Toward a Game Theoretic Model of Information Release in Social Media with Experimental Results

Christopher Griffin
Applied Research Laboratory
Penn State University
University Park, PA 16802
E-mail: griffinch@ieee.org

Anna Squicciarini
College of Information Science and Technology
Penn State University
University Park, PA 16802
E-mail: asquicciarini@psu.edu

***Abstract*—Social sites frequently ask for rich sets of user identity properties before granting access. Users are given the freedom to fail to respond to some of these requests, or can choose to submit fake identity properties, so as to reduce the risk of identification, surveillance or observation of any kind. However, this freedom has led to serious security and privacy incidents, due to the role users' identities play in establishing social and privacy settings. In this paper, we take a step toward addressing this open problem, by analyzing the dynamics of social identity verification protocols. Based on some real-world data, we develop a deception model for online users. The model takes a game theoretic approach to characterizing a user's willingness to release, withhold or lie about information depending on the behavior of individuals within the user's circle of friends. We provide an illustrative example and conjecture a relationship between the qualitative structure of Nash equilibria in the game and the automorphism group of the social network.**

***Index Terms*—Deception, Game Theory, Social Networks**

I. Introduction

Online social networks (OSNs) have experienced explosive growth in recent years and are nowadays de facto portal for hundreds of millions of Internet users. As the popularity of OSNs continues to grow, a huge amount of sensitive and private information is consistently uploaded to OSNs. As a result, users in OSNs are now content creators and managers, rather than just being content consumers. A typical OSN allows users to create connections to friends, thereby sharing with them a wide variety of personal information. These connections, however, are often based on the alleged identities and properties of social users populating the OSN. Users of social media sites can, however, generate accounts containing unverified attributes. On the one hand, this allows the users to avoid identification and surveillance or observation of any kind. On the other one, the ability to generate unverified accounts on most of these sites, which is crucial to preserving the privacy of honest citizens, renders social relationships weak and hard to control, if based on fake identities. Further, unverified accounts may and are often used by malicious users to carry out disruptive activities hidden behind fake identities. The intuitive solution to these issues is to provide stronger controls on the information provided by users on OSNs. This is generally impractical, due to the sheer volume of data consistently raised by social users. Further, users may rightfully choose to misrepresent, as a mechanism to protect their privacy. Finally, verifying an online identity is known to be a complex task, for a number of reasons. First off, as in the real world, online identity is a complex combination of several pieces of information of heterogeneous nature. Typically, a social user's identity includes social features (e.g., number of connections, preferences, activities), identifying attributes (e.g., name, location, age, gender), and the sum of the users activities. While some of these properties are inexpensive to verify (i.e. account or valid phone numbers), others are more complex, and harder to track (e.g., current employment status).

Further, while some work has studied the incentives behind information disclosures in OSNs [6], [7], little is known about identity misrepresentations. To date, Li and colleagues have shown that users' information disclosure is the result of the competing influences of exchange benefits and two type of beliefs: privacy protection belief and privacy risk belief [5]. Users are more likely to disclose personal information if the risks can be offset by benefits. Whether this information is truthful or not, and when users choose to deceive rather than withholding information, it has not been thoroughly studied.

In this paper, we speculate that information revelation in OSN is a complex process where multiple contrasting influences are in play: not only privacy attitudes, but also social pressure and personal attitudes are at stake. Focusing on three types of users' behavior related to information revelation: truthful information disclosure, withhold information and deception, we study the effect of misrepresentation in these environments. To ground our hypothesis, we conduct a preliminary survey collecting data about users' common behavior and their attitude toward personal information disclosure. Our study involves over 200 subjects. Based on the analysis of the responses, we design a preliminary model of deception using a generalized game theoretic model. The model presupposes that individuals release, withhold or lie about certain classes of information based on individual payoff functions whose output is affected by the behavior of a circle of close friends. We provide an example model as well as results on the Nash equilibria in this model and a conjecture on the nature of Nash equilibria and their relationship to the automorphism class of the social network.

DOI 10.1109/SPW.2012.24

II. Exploratory study

In order to understand and model what is typical of one's identity in social platforms, we proceeded with collecting data from real users through survey-based methods.

We notice that it is generally extremely difficult to obtain large amounts of valid and detailed personal information about the users of social sites. Even more challenging is ascertaining the truthfulness and provenance of the information provided. Specifically, it is difficult to determine which identities are duplicate profiles belonging to the same actual user, and which properties of a profile are fake. Our surveys allows us to analyze small high-quality data on questions related to these issues. The aim of our exploratory study is to gain a deeper understanding of users identity-revealing actions, the peculiar features of average users, and the perceived understanding of identity on social sites. Specifically, we conducted an initial survey of a group of 200 participants, aged between 20 and 35 (avg 24, sd=2.34). The respondents were 65% women and 35% male. 99.3% of them declared to have at least one account on social sites, and 12% declared to have more than one account on a same OSN. 83.6% of the participants declare to browse their favorite OSN sites one or more times a day, and 73.7% indicate that their OSN profile is comprehensive, and it is easy for them to be identified by it.

The survey was constructed to collect information about three specific aspects of users' behavior (1) privacy awareness, (2) attitude toward information withholding and practices (3) attitude toward lies and misrepresentation. We here summarize some of the most interesting findings used to inform our model.

Our initial findings reveal that a social network user's tendency to lie is highly correlated with his or her desire to portray a successful social image, and only weakly related to privacy concerns. In other terms, the perceived usefulness of the social network service increases online users willingness to disclose their personal information. Service providers also try to promote the idea of successful "social image": rich user profiles have a significant economic value for providers, who therefore consistently increase the amount of user information requested. Our respondents appear aware of the pressure imposed by the social network sites, but are still heavily committed to information disclosure.

In our survey, most of the participants claimed to misrepresent only specific pieces of information, demonstrating that peer pressure leads to truthful revelation of basic identity properties, such as gender, age, etc., for which lying would be useless. This also suggests that lying is considered convenient only for certain classes of information. For example, it is a huge disadvantage to lie about an attribute, such as a personal website address, that is revealed in the course of social interaction with other connected users, but it may be convenient to lie about an attribute like dating status that is less easily discovered through online interaction. The results also show strong correlation between peer-pressure and attitude to lie. Users who feel peer-pressured are more likely to lie about some of their information, especially their whereabouts. Surprisingly, we found no significant statistical correlation among users' attitude toward withholding information and privacy. Also, users do not connect lying or withholding information on a social networking site with actual lying associated to morality. Rather, this is perceived as a mechanism for boundary control.

Some other interesting findings were related to the existence and importance of inner circles. Despite the complex social connections tying users together, users are most strongly influenced by a small set of connections whom they interact with regularly and whose opinion counts to them. Most of the actions (e.g., comments and feedback) users perform involve inner-circle users, who are the ones influencing users decisions about lying and not lying.

It also appears that awareness of privacy controls is not directly reflected in users actions. The responses from our participants confirm the well-known privacy paradox [1]: users are aware of privacy risks and possible information leakage [3], but there is no correlation between privacy awareness and the amount of personal information they choose to reveal. Instead, peer pressure and the need to establish a successful image are what leads a user to reveal detailed information. This is also confirmed by the motives given for lying. There is no strong correlation between lies and privacy awareness: rather, it appears that the strongest motive is social image.

III. Initial Model

Based on our experimental results, we have identified that users treat types of information differently with respect to whether they disclose, withhold or deceive. Furthermore, we know that the behavior of users is highly dependent on the behavior of a small group of their immediate network neighbors. Let $G = (V, E)$ be a user graph for a social network and suppose we have several classes of information $\mathcal{I} = \{1, \ldots, m\}$. Let $r_i^{(j)} \in [0, 1]$ be the proportion of information type i that Player j will release and let $q_i^{(j)} \in [0, 1]$ be the proportion of information type i about which Player j deceives. Then $w_i^{(j)} \in [0, 1]$ is the proportion of information type i that Player j withholds. Then we have:

$$r_i^{(j)} + q_i^{(j)} + w_i^{(j)} = 1 \tag{1}$$

Let:

$$\bar{w}_i^{(j)} = \frac{1}{|N(j)|} \sum_{k \in N(j)} w_i^{(k)} \tag{2}$$

where $N(i)$ is the neighborhood of Player j in G. We make similar definitions for $\bar{r}_i^{(j)}$ and $\bar{q}_i^{(j)}$. For Player j and information type i, we assume there is a function $f_i^{(j)}(r^{(j)}, q_i^{(j)}, \bar{r}_i^{(j)}, \bar{q}_i^{(j)})$ that returns the benefit for releasing a certain amount of information and that this function is dependent on the average proportion of information released so that if $q^{(j)} + r^{(j)} < \bar{q}_i^{(j)} + \bar{r}_i^{(j)}$ then some penalty is incurred and presumably there are diminishing returns to information disclosure when $q^{(j)} + r^{(j)} > \bar{q}_i^{(j)} + \bar{r}_i^{(j)}$. This function could

be piecewise linear with e.g., a slope change point at the group average.

In addition, we suppose that for each Player j and each information type i there is a function $g_i^{(j)}(q^{(j)}, \bar{q}_i^{(j)})$ that determines the cost of deceiving the group and that it is dependent on the average quantity of deception so that a player suffers a greater cost if $q^{(j)} > \bar{q}_i^{(j)}$ or perhaps if $q^{(j)} < \bar{q}_i^{(j)}$. This function may also be piecewise linear with a slope change at the group average.

Lastly, we assume there is a function $h_i^{(j)}(q^{(j)})$ that is an increasing cost function for deceiving that the player suffers as a result of his morals. Then the net payoff function for Player j is:

$$\pi^{(j)} = \sum_i \Big(f_i^{(j)}(r^{(j)}, q_i^{(j)}, \bar{r}_i^{(j)}, \bar{q}_i^{(j)}) - g_i^{(j)}(q^{(j)}, \bar{q}_i^{(j)}) - h_i^{(j)}(q^{(j)}) \Big) \quad (3)$$

Since $r_i^{(j)}$, $w_i^{(j)}$ and $q_i^{(j)}$ must be between 0 and 1 and sum to 1, any Nash equilibrium of this game is constrained in the convex polytope:

$$\Omega = \prod_i \Big\{ (r_i^{(j)}, q_i^{(j)}, w_i^{(j)}) \in [0,1]^3 : r_i^{(j)} + q_i^{(j)} + w_i^{(j)}) = 1 \ , \ r_i^{(j)}, q_i^{(j)}, w_i^{(j)} \geq 0 \Big\} \quad (4)$$

Proposition III.1. *Suppose that $f_i^{(j)}(r^{(j)}, q_i^{(j)}, \bar{r}_i^{(j)}, \bar{q}_i^{(j)})$ is concave for all i and j, $g_i^{(j)}(q^{(j)}, \bar{q}_i^{(j)})$ and $h_i^{(j)}(q^{(j)})$ are convex for all i and j then there is a Nash equilibrium in Ω for this game.*

Proof: See Theorem 1 of [4]. ■

The uniqueness of a Nash equilibrium in this case is completely a function of the structure of specific objective functions.

IV. Example

For simplicity of analysis, we look at a simple example in which individuals must simply choose how much information to withhold vs. how much (true) information to divulge, since our experience with actual students suggests that there is more motivation to simply withhold information rather than to present false information. Additionally, we will assume that there is exactly one type of information, since information types do not appear coupled in our empirical surveys.

Let $r_i \in [0,1]$ be the proportion of available information that Player i divulges. Define:

$$\bar{r}_i = \frac{1}{|N(i)|} \sum_{j \in N(i)} r_j \quad (5)$$

Then the payoff function for Player i is:

$$\pi^{(i)}(\mathbf{r}) = \log(r_i - \alpha \bar{r}_i) - \beta (r_i - \kappa \bar{r}_i)^m - \gamma r_i + M \quad (6)$$

Here:

$$f_i^{(j)}(r^{(j)}, q_i^{(j)}, \bar{r}_i^{(j)}, \bar{q}_i^{(j)}) = \log(r_i - \alpha \bar{r}_i) \quad (7)$$

$$g_i^{(j)}(q^{(j)}, \bar{q}_i^{(j)}) = \beta (r_i - \kappa \bar{r}_i)^m \quad (8)$$

$$h_i^{(j)}(q^{(j)}) = \gamma r_i + M \quad (9)$$

In short, each player receives a benefit for divulging information given by $\log(\alpha r_i + \rho \bar{r}_i)$ that has decreasing marginal return as r_i increases and is positively affected by the average amount his friends divulge. Each player incurs a cost of $\beta (\kappa r_i + \lambda \bar{r}_i)^m$ associated with divulging information with increasing marginal cost as r_i increases and that is affected by the average information released.

For $N = 3$, assuming $r_1 = r_2 = r_3$, the payoff function to each player has shape given in Figure 1. In this figure, $\alpha = \kappa = 0.1$, $\beta = 1$, $\gamma = 2$ and $M = 3$.

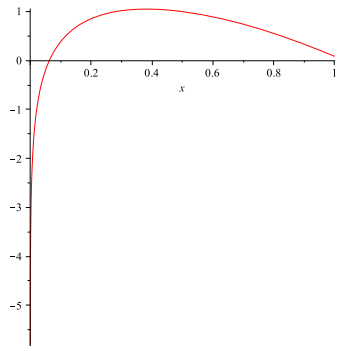

Fig. 1. The payoff function for Player i in a three player game in which $\alpha = \kappa = 0.1$, $\beta = 1$, $\gamma = 2$ and $M = 3$.

Theorem IV.1. *If $1 - \alpha > 0$ and $\alpha > 0$ and*

$$1 - \alpha - \beta (1 - \kappa)^{m-1} m (1 - \kappa)(1 - \alpha) - \gamma (1 - \alpha) < 0 \quad (10)$$

then there is a Nash equilibrium $r_1^* = r_2^* = \cdots = r_N^*$ with $r_i \in [0,1]$ for $i = 1, \ldots, N$ irrespective of the graph structure G.

Proof: Suppose that $r = r_1 = r_2 = \cdots = r_N$. Then

$$\frac{\partial \varphi(\mathbf{r})}{\partial r_i}|_{r_i = r} = \frac{1 - \alpha}{r - \alpha r} - \beta (r - \kappa r)^{m-1} m (1 - \kappa) - \gamma \quad (11)$$

Setting Equation 11 to zero and solving yields:

$$z(r) = 1 - \alpha - 2\beta (r - \kappa r)(r - \alpha r)(1 - \kappa) - \gamma (r - \alpha r) \quad (12)$$

We have $z(0) = 1 - \alpha$ and $z(1)$ is given by:

$$1 - \alpha - \beta (1 - \kappa)^{m-1} m (1 - \kappa)(1 - \alpha) - \gamma (1 - \alpha) \quad (13)$$

Given the assumptions, the intermediate value theorem yields the desired result. ■

V. Asymmetric Equilibria

What the result in the previous section tells us is that in a completely homogenous population, the graph structure is completely irrelevant. Individuals may converge to a unique population equilibrium. This is not the case when the population is not homogeneous. Consider the graph shown in Figure 2. If the players on the right set $\beta = 4$, meaning

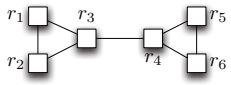

Fig. 2. A simple graph with assymetric payoff.

they are more sensitive to the cost of disclosing information in the system, while player on the left are less sensitive to information disclosure and $\beta = 1$, then an equilibrium solution is:

$$\begin{aligned} r_1^* = r_2^* &= 0.4064752728 \\ r_3^* &= 0.4025221076 \\ r_4^* &= 0.2819505635 \\ r_5^* = r_6^* &= 0.2779973982 \end{aligned}$$

These values indicate the proportion of an information heap to disclose, with the players on the left divulging more information, consistent with their insensitivity to privacy concerns. Note also that even though Player 3 is less sensitive to disclosing information, he still discloses less information than the other member of his clique. An opposite fact is true for Player 4. There are still quasi-diagonal solutions to this problem. This can be explained by the following conjecture.

Conjecture V.1. *There is a class of payoff functions Π_{AUT} so that given a set of payoff functions $\pi^{(i)}$ drawn from Π_{AUT}, if there is an automorphism of G, $\xi : V \to V$, with the property that $\pi^{(i)} \equiv \pi^{\xi(i)}$ for all $i = 1, \ldots, N$. Then there is an equilibrium solution in which $r_i^* = r_{\xi(i)}^*$.*

A critical problem for future research is qualifying the class Π_{AUT}. Likewise, we believe that properties like the one given in the foregoing conjecture can aid in determining payoff functions. For example, if a (small portion) of a social network is known to have an automorphism that does not preserve observed equilibrium values, then the payoffs of the players are either not elements of Π_{AUT} *or* they are not preserved under the automorphism. This gives us insight into the payoff functions of the individuals players.

VI. Future Directions

In this work, we have shown our initial work on deception and misrepresentation in OSNs.

Based on a user centered study, we have formulated a model for deception and identity representation in these environments.

This work is still at an infancy stage. We are left with a number of unsolved questions, that we plan to explore in the near future. First, we are interested in collecting more detailed data from real-world users, to deepen our understanding of users' interactions and identity revelation processes. For example, in the current study we did not focus on the users' actions, that are used as a vectors for identity disclosure. What are the typical passive social transactions (post an item on your page which may be silently consumed by those who've been given access to it) or active (sharing, comments, likes), and how do different outcomes of such transactions affect social images, and therefore lead toward truthful and untruthful behavior? How are secondary (friend-of-friend, triad) relationships learned and what impact do they have on identity revelation, if any? What role does context (communities of interest) play in the above? Results obtained from these studies will guide the next step of our research, to construct informed and therefore accurate authentication models.

Lastly, we are interested in studying some of the theoretical interactions between graph theory and game theory in the presentation of various equilibria. We hypothesize that the observation of actual behavior viz. information withholding and deception along with an observation of the social network graph structure may give substantial insight into the nature of the objective functions used by social network users.

Acknowledgement

Portions of Dr. Griffin's work were supported by the Army Research Office under Grant W911NF-11-1-0487.

References

[1] Susan B. Barnes. A privacy paradox: Social networking in the United States. *First Monday*, 11(9), September 2006.

[2] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 551–560, New York, NY, USA, 2009. ACM.

[3] Hiawatha Bray. Privacy still a nagging concern on Facebook, 2010.

[4] J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534, 1963.

[5] Li, H. and Sarathy, R. and Xu, H. Understanding situational online information disclosure as a privacy calculus, Journal of Computer Information Systems, 51(1):52–71, 2010.

[6] J. DiMicco and D. Millen. Identity management: multiple presentations of self in Facebook. In *GROUP '07: Proceedings of the 2007 international ACM conference on Supporting group work*, pages 383–386, New York, NY, USA, 2007. ACM.

[7] F Stutzman. An evaluation of identity-sharing behavior in social network communities. *iDMAa Journal*, 3(1), 2006.