Privacy and Social Capital in Online Social Networks

Jin-Hee Cho
US Army Research Laboratory
Adelphi, MD
Email: jinhee.cho@us.army.mil

Izzat Alsmadi
University of Texas A&M
San Antonio, TX
Email: ialsmadi@tamusa.edu

Dianxiang Xu
Boise State University
Boise, ID
Email: dianxiangxu@boisestate.edu

***Abstract*—In online social networks (OSNs), individual users have a strong desire to expand their social networks through OSN activities and try to maximize the benefits from the social relationships, called *social capital*. However, with a large-scale social network, their privacy rights have been significantly intruded by adversarial users that perform social attacks including false / illegal private information dissemination or the use of fake identities. In this work, we study how individual users can expand their social networks by making trustworthy friends while not leaking their private information out to unauthorized parties or social attackers. We adopt the concepts of trust and reputation in order to preserve users' privacy while enhancing their social capital in OSNs. Given a social network topology from the Facebook, we model an individual user's interactions with other users based on feeding (e.g., posting information) and feedback behaviors (e.g., providing likes or comments). Our results show that there exists a tradeoff between social capital and privacy preservation. In addition, we show there exists a balance point of social capital and privacy thresholds that maximizes correct information diffusion while minimizing illegal private information leakout, given users' risk appetite for preserving privacy.**

***Index Terms*—online social networks, trust, privacy, social capital**

I. Introduction

With the proliferation and popularity of online social networking sites, online social networks (OSNs) became more common places for people to interact with others and to share their opinions, preferences (e.g., political leanings), social pleasure, and their own emotional feelings or experiences than ever [1]. In OSNs, connecting with other people based on common interests, similarities, or opinions is regarded as an important asset for people who desire to expand their network for practical benefits [2]. However, due to the voluntarily disclosed private lives of many people, social networking sites also reveal high vulnerabilities exposed by adversarial users such as exploiting naive people's private information to meet their own selfish or malicious needs.

Although users are aware of risk associated with privacy issues from online activities, millions of social network users do share their thoughts, news, photos, or videos. These can expose the users to vulnerabilities that may be exploited by malicious users. This is called *privacy paradox* [3]. We human beings have a strong desire to expand our network in order to maximize productivity through relationships (e.g., social and emotional support, knowledge sharing, acknowledgement by others, social movement, marketing). However, these social activities should be alerted by serious risks and consequences when the private information is exposed to or exploited by malicious entities, such as serious financial loss, identity theft, or social issues (e.g., exposing private lives of celebrities).

We adopt the concept of trust and reputation to achieve the dual conflicting goals of minimizing privacy intrusions and maximizing social capital. Although privacy preservation mechanisms [10, 11] have been proposed in the literature, little studies have been conducted to investigate the relationship between privacy preservation and social capital. We propose a trust and reputation model, called `p2ReMon`, representing Privacy-Preserving tRust and rEputation Model for Online social Networks, that can control and identify an optimal setting to meet a user's privacy right and enhance the social capital in a social network. This work has the following **unique contributions**:

1) To minimize privacy intrusions and maximize social capital in an OSN, `p2ReMon` uses privacy and social capital thresholds to limit access to a user's posted information while maximizing the user's social capital by making trustworthy friends. The privacy thresholds are set based on a friend's trust estimated by the user. The social capital threshold is used to make friending decisions (e.g., accepting / declining a friend request or sending a friend request) based on a potential friend's reputation and similarity between the user and the potential friend. This is the first work that considers both privacy and social capital of users in OSNs while most existing works only concern the privacy issue.
2) We investigate a critical trade-off between a user's social capital and privacy preservation in which the large size of a social network increases social capital while exposing high potential privacy intrusion. We identify an optimal setting that meets a user's desired levels of both conflicting goals.
3) Through extensive simulation experiments using a Facebook dataset [4], we show the existence of an optimal setting that meets both goals under various degrees of hostility in OSNs in terms of correct information diffusion and privacy information leakout.

978-1-5090-1328-9/16/$31.00 ©2016 IEEE

The rest of this work is organized as follows. Section II gives an overview of related work. Section III describes network, interaction, and social adversarial models. Section IV gives the overview of p2ReMon including trust, reputation, social capital, and privacy models. Section V explains performance metrics, experimental setup, and the results of sensitivity analysis. Finally, Section VI concludes this work and suggests future research directions.

II. Related Work

This section discusses the concepts of trust and social capital, and existing models of privacy, trust, and reputation in OSNs.

Concept of Trust: Trust indicates a relationship between two entities called a trustor and trustee in which the trustor assesses its subjective opinion towards the trustee based on given criteria [5]. Trust plays a role of enabling relationships and interactions to take place within OSNs. Although users are aware of risks and threats by participating in activities of OSNs, they seem to trust other users by posting their private information including messages, images, or videos [6].

Concept of Social Capital: Through the interactions and relationships formed in OSNs, people want to obtain a benefit from them, called *social capital*. Although there are different definitions of social capital, most researchers agree with its common role as the social structure to facilitate achieving individual or collective goals through the investment in personal relationships [7]. Putnam [8] defines *social capital* in terms of two aspects: *bridging* and *bonding*. *Bridging social capital* indicates new information can be obtained through weak ties, promoting the diffusion of non-redundant information. On the other hand, *bonding social capital* is derived from strong ties between intimate people with high trust, reciprocity, and emotional support.

Privacy Models: Before we enter the generation of this OSNs, privacy originally meant "one's right to be alone" [9]. To be specific, privacy refers to an entity's request to decide when, how, and to what extent its own private information may be revealed to a third party where the entity can be an individual person, a group, or an organization. Thus, privacy means *controlled access to private information*, not just hiding information. Due to the importance of privacy issues in OSNs, many researchers have proposed and developed privacy-preserving mechanisms. Guo et al. [10] design privacy-preserving mechanisms using user-centric private matching that considers social attributes in e-Health networks. Later Guo et al. [11] also propose a privacy-preserving scheme to disseminate users' information on social relationships for a delay tolerant network, namely *PSaD*. The above works [10, 11] use a traditional encryption / decryption technique to ensure privacy and confidentiality. Different from [10, 11], our work uses a user's individual behavior and location in social structure to derive the user's trust, reputation and social capital to make decisions on selecting friends or setting privacy levels for friends' access to his or her private information.

Trust and Reputation Models: While the concept of trust is used to indicate the relationship between two entities, *reputation* is used to refer to a more general sense of trust towards a particular entity based on opinions by multiple entities [5]. The concept and properties of trust have been utilized in various types of OSN applications. Grabner-Krauter [12] treats trust as an indicator of an individual's confidence in dynamic decision making context in which trust is used as an instrumental support emphasizing its practical use as well as subjective trust based on emotional nature.

Although the concept of reputation is overlapped with that of trust in terms of subjective perception, expectation, and/or belief about capability, honesty, and/or reliability of something or someone, it has a more aspect of the objective concept than that of trust due to the nature of aggregated opinions by multiple third parties [13].

III. Preliminaries

In this section, we describe the network model, user interaction types, and social adversarial model.

A. Network Model

This work concerns an OSN where users interact with other users, make friends by sending or receiving friend requests, exchange their thoughts or personal lives by posting photos or videos or use social media tools for practical purposes such as marketing. A social network can be represented as a graph, G, where each user is a vertex, v_i, for any user i, and the relationship between two users is indicated as an edge, e_{ij}, for friends of two users i and j. A weight of the edge, w_{ij}, represents the depth of trust relationships such as a social tie where w_{ij} can be calculated based on the interactions between i and j. Note that w_{ij} may not be same as w_{ji} because trust is a subjective opinion and may not be symmetric. We describe how to compute w_{ij} as the measure of trust relationships between two users. Note that we use a notation $T_{ij}(t)$ to indicate the degree of trust user i has towards friend user j at time t, which can be treated as same as w_{ij} at an instantaneous time, as described in Section IV-A.

B. User Interaction Types

A user can communicate through the following interactions:

- *Posting information*: A user may upload real-time postings as well as his/her personal profile including affiliations, education, profile pictures, or preferences.
- *Sharing information*: A user may share information originated from third parties (e.g., news, pictures, videos).
- *Providing preferences*: A user can express preferences as the evidence of positive feedback towards friends' posts (e.g., 'likes' in the Facebook or 'Favorites' in the Twitter).
- *Leaving comments*: A user can leave comment(s) towards any posts by friend(s).

To restrict accessibility to posts by allowing various activities such as view, comment, share, or propagate, p2ReMon provides the automatic privacy setting based on the estimated trust of friends while allowing the user to adjust the privacy setting based on a user's specific need.

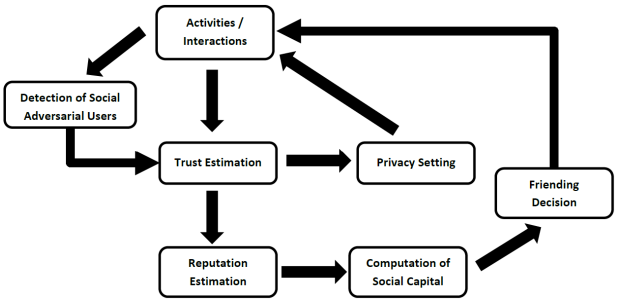


Fig. 1: The overview of p2ReMon

C. Social Adversarial Model

We consider the following social adversarial behaviors:

- *Self-promotion / use of fake identity*: Adversarial users can make up and post fake profiles to promote themselves using a fake identity and false background (e.g., education, affiliation, gender, picture). Their goal is to promote themselves by impressing other users in order to obtain a high reputation. This self-promotion with a high reputation can be exploited by an attacker aiming to expand the social network by sending friend requests to other users, particularly targeted users. A fake account using a fake identity is often used by attackers who aim to propagate false information, which is described next.
- *False information dissemination*: Information, misinformation (i.e., false information to deceive), or disinformation (i.e., false information to mislead public opinions such as propaganda) can spread out quickly in social media due to its viral nature of fast information diffusion [14]. An adversarial user can also provide false feedback by leaving a large amount of 'likes' or 'dislikes' in order to mislead public opinions.
- *Illegal private information leak-out attack*: An adversarial user can leak another user's private information out to unauthorized parties.

To model an adversarial user (or attacker), we consider a certain percentage of adversarial users among all users in a network, denoted as P_c assuming that they perform attacks persistently. We discuss the countermeasures embedded in p2ReMon to deal with these attacks in Section IV-F.

IV. P2REMON

Fig. 1 describes the connections of multiple models embedded in p2ReMon. First, each user will compute a peer-to-peer trust based on the interaction patterns with their friends. The trust values are used as evidence to compute a single reputation value of each user. Each user's trust towards a friend is different from the reputation of the friend. Trust is subjective as individuals estimate other user's trust based on their personal interactions. However, reputation represents a global, single opinion towards each user due to the nature of aggregating multiple parties' opinions. In the proposed privacy model, a user's trust estimated by another user (i.e., a friend of the user) is used to determine the user's access right to the information provided by the friend. For example, when A and B are friends with each other, A determines B's access right to A's posts. On the other hand, we use a user's reputation to calculate the user's social capital. Since the social capital represents how much benefit a friend user brings to the user, it needs an objective value of the friend. We calculate a user's social capital based on the quality of friends using a certain reputation threshold, as discussed in Section IV-C. A user's social capital value is used to make friending decisions (i.e., whom to select as a friend) in this work. After a user makes friending decisions, the user's trust and reputation are updated based on the added relationship. Accordingly the updated trust and reputation affect the privacy setting and social capital of the user.

A. Trust Model

A user's overall trust is estimated by the following three behavior features:

- *Feeding*: This measures the quality and frequency of posted information.
- *Feedback*: This measures the quality and frequency of leaving comments or providing feedback (i.e., how often a user provides positive feedback such as *likes* or *comments*).
- *Common Friends*: This measures the similarity of two users based on common friends.

$T_{ij}(t)$ indicates trust of user i in user j at time t, and is computed by:

$$T_{ij}(t) = \sum_{x \in X} t_x T_{ij}^x(t) \tag{1}$$

where X is a set of trust dimensions and x indicates a particular dimension of trust. Our trust model is generic in that different dimensions of trust can be used based on domain specific context. In this work, we select three trust dimensions including *feeding*, *feedback*, and *friends similarity*. t_x indicates a weight for trust in x where $\sum_{x \in X} t_x = 1$. The weight can be selected based on a domain specific requirement of the priority setting in system trust. $T_{ij}(t)$ is scaled in the range of $[0, 1]$ where $T_{ij}^x(t)$ is in $[0, 1]$ as a real number. $T_{ij}^x(t)$ is computed by:

$$T_{ij}^x(t) = \frac{\sum_{t=0}^{T} I_{ij}^x(t)}{\max[\sum_{t=0}^{T} I_{ik}^x(t) \text{ for } k \in A_i]} \tag{2}$$

$I_{ij}^x(t)$ is the number of positive interactions user i received from user j. T is the latest time point user i had an interaction with user j. $\max[\sum_{t=0}^{T} I_{ik}^x(t)]$ indicates the maximum number of accumulated positive interactions from $t = 0$ to $t = T$ user i had with any user k in which A_i is a set of i's friends, k's.

B. Reputation Model

User i's reputation, R_i, is computed based on the aggregations of i's friends, j's trust towards i, over the total number of i's friends and given by:

$$R_i = \frac{\sum_{j \in A_i} T_{ji}(t)}{|A_i|} \tag{3}$$

where A_i is a set of user i's friends. Note that having a large number of friends may affect R_i because we consider the number of common friends between two users. However, only having a lot of friends does not lead to a user's high reputation when there are little interactions between them.

Filtering evidence for positive interactions can be achieved by using sentimental analysis, a popular data mining technique for opinion mining [15].

C. Reputation-based Social Capital Model

We adopt the concept of *social capital* by Putnam [8] in that a user can achieve dual goals of increasing social capital in terms of *bridging* and *bonding*. In order for the user to obtain new, non-redundant information or new opportunities through the network, the user can accept or send a friend request to expand the social network through weak ties for *bridging*. However, if the new friend is actually not trustworthy, exhibiting social adversarial behavior patterns, the friending decision may introduce privacy intrusions (or vulnerabilities). This supports why a user needs to choose a trustworthy friend based on *bonding*.

We use the ratio of the *neighborhood overlap* of an edge connecting two users [16] as the basis of a friending decision. We modify the concept of the neighborhood overlap of two users by defining user i's index based on reputation-based social capital (RSC) of user j, RSC_{ij}, which is given by:

$$RSC_{ij} = \frac{R_{TA_i \cap TA_j}}{R_{TA_i \cup TA_j}} \tag{4}$$

where $R_{TA_i \cap TA_j}$ is the sum of reputation values of users in a set $TA_i \cap TA_j$ and $R_{TA_i \cup TA_j}$ is the sum of reputation values of users in a set $TA_i \cup TA_j$. TA_i and TA_j represent the sets of user i's friends and j's friends who have their reputation levels in their clique no less than a reputation threshold ρ, respectively. TA_i is defined by:

$$TA_i = \{k \in A_i \mid R_k \geq \rho\} \tag{5}$$

where A_i is a set of user i's friends and k is a friend of user i and so belongs to A_i. R_k is user k's reputation in k's clique where the clique consists of k and k's friends whose reputations are no less than the reputation threshold, ρ. R_k is computed as shown in Eq. (3). $R_{TA_i \cap TA_j}$ and $R_{TA_i \cup TA_j}$ are computed by:

$$R_{TA_i \cap TA_j} = \sum_{k \in (TA_i \cap TA_j)} R_k \tag{6}$$

$$R_{TA_i \cup TA_j} = \sum_{k \in (TA_i \cup TA_j)} R_k \tag{7}$$

In Eq. (4), higher ρ implies making a less number of users but more trustworthy users as friends while lower ρ means a more number of users but less trustworthy users as friends. A large size of a social network with low ρ does not necessarily ensure a high social capital of an individual user because the relationships with bad friends may ruin the user's reputation or introduce privacy intrusions.

D. Trust-based Privacy Model

When a user posts something, an automatic privacy setting can be provided with several levels of access rights based on an original poster's peer-to-peer trust estimation towards each friend. If the user wants to manually specify the privacy setting to a particular friend (e.g., family members), it can be allowed. Otherwise, a peer-to-peer trust estimation will be used to determine a privacy setting of an individual user. We consider the four types of access rights to each post:

- *View*: A friend can view a post but the friend's friends cannot view the post.
- *Comment*: A friend can comment or leave 'like' or 'dislike' on a post, but the friend's friends cannot view or comment on the post.
- *Share*: A friend can share a post and the friend's friends can view and comment on the post but cannot share the post.
- *Propagate*: A friend can determine whether the friend's friends can share or propagate the original post.

Note that this four types of access rights are proposed in this work as a novel approach which was not considered in the current OSN applications. An original poster i can determine whether his / her friend j can have a right to view, comment, share, or propagate based on j's trust estimated by i as:

$$a_x^{ij} = \begin{cases} 1 & \text{if } T_{ij}(t) \geq r_x \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where a_x^{ij} refers to an access right including view (a_v^{ij}), comment (a_c^{ij}), share (a_s^{ij}), or propagate (a_p^{ij}). a_x^{ij} is binary indicating $a_x^{ij} = 1$ for 'permitted' and $a_x^{ij} = 0$ for 'not permitted'. r_x is the minimum trust threshold user i can apply towards user j to determine whether user j can have a right in x, denoted as r_v, r_c, r_s, and r_p, respectively. In our experiments of Section V, we use same values for r_x for all trustors and trustees to investigate their impact on performance. However, r_x can be used as r_x^{ij} to indicate a different threshold value for each user to determine privacy levels based on his / her own risk appetite.

E. Friending Decision based on Social Capital

A user can employ the social capital threshold, τ, to make friending decisions by selecting friend j where $RSC_{ij} \geq \tau$ as shown in Eq. (4). With a linking probability p, user i will choose user j as a friend. This can be represented by:

$$e_{ij} = \begin{cases} 1 & \text{if } (RSC_{ij} > \tau) \wedge (P_r \leq p) \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where e_{ij} refers to an edge between users i and j indicating users i and j became friends when $e_{ij} = 1$; no friends otherwise. Since the network considered in this work is undirected, this leads to $e_{ij} = e_{ji} = 1$. RSC_{ij} is the value of social capital user i has towards user j. P_r is a random probability where p is a linking probability user i is connected to user j. That is, user i is connected with user j where user j's social capital to user i is higher than the threshold τ with the linking probability

p. For simplicity, we omit time t in e_{ij} and RSC_{ij} which are time-varying values changing based on the evolution of relationships. Note that lower τ reflects *bridging* by extending a social network with more dissimilar friends while higher τ considers *bonding* by making more similar friends.

F. Countermeasures Against Attacks

An adversarial user can exhibit malicious behaviors as described in Section III-C. We discuss how p2ReMon countermeasures those adversarial behaviors respectively as follows:

- *Self-promotion / use of fake identity*: An attacker cannot achieve high trust only with fake identity and promotion by using fake personal background because an overall trust is computed based on multiple types of interaction-based trust (e.g, positive evidence on feeding and feedback behavior) and similarity score based on common friends. The attacker may establish high reputation with a large number of other attackers as friends. However, by using a social capital threshold, τ, a user can make friending decisions by filtering out complete strangers or attackers who have no common friends with reputation no less than τ.
- *False information dissemination*: Our trust model considers positive interactions in feeding and feedback behaviors between two users. An attacker propagating false information can be easily isolated or evicted from the network by misbehavior report leading to account termination or terminating the relationship.
- *Illegal private information leak-out attack*: This is significantly mitigated by using a set of privacy thresholds, r_x. Since friending decision is made by using the social capital threshold, τ, vulnerability to the attacker can be minimized. However, upon the occurrence of this tragedy due to a user's detection error, the misbehavior is reported to the system which leads to the suspension of the attacker's account or the victim immediately quits the relationship (i.e., unfriending) which also lowers the attacker's trust, reputation, and social capital.

V. Results and Analysis

This section discusses the metrics, experimental setup, and simulation results with the interpretation of their trends.

A. Metrics

Correct Information Diffusion ($\mathcal{C}_f$): This measures the ratio of correct messages received by all users over all messages received by all users in a network and is computed by:

$$\mathcal{C}_f = \frac{\sum_{i \in U'} \sum_{m \in M} N_{i,m}}{\sum_{j \in U} \sum_{m \in M} N_{j,m}} \tag{10}$$

where U is a set of users i's in the network, U' is a set of users j's who received messages m's correctly. M is a set of messages m's propagated over the network. $N_{i,m}$ is the number of correct messages received by users i's while $N_{j,m}$ is the number of messages received by users j regardless of whether the message m is received correctly or not. In order to give an absolute sense of performance in the correct message delivery, we also show the average number of correct messages delivered to each user, denoted as N_f.

Private Information Leak-out ($\mathcal{P}_v$): This measures the ratio that users' information is leaked out to unauthorized parties or attackers attempting to perform illegal or false information propagation over the total number of messages received by users in the network, and is estimated by:

$$\mathcal{P}_v = \frac{\sum_{l' \in L \ \& \ l' \in C} \sum_{m \in M} N_{l',m}}{\sum_{l \in L} \sum_{m \in M} N_{l,m}} \tag{11}$$

where M is a set of messages m's posted by i's. L is a set of users who have a legitimate access right to message m. C is a set of adversarial users. $N_{l',m}$ is the number of messages m's received by users l''s who belong to both L and C while $N_{l,m}$ includes all users who received messages m's regardless of whether they are adversarial or not. Similar to N_f, we also show the average number of private information leaked out to adversarial users or unauthorized parties, denoted as N_v.

TABLE I: Scenarios of privacy thresholds

r_x	$\{r_v, r_c, r_s, r_p\}$	r_x	$\{r_v, r_c, r_s, r_p\}$
r_1	$\{0.1, 0.2, 0.3, 0.4\}$	r_4	$\{0.4, 0.5, 0.6, 0.7\}$
r_2	$\{0.2, 0.3, 0.4, 0.5\}$	r_5	$\{0.5, 0.6, 0.7, 0.8\}$
r_3	$\{0.3, 0.4, 0.5, 0.6\}$	r_6	$\{0.6, 0.7, 0.8, 0.9\}$

TABLE II: Key parameters and their default values

param.	val.	param.	val.	param.	value	param.	val.
N	332	t_x	1/3	$\|M\|$	10	ρ	0.5
P_{low}	0.5	P_c	0.2	p	0.1	T	1000

B. Experimental Setup

This work uses an ego-Facebook dataset [4], for network topology as the state of an initial network deployment. For this simulation study, we considered the total number of users, $N = 332$. To model the user's interactions with other users based on Sections III-B, we use behavior seeds as probabilities to trigger those interactions, denoted as $\mathcal{P}_i^f$ (feeding seed) and $\mathcal{P}_i^b$ (feedback seed) for each user i where they are selected as a random real number in $[P_{low}, 1]$. We vary the social capital threshold, τ, and the percentage of attackers, P_c, to study their impact on the performance metrics assuming that they consistently perform attacks whenever they show feeding or feedback behaviors based on their behavioral seeds, $\mathcal{P}_i^f$ and $\mathcal{P}_i^b$.

To consider six different scenarios, we vary a set of privacy thresholds, $\{r_v, r_c, r_s, r_p\}$, with 0.1 difference between thresholds as shown in Table I. In r_x for $x = 1-6$, r_1 is the lowest privacy level allowing most of users to access a user's posted information while r_6 is the highest privacy level allowing only users with high trust to access the information. Table II summarizes the key design parameters and their default values used for this experiment.

C. Sensitivity Analysis

In this section, we show our simulation results based on the metrics discussed in Section V-A including C_f, N_f,

P_v, and N_v. The experiment varies key design parameters, $\{r_v, r_c, r_s, r_p\}$, τ, and P_c to examine their impact on the performance metrics above.

Effect of social capital threshold (τ): Fig. 2 shows the performance of p2ReMon when varying τ and fixing the percentage of attacks with $P_c = 0.2$, with respect to a different set of privacy thresholds for r_x with $r = 1-6$. Recall that τ is used to determine who can be in a user's clique as explained in Section IV-C. Higher τ means that a user includes a friend who has a high social capital while lower τ implies that the user includes a friend with a low social capital. Using a set of low private thresholds (e.g., r_1 or r_2) means that most of a user's friends are allowed to view, comment, share, or propagate posted information by the user. On the other hand, using a set of high private thresholds (e.g., r_5 or r_6) means that only a small set of the user's friends who have high trust estimated by the user can access the posted information.

Fig. 2 (a) shows the correct messages over all messages received by all users who feed or provide feedback to their friends. We can observe a low performance particularly when r_x is at a medium level such as r_3. Under r_1 or r_2, more untrustworthy users are in a user's clique, and accordingly the user can detect the misbehavior of the untrustworthy friends through interactions, leading to filtering the untrustworthy users out by unfriending them (i.e., quitting the relationships) right after detecting them. Under r_3 - r_6, untrustworthy users do not have access to view, comment, share, or propagate posted information. In that case, there is a less chance for the posted information to be disseminated to untrustworthy users. However, when r_x is at a medium level such as r_3, users receiving the posted information may not have a sufficiently high trust and can be adversarial users who can modify and propagate false information over the network. This causes a low performance in disseminating correct messages over the network as shown in Fig. 2 (a).

While Fig. 2 (a) shows a relative performance over the total messages received by all users, Fig. 2 (b) shows the actual number of correct messages received by each user. This shows particularly when r_x is at r_4 or r_5, we notice a significantly high performance in correct information diffusion. Figs. 2 (c)-(d) show the results related to the private information propagated to adversarial or unauthorized parties. In Fig. 2 (c), since the absolute number of messages received by all users is relatively small at r_3, we observe the fraction of privacy leakout is maximized at r_3. But as shown in Fig. 2 (d), the actual number of privacy leakout occurred is maximized at r_4. From this result, a set of privacy thresholds, r_x, should be selected in order to balance the conflicting goals of maximizing correct information diffusion and minimizing illegal privacy information leakout. r_5 would be a good choice to have a sufficiently high performance in correct information diffusion (see Fig. 2 (b)) while reducing privacy leakout significantly (see Fig. 2 (d)).

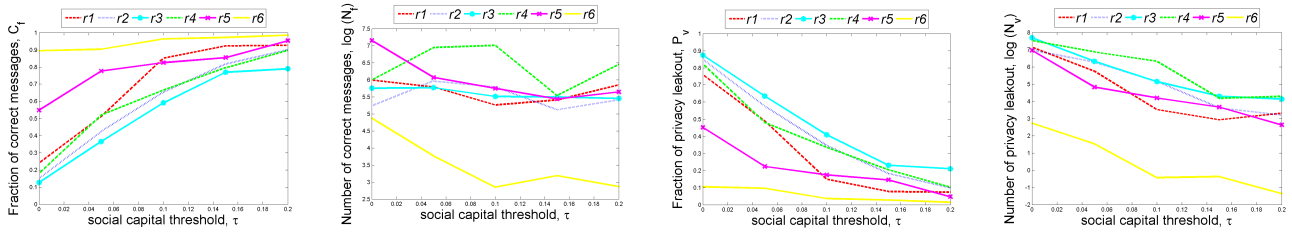
In Figs. 2 (a)-(d), too high τ minimizes privacy leakout while also hurting the performance in correct information diffusion. On the other hand, too low τ significantly increases privacy leakout and does not help correct information diffusion either. Thus, using $\tau = 0.05$ or 0.1 can make a balance between these two goals to achieve a sufficiently high correct information diffusion while not incurring too high privacy leakout.

Effect of attack density (P_c): Figs. 3 (a)-(d) show how the percentage of attackers, P_c, affects correct information diffusion and privacy leakout under varying a set of privacy thresholds, r_x. We fixed the social capital threshold $\tau = 0$ in Fig. 3. Similar to Fig. 2, there exists a clear point of r_x that maximizes the number of correct messages received by a user and privacy leakout in Fig. 3. Again, we need to choose right r_x such as r_5 which produces a sufficiently high correct information diffusion while not incurring too high privacy leakout as shown in Figs. 3 (b) and (d). This may be related to a user's risk appetite towards whether to take risk to have more benefits from the social network (i.e, social capital) or to take less risk to preserve his / her privacy right. As the effect of P_c, we can clearly notice that higher P_c reduces the fraction of correct information diffusion while introducing higher privacy leakout as in Figs. 3 (a) and (c). However, in terms of the actual number of correct messages delivered to a user and privacy leakout occurrences as shown in Figs. 3 (b) and (d), when P_c is low enough such as 0.1 or 0.2, there are relatively high privacy leakout occurrences in addition to a higher number of correct messages delivered. This is because more interactions can occur when there are more trustworthy users (i.e., not adversarial) and the interactions can lead to more situations that can face privacy leakout. When P_c is too high meaning more attackers in the network, they are mostly detected by their friends and the relationships are ended immediately after being detected. This case leads to less vulnerability to privacy leakout but also lowering correct information diffusion.

VI. Conclusion

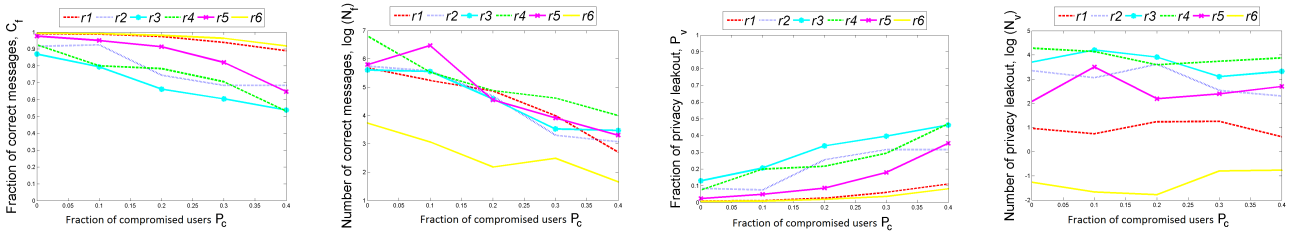
We proposed a trust and reputation model, named p2ReMon, that minimizes a user's privacy intrusion while maximizing social capital in an OSN. p2ReMon consists of the peer-to-peer trust-based privacy model and reputation-based social capital model. We devised a set of privacy thresholds to determine a friend's access right to the user's posts in the OSN. We also designed a social capital threshold used for friending decisions to maximize the user's social capital while concerning privacy intrusions. Our experimental results show that there exists a critical point of privacy thresholds or social capital thresholds that can maximize correct information diffusion and privacy leakout. In order to make a good balance between the two conflicting goals, we need to select a right set of privacy and social capital thresholds that can meet both requirements determined by a user's risk appetite.

As our future work, we plan to: (1) use a real dataset as a basis to derive the behavioral patterns of dynamic interactions between users using the NetLogo [17] reflecting the evolution of user's behavior and network structure; (2) consider other types of attacks including random attacks (i.e., varying attack intensity) to investigate its impact; (3) use a larger network



(a) C_f under τ vs. r_x (b) N_f under τ vs. r_x (c) P_v under τ vs. r_x (d) N_v under τ vs. r_x

Fig. 2: Effect of social capital threshold τ with respect to privacy threshold r_x for $k = 1 - 6$ on C_f, N_f, P_v, and N_v.



(a) C_f under P_c vs. r_x (b) N_f under P_c vs. r_x (c) P_v under P_c vs. r_x (d) N_v under P_c vs. r_x

Fig. 3: Effect of attacker density P_c with respect to privacy threshold r_x for $k = 1 - 6$ on C_f, N_f, P_v, and N_v.

size with more users; and (4) conduct comparative analysis with other existing trust or reputation counterparts.

REFERENCES

[1] J. Hart, C. Ridley, F. Taher, C. Sas, and A. Dix, "Exploring the facebook experience: A new approach to usability," in *5th Nordic Conference on Human–Computer Interaction: Building Bridges*, Lund, Sweden, 2008.

[2] Y. Zhang, L. Tang, and L. Leung, "Gratifications, collective self-esteem, online emotional openness, and traitlike communication apprehension as predictors of facebook uses," *Cyberpsychology, Behavior, and Social Networking*, vol. 14, no. 12, pp. 733–739, 2011.

[3] P. B. Brandtzaeg, M. Luders, and J. H. Skjetne, "Too many facebook "friends"? content sharing and sociability versus the need for privacy in social network sites," *Journal of HumanComputer Interaction*, vol. 26, no. 11–12, pp. 1006–1030, 2010.

[4] J. Leskovec. (2012) Social circles: Facebook. [Online]. Available: https://snap.stanford.edu/data/egonets-Facebook.html

[5] J.-H. Cho, K. S. Chan, and S. Adali, "A survey on trust modeling," *ACM Computing Survey*, vol. 48, no. 2, p. Article No. 28, Nov. 2015.

[6] S. Grabner-Krauter and S. Bitter, "Trust in online social networks: A multifaceted perspective," *Forum for Social Economics*, vol. 44, no. 1, pp. 48–68, 2015.

[7] J. L. Glanville and E. J. Bienenstock, "A typology for understanding the connections among different forms of social capital," *American Behavioral Scientist*, vol. 52, no. 11, pp. 1507–1530, 2009.

[8] R. D. Putnam, *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, 2000.

[9] M. Turculeta, "Ethical issues concerning online social networks," *Procedia-Social and Behavioral Sciences*, vol. 149, pp. 967–972, 2014.

[10] L. Guo, X. Liu, Y. Fang, and X. Li, "User-centric private matching for ehealth networks-a social perspective," in *IEEE Global Communications Conference*, 2012, pp. 732–737.

[11] L. Guo, C. Zhang, H. Yue, and Y. Fang, "Psad: A privacy-preserving social-assisted content dissemination scheme in dtns," vol. 13, no. 12, pp. 2903–2918, Dec. 2014.

[12] S. Grabner-Krauter, "Web 2.0 social networks: The role of trust," *Journal of Business Ethics*, vol. 90, pp. 505–522, 2009.

[13] F. K. Hussain and E. Chang, "An overview of the interpretations of trust and reputation," in *3rd Advanced International Conference on Telecommunications*, 2007.

[14] K. K. Kumar and G. Geethakumari, "Detecting misinformation in online social networks using cognitive psychology," *Human-Centric Computing and Information Sciences*, vol. 4, no. 14, pp. 1–22, 2014.

[15] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, Jan. 2008.

[16] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[17] U. Wilensky. (2016) Letlogo. [Online]. Available: https://ccl.northwestern.edu/netlogo/