Determining a Discrete Set of Site-Constrained Privacy Options for Users in Social Networks Through Stackelberg Games

Sarah Rajtmajer[1](✉), Christopher Griffin[2], and Anna Squicciarini[3]

[1] Department of Mathematics, The Pennsylvania State University,
State College, USA
smr48@psu.edu

[2] Mathematics Department, United States Naval Academy, Annapolis, USA
griffinch@ieee.org

[3] College of Information Sciences and Technology,
The Pennsylvania State University, State College, USA
acs20@psu.edu

Abstract. The privacy policies of an online social network play an important role in determining user involvement and satisfaction, and in turn site profit and success. In this paper, we develop a game theoretic framework to model the relationship between the set of privacy options offered by a social network site and the sharing decisions of its users within these constraints. We model the site and the users in this scenario as the leader and followers, respectively, in a Stackelberg game. We formally establish the conditions under which this game reaches a Nash equilibrium in pure strategies and provide an approximation algorithm for the site to determine a discrete set of privacy options to maximize payoff. We validate hypotheses in our model on data collected from a mock-social network of users' privacy preferences both within and outside the context of peer influence, and demonstrate that the qualitative assumptions of our model are well-founded.

1 Introduction

At its core, an online social network (SN) is an infrastructure for user-generated shared content. Users have the ability to exercise control over their individual channels in the network, by deciding which content to share and with whom to share it. The SN site benefits from shared content in important ways. Shared content attracts new users, deepens the involvement of existing users, strengthens the community, and can be leveraged for monetization.

Individual behavior online, like individual behavior offline, is also subject to social norms and peer influence [12,15,24]. Notions of what is appropriate in content sharing online is defined comparatively, so that subtle shifts in local behavior may have much farther-reaching consequences for the network as a whole. In sum, unlike the SN site which is ultimately a business operating with

MHR Khouzani et al. (Eds.): GameSec 2015, LNCS 9406, pp. 208–227, 2015.
DOI: 10.1007/978-3-319-25594-1_12

a business model, users are individuals with more complex incentives, concerns and considerations operating voluntarily within the constraints of the SN.

Questions related to privacy in SNs have gained increasing interest over the last few years as the ubiquity of social media has become apparent and anecdotes of repercussions for over-disclosure more available. Many users are now aware of the risks associated with revelation online and concerned with protecting personal information from widespread dissemination. Advocates of fine-grained privacy policies argue that detailed user management of privacy settings for shared content can avert some of the potential risks users face in online SNs [20,28]. Users can sort their data into categories to be shared with certain individuals in the network (i.e., friends, friends of friends, groups, everyone). SNs like Facebook and Google+ have implemented this model, allowing users to create narrower social circles from among their list of friends and to define which content is shared with whom. Unfortunately, studies have also shown that users often do not take advantage of finely-tuned options available to them. The majority of users on both Facebook and Twitter maintain the default privacy settings established by the site [12,19], which tend to be more permissive than users would like [23].

In this work, we focus on the topic of privacy, from the perspectives of both the SN site and its users. We seek to determine an optimal discrete set of privacy options to be made available to users for content sharing. We define optimality here from the perspective of the site, taking into account user satisfaction. Intuitively, the site is to choose a set of options for users' shared content in order to maximize sharing. Yet, the site should allow users to maintain a level of control over their content without being overwhelmed by too many or too complex privacy settings from which to choose.

We model the conflicting yet complementary goals of the SN site and its users as a Stackelberg game whereby the leader (the site) moves first in setting the privacy options to be made available to user-members for shared content. Followers (users) respond by selecting privacy settings from among these options. Payoff to the site can be expressed in terms of amount of shared content and total user happiness. Payoff to each user depends on how closely the available options approximate his ideal sharing preferences, which is in turn a function of an inherent comfort and peer influences. We formally present this two-level game as well as a characterization of its convergence to a Nash equilibrium in pure strategies under certain simplifying assumptions. We develop an agent-based model to approximate optimal strategies on arbitrary network graphs and validate model assumptions with a study of 60 individuals, run over a mock-SN.

The remainder of this paper is organized as follows. The next section reviews related work, followed by our problem statement, succeeded by an overview of our model in Sect. 4. Section 5 presents approximation algorithms, and Sect. 6 discusses the experimental study we carried out. We conclude the paper in Sect. 7.

2 Related Work

The scale and gravity of privacy and security risks associated with online social networks have led to a rich body of work addressing a wide spectrum of these issues. By sharing their personal information, users in SNs become vulnerable to attacks from other users, the SN itself, third-party applications linked to their SN profiles, or other outside attackers able to de-anonymize user data published by the SN site. See [2,18] for recent reviews. These attacks may take the form of identity theft [12], scraping and harvesting [21], social phishing [17], or automated social engineering [3]. The risk of a breach of privacy in some form is particularly salient for users who are not closely monitoring their privacy settings or leaving privacy settings at their default values.

As a means of mediating some of these risks, there is a growing literature using machine learning to determine individual default privacy settings. *PriMa* [31] and *Privacy Wizard* [8] are examples of supervised learning algorithms which look at the behavior and preferences of a user, the behavior and preferences of his peer group or related users, and offer a classification of default settings for different types of shared content. We see this work as complementary to ours in that it does not suggest a method for the determining the privacy settings from which a user may choose, but rather once these options are in place, gives a method for selecting defaults amongst them which may most closely match a user's preferences.

This work is related in general to the body of work on game theory in social networks, both offline and online. Fundamental research efforts exploring cooperation in structured human populations include [23,26,38]. In the realm of online social networks, game theoretic models have been implemented for the study of the evolution of various social dilemmas and associated changes in network structure [9,16,25].

Most closely related to our work is the subset of this research concerning agent-based decision-making related to privacy and security in online social networks. Chen and colleagues model users' disclosure of personal attributes as a weighted evolutionary game and discuss the relationship between network topology and revelation in environments with varying level of risk [5].

In a series of papers considering the circumstances of deception in online SNs, Squicciarini et al. characterize a user's willingness to release, withhold or lie about information as a function of risk, reward and peer pressure within different game-theoretic frameworks [29,33]. They describe the relationship between a site and its users, determining that in the in the presence of a binding agreement to cooperate (strong guarantees on privacy), most users will agree to share real identifying information in return for registration in the system [34]. Authors also use a game theory to model of collective privacy management for photo sharing in SNs [32,35]. Their approach proposes automated privacy settings for shared images based on an extended notion of content co-ownership.

To the best of our knowledge, a game-theoretic approach to determining the privacy policy of an online SN has not been considered before in the literature.

In a previous work [11], we tackled the simpler question of determining a mandatory lower-bound on shared content. That is, we have addressed the SN site's decision of selecting the minimum amount of shared personal information which should be required of user with an active account in the network. For example, Facebook requires all users with a personal account to give a first name, last name, valid email address, password, gender and birth date. In fact, Facebook institutes further sharing requirements on various elements of a user's profile, e.g., a user's cover photo is always public [6].

3 Problem Statement

We assume a captive social network site, wherein users share pieces of personal content freely within the network and possibly with selected subgroups of network users, according to a set of privacy options for shared content made available by the site to its users.

We assume the site benefits when users share as freely as possible and it is of course incentivized to create options that promote the widest distribution of posted content. The site, however, must also be wary to consider users who are inherently more cautious about public sharing. A site requiring all shared content to be public, for example, may lure some users to post publicly who might otherwise have only shared with a narrower group, i.e., "friends only". But in other cases, this policy might have a detrimental effect for the site, as users may choose not to post at all. In any case, if the privacy setting a user would prefer for a piece of content is not presented the user will experience some degree of dissatisfaction in having to select an alternative. Figure 1 illustrates the problem space.

Users react to the options offered by choosing what to disclose and with whom. Examples of these settings in practice may include "visible to only me", "share with specific individuals", "share with friends", "share with my network"and "public". We abstract away from the details of how privacy options

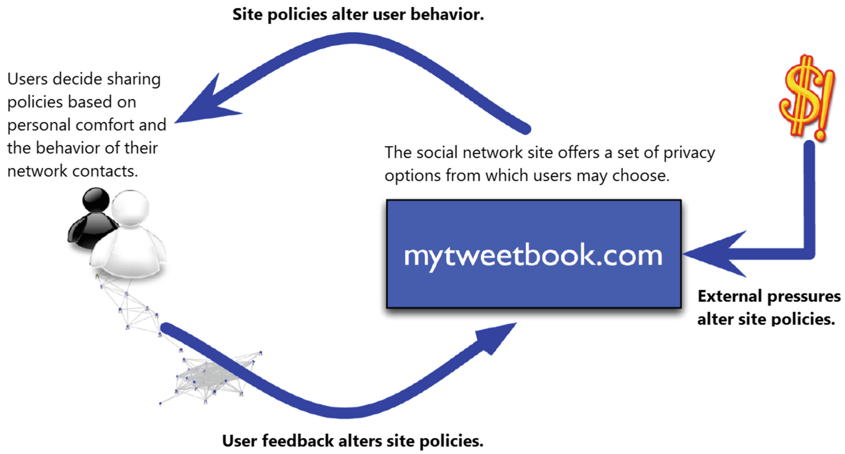

Site policies alter user behavior.
Users decide sharing policies based on personal comfort and the behavior of their network contacts.
$!
The social network site offers a set of privacy options from which users may choose.
mytweetbook.com
External pressures alter site policies.
User feedback alters site policies.

Fig. 1. There is a natural push and pull between a SN site and its users with regard to sharing policies.

are presented to users, and map them to real values on the interval $[0, 1]$. The granularity of these options should be fine enough to meet users' needs, but coarse enough to be manageable in implementation for both the users and the SN site.

We formulate the site's utility as a function of user happiness and shared content, so that minimally the site would like to make sure that no user is unable to share content as freely as he would like due to a lack of available sharing options. In fact, the site would stand to profit by pushing users toward the upper boundary of their sharing comfort, and having a carefully chosen set of options may enable this to happen.

We model each user's utility function as a weighted sum of discomfort and peer pressure. Specifically, each user will act to minimize the difference between his selected privacy setting and his personal comfort level, and the difference between his selected privacy setting and the average privacy settings of his peers. The intuition is that users have an inherent degree of disclosure they feel most comfortable with, but are also influenced by their peers when making sharing decisions [7,14]. Since these two dimensions may not be considered equally for all users, we introduce weights to capture interpersonal differences in susceptibility to peer pressure. Precisely, we offer the option of including weights on either the peer pressure or personal comfort components of the user's utility function allowing customization of the model for non-homogeneous users and an opportunity to strengthen the model in the presence of additional information on user behavior, which the site may learn through observation.

4 Model Overview

We define two optimization problems: one for the SN user and one for the SN site. The optimal solutions to these problems determine the behavior of the user and site regarding privacy policies.

4.1 User Model

Our user model extends the model presented in [11] for the modeling of a lower-bound on information disclosure for membership in the SN. The motivations and actions of users with respect to content sharing in this framework are consistent with this prior work, but will be enacted within the constraints of the site's problem which is significantly different.

Assume a SN is represented by a graph $G = (V, E)$, where V is a set of users (represented by vertices in the graph) and E is the set of social connections (edges) between them. For the remainder of this paper, assume $|V| = N$. Users post information to the SN for reasons known only to themselves. Unlike in [30], we assume users who are perfectly honest, but may choose to omit (or keep private) a certain amount of information. Previous work has observed [10,30] that users have distinct sharing behaviors for different types of information, depending on the "social" value of such information (e.g., users are more willing to share their gender than their phone number). Assume there are M types of information. Since it is nontrivial to specify what a piece of information corresponds to in a SN, we will abstract away from any specific characterization of information, and assume User $i \in V$ accumulates *postable* information of type j at a rate of $\beta_i^j(t)$ (given in bits per second). Each user chooses a proportion (probability) of information of type j to share, denoted by $x_i^j(t) \in [0, 1]$.

In general, users do not change their privacy policy frequently [22], and thus we can consider a simplified problem in which we attempt to find optimal values for (fixed) x_i^j ($i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, M\}$). To do this, we define optimality in terms of:

1. Peer Pressure (and reputation),
2. Comfort level

Comfort level in the context of privacy and information disclosure refers to the degree of disclosure users feel comfortable with. This notion, often used to characterize information sharing in online sites (e.g. [1,7]), is also adopted in our model. Users reaching their optimal comfort level wish not to change any of their information sharing practices. Reputation and peer pressure are self-explanatory, and are combined in a single dimension as they are highly correlated [30].

Without loss of generality, focus on *one* information type, $x_i \in [0, 1]$. To model peer pressure, we assume that individuals are encouraged to behave in accordance with the norms of their social group. Thus for User i, we define:

$$\bar{x}_{-i} = \frac{\sum_{j \in N_G(i)} v_{ij} x_j}{V_G(i)}$$

where $v_{ij} \geq 0$ and

$$V_G(i) = \sum_j v_{ij} \tag{1}$$

is the weighted neighborhood size of i in G. If $v_{ij} = 1$ for all j, then $V_G(i) = |N_G(i)|$, the size of the neighborhood of i in G. The neighborhood may be defined in terms of the social graph of the user, or it may be a more restrictive subset of peers with whom the user actively interacts. Let the peer pressure function for User i be given by:

$$P_i(x) = v_i f_P(x - \bar{x}_{-i}) \tag{2}$$

where f_P is a concave function with maximum at 0 and $v_i \geq 0$ is the subjective weight User i places on the peer pressure function. Thus, the payoff $P_i(x)$ is maximized as x_i approaches $\bar{x}_{-i}$.

We note that an alternate and equally reasonable approach to defining $P_i(x)$ is as:

$$\tilde{P}_i(x) = \sum_{j \in N_G(i)} v_{ij} f_P(x - x_j) \tag{3}$$

where $v_{ij} \geq 0$. In this case, User i attempts to minimize a weighted function of the difference in privacy levels from all of his neighbors simultaneously.

Estimated weights on the link between User i and User j might be obtained, for example, as a function of the frequency and type of online interactions between them. This formulation increases the complexity of the problem and ultimately makes computation more cumbersome, but allows a richer model when more detailed information about users' relationships and peer influence is present.

By similar argument, assume that User i has a sharing level x_i^+ at which he is happiest. The comfort function $f_C(z)$ for User i is given by:

$$C_i(x) = w_i f_C(x - x_i^+)$$

for $w_i \geq 0$, which can be thought of as a user's tendency to act in preference to his own comfort rather than in response peer pressure. Here again, f_C is concave with maximum at 0, so that the comfort of User i is maximized as x_i approaches x_i^+.

In practice x_i^+ may be difficult to determine for an unknown User i. However, we assume that based on user demographics, as well as observed overall user behavior for a mass of users, either at the individual or group level, it is possible to infer of x_i^+, or at least an expected value $E[x_i^+]$ within a tolerated window of error.

Thus, the total objective function for User i is:

$$J_i(x_i; x_{-i}) = P_i(x_i) + C_i(x_i) = v_i f_P\left(x_i - \frac{\sum_{j \in N_G(i)} x_j}{|N_G(i)|}\right) + w_i f_C(x_i - x_i^+) \quad (4)$$

or, the weighted variant:

$$\tilde{J}_i(x_i; x_{-i}) = \tilde{P}_i(x_i) + C_i(x_i) = \sum_{j \in N_G(i)} v_{ij} f_P(x_i - x_j) + w_i f_C(x_i - x_i^+). \quad (5)$$

Here, x_{-i} indicates the privacy choices of all other users besides i and we write $J_i(x_i; x_{-i})$ to indicate that User i's utility is a function not only of his own decisions, but also of the decisions of the other users.

When f_P and f_C are concave, the following proposition holds [27]:

Proposition 1. *Assume that each x_i is constrained to lie in a convex set $X_i \subseteq [0, 1]$ for $i = 1, \ldots, N$. There is at least one value x_i^* for each User i so that every user's objective function is simultaneously maximized and $(x_1^*, \ldots, x_N^*)$ is a Nash Equilibrium for the multi-player game defined by any combination of objective functions $J_1, \ldots, J_N$ or $\tilde{J}_1, \ldots, \tilde{J}_N$.* □

By similar reasoning, the preceding proposition can be extended to the case of multiple independent information types. In this case for each $j = 1, \ldots, M$ there is an equilibrium solution $x_i^{j^*}$ $i = 1, \ldots, N$. Correlated payoffs for information sharing among information types are beyond the scope of the current work.

In general, in this case, each user would have an information sharing strategy $\mathbf{x}_i \in [0,1]^M$ and a corresponding multi-dimensional payoff function. The existence of a Nash equilibrium would be guaranteed for convex functions with convex constraints.

4.2 Site Model for the Determination of a Discrete Set of Privacy Options for Shared Content

For the remainder of this paper, we will assume a user objective function of the form $\tilde{J}_i$ and fix $f_C(z) = f_P(z) = -z^2$, which is concave with maximum at zero. Furthermore, and for notational simplicity, we will consider the minimizing form of the problem in which User i minimizes $-\tilde{J}_i$.

Assume the site offers a discrete set of privacy settings $l_1, \dots, l_K \in [0,1]$. Each user must choose from among these options for each piece of shared content. This is equivalent to choosing a generic privacy policy within a social network. Let $\mathbf{l}$ be the vector of these options. Define:

$$y_{ij} = \begin{cases} 1 & \text{Player } i \text{ chooses privacy level } j \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

these binary variables indicate the privacy levels of each player. Naturally we require:

$$\sum_j y_{ij} = 1 \tag{7}$$

Let $\mathbf{y}$ be the matrix of y_{ij} values. Furthermore:

$$x_i(\mathbf{y};\mathbf{l}) = \sum_{j=1}^{K} y_{ij} l_j$$

For given values y_{ij} ($i = 1, \dots, N$ and $j = 1, \dots, K$), the payoff to Player i is:

$$H_i(\mathbf{y};\mathbf{l}) = \sum_{j \in N(i)} v_{ij}(x_i - x_j)^2 + w_i(x_i - x_i^+)^2 \tag{8}$$

Note, this is simply $-\tilde{J}_i$. Then the net payoff to the site is:

$$J(\mathbf{y};\mathbf{l}) = \sum_i \left(\sum_j \pi_j y_{ij} - \lambda H_i \right), \tag{9}$$

where π_j is the benefit the site receives for a piece of content shared with privacy setting j and λ is the weight applied to the payoff of the users; i.e., the weight the site places on user happiness. When $\mathbf{y}$ is determined endogenously by the players, then the site's bi-level combinatorial optimization problem is:

$$\begin{cases} \max_{\mathbf{l}} \ J(\mathbf{y};\mathbf{l}) = \sum_i \left(\sum_j \pi_j y_{ij} - \lambda H_i \right) \\ \quad s.t. \ l_1, \ldots, l_K \in [0,1] \\ \qquad l_j \le l_{j+1} \quad j = 1, \ldots K-1 \\ \qquad \forall i \begin{cases} H_i(\mathbf{y};\mathbf{l}) = \min_{\mathbf{y}_i} \ \sum_{j \in N(i)} v_{ij}(x_i - x_j)^2 + w_i(x_i - x_i^+)^2 \\ \qquad s.t. \ x_i = \sum_{j=1}^{K} y_{ij} l_j \\ \qquad \sum_j y_{ij} = 1 \\ \qquad y_{ij} \in \{0,1\} \end{cases} \end{cases} \tag{10}$$

In this problem, each User i must decide the value of y_{ij} independently of all other users, while being simultaneously affected by her choice. It is clear that the sub-game has a solution in mixed strategies from Proposition 1, but what is less clear is whether it has a solution in pure strategies.

Consider the user game-theoretic sub-problem:

$$\forall i \begin{cases} H_i(\mathbf{y};\mathbf{l}) = \min_{\mathbf{y}_i} \ \sum_{j \in N(i)} v_{ij}(x_i - x_j)^2 + w_i(x_i - x_i^+)^2 \\ \qquad s.t. \ x_i = \sum_{j=1}^{K} y_{ij} l_j \\ \qquad \sum_j y_{ij} = 1 \\ \qquad y_{ij} \in \{0,1\} \end{cases}$$

Define the energy function:

$$H_0(\mathbf{y};\mathbf{l}) = \sum_{i \in V} \sum_{j \in N(i)} v_{ij}(x_i - x_j)^2 + w_i(x_i - x_i^+)^2 \tag{11}$$

It is straightforward to see there is a $\mathbf{y}^*$ that minimizes $H_0(\mathbf{y};\mathbf{l})$. We characterize the conditions under which this $\mathbf{y}^*$ is a Nash Equilibrium in pure strategies for the players. Suppose the optimal solution $\mathbf{y}^*$ yields x_i^* with $x_i^* = l_j$ for some $j \in \{1, \ldots, K\}$. If User i chooses to deviate from this strategy, then her change in payoff is:

$$\Delta H_i = H_i^* - H_i = \sum_{j \in N(i)} v_{ij} \left[(x_i^* - x_j^*)^2 - (x_i - x_j^*)^2 \right] + w_i \left[(x_i^* - x_i^+)^2 - (x_i - x_i^+)^2 \right] \tag{12}$$

while:

$$\Delta H_j = H_j^* - H_j = v_{ji}\left[(x_j^* - x_i^*)^2 - (x_j^* - x_i)^2\right] = v_{ji}\left[(x_i^* - x_j^*)^2 - (x_i - x_j^*)^2\right] \tag{13}$$

for each $j \in N(i)$. Under a symmetric weight assumption (i.e., $v_{ij} = v_{ji}$), we have:

$$\Delta H_0 = \sum_{i\in V} \Delta H_i = 2 \sum_{j\in N(i)} [v_{ij}(x_i^* - x_j^*)^2 - (x_i - x_j^*)^2] + w_i[(x_i^* - x_i^+)^2 - (x_i - x_i^+)^2] \tag{14}$$

Let:

$$A = \sum_{j\in N(i)} v_{ij}\left[(x_i^* - x_j^*)^2 - (x_i - x_j^*)^2\right]$$
$$B = w_i\left[(x_i^* - x_i^+)^2 - (x_i - x_i^+)^2\right]$$

Then $\Delta H_i = A + B$ and $\Delta H_0 = 2A + B$. The fact that $\mathbf{y}^*$ is a minimizer for H_0 implies that $\Delta H_0 \leq 0$ otherwise, $\mathbf{y}^*$ could not have been a minimizer. Thus $2A + B \leq 0$. For a rational Player i a change in strategy make sense if (and only if) $A + B > 0$. There are four cases to consider:

Case 1: If $A \leq 0$ and $B \geq 0$, and since $2A + B \leq 0$ and $A + B > 0$, we have $|A| < |B| \leq 2|A|$. That is, Player i has benefitted by moving closer to her comfort value, sacrificing reputation. If this is not the case, then there is no rational reason for Player i to change strategies.

Case 2: If $A, B \leq 0$, then immediately $\Delta H_i \leq 0$ and Player i has not benefitted from changing.

Case 3: If $A \geq 0$ and $B \leq 0$, then $2A + B \leq 0$ implies $|B| \geq |A|$ which implies $A + B \leq 0$ and thus Player i would not change to this alternate strategy.

Case 4: If $A, B \geq 0$, then $2A + B \geq 0$ and $\mathbf{y}^*$ was either not a minimum or (in the case when $A = B = 0$) not a unique minimum.

It follows that only Case 1 prevents a global minimizer for H_0 from being a Nash equilibrium. For $w_i \approx 0$ we have $|B| \approx 0$ and in this case, we see necessarily that $A \leq 0$. Thus the energy minimizing solution is a Nash equilibrium. The following theorem follows naturally from this analysis:

Theorem 1. *For any set of comfort values* $\left\{x_i^+\right\}_{i=1}^N$ *and fixed privacy levels* $\mathbf{l} = \langle l_1, \ldots, l_K\rangle$ *there is an* $\epsilon \geq 0$ *so that if* $w_i \leq \epsilon$ *for* $i = 1, \ldots N$*, then there is a pure strategy Nash equilibrium for the following game:*

$$\forall i \begin{cases} H_i(\mathbf{y};\mathbf{l}) = \min\limits_{\mathbf{y}_i} \quad \sum\limits_{j\in N(i)} v_{ij}(x_i - x_j)^2 + w_i(x_i - x_i^+)^2 \\ \qquad\qquad s.t. \;\; x_i = \sum\limits_{j=1}^{K} y_{ij} l_j \\ \qquad\qquad\qquad \sum\limits_j y_{ij} = 1 \\ \qquad\qquad\qquad y_{ij} \in \{0,1\} \end{cases} \tag{15}$$

□

Remark 1. The results in Theorem 1 can be generalized to a game of the form:

$$\forall i \begin{cases} H_i(\mathbf{y};\mathbf{l}) = \min\limits_{\mathbf{y}_i} \quad \sum\limits_{j\in N(i)} v_{ij} f_P(x_i - x_j) + w_i f_C(x_i - x_i^+) \\ \qquad\qquad s.t. \;\; x_i = \sum\limits_{j=1}^{K} y_{ij} l_j \\ \qquad\qquad\qquad \sum\limits_j y_{ij} = 1 \\ \qquad\qquad\qquad y_{ij} \in \{0,1\} \end{cases}$$

for appropriately chosen convex functions f_C and f_P with minima at 0. Moreover, for $w_i \approx 0$ the bi-level problem is simply a bi-level combinatorial optimization problem.

Remark 2. If $w_i \gg 0$, then the player will conform more closely to her comfort level and for extremely high values of w_i (for $i = 1, \ldots, N$) there is again a pure strategy Nash equilibrium computed by finding the l_k value as close as possible to Player i's comfort level. Thus, settings with no pure strategy equilibria occur when the Players have values w_i large enough to prevent a pure strategy equilibrium consistent with social conformity, but not large enough to cause all players to follow their own comfort signal.

5 An Approximation Algorithm for Arbitrary Graphs - A Simulation

We have characterized the circumstances under which there exists a pure strategy Nash equilibrium for the bi-level optimization problem which describes the site's task of choosing a discrete set of privacy settings to optimize its payoff. Namely, this equilibrium exists in cases of extremely weak or extremely strong comfort level effects. Even in the case that such an equilibrium exists, we anticipate that finding the solution explicitly is NP-hard. Bi-level optimization problems are NP-hard [13], and even evaluating a solution for optimality is NP-hard [36]. Accordingly, an alternate approach in which we find an approximate solution is needed.

We argue that an approximation algorithm is also a more realistic approach in practice, since real SNs do not typically have the sharing comfort level for each individual user or potentially weighted influences amongst users' peers *a priori*. These parameters of the model are inferred through observation of user behavior under varying constraints, often using similar techniques to those we employ in the sequel; that is a site analyzes users' responses to minor alterations in its policies and recalibrates accordingly.

Here, we present a two-part algorithm for approximately computing the users' and site's utility functions on an arbitrary graphs in order to determine a discrete set of privacy settings beyond the determined lower bound to be made available to users in the SN. The Player Algorithm uses fictitious play simulating the convergence of the players' strategies to a strategy vector dependent on the players' personal comfort levels and the fixed set of privacy options determined by the SN site. Note, from Theorem 1, this may in fact be a pure strategy Nash equilibrium under appropriate assumptions.

To determine the full set $\mathbf{l}$ of privacy settings to be offered to users, the Site Algorithm wraps around the Player Algorithm as follows. The site lets $l_1 = 0$. Since players are captive to the site in this model, all players adopt strategy l_1. The level of unhappiness each player experiences for being forced to choose l_1 is calculated. Next, the site makes available a second option $\mathrm{l}_2 = \mathrm{l}_1 + \delta$. The Player Algorithm uses fictitious play to simulate the convergence of each player's strategy to either l_1 or l_2. A corresponding payoff for the site is calculated. Provided that there is at least one user whose comfort level for sharing is greater than l_1 and δ is small enough, the addition of option l_2 will increase the site's payoff. The site moves l_2 up by increments of δ, monitoring users' responses at each move, recalculating the corresponding site payoff and stopping when this payoff starts to decrease. Intuitively, when l_2 moves too far above individuals' comfort levels, users will become increasingly unhappy and eventually revert back to sharing at l_1 rather than l_2. The local optimum achieved here is taken as $l_2 \in \mathbf{l}$. Following this, the site makes available a third option $l_3 = l_2 + \delta$ and allows players to converge on strategies from the set of three options available, incrementing l_3 as before until a local optimum is achieved. At this time, l_3 is added to $\mathbf{l}$. This heuristic is repeated and the set $\mathbf{l}$ of privacy options grows by one as each local optimum is discovered until no further gains in site payoff or user happiness can be achieved, which is guaranteed to occur at a value no higher than the comfort level of the site's most privacy-lenient user. Pseudocode for the Player Algorithm and Site Algorithm are given in Figs. 2 and 3, respectively.

Figure 4 visualizes a well-known, real-world social network of members of a karate club [39]. In the absence of any constraints instituted by the site, equivalently in the case that each user may select his optimal privacy setting for a given piece of content, the trajectories of users' selections are guided by inherent personal comfort with sharing and the influence of their peers. Immediate neighbors in the graph are considered peers. We simulate the trajectory of privacy selections for member-users of the karate club network, first given the player algorithm described above in the unconstrained case, namely assuming that users have access to the complete set of options on the interval $[0, 1]$. Figure 5

```
Player Algorithm

Initialize $x_i(0)$ $(i = 1, \dots, N)$, $t = 0$
while $t < T_{\max}$
    t:=t+1
    for each $i \in \{1, \dots, N\}$
        Minimize $H_i(\mathbf{y}; \mathbf{l})$ to obtain $y_i^*$
        $x_i^*(t + 1) := y_i^*(t)$
    end
end
```

Fig. 2. PlayerAlgorithm

```
Site Algorithm

Initialize $l_1 = 0$,$\mathbf{l} = \{l_1\}$
Run the Player Algorithm to obtain $\mathbf{x}^*$
$i = 1$
while $l_i < 1$
    $l_i = l_{i-1} + \delta l_{i-1}$
    Run the Player Algorithm $\{\mathbf{l}, l_i\}$ to obtain $\mathbf{y}^*$
    if $J_S(\{\mathbf{l}, l_i\}, \mathbf{y}^*) - J_S(\{\mathbf{l}, l_{i-1}\}, \mathbf{x}^*) < 0$
        Add $l_i$ to $\mathbf{l}$
    else
        $i = i + 1$
    end
end
```

Fig. 3. Site Algorithm

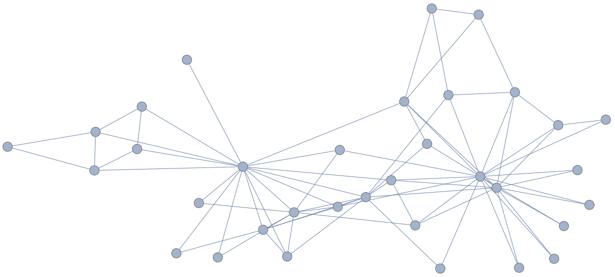


Fig. 4. A visualisation of the karate club network.

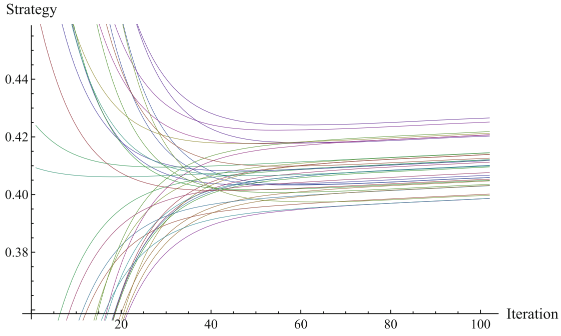


Fig. 5. A visualisation of players' strategies over time, initialised randomly, according to the user model

illustrates players' strategies over time. Strategies are initialized as user's individual sharing comfort levels and comfort levels are selected uniformly randomly from the interval $[0, 1]$. Notice that in this case, the vector of user strategies converges to equilibrium, as guaranteed by Proposition 1.

As described, the site's approximation algorithm influences the user model by iteratively choosing a discrete set of options to be made available to users, simulating user behavior given those constraints, and then adjusting the set of options by small increments until local optima are discovered. A visualization of site payoff during this process simulated over the karate club network is given in Fig. 6. Local optima occur at $x = \{0.4, 0.6, 0.72, 0.88\}$, so the site determines the set of privacy options as $l_1 = 0.4$, $l_2 = 0.6$, $l_3 = 0.72$, $l_4 = 0.88$ and $l_5 = 1$. User comforts are the same as those given in Fig. 5, and we choose $\delta = 0.04$. Note that the choice of δ may indicate a site's willingness to offer a finer granularity of privacy options to its users. A greater value of δ will lead to the discovery of fewer local optima, while smaller delta will yield more. This choice may also depend on the initial set of user comforts and the site business model. To this extent, the general algorithm we present here is the framework for a more personalized algorithm representative of a site's policies, practices and user base.

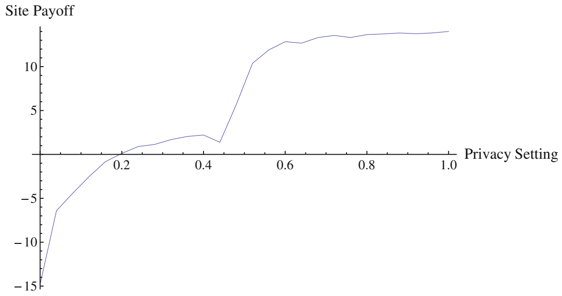
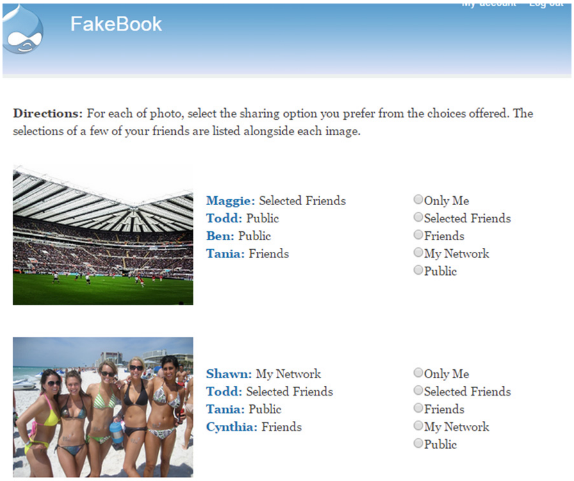


Fig. 6. Site payoff as privacy options are iteratively made available

6 Experimental Results

We designed and executed experiments to evaluate two of our key assumptions with a user study involving 60 participants in a simulated social network. First, our core model assumes that users' sharing decisions are influenced by a weighted sum of peer influence and personal comfort. We aim to determine whether postulated effects peer influence may be observed, even in a simulated context. Second, we seek to determine whether the iterative approach we take in our approximation algorithm may be assumed to in fact approximate the optimal discrete set of privacy options offered by the site. Hypothesizing that it will, we expect to rule out the notion that iterative presentation of an increased number



Directions: For each of photo, select the sharing option you prefer from the choices offered. The selections of a few of your friends are listed alongside each image.

Maggie: Selected Friends
Todd: Public
Ben: Public
Tania: Friends

○Only Me
○Selected Friends
○Friends
○My Network
○Public

Shawn: My Network
Todd: Selected Friends
Tania: Public
Cynthia: Friends

○Only Me
○Selected Friends
○Friends
○My Network
○Public

Fig. 7. Sample screenshot from Phase 2

of sharing options will significantly alter or confuse optimal individual preferences. Put more simply, users will not change their decision if they are offered an (optimal) set of privacy options **l** in one shot verses if **l** is iteratively built. These two assumptions are at the core of our user model and site model, respectively, and therefore validating them gives confidence in theoretical findings.

Subjects were presented with a series of images and asked to select a privacy setting for each, to be uploaded to social media. We organized the study in three distinct "phases".

1. In Phase 1 of our experiment, subjects were shown 15 images and given five sharing options from which to choose for each, i.e., "only me", "selected friends", "friends", "my network" and "public".
2. In Phase 2, subjects were shown the same images again and asked to choose from amongst the same options, but with the addition of the privacy selections of four of the subject's"friends" listed next to each image (see Fig. 7 for a sample screenshot). In attempt to create a more realistic sense of friendship between the subject and the simulated users, we endowed each simulated user with a profile page including demographic information, photos and other personal details and hyperlinked these profile pages throughout. Subjects were divided into several subgroups and treated to three variations of peer pressure in which friends' selections were skewed towards more private (skew-down), more public (skew-up) or random. In Sect. 6.1, we compare the selections of each user in Phase 1 (which we take as a baseline) with their selections in Phase 2. We expect that users may be influenced to increase their privacy restrictions when seeing that their peers are sharing more conservatively than they are, while on the other hand users may feel comfortable sharing more freely when their friends do the same.

3. Phase 3 was designed to test whether the iterative addition of privacy options (see Sect. 5) would influence users' ultimate privacy selections. Assuming a fixed set of options (i.e., l_1 = "only me", l_2 = "selected friends", l_3 = "friends", l_4 = "my network", l_5 = "public"), we iteratively presented subjects with a subset of photos from Phase 1 and Phase 2. At first, subjects were offered only l_1 and l_2 as privacy settings, next l_1, l_2 and l_3, subsequently l_1 through l_4, and finally l_1 through l_5. Variants of Phase 3 incorporating skew-down, skew-up and random peer pressure, implemented identically as in Phase 2, were also included for subsets of participants. In Sect. 6.2, we compare the selections of each user in Phase 2 with the their selections in the final iteration of Phase 3.

Participants in our study were 68 % female and 32 % male, with mean age 25.6 and standard deviation 2.98. In an initial survey preceding the experiment 100 % of subjects claimed to have an account with at least one social media site, with 92 % asserting that they maintain at least one "comprehensive" social media profile. On average, subjects claimed to participate in 3.4 different social networks, including Facebook, Instagram, Twitter, LinkedIn, Pinterest, Google+ and Vine.

6.1 Experimental Results: Peer Pressure Effects on Privacy Preferences

With respect to peer pressure, subjects were queried during the initial survey on several points related to privacy and peer pressure in content sharing. Over half (54.7 %) of subjects admitted to sometimes, often or always posting content with one privacy setting and later changing their mind and revising this setting, with 70 % of these subjects citing peer pressure as the reason for the revision.

In Table 1, we present the results of a one-factor analysis of variance (ANOVA) on change from baseline privacy selections for users treated with skew-down, skew-up or random peer influence in Phase 2. To quantify privacy options, we let $l_1 = 1$, $l_2 = 2$, $l_3 = 3$, $l_4 = 4$ and $l_5 = 5$. For each subject, for each image, we let change from baseline be defined as *(value of selection in Phase 2)-(value of selection in Phase 1).* Note that a significant change in user sharing is detected in both subgroups subjected to a consistent peer pressure in either direction of more or less sharing. As might be expected, no significant change in sharing is detected in the random pressure control group. Of note, the most statistically significant change is observed when users are exposed to skew-down peer pressure, that is, when participants observe a change of their friends' privacy settings toward more conservative choices. This finding is consistent with the participants response of change of settings mentioned above, and also in line with existing research in this field [4,37], which has shown how users may change their mind with respect to sharing and may tend to be more conservative once they see the "network" behavior or reactions to their choices.

Follow-on ANOVA analyses blocking on subjects and images also give insight into more subtle user behavior dynamics. In both the skew-up and skew-down groups, subject effects (i.e., the affect a subject's identity had on output privacy settings) were highly significant ($p \approx 0$). This finding is intuitive and serves as strong justification for the inclusion of the parameter v_i in Eq. 2. That is, we must consider individual differences in susceptibility to peer pressure when implementing this type of model. Interestingly, an image effect was present in one of the experimental groups as well. Specifically, a significant effect was observed when image number was treated as an input in the skew-up group ($p = 0.0013$) but not for skew-down ($p = 0.1887$). When considered alongside the strength of skew-down peer pressure effects noted in Table 1, we suggest that these finding may again indicate users' readiness to make more conservative sharing choices for all photos, but hesitance to share more freely for specific images they would prefer to keep private, even when influenced to do so.

Table 1. Change from baseline after exposure to peer influence (Phase 2)

	Subjects	Average Change	p-Value
Skew-Down	17	-0.305	**0.0067**
Skew-Up	19	+0.192	**0.049**
Random	17	-0.086	0.375

6.2 Experimental Results: Iterative Approximation of Privacy Preferences

We have argued that using an approximation algorithm is both necessary and realistic, in the context of our bi-level optimization problem describing the site's task of choosing an optimal set of privacy options to offer its users. We here seek to validate the notion that an iterative approach like the one we take in our proposed algorithm does not disturb players' optimal privacy selections as determined in the theoretical case. Following we present the results of Phase 3 of the experiment, as described above.

For this analysis, we again separate study participants into subgroups by the peer pressure to which they were exposed, if any. Table 2 gives the results of a one-factor analysis of variance (ANOVA) on change from Phase 2 privacy selections for users treated with skew-down, skew-up or random peer influence. As a control group for this Phase, we keep a subset of subjects away from any exposure to peer pressure (that is, these subjects did not participate in Phase 2) and compare their results for Phase 3 with their Phase 1 baseline selections. Findings here indicate no significant change in users' final privacy selections due to the iterative nature of presentation of the options in any of the experimental groups, validating the approximation-algorithm approach as a reasonable alternative for modelling user behavior in cases that closed-form solutions are intractable.

We note here that Phase 3 studies user behavior given that options l_1, l_2 and so forth are presented additively one by one. The approximation algorithm as it

Table 2. Change from Phase 2 selections in the iterated model (Phase 3)

	Subjects	Average Change	p-Value
No Peer Pressure	7	0.086	0.774
Skew-Down	17	-0.28	0.19
Skew-Up	19	-0.2	0.282
Random	17	-0.117	0.527

presented is deployed accordingly, but also includes a routine for the selection of the set of values $\{l_i\}$ making very small, incremental changes to each l_i and monitoring users' responses throughout.

7 Conclusion

In this paper, we have presented a model for privacy decision-making in the context of online social networks. We have modeled the site's role in setting privacy policies that can help to retain users while also optimizing the site's payoff. Our work lays the foundation for further game-theoretic modeling of privacy-related behaviors in online SNs toward the better understanding of the interplay and repercussions of site and user choices.

As future work, we will refine the outlined approximation algorithm, with particular focus on how incremental privacy boundaries could actually be offered to end users. We also plan to investigate how changes to the social network topology and user attitudes towards privacy over time may affect this game. Finally, we plan to carry out more extensive user studies to validate our findings.

Acknowledgements. Portions of Dr. Rajtmajer's, Dr. Griffin's and Dr. Squicciarini's work were supported by the Army Research Office grant W911NF-13-1-0271. Portions of Dr. Squicciarini's work were additionally supported by the National Science Foundation grant 1453080.

References

1. Ackerman, M.S., Cranor, L.F., Reagle, J.: Privacy in e-commerce: Examining user scenarios and privacy preferences. In: Proceedings of the 1st ACM Conference on Electronic Commerce, EC 1999, pp. 1–8. ACM, New York (1999)
2. Acquisti, A., Brandimarte, L., Loewenstein, G.: Privacy and human behavior in the age of information. Science **347**(6221), 509–514 (2015)
3. Bilge, L., Strufe, T., Balzarotti, D., Kirda, E.: All your contacts are belong to us: automated identity theft attacks on social networks. In: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, pp. 551–560. ACM, New York (2009)
4. Caliskan-Islam, A.: How do we decide how much to reveal? SIGCAS Comput. Soc. **45**(1), 14–15 (2015)

5. Chen, J., Brust, M.R., Kiremire, A.R., Phoha, V.V.: Modeling privacy settings of an online social network from a game-theoretical perspective. In: 2013 9th International Conference on Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), pp. 213–220. IEEE, October 2013
6. Chron. What is the minimum information needed to create a facebook account? (2013). http://smallbusiness.chron.com/minimum-information-needed-create-facebook-account-27690.html
7. Cranor, L.F., Reagle, J., Ackerman, M.S.: Beyond Concern: Understanding Net Users' Attitudes About Online Privacy. MIT Press, Cambridge (2000)
8. Fang, L., LeFevre, K.: Privacy wizards for social networking sites. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 351–360. ACM, New York, April 2010
9. Fu, F., Chen, C., Liu, L., Wag, L.: Social dilemmas in an online social network: The structure and evolution of cooperation. Phys. Lett. A **371**(1–2), 58–64 (2007)
10. Griffin, C., Squicciarini, A.: Toward a game theoretic model of information release in social media with experimental results. In: Proceedings of the 2nd Workshop on Semantic Computing and Security, San Francisco, CA, USA, 28 May 2012
11. Griffin, C., Squicciarini, A., Rajtmajer, S., Tentilucci, M., Li, S.: Site-constrained privacy options for users in social networks through stackelberg games. In: Proceedings of Sixth ASE International Conference on Social Computing, May 2014
12. Gross, R., Acquisti, A.: Information revelation and privacy in online social networks. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, WPES 2005, pp. 71–80. ACM, New York (2005)
13. Hansen, P., Jaumard, B., Savard, G.: New branch-and-bound rules for linear bilevel programming. SIAM J. Sci. Stat. Comput. **13**(5), 1194–1217 (1992)
14. Hoadley, C.M., Xu, H., Lee, J.J., Rosson, M.B.: Privacy as information access and illusory control: The case of the facebook news feed privacy outcry. Electron. Commer. Res. Appl. **9**(1), 50–60 (2010)
15. Hui, P., Buchegger, S.: Groupthink and peer pressure: social influence in online social network groups. In: 2009 International Conference on Advances in Social Network Analysis and Mining (ASONAM), pp. 53–59. IEEE, Los Alamitos, July 2009
16. Immorlica, N., Lucier, B., Rogers, B.: Emergence of cooperation in anonymous social networks through social capital. In: Proceedings of the 11th ACM Conference on Electronic Commerce EC (2010)
17. Jagatic, T.N., Johnson, N.A., Jakobsson, M., Menczer, F.: Social phishing. Commun. ACM **50**(10), 94–100 (2007)
18. Kayes, I., Iamnitchi, A.: A survey on privacy and security in online social networks (2015). arXiv preprint arXiv:1504.03342
19. Krishnamurthy, B., Gill, P., Arlitt, M.: A few chirps about twitter. In: Proceedings of the First Workshop on Online Social Networks, WOSN 2008, pp. 19–24. ACM, New York (2008)
20. Krishnamurthy, B., Wills, C.E.: Characterizing privacy in online social networks. In: Proceedings of the First Workshop on Online Social Networks, WOSN 2008, pp. 37–42. ACM, New York (2008)
21. Lindamood, J., Heatherly, R., Kantarcioglu, M., Thuraisingham, B.: Inferring private information using social network data. In: WWW 2009 Proceedings of the 18th International Conference on World Wide Web, pp. 1145–1146. ACM, New York, April 2009

22. Liu, Y., Gummadi, K.P., Krishnamurthy, B., Mislove, A.: Analyzing facebook privacy settings: user expectations vs. reality. In: Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, pp. 61–70 (2011)
23. Ohtsuki, H., Hauert, C., Lieberman, E., Nowak, M.A.: A simple rule for the evolution of cooperation on graphs and social networks. Nature **441**(7092), 502–505 (2006)
24. Park, N., Kee, K., Valenzuela, S.: Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. CyberPsychol. Behav. **12**(6), 729–733 (2009)
25. Rajtmajer, S.M., Griffin, C., Mikesell, D., Squicciarini, A.: A cooperate-defect model for the spread of deviant behavior in social networks. CoRR, abs/1408.2770 (2014)
26. Rand, D.G., Arbesman, S., Christakis, N.A.: Dynamic social networks promote cooperation in experiments with humans. Proc. Nat. Acad. Sci. **108**(48), 19193–19198 (2011)
27. Rosen, J.B.: Existence and uniqueness of equilibrium points for concave n-person games. Econometrica **33**(3), 520–534 (1965)
28. Simpson, A.: On the need for user-defined fine-grained access control policies for social networking applications. In: Proceedings of the Workshop on Security in Opportunistic and SOCial Networks, SOSOC 2008. ACM, New York (2008)
29. Squicciarini, A., Griffin, C.: An informed model of personal information release in social networking sites. In: Proceedings of the Fourth IEEE International Conference on Information Privacy, Security, Risk and Trust, September 2012
30. Squicciarini, A., Griffin, C.: An informed model of personal information release in social networking sites. In: 2012 ASE/IEEE Conference on Privacy, Security, Risk and Trust, Amsterdam, Netherlands, September 2012
31. Squicciarini, A., Paci, F., Sundareswaran, S.: PriMa: an effective privacy protection mechanism for social networks. In: Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, pp. 320–323 (2010)
32. Squicciarini, A., Shehab, M., Wede, J.: Privacy policies for shared content in social network sites. VLDB J. **19**(6), 777–796 (2010)
33. Squicciarini, A.C., Griffin, C.: Why and how to deceive: game results with sociological evidence. Soc. Netw. Anal. Min. **4**(1), 161 (2014)
34. Squicciarini, A.C., Griffin, C., Sundareswaran, S.: Towards a game theoretical model for identity validation in social network sites. In: PASSAT/SocialCom 2011, Boston, MA, USA, 9–11 October, 2011, pp. 1081–1088 (2011)
35. Squicciarini, A.C., Shehab, M., Paci, F.: Collective privacy management in social networks. In: WWW 2009, pp. 521–530. ACM, New York (2009)
36. Vicente, L., Savard, G., Jdice, J.: Descent approaches for quadratic bilevel programming. J. Optim. Theory Appl. **81**(2), 379–399 (1994)
37. Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P.G., Cranor, L.F.: "I regretted the minute I pressed share": a qualitative study of regrets on Facebook. In: Proceedings of the Seventh Symposium on Usable Privacy and Security, SOUPS 2011, pp. 10:1–10:16. ACM, New York (2011)
38. Wu, Z.-X., Rong, Z., Yang, H.-X.: Impact of heterogeneous activity and community structure on the evolutionary success of cooperators in social networks. Phys. Rev. E **91**, 012802 (2015)
39. Zachary, W.W.: An information flow model for conflict and fission in small groups. J. Anthropol. Res. **33**(4), 452–473 (1977)