Rational Privacy Disclosure in Social Networks

Josep Domingo-Ferrer

Universitat Rovira i Virgili
UNESCO Chair in Data Privacy
Department of Computer Engineering and Mathematics
Av. Països Catalans 26, E-43007 Tarragona, Catalonia
josep.domingo@urv.cat

Abstract. Social networking web sites or social networks for short (SNs) have become an important web service with a broad range of applications. In an SN, a user publishes and shares information and services. We propose a utility function to measure the rational benefit derived by a user from her participation in an SN, in terms of information acquired vs information provided. We show that independently and selfishly maximizing this utility leads users to "free-riding", *i.e.* getting information about other users and offering no information about themselves. This results in SN shutdown (no functionality). We then propose protocols to achieve a correlated equilibrium between users, in which they coordinate their disclosures in view of jointly maximizing their utilities. The proposed protocol can be used to assist an SN user in making rational decisions regarding which of her attributes she reveals to other users.

Keywords: Social networks, Data privacy, Game theory.

1 Introduction

Social networking web sites or social networks for short (SNs) have become an important web service with a broad range of applications: collaborative work, collaborative service rating, resource sharing, friend search, etc. Facebook, MySpace, Xing, LinkedIn, etc., are well-known examples. In an SN, a user publishes and shares information and services.

There are two types of privacy in SNs:

– *Relationship privacy.* In some SNs, the user can specify how much it trusts other users, by assigning them a trust level. It is also possible to establish several types of relationships among users (like "colleague of", "friend of", etc.). The trust level and the relationship type are used to decide whether access is granted to resources and services being offered. The availability of information on relationships (trust level, relationship type) has increased with the advent of the Semantic Web and raises privacy concerns: knowing who is trusted by whom and to what extent discloses a lot about the users thoughts and feelings; in fact, knowing relationships discloses the social network topology and this can allow re-identification of users even if

V. Torra, Y. Narukawa, and M. Daumas (Eds.): MDAI 2010, LNAI 6408, pp. 255–265, 2010.

they pretend to stay anonymous [5]. Relationship privacy is about allowing the normal operation of the SN while allowing users to preserve their relationships and trust levels as private as possible (see [2,3]).

– *Content privacy.* This type of privacy applies to *all* SNs and *is* the subject of this paper. The information content a user publishes clearly affects her privacy. Recently, a privacy risk score [4] has been proposed for the user to evaluate the privacy risk caused by the publication of a certain information. Let the information attributes published by the users in an SN be labeled from 1 to n. Then the privacy score risk of user j is

$$PR(j) = \sum_{i=1}^{n} \sum_{k=1}^{\ell} \beta_{ik} V(i,j,k) \tag{1}$$

where $V(i,j,k)$ is the visibility of user j's value for attribute i to users who are k links away from j and β_{ik} is the privacy sensitivity of attribute i (how embarrassing it is for a user to reveal attribute i to people k links away). The visibility $V(i,j,k) = 1$ if user j makes her attribute j visible to those users k links away from j; it is zero otherwise. An interesting special case is the dichotomous case, in which an attribute is either kept hidden or published for everyone; in the dicotomous case, $V(i,j) = 1$ means that user j publishes her attribute i and $V(i,j) = 0$ means user j keeps her attribute i secret. The dichotomous privacy score is

$$PR_2(j) = \sum_{i=1}^{n} \beta_i V(i,j) \tag{2}$$

Regarding the above privacy risk score, note that the greater it is for a user, the lower is the privacy preservation utility for that user. On the other hand, $PR(j)$ is a monotonically increasing function of the *sensitivity* of the user's attributes and the *visibility* these attributes get. Also, as noted in [4], the sensitivity β_{ik} is monotonically increasing with k, that is, if $k < k'$ then $\beta_{ik} \leq \beta_{ik'}$.

1.1 Contribution and Plan of This Paper

The aim of this paper is to provide protocols to assist a user in an SN in making rational decisions regarding which of her attributes she reveals to other users.

In Section 2, we define the utility a user derives from participating in an SN as the functionality the user gets divided by privacy risk score the user incurs. By functionality, we mean what the user can see about other users in the SN (we do not mean performance or similar issues). In terms of that privacy-functionality utility, we show in Section 3 that, if users independently choose their disclosure strategies, the dominant strategy (and hence the Nash equilibrium) is for SN users to "free-ride", *i.e.* to try to learn as much as possible from other users and disclose nothing about themselves, which leads to shutting down the SN. This zero-utility outcome can be improved for all users if they coordinate their strategies. In Section 4, we propose protocols to assist users in achieving correlated

equilibrium, that is, to help them to jointly maximize their utilities by revealing their attributes to each other in a correlated way. Simulation results are given in Section 5. Finally, conclusions and future research directions are summarized in Section 6.

2 A Privacy-Functionality Score

As mentioned above, our definition of user utility can be roughly summarized as the amount of information the user can see about other users in the SN divided by the amount of information the user shows about herself. This "rational" utility does not probably explain the attitude of the typical Facebook user, who tends to tell her friends a lot about herself, without caring much what she gets in return. Our definition of utility is more adapted to social networks for professionals, like Xing or LinkedIn: in those networks, employers and job applicants tend to disclose their information in a more targeted and cautious way.

We quantify the above idea of the utility a user j derives from participating in an SN by using the following privacy-functionality score

$$PRF(j) = \frac{\sum_{j'=1,j'\neq j}^{N} \sum_{i=1}^{n} \sum_{k=1}^{\ell} \beta_{ik} V(i,j',k) I(j,j',k)}{1 + PR(j)}$$

$$= \frac{\sum_{j'=1,j'\neq j}^{N} \sum_{i=1}^{n} \sum_{k=1}^{\ell} \beta_{ik} V(i,j',k) I(j,j',k)}{1 + \sum_{i=1}^{n} \sum_{k=1}^{\ell} \beta_{ik} V(i,j,k)} \quad (3)$$

where $I(j,j',k)$ is 1 if j and j' are k links away from each other, and it is 0 otherwise.

Note that:

- $PRF(j)$ decreases as the privacy score $PR(j)$ in its denominator increases, that is, as user j discloses more of her privacy.
- $PRF(j)$ increases as its numerator increases; this numerator adds up the components of privacy scores of users $j' \neq j$ due to those users disclosing attribute values to j.

The dichomotous version of the privacy-functionality score is simply:

$$PRF_2(j) = \frac{\sum_{j'=1,j'\neq j}^{N} \sum_{i=1}^{n} \beta_i V(i,j')}{1 + PR(j)}$$

$$= \frac{\sum_{j'=1,j'\neq j}^{N} \sum_{i=1}^{n} \beta_i V(i,j')}{1 + \sum_{i=1}^{n} \beta_i V(i,j)} \quad (4)$$

3 The SN Functionality-Privacy Game with Independent Strategies

If we regard $PRF(j)$ as a game-theoretic utility function [6], the higher $PRF(j)$, the higher the utility for user j. Let us first deal with the dichotomous case, for simplicity.

The set of possible strategies S_j available to user j are the numbers from 0 to $2^n - 1$. In the binary expression of a strategy $s_j \in S_j$, a 1 in position $i \in \{0, \cdots, n-1\}$ means that, under s_j, j publishes attribute $i+1$ ($V(i+1, j) = 1$), whereas a 0 means that, under s_j, j keeps attribute $i+1$ secret ($V(i+1, j) = 0$).

Now, consider a strategy vector $s = (s_1, \cdots, s_N)$ formed by the strategies *independently and selfishly* chosen by all users. When user j chooses s_j, denote by s_{-j} the $N-1$ dimensional vector of the strategies chosen by the other users. If we use $PRF_2(j)$ to quantify the utility $u_j(s)$ incurred by user j, we have

$$u_j(s) = \frac{\sum_{j'=1, j' \neq j}^{N} \sum_{i=1}^{n} \beta_i V(i, j')}{1 + \sum_{i=1}^{n} \beta_i V(i, j)} \tag{5}$$

where the values $V(i, j)$, $i = 1, \cdots, n$, are those specified by the binary expansion of s_j.

It turns out that the strategy vector all zeros, *i.e.* $s^0 = (0, 0, \cdots, 0)$ is dominant, because, for any user j and each alternate strategy vector s', we have

$$u_j(s_j^0, s'_{-j}) \geq u_j(s'_j, s'_{-j}) \tag{6}$$

Disclosing no information is better for user j than disclosing some information, assuming that each user chooses her strategy independently. Since a dominant strategy is also a Nash equilibrium [6], the strategy vector s^0 is also a Nash equilibrium; this can be checked directly, because Inequality (6) implies

$$u_j(s_j^0, s_{-j}^0) \geq u_j(s'_j, s_{-j}^0)$$

Thus, it turns out that rational and independent choice of strategies leads to no user offering any information on the SN, which results in the SN being shut down. It is easy to show that this also holds in the general case ($k > 1$).

A similar pessimistic result is known for the P2P file sharing game, in which the system goal is to leverage the upload bandwidth of the downloading peers: the dominant strategy is for all peers to attempt "free-riding", that is, to refuse to upload [1], which causes the system to shut down.

Example 1. The simplest version of the above game is one with two users having each one attribute, which they may decide to keep hidden (a strategy denoted by H, which implies visibility 0 for the attribute) or publish (a strategy denoted by P, which implies visibility 1). Assuming a sensitivity $\beta = 1$ for that attribute and using Expression (5), the user utilities for each possible strategy vector are as follows:

$$u_1(H, H) = 0; u_1(H, P) = 1; u_1(P, H) = 0; u_1(P, P) = 1/2$$

$$u_2(H, H) = 0; u_2(H, P) = 0; u_2(P, H) = 1; u_1(P, P) = 1/2$$

This simple game can be expressed in matrix form:

User 2 / User 1	H	P
H	0 0	0 1
P	1 0	1/2 1/2

The above matrix corresponds to the Prisoner's Dilemma [6], perhaps the best-known and best-studied game. Consistently with our argument for the general case, it turns out that (H, H) is a dominant strategy, because:

$$u_1(H,P) = 1 \geq u_1(P,P) = 1/2; u_1(H,H) = 0 \geq u_1(P,H) = 0$$

$$u_2(P,H) = 1 \geq u_1(P,P) = 1/2; u_2(H,H) = 0 \geq u_2(H,P) = 0$$

The second and fourth equations above guarantee that (H, H) is a Nash equilibrium (in fact, the only one). The Prisoner's Dilemma with $N > 2$ users is known as the Pollution Game [6] and corresponds to the dichotomous SN game considered above.

4 The SN Functionality-Privacy Game with Correlated Strategies

The outcome of independent rational behavior by users, provided by Nash equilibria and dominant strategies, can be inferior to a centrally designed outcome. This is clearly seen in Example 1: the strategy (P, P) would give more utility than (H, H) to *both* users. However, usually no trusted third-party accepted by all users is available to enforce correlated strategies; in that situation, the problem is how User 1 (resp. User 2) can guess whether User 2 (resp. User 1) will choose P.

Using a solution based on cryptographic protocols for bitwise fair exchange of secrets would be an option, but it seems impractical in current social networks, as it would require a cryptographic infrastructure, unavailable in most SNs.

A more practical solution to this problem may be based on direct reciprocity (*i.e.* tit-for-tat) or reputation, two approaches largely used in the context of P2P file-sharing systems. We describe below two correlated equilibrium protocols based on tit-for-tat and reputation, respectively. They are intended as "assistants" to the human user of the SN in deciding whether to disclose an attribute to another user; however, the ultimate decision belongs to the human, who may quit and renounce to reach the equilibrium. In particular, in both protocols, User 1, as the initiator, first takes the risk of not being corresponded by User 2. However, the "loss" of User 1 will be limited to those attributes she disclosed in the last iteration. User 1 will not disclose to User 2 her remaining, more sensitive attributes.

In both protocols, we introduce user-dependent attribute sensitivities. For example, whereas some people boast their religion (and even dress according to it), for other people this is a very sensitive attribute. Also, when a user evaluates the sensitivity of the attributes received from the other user, the evaluating user is forced to use her own sensitivity scale, because she cannot be assumed to know the evaluated user's sensitivity scale. The real sensitivity scale of an individual is normally highly confidential; for that same reason, if someone discloses her sensitivity scale, there is no guarantee that it is her real scale.

4.1 Adaptation of the Dichotomous Game to Tit-for-Tat

In the protocol below, β_{ij} denotes sensitivity of attribute i according to User j's sensitivity scale. We assume that disclosing an attribute means making it visible to the other user in the protocol (not to all users): therefore, we write $V(i, j, j')$ to denote the visibility of attribute i granted by User j to User j'. Initially, all visibilities are assumed to be zero.

Protocol 1 (Tit-for-tat correlated equilibrium)
User 1 does:

1. *Set Quit* $:= 0$.
2. *While Quit* $= 0$ *do:*
   (a) *If User 1 has already disclosed all her attributes* ($V(i, 1, 2) = 1$ *for all* i) *then set Quit* $:= 0$.
   (b) *Disclose to User 2 the attribute* i^* *such that*

$$i^* = \arg \min_{i:V(i,1,2)=0} \beta_{i,1}$$

   that is, the least sensitive attribute among those not yet disclosed. Disclosure implies setting $V(i^*, 1, 2) := 1$.
   (c) *Request User 2 to disclose to User 1 the same attribute* i^* *disclosed by User 1 to User 2.*
   (d) *If User 1 does not receive User 2's value for the same attribute* i^*, *then set Quit* $:= 1$.

While simple, Protocol 1 has the shortcoming of requiring that the ordering of attributes by sensitivity be the same for User 1 and User 2. Indeed, after User 1 discloses her least sensitive attribute i^*, she expects User 2 to disclose exactly that same attribute i^*. This will only happen if User 2 also considers i^* as her least sensitive undisclosed attribute. In case User 1 does not get i^*, she will consider there is no reciprocity and she will quit the protocol; thus, the protocol lacks robustness. One could change the protocol so that the exchange of attributes is groupwise (several attributes exchanged at a time). However, the issue arises as to which is a reasonable group size: *e.g.* disclosing all attributes in a single iteration is very robust but it is quite risky for User 1, who makes the first move. The reputation of User 2 is a way to decide on the group size. This the idea of the protocol in the next section.

4.2 Adaptation of the Non-dichotomous Game to Reputation

We adapt the non-dichotomous game as follows:

- The parameter $k \in \{1, \cdots, \ell\}$ will be used as an *intimacy level* rather than as a link distance; the greater k, the lower is intimacy. When a User j first interacts with another User j', User j admits User j' in the lowest intimacy level $k \in \ell$ (that of first-time acquaintances). Subsequent interactions may result in User j' being admitted by User j into higher intimacy levels (with smaller k).
- Attribute sensitivities β_{ijk} will now depend on the specific attribute i, the sensitivity scale of User j and the intimacy level k.
- Each User j assigns to each other User j' a reputation $v_{jj'}$ defined as the maximum sensitivity of the attributes User j is willing to show to User j'. Note that reputation is different from intimacy level: a user is probably less intimate with her psychotherapist than with an office colleague, but she surely assigns a greater reputation to her psychotherapist.
- The visibility $V(i, j, j', k)$ denotes whether attribute i is *first* made visible by User j to User j' at intimacy level k. That is, $V(i, j, j', k) = 1$ means that k is the greatest value (the lowest intimacy level) for which attribute i is made visible by User j to User j'; on the other hand, $V(i, j, j', k) = 0$ may mean that either the attribute is not visible to j' at level k or that it is visible and was first made visible by j to j' for some $k' > k$.

In this way, the utility for User j becomes

$$PRF(j) = \frac{\sum_{j'=1, j' \neq j}^{N} \sum_{i=1}^{n} \sum_{k=1}^{\ell} \beta_{ijk} V(i, j', j, k)}{1 + \sum_{i=1}^{n} \sum_{k=1}^{\ell} \beta_{ijk} V(i, j, j', k)} \tag{7}$$

The above definition of visibility ensures that disclosure of attribute i by User j' to User j is counted in $PRF(j)$ only for one intimacy level, the lowest one (that is, the greatest k) at which attribute i is disclosed by User j' to User j.

Note that, in Expression (7), the sensitivities in the numerator are those corresponding to User j, because User j does not know the sensitivity scale of the other users j'.

We next specify a protocol for correlated equilibrium in a game with two users, each of which have values for n attributes numbered from 1 to n, with the sensitivity of user j's attribute i vs level k users being β_{ijk}. Reputation v_{12} in the protocol below is taken in the same range as attribute sensitivities. All visibilities are initially zero.

Protocol 2 (Reputation-based correlated equilibrium)
User 1 does:

1. *If available, use a previous value v_{12} for the reputation of User 2. If none is available, initialize v_{12} with a prior estimate (this guess may be human-assisted, if the human behind User 1 wishes it). Set $k := \ell$ and Quit $:= 0$.*
2. *While $k \geq 1$ and Quit $= 0$ do:*

(a) *Disclose to User 2, that is, set* $V(i,1,2,k) := 1$*, all attributes* i *such that* $V(i,1,2,k) = 0$ *and* $\beta_{i,1,k} \leq v_{12}$*; if there are no disclosable attributes, set Quit* $:= 1$.
(b) *Request User 2 to disclose to User 1 the same attributes disclosed by User 1 to User 2.*
(c) *If User 1 does not receive User 2's values for the same attributes User 1 disclosed, then*
  – *Set Quit* $:= 1$.
  Else
  – *Call* UPDATE(v_{12}).
  – *Set* $k := k - 1$.

Protocol 2 is more robust than Protocol 1 because in the former the users exchange several attributes at a time, not just one, so that some differences in the attribute sensitivity ordering can be tolerated. In Protocol 2, the procedure UPDATE(v_{12}) is used by User 1 to decide whether:

– The reputation v_{12} is kept unaltered in the next iteration;
– The reputation v_{12} is increased to the maximum between v_{12} and the maximum sensitivity of the attributes received from User 2, according to User 1's own sensitivity scale;
– The reputation v_{12} is decreased due to the content of the attributes disclosed by User 2 in the current iteration (*e.g.* if User 2 reveals that she has been in jail, this may be a very sensitive attribute, but probably it will cause her reputation vs User 1 to decrease). In particular, decreasing v_{12} to 0 is a way for User 1 to quit Protocol 2.

Clearly, updating someone's reputation is a procedure that is likely to need the intervention of the human behind User 1, as it involves subjective judgment. Specifically, User 1 (or rather the human behind her) might decide *not* to increase the reputation of User 2 to the the maximum sensitivity of the attributes received from User 2 if User 1 does not wish to correspond to the overtures of User 2. However, even if the reputation stays the same or decreases, User 2 might learn new attributes from User 1 in the next intimacy level $k - 1$, because of the monotonicity of the sensitivities, *i.e.* because $\beta_{ijk'} \leq \beta_{ijk}$ for all $k' < k$. If User 1 wants to make sure that User 2 will not learn any further attribute, User 1 should set

$$v_{12} < \min_i \beta_{i,1,k-1}$$

which will cause the protocol to be quit in the next iteration.

In Protocol 2, User 2, when requested at Step 2b for the first time, acts in slave mode but proceeds much like User 1:

– User 2 assigns an initial reputation v_{21} to User 1 (maybe in a human-assisted way).
– User 2 uses the same value of k as User 1 in every iteration (starting with $k = \ell$).
– User 2 updates v_{21} (similarly to the way described above for User 1 vs v_{12}).
– User 2 decides whether she can disclose the attributes i requested by User in Step 2b above (by setting $V(i,2,1,k) := 1$) by checking whether $\beta_{i,2,k} \leq v_{21}$.

5 Simulation Results

In this section, we report on experimental results. We simulated Protocol 2 ($N = 2$ users). We took the number of attributes to be $n = 10$ and the number of intimacy levels to be $\ell = 16$. We did the following 1000 times:

- Generate attribute sensitivities $\beta_{ij\ell}$, for $i = 1, \cdots, 10$ and $j = 1, 2$ by randomly and uniformly drawing from $[0, 1]$.
- For $k = \ell - 1$ down to 1 generate β_{ijk}, for $i = 1, \cdots, 10$ and $j = 1, 2$ by randomly and uniformly drawing from $[0, \beta_{i,j,k+1}]$.
- Run Protocol 2 for the previous attribute sensitivities. Initial reputations are randomly and uniformly drawn from $[0, 1]$. The human-made decision in UPDATE(v) about the other user's reputation was simulated as follows:
  - Leave v unaltered with probability 0.45.
  - Increase v with probability 0.45 to the maximum between v and the maximum sensitivity of the attributes received from the other user (according to the decision-maker's sensitivity scale).
  - Decrease v with probability 0.1 to a value uniformly and randomly drawn from $[0, v]$.

The average number of iterations performed in one run of Protocol 2 was 10.99, that is, the protocol was quit on average after 11 iterations, out of the maximum 16 iterations.

Figure 1 shows the growth of User 1 and User 2's utilities as Protocol 2 progresses. Utilities are measured with Expression (7). It can be seen that utilities start growing for both users (improving the utility for both users is the purpose of correlated equilibrium) and they stabilize after the first four iterations.

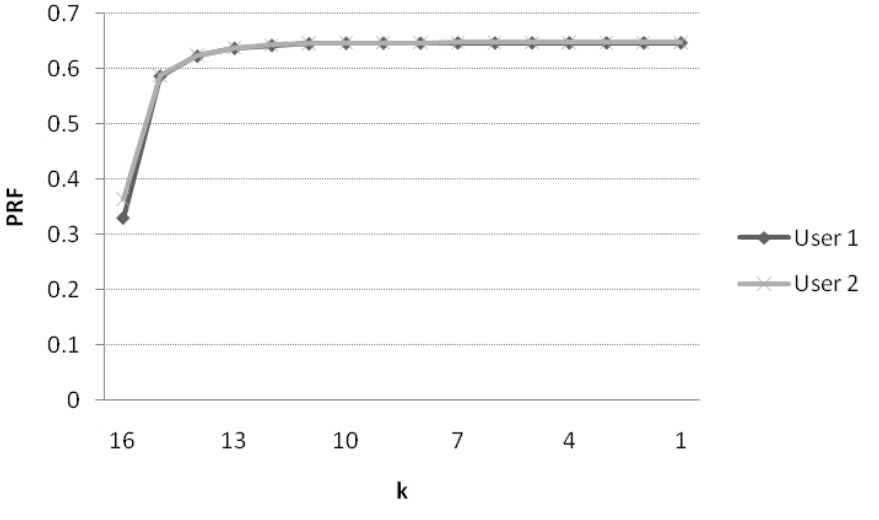

Fig. 1. Evolution of the user utilities measured with Expression (7)

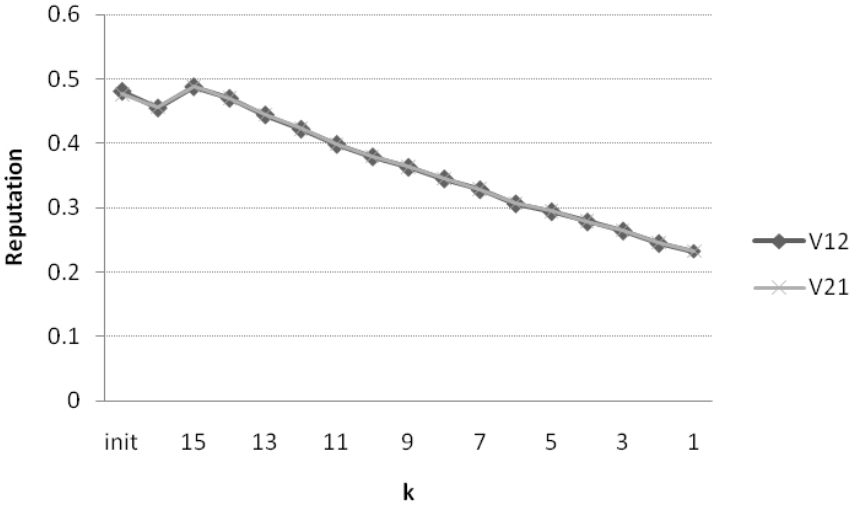

Fig. 2. Evolution of the user reputations

Figure 2 shows the evolution of the reputations User 1 and User 2 assign to each other. Users start with an initial reputation, then this reputation increases because they reveal some highly sensitive attributes; in the next higher intimacy levels, attribute sensitivities are lower because of monotonicity, hence reputation cannot grow due to new high-sensitivity disclosures; it can only decrease, according to the simulated UPDATE(v).

6 Conclusions

We have characterized the utility a user derives from an SN as the information she learns on other users divided by the information she discloses on herself. In terms of this utility, we have shown that, if a user must choose a disclosure strategy without knowing the strategies of other users, her best option is to reveal nothing, which renders the SN useless and provides zero utility to all users. However, better outcomes are possible if users coordinate their disclosure strategies, that is, if they attempt to achieve a correlated equilibrium. We have provided protocols to pursue such an equilibrium by assisting the user in rationally deciding which of her attributes she reveals. Empirical results show that our second protocol results in a utility increase for the two users participating in it.

Future research will include:

– Extending our protocols from pairwise correlated equilibria (taking the users of the SN two by two) to groupwise correlated equilibria (simultaneously correlating strategies of $N > 2$ users);

– Investigating other utility functions which could reasonably model the disclosure attitude of users of SNs for personal contact like Facebook;
– Incorporating concepts such as security and user authentication, which are quite challenging due to the very nature of social networks.

Acknowledgments and Disclaimer

Thanks go to Úrsula González-Nicolás for carrying out the simulations. This work was partly funded by the Spanish Government through projects TSI2007-65406-C03-01 "E-AEGIS" and CONSOLIDER INGENIO 2010 CSD2007-00004 "ARES", and by the Government of Catalonia through grant 2009 SGR 1135. The author is partly supported as an ICREA-Acadèmia researcher by the Government of Catalonia. He holds the UNESCO Chair in Data Privacy, but the views expressed in this paper are his own and do not commit UNESCO.

References

1. Babaioff, M., Chuang, J., Feldman, M.: Incentives in peer-to-peer systems. In: Nisan, N., Roughgarden, T., Tardos, É., Vazirani, V.V. (eds.) Algorithmic Game Theory, pp. 593–611. Cambridge University Press, Cambridge (2007)
2. Carminati, B., Ferrari, E., Perego, A.: Private relationships in social networks. In: Proceedings of ICDE 2007 Third International Workshop on Privacy Data Management, pp. 163–171. IEEE Computer Society, Los Alamitos (2007)
3. Domingo-Ferrer, J., Viejo, A., Sebé, F., González-Nicolás, Ú.: Privacy homomorphisms for social networks with private relationships. Computer Networks 52, 3007–3016 (2008)
4. Liu, K., Terzi, E.: A framework for computing the privacy scores of users in online social networks. In: Proc. of ICDM 2009-The 9th IEEE International Conference on Data Mining, pp. 288–297 (2009)
5. Narayanan, A., Shmatikov, V.: De-anonymizing social networks. In: 30th Symposium on Security and Privacy, Oakland, California, pp. 173–187 (2009)
6. Tardos, É., Vazirani, V.V.: Basic solution concepts and computational issues. In: Nisan, N., Roughgarden, T., Tardos, É., Vazirani, V.V. (eds.) Algorithmic Game Theory, pp. 3–28. Cambridge University Press, Cambridge (2007)