *vs.* self-disclosure, and collected data from Twitter to show significant harms caused by co-disclosure and illustrate the interesting differences between different identity elements revealed through self- and co-disclosure. The current study contributes to privacy practices in two important ways: First, an understanding of what information tend to get co-disclosed helps OSN users with properly regulating self-disclosures, to prevent the inference of sensitive data from a combination of self- and co-disclosures. Second, an understanding of how co-disclosure contributes to the identification of an individual also changes how businesses harvest customer information, and in turn prompts OSNs to revisit their privacy policies and access control practices. It is our hope that this work will inspire and motivate more data-driven studies to measure real privacy harms with observational data in the future.

9. Acknowledgement

The authors were supported in part by NSF under 0953749, 1343976, 1624074, ARO under W911NF-15-1-0020. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the sponsors.

10. References

Acquisti, A. 2004. "Privacy and security of personal information: Economic incentives and technological solutions," In The Economics of Information Security.

Acquisti, A. and Gross, R. 2006. "Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook," *Proceedings of 6th Workshop on Privacy Enhancing Technologies* (pp. 36–58).

Acquisti, A., Brandimarte, L., and Loewenstein, G. 2015. "Privacy and human behavior in the age of information," Science, 347(6221), 509-514.

Alsarkal, Y., Zhang, N., and Zhou, Y. 2015 "Linking virtual and real-world identities," IEEE ISI 2015.

Ananthula, S., Abuzaghleh, O., Alla, N. B., Chaganti, S. B., Kaja, P. C., & Mogilineedi, D. (2015). "Measuring privacy in online social networks," International Journal of Security, Privacy and Trust Management, 4(2), 1-9.

Belanger, F. and Crossler, R. E. 2011. "Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems," *MIS Quarterly*, (35: 4) pp.1017-1041.

Bertino, E., Lin, D., and Jiang W. 2008. "A survey of quantification of privacy preserving data mining algorithms," Privacy-preserving data mining. Springer US, 2008. 183-205.

Bethlehem, J.G., Keller, W.J. and Pannekoek, J. 1990. "Disclosure Control of Microdata," Journal of the American Statistical Association, Vol. 85, pp. 38-45.

Bhattacharya, I. and Getoor, L. 2007. "Collective entity resolution in relational data," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2007.

Böhme, R., and Grossklags, J. 2013. "Trading Agent Kills Market Information: Evidence from Online Social Lending," in WINE 2013.

Biczók, G. and Chia, P.H. 2013 "Interdependent privacy: Let me share your data," In FC 2013, pp. 338–353.

Brickel, J. and Shmatikov, V. 2008. "The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing," in KDD 2008.

Choi, B.C.F., Jiang, Z., Xiao, B., and Kim S.S. 2015. "Embarrassing Exposures in Online Social Networks: An Integrated Perspective of Privacy Invasion and Relationship Bonding," Information Systems Research 26(4):675-694.

Cover, T. and Thomas, J. 2006. Elements of Information Theory. John Wiley and Sons.

Creese, S., Gibson-Robinson, T., Goldsmith, M., Hodges, D., Kim, D., Love, O., Nurse, J., Pike, B., and Scholtz, J. 2013. "Tools for Understanding Identity," *HST 2013.*

Culnan, M. and Armstrong, P. 1999. "Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation," *Organization Science*, *10*(1), 104-115.

Dinev, T and Hart, P. 2006. "An extended privacy calculus model for e-commerce transactions," European Journal of Information Systems 17(1):61–80

Eisenberger, R., Fasolo, P.M., and Davis-LaMastro, V. 1990. “Perceived organizational support and employee diligence,” commitment, and innovation. Journal of Applied Psychology. 75(1):51–59.

Ellison, N.B., C. Steinfeld and C. Lampe. (2007). “The benefits of Facebook “friends:” Social capital and college students use of online social network sites.” Journal of Computer-Mediated Communication, 12(4), 1143–1168.

Gavison, R., 1980, “Privacy and the Limits of Law,” *Yale Law Journal*, 89: 421–71

Erlich, Y. and Narayanan, A. 2014. “Routes for breaching and protecting genetic privacy,” Nature Reviews Genetics 15(8), June 2014, 409-421.

Homans, G.C. 1958. “Social Behavior as Exchange,” American Journal of Sociology 63: 597-606.

Humbert, M., Ayday, E., Hubaux, J.P., and Telenti A. 2013. “Addressing the concerns of the Lacks family: quantification of kin genomic privacy,” In *CCS 2013*.

Jia H. and Xu H. 2015. “Measuring Individuals’ Concerns over Collective Privacy on Social Networking Sites,” In ICIS 2015.

Jin, X., Zhang, N., and Das, G. 2010. “Algorithm-safe privacy-preserving data publishing,” In EDBT.

Krasnova, H., Spiekermann, S., Koroleva, K., Hildebrand, T. 2010. “Online Social Networks: Why We Disclose,” Journal of Information Technology, 25 (2), 109-125.

Li, J. and Wang, A.G. 2015. “A framework of identity resolution: evaluating identity attributes and matching algorithms,” Security Informatics (2015).

Li, J. and Wang, G.A. 2011 Criminal identity resolution using social behavior and relationship attributes. In IEEE ISI 2011.

Li, J., Wang, G.A., Chen H. 2010 Identity matching using personal and social identity features. Inf. Syst. Front. 13, 101–113 (2010)

Li, T. and Li, N., 2009. “On the tradeoff between privacy and utility in data publishing,” in KDD 2009.

Li, X., Liu X., Motiwalla L. 2015. “Valuation of Personal Information,” In ICIS, 2015.

Madden, M., Fox, S., Smith, A., and Vitak, J. 2007. "Digital Foot- prints: Online Identity Management and Search in the Age of Transparency," PEW Research.

Malhotra, A., Totti, L., Meira, Jr. W., Kumaraguru, P., and Almeida, V. 2012. “Studying User Footprints in Different Online Social Networks,” ASONAM 2012.

Mathews, R. 2010. “The Social and psychological impact of online social networking,” Australian Psychological Society.http://www.psychology.org.au/publications/inpsych/2010/december/social

Metzger, M. J. 2004. “Privacy, trust and disclosure: Exploring barriers to electronic commerce,” *Journal of Computer-Mediated Communication, 9*(4).

Narayanan, A. and Shmatikov, V. 2008. “Robust De-anonymization of Large Sparse Datasets,” In Proc. 2008 IEEE Symposium on Security and Privacy.

Narayanan, A. and Shmatikov, V. 2009. “De-anonymizing Social Networks,” *IEEE Symp. Security and Privacy*.

Nissenbaum, H. 2011, "A Contextual Approach to Privacy Online", *Daedalus*, vol. 140, no. 4, pp. 32-48.

Ohm P. 2010. “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,” UCLA Law Review, Vol. 57, p. 1701, 2010.

Park, S.H. and Huh, S.Y. 2012. “A Social Network-Based Inference Model For Validating Customer Profile Data,” MIS Quarterly, 36 (4), 1217-1237.

Pu, Y. and Grossklags, J. 2015. “Using Conjoint Analysis to Investigate the Value of Interdependent Privacy in Social App Adoption Scenarios,” In ICIS 2015.

Roloff, M. E. 1981 Interpersonal communication: The social exchange approach. Beverly Hills, CA: Sage Publications, Inc.

Schofield, P. and Joinson, A.N. 2008. “Privacy, trust, and disclosure online,” In A. Barak (Ed.), *Psychological aspects of cyberspace: Theory, research, applications* (pp. 13-31). Cambridge University Press.

Schneier, B. 2010. A taxonomy of social networking data. *IEEE Security & Privacy*, *8*, 4, 88.

Schwartz, P. M. and Solove, D. J. 2011. The PII Problem: Privacy and a New Concept of Personally Identifiable Information, 86 N.Y.U. L.Q. Rev. 1814 (2011)

Shah, K. 2014 “A Developer's Guide: Popular Uses for WhitePages PRO API,” Whitepages Pro. [Online].

Shi, P., Xu, H., and Chen, Y. 2013. “Using Contextual Integrity to Examine Interpersonal Information Boundary on Social Network Sites,” in *CHI*, 35–38.

Smith, H.J., Milberg, J.S., and Burke, J.S. 1996. "Information Privacy: Measuring Individuals' Concerns About Organizational Practices," *MIS Quarterly* (20:2), pp 167-196.

Smith, H.J., Dinev, T., Xu, H. 2011. “INFORMATION PRIVACY RESEARCH: AN INTERDISCIPLINARY REVIEW,” *MIS Quarterly*, vol. 35, no. 4, 2011, pp. 989–1015.

Stutzman, F., and Kramer-Duffield, J. 2010. “Friends Only: Examining a Privacy-Enhancing Behavior in Facebook,” CHI 2010, pp. 1553-1562.

Stutzman, F., Gross, R., and Acquisti, A. 2013 Silent Listeners: The Evolution of Privacy and Disclosure on Facebook. *J. Privacy and Confidentiality*: 4(2).

Sweeney, L. 2001. “Information Explosion, Confidentiality, Disclosure, and Data Access: Theory and Practical, Applications for Statistical Agencies,” Urban Institute,Washington, DC, 2001.

Sweeney, L. 2000. “Simple Demographics Often Identify People Uniquely,” Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

Tanis, M., and Postmes, T. 2007. Two faces of anonymity: Paradoxical effects of Cues to identity in CMC. *Computers in Human Behaviour, 23*, 955–970.

Tobina, S., Vanmana, E., Verreynnea, M., and Saeria A.. 2014. “Threats to belonging on Facebook: lurking and ostracism,” Social Influence, 10, 1, 31-42

Tufekci, Z. 2008. “Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites,” Bulletin of Science, Technology & Society, 2008

Yu, J., Hu, P.J., and Cheng, T. 2015. “Role of Affect in Self-Disclosure on Social Network Websites: A Test of Two Competing Models,” Journal of Management Information Systems , 32(2), 239-277.