An Ultimatum Game Model for the Evolution of Privacy in Jointly Managed Content

Sarah Rajtmajer[1], Anna Squicciarini[1], Jose M. Such[2], Justin Semonsen[3] and Andrew Belmonte[1]

[1] Pennsylvania State University, University Park, PA, USA
[2] King's College London, Strand, London, UK
[3] Rutgers University, Breswick, NJ USA

Abstract. Content sharing in social networks is now one of the most common activities of internet users. In sharing content, users often have to make access control or privacy decisions that impact other stakeholders or co-owners. These decisions involve negotiation, either implicitly or explicitly. Over time, as users engage in these interactions, their own privacy attitudes evolve, influenced by and consequently influencing their peers. In this paper, we present a variation of the one-shot Ultimatum Game, wherein we model individual users interacting with their peers to make privacy decisions about shared content. We analyze the effects of sharing dynamics on individuals' privacy preferences over repeated interactions of the game. We theoretically demonstrate conditions under which users' access decisions eventually converge, and characterize this limit as a function of inherent individual preferences at the start of the game and willingness to concede these preferences over time. We provide simulations highlighting specific insights on global and local influence, short-term interactions and the effects of homophily on consensus.

1 Introduction

We aim to investigate the impact of multi-party decision sharing in a social network. In highly connected networks, content sharing is frequent and users make decisions about the amount and type of content they choose to share, as well as their preferred privacy preferences. Previous work has largely investigated how to reconcile users' (possibly conflicting) privacy preferences with respect to commonly owned (or jointly managed) content [34, 16]. For instance, the typical example used in the literature is that of a photo in which multiple users are depicted, they have conflicting privacy preferences as to with whom the photo would be shared in a social network, and they use a (technology-aided) reconciliation method to resolve the conflicts. Despite the amount of work in this area, the impact of these interactions over time - both on users and on the content shared - regardless of the reconciliation method, is largely unexplored. In particular, we are yet to understand how individuals' sharing decisions change over time, who are the most influential users, how they benefit from it, and the privacy gains and losses from a collective perspective.

This "research gap" is possibly due to two (related) reasons. First, to our knowledge, proposed content sharing models to date have not been translated into practical features or applications: social networks provide minimal support for multi-party decision making tools. Hence, an exploration in the wild of the effects of multi-party sharing is fundamentally hard. Second, to date, work that focuses on multi-party sharing has adopted a micro-scale view of the interactions among users (i.e., one-on-one and one-shot interactions), in an attempt to minimize discomfort and other security properties from a one-interaction at a time standpoint.

In this paper, we aim to answer a broader and, we believe, more important set of questions about the potential longitudinal effects of repeated negotiations over jointly managed content among users in a social network. We assume, consistent with reality [39, 20, 3], that users wish to reach agreement and share content jointly. Over time, this will lead users to feel pressure to move away from their individual preferred settings and toward the preferences of their peers. In doing so, some users will experience sharing loss, while others will experience privacy loss. In this setting, our specific questions are:

- How does multi-party involvement in access control decisions affect the individual behaviors of social network users?
- What are the collective privacy gains and losses associated with multi-user sharing?
- Bearing in mind that users adopt individual strategies to respond to access decisions for shared content, which users are more likely to drive group decisions? Likewise, which users are most likely to benefit from repeated interactions?

We model user interactions through a repeated game. Specifically, evidence indicates that one-shot decisions for multi-party access control may be well-described using the language of the Ultimatum Game, specifically a natural tension between selfish preferences (i.e., maximizing a personal utility function) and a less-tangible desire to cooperate [34, 39, 20, 3]. That is, empirical studies about multi-party access control showed that users are naturally selfish and seek to impose their preferences as much as they can even when they know other stakeholders may not be happy about it [34], but at the same time users do collaborate [39] as they do not want to cause any deliberate harm to other stakeholders and would normally consider their preferences and potential objections in a more cooperative way [20, 3].

Accordingly, we present a variation of the one-shot Ultimatum Game, wherein individuals interact with peers to make a decision on a piece of shared content. The outcome of this game is either success or failure, wherein success implies that a satisfactory decision for all parties is made and failure instead implies that the parties could not reach an agreement. This approach was inspired by recent work of fairness in the Ultimatum Game [42].

Our proposed game is grounded on empirical data about individuals' behaviour in one-shot, multi-party access control decisions [35, 34, 39] mentioned above to structure repeated pairwise negotiations on jointly managed content in a social network. We theoretically demonstrate that over time, the system converges towards a “fair” state, wherein each individual's preferences are accounted for. In this state, users' preferred privacy values approach a constant value that is dependent on how stubborn individual users are, until all values are within a window of compromise (which in turn depends on the structure of the network). We also carry out a series of numerical experiments on simulated data, and provide insights on a number of interesting cases, e.g., when a number of perfectly stubborn users (i.e. users unwilling to compromise or adapt to other users' preferences) are at play, when highly connected users exist in the network, and when networks are homogeneous.

The paper is organized as follows. In the next section, we highlight our assumptions and the problem statement. In Section 3, we present our theoretical model. We discuss theoretical results in Section 4 and provide experimental insights in Section 5. We overview related work in Section 6. Finally, we conclude the paper with a discussion of limitations and future work in Section 7.

2 Problem Statement

We consider an online social network wherein linked users, i.e., two users connected by an ”edge” in the social network graph, may jointly manage content. While one user is typically first to share a given piece of content, henceforth the ”poster”, other users, henceforth the ”stakeholders”, may also be affected by the content (e.g. a photo in which she is depicted). Users, both posters and stakeholders, likely differ in both structural and inherent qualities. Structurally, they have variable numbers of friends, i.e., degree ($deg(n)$), and variable (closeness, betweenness) centrality. Inherently, users may differ in propensity for sharing [22] and stubbornness [2, 40].

As a piece of jointly managed content is considered, the stakeholder has the opportunity to accept or decline the privacy settings selected by the poster — a decision that is made based on a joint effect of inherent sharing preference, stubbornness, the personal relationship between the two users and the nature of the content iteself. Access settings, then, are co-determined by posters and stakeholders using a one-round negotiation, which we model as a one-shot Ultimatum Game.

An important assumption underlying this game is that the proposer and responder *would like* to reach agreement. First, the underlying social network structure implies that the proposer and responder are friends, acquaintances or members of a social cohort. Reaching agreement represents social harmony that is preferable, and empirical evidence tells us that both posters and stakeholders listen to and consider each others' preferences and objections [39]. In some cases, agreement may be required for content to successfully be posted. In other cases, the proposer may have authority to post content at his desired privacy level without consent of other stakeholders, but she hesitates to do so understanding that her cohort may take the same liberty with future content, or because they put themselves in the position of stakeholders and understand they may not be happy with the content shared [34]. In order to reach agreement, both proposer and stakeholders understand they must concede (part of) their preferences and move toward some compromise privacy setting [39, 20, 3]. However, the amount each party shifts (or concedes) may not be the same, and its likely influenced by their individual propensity for sharing [22] and stubbornness [2, 40] as stated above.

We study the impact of this variant of one-shot ultimatum games over time, and specifically, the extent to which these one-shot interactions, wherein users must compromise (as much as they feel comfortable) in order for content to be shared, is conducive of a "fair" system. Here, by fair system, we refer to a system wherein each user is given an equal opportunity to participate in an interaction, based on his/her current degree in the network graph. Furthermore, each user is free to respond based on his own preferences and inclinations, and each user's response for each game equally influences system dynamics. Given these equitable rules of the game, answers to the three research questions posed above may shed light on the ways in which outcomes are and are not as equitable.

Following, we discuss the model and its outcomes with focus on the case of one poster and one stakeholder, for simplicity of presentation. Note however that this is not a loss of generality, as k asynchronous players are essentially a specific ordering of 2-player interactions.

3 The Model

We play a variant of the one-shot ultimatum game [42], repeatedly, amongst pairs of individuals situated within a social network graph. The rules of the game, which are formally specified below, reflect the real-world scenario of multi-party sharing, namely determining access settings for content associated with multiple stakeholders [33, 15, 25]. These rules formally capture empirical evidence of concession behaviour in multi-party sharing [33], like being generally accommodating to the preferences of others to reach agreement [39, 20, 3].

Consider a social network graph $G = \{V, E\}$ where V is the set of users, represented as nodes in the graph. The set E of pairwise links between nodes represents relationships, or more generally, users with some connection who may both be party to the same content. Links may be weighted according to a weigh function W_{ij}, where weights between users i and j indicate strength of relation, or strength of social influence.

Each user i has an inherent, personal *comfort* C_i with sharing and an inherent *stubbornness* T_i that do not change over the lifespan of the game. Both are represented as value in $[0, 1]$. In the case of comfort, 0 indicates private and 1 public[4]; likewise for stubbornness, 0 is least stubborn and 1 most stubborn. Each

[4] Note that we abstract ourselves from the actual privacy settings or access control paradigm used by the online social network provider. For each social media infrastructure or privacy policy language used, a mapping could be defined that turns available settings into values in $[0, 1]$ and vice versa.

user is also perpetually endowed with two dynamic values – a “proposal” value $P_i(t)$, and a “response” value $R_i(t)$. These values represent the user’s preferred settings when acting as the content owner (”poster”) or when party to content posted by someone else, respectively, which is aligned with empirical evidence that shows that the perceptions and behaviours of users are significantly different when they are playing the role of poster or stakeholder [34]. Changes in these values over time are governed by the set of rules of the game, detailed as follows.

We initialize the proposal value and the response values for each user as his comfort value, i.e., $P_i(0) = R_i(0) = C_i$. The intuition here is that, without the influence of peers (i.e., without playing the game) a user is inclined to both offer and accept the sharing level for a piece of content that most closely matches his comfort level.

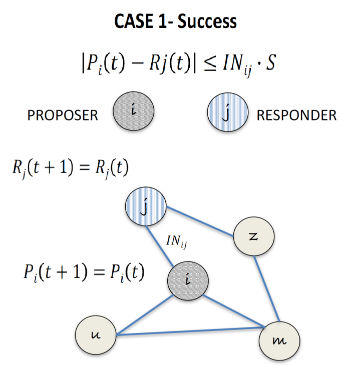

Fig. 1: Example of successful interaction and update rule

The game is played for some fixed number of iterations. At each iteration, a “proposer” is chosen at random. Intuitively, this is the owner/poster of a piece of content in which other users have a stake. A “responder” is selected at random from among his contacts, namely those users adjacent on the social network graph. The proposer offers his proposal value to the responder, i.e. the privacy level or disclosure setting for the co-owned content to be shared. The responder in turn accepts or declines this offer. Intuitively, the decision to accept or decline represents the responder’s approval or disapproval of the proposed privacy setting. This decision is made based on the responder’s *willingness to compromise*, which in turn relies primarily on two factors: (1) the *strength of influence* of the proposer on the responder, i.e., their relationship strength (possibly asymmetric) [11, 10], and (2) the *sensitivity* of the content in question — if a user feels that an item is very sensitive for her, she will be less willing to approve sharing [38, 30]. Conditions for acceptance and success of an interaction are given in the next definition and examples of successful and unsuccessful interactions are depicted in Figure 1 and Figure 2.

Definition 1 (Successful Interaction Conditions). *Let the strength of influence of user i on user j be represented by a value in $IN \in [0, 1]$, with 0 indicating most weak and 1 most strong. Likewise, let the sensitivity of the content be denoted $S \in [0, 1]$, with 0 most sensitive and 1* least sensitive. *An interaction is successful, i.e., the responder j accepts the proposer’s (i) proposal if*

$$|P_i(t) - R_j(t)| < IN(i, j) \times S$$

and a failure otherwise.

After each interaction, the involved players' proposal and response values are updated [42], as follows:

– *If the interaction is successful*, the proposer and responder do nothing. Specifically, P_i and R_j remain the same moving forward in time.
– *If the interaction is unsuccessful*, the proposer and responder move their proposal and response values, respectively, by some amount modulated by the *stubbornness* of each individual user *toward the midpoint* of the two as a way of conceding, so that future interactions are more likely to be successful.

$$P_i(t+1) = P_i(t) \times T_i + \frac{P_i(t) + R_j(t)}{2} \times (1 - T_i) \tag{1}$$

and

$$R_j(t+1) = R_j(t) \times T_j + \frac{P_i(t) + R_j(t)}{2} \times (1 - T_j). \tag{2}$$

The rules above capture notions of social influence and empirical evidence of multi-party access control decisions. In particular, informed by Fredkin's social influence theory [13], stating that strong ties are more likely to affect users' opinions and result in persuasion or social influence, in both Equation (1) and Equation (2) users will move toward their peers values. This is consistent with empirical evidence about

multi-party access control decisions that showed that both proposer and stakeholders are willing to collaborate and make concessions toward some compromise privacy setting [35, 39, 20, 3]. The amount each party shifts (or concedes) may not be the same, as each party may be influenced by peers only to a certain point [36] driven by their stubbornness [2, 40] and degree of selfishness [34].

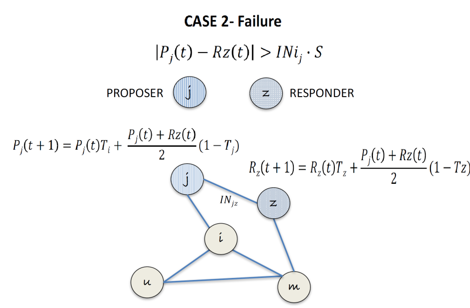

Fig. 2: Example of failed interaction

Of note, the proposer i does not change his response value and the responder j does not change his proposal value moving forward, i.e., $R_i(t+1) = R_i(t)$ and $P_j(t+1) = P_j(t)$. Likewise, all players in the game who were not involved in the interaction undergo no change in either proposal or response

4 Theoretical Findings

In this section, we present our theoretical findings for the proposed Ultimatum Game. We demonstrate that, unless trivially impossible, the system converges towards a consensus state, wherein each individual's preferences are accounted for. In this state, both proposal and response values approach a constant c that is dependent on the stubbornness values associated with individual users, until all values are within a window of compromise which depends on the structure of the network.

4.1 Energy Conservation on Repeated Iterations

We first derive the following technical lemma on energy conservation, which will help determine conditions and value of convergence.

Lemma 1. *In an ultimatum game, let P_i and R_i be proposer's and offeror's value for user i, with P_i and R_i defined according to Equations (1) and Equation (2), respectively. When $T_i \neq 1 \, \forall \, i$, the quantity $\sum_i \frac{P_i(t)+R_i(t)}{1-T_i}$ is conserved.*

Proof. Consider a single Ultimatum game, with proposer i and responder$j$.

If the proposal is accepted, neither P_i nor R_j change, and no other proposal or response values are affected, so every value remains the same, and thus the weighted sum is unaffected.

If the proposal is rejected, then the new proposal value becomes $P_i' = T_i P_i + (1-T_i)\frac{P_i+R_j}{2}$ and the new response value becomes $R_j' = T_j R_j + (1-T_j)\frac{P_i+R_j}{2}$. Since no other values are changed, then:

$$\sum_i \frac{P_i' + R_i'}{1-T_i} - \sum_i \frac{P_i + R_i}{1-T_i} = \frac{P_i' - P_i}{1-T_i} + \frac{R_j' - R_j}{1-T_j} = \frac{P_i + R_j}{2} - P_i + \frac{P_i + R_j}{2} - R_j = 0$$

This means that regardless of which proposals are given or whether or not they is accepted, the quantity $\sum_i \frac{P_i(t)+R_i(t)}{1-T_i}$ remains constant. $\diamond$

We will show in the next subsection that the $P_i(t)$ and $R_i(t)$ converge to a given constant c. Using the relation obtained in Lemma 1, we posit that the constant c must be the unique constant for which this sum is conserved. Therefore

$$c = \frac{\sum_i \frac{P_i(0)+R_i(0)}{1-T_i}}{\sum_i \frac{2}{1-T_i}} \tag{3}$$

Next, we define the following vector $\mathbf{d}(t)$. We compute $|P_i(t)-c|$ and $|R_i(t)-c|$ for each i, and sort each difference in non-increasing order. We show that $\mathbf{d}(t)$ constructed in this way decreases in lexicographical order over time, and therefore $P_i(t)$ and $R_i(t)$ both approach c. The following Lemma holds.

Lemma 2. *At each time step t, the inequality $\mathbf{d}(t+1) \leq_{lex} \mathbf{d}(t)$ is verified. In particular, at time $t+1$, $\mathbf{d}(t+1) <_{lex} \mathbf{d}(t)$ when the conditions of acceptance per Def. 1, are not met, and P_i is rejected by j.*

Proof. Consider a single ultimatum game taking place at time t, with proposer i and responder j. If the proposal P_i is accepted (i.e. acceptance condition per

Definition 1 hold true), no changes are made to P_i and R_j. Since this proposal does not affect any other proposal or responder values, $\mathbf{d}(t+1) = \mathbf{d}(t)$.

If the proposal is rejected, then let $a = \max\{|P_i(t) - c|, |R_j(t) - c|\}$ and $b = \min\{|P_i(t) - c|, |R_j(t) - c|\}$. Note that rejection means that $a > b$.

Since we sort these differences (including a and b) in non-increasing order, let k be the index of the last occurrence of a in $\mathbf{d}(t)$. We note that:

$$
\begin{aligned}
|P_i(t+1) - c| &= \left| T_i P_i + (1 - T_i)\frac{P_i + R_j}{2} - c \right| \\
&\leq T_i |P_i - c| + (1 - T_i)\left| \frac{P_i + R_j}{2} - c \right| \\
&\leq \frac{1 + T_i}{2} |P_i - c| + \frac{1 - T_i}{2} |R_j - c|
\end{aligned}
$$

Assuming $T_i < 1$, $|P_i(t+1) - c| < a$. Similarly, $|R_j(t+1) - c| < a$. Since no other values in $\mathbf{d}(t)$ change, this means that $\mathbf{d}_k(t+1) = \max\{|P_i(t+1) - c|, |R_j(t+1) - c|, \mathbf{d}_{k+1}(t)\}$. All of these possibilities are strictly smaller than $\mathbf{d}_k(t) = a$. Since none of indices preceding k are affected, the inequality $\mathbf{d}(t+1) <_{lex} \mathbf{d}(t)$ holds. $\diamond$

The lemma essentially shows that so long as there is a positive probability that a proposal will fail to be accepted, $\mathbf{d}(t)$ will converge towards $\mathbf{0}$, meaning that P_i and R_i will all converge to c.

Next, we formally identify conditions under which failure has to be possible. In this case, unlike Lemma 1 and Lemma 2, the results are influenced by the structure of G, the sensitivity of content S and influence between players IN_{ij}.

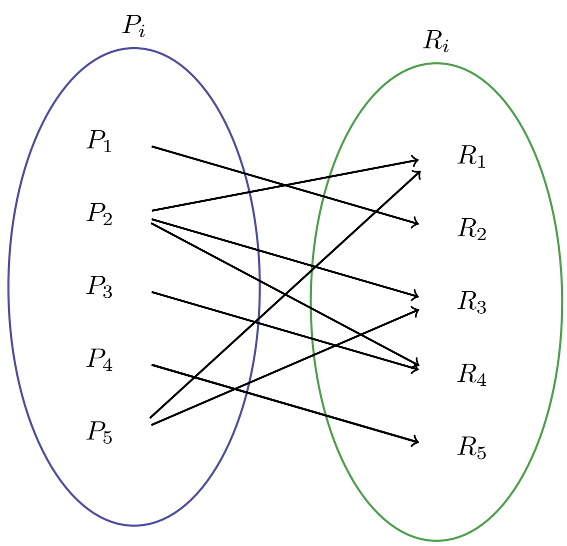

Fig. 3: Example of auxiliary graph

4.2 Convergence Results

We first create an auxiliary graph wherein we split apart the P_i and the R_i values for every user i. In this auxiliary graph, each P_i and R_i is associated with its own vertex. Because every game iteration involves one P_i and one R_j value (and never a P_i with a R_i or even another P_j value), this graph will be bipartite. An example of this type of graph is reported in Figure 3.

Definition 2 (Auxiliary Graph). *Let (G, V) be a connected graph, wherein each $i \in G$ is associated with values (P_i, R_i). H is the auxiliary graph obtained* *by taking 2 copies of the vertices of G. Label the vertices by i^1 and i^2 respectively. Then, let $i^1 \sim_H j^2, i^2 \sim_H j^1 \iff i \sim_G j$. We will associate i^1 with P_i and i^2 with R_i.*

In the general case (i.e. G is connected and not bipartite), there is an odd cycle $i_1 i_2 \dots i_k i_1$ in G. This means that i_1^1 is connected to i_1^2 in H by the path $i_1^1 i_2^2 \dots i_k^1 i_1^2$. Because G is connected, this means that H is also connected.

We will use the notation $\text{diam}_{IN}(G)$ to denote the usual diameter of G with edge weights given by IN. The same is true for H, where the weight of an edge (i^1, j^2) is the same as the weight in G of (i, j).

Lemma 3. *When G is not bipartite, while any of $|P_i - R_j|, |P_i - P_j|$, or $|R_i - R_j| > \inf\{s : s \in S\} \cdot \text{diam}_{IN}(H)$, there is a positive probability that $\mathbf{d}(t+1) <_{lex} \mathbf{d}(t)$.*

Proof. Assume that $\mathbf{d}(t+1) = \mathbf{d}(t)$ with probability 1. By Lemma 2, this means that every possible ultimatum game (each edge that can be chosen with positive probability) results in acceptance. This means that for any i and j adjacent, $|P_i - R_j| < s \cdot IN(i, j)$ for any $s \in S$, so $|P_i - R_j| \leq \inf\{s : s \in S\} \cdot w_{(i^1, j_2)}$.

Since there is a path between vertices i^1 and j^2 in H (H constructed according to Definition 2), then $|P_i - R_j| \leq \inf\{s : s \in S\} \cdot d_{IN}(i^1, j^2)$, and thus $|P_i - R_j|, |P_i - P_j|, |R_i - R_j| \leq \inf\{s : s \in S\} \cdot \text{diam}_{IN}(H)$ for any i and j. $\diamond$

Per the above lemma, given enough iterations of the game, every value of P_i and R_i for all vertices in V will converge in a window of size $\inf\{s : s \in S\} \cdot \text{diam}_{IN}(H)$. Note that since c is a weighted average of these values (see Equation 3), c is in the window.

Special considerations must be made for the (rare) case of G being bipartite. If G is bipartite, then let the partition of the vertices be V_1 and V_2. Note that if $i \sim j$ then $i^1 \sim j^2$, so $\{v^1 : v \in V_1\}$ and $\{v^2 : v \in V_2\}$ are connected in H, and in fact form a subgraph H' of H that is isomorphic to G. Similarly $\{v^2 : v \in V_1\}$ and $\{v^1 : v \in V_2\}$ also form a subgraph H'' of H that is isomorphic to G.

To analyze this case, we use the technique from Lemma 3 on each part of the disconnected H:

Lemma 4. *When G is bipartite, while any of $|P_i - R_j|, |P_i - P_j|$, or $|R_i - R_j| > 2 \cdot \inf\{s : s \in S\} \cdot \text{diam}_{IN}(G)$, there is a positive probability that $\mathbf{d}(t+1) <_{lex} \mathbf{d}(t)$.*

Proof. Per Lemma 3, let the partition of the vertices be into sets V_1 and V_2. Considering only the $\{P_i : i \in V_1\}$ and $\{R_i : i \in V_2\}$, we can use Lemmas 1 and 2 to define c' and $\mathbf{d}'$ to only consider those values. We can also define c'' and $\mathbf{d}''$ to be defined using only $\{R_i : i \in V_1\}$ and $\{P_i : i \in V_2\}$.

Because we consider only one of each P_i and R_i, algebraic manipulation shows that $c = \frac{c' + c''}{2}$. However, since $R_i(0) = P_i(0)$ for every i, we note that $c' = c''$, and thus both are equal to c. This means that vector $\mathbf{d}$ is simply a reordering of the entries of $\mathbf{d}'$ and $\mathbf{d}''$.

Using the same techniques used to prove Lemma 3, it is easy to show that if $\mathbf{d}'(t+1) = \mathbf{d}'(t)$ with probability 1, then all $P_i(t)$ and $R_i(t)$ in H' fall within a window of diameter $\inf\{s : s \in S\} \cdot \text{diam}_{IN}(H')$ that contains c.

In an identical manner, all $P_i(t)$ and $R_i(t)$ in H'' fall within another window of diameter $\inf\{s : s \in S\} \cdot \text{diam}_{IN}(H'')$ that also contains c. However, since both H' and H'' are isomorphic to G including edge weights, the union of these windows is a window of diameter less than $2 \cdot \inf\{s : s \in S\} \cdot \text{diam}_{IN}(G)$. $\diamond$

By relying on the two lemmas above, we now derive the following theorem for general values of T_i:

Theorem 1. *For any graph G, one of the following will occur:*

1. *If $T_i \neq 1 \, \forall \, i \in V$, all P_i and R_i will eventually converge to a window of size $2 \cdot \inf\{s : s \in S\} \cdot \text{diam}_{IN}(G)$ or $\inf\{s : s \in S\} \cdot \text{diam}_{IN}(H)$ as appropriate around c.*
2. *If $\exists \, f$ such that $P_i(0) = R_i(0) = f \, \forall \, i : T_i = 1$, all P_i and R_i will eventually converge to a window of size $2 \cdot \inf\{s : s \in S\} \cdot \text{diam}_{IN}(G)$ or $\inf\{s : s \in S\} \cdot \text{diam}_{IN}(H)$ as appropriate around f.*
3. *Otherwise, consensus is impossible.*

Proof. The proof considers each case. For case 1: If $T_i \neq 1 \, \forall \, i \in V$, then by Lemmas 3 and 4, there is a positive probability of decrease if this window is larger than $2 \cdot \inf\{s : s \in S\} \cdot \text{diam}_{IN}(G)$ or $\inf\{s : s \in S\} \cdot \text{diam}_{IN}(H)$. This means that eventually the size of the window will decrease.

With respect to case 2: If $\exists \, f$ such that $P_i(0) = R_i(0) = f \, \forall \, i : T_i = 1$ Let $f' = \frac{\sum_{i:T_i \neq 1} \frac{P_i + R_i}{1 - T_i}}{\sum_{i:T_i \neq 1} \frac{2}{1 - T_i}}$ be the weighted average value over all the less stubborn players.

In the same manner as in Lemma 1, f' is conserved for any outcome of any ultimatum game between two vertices from the set $\{i : T_i \neq 1\}$. Note that the same is true for any outcome of of any ultimatum game between two vertices from the set $\{i : T_i = 1\}$, as well as a game with a vertex from each set where the ultimatum is accepted.

If we have a game with a vertex from each set where the ultimatum is not accepted, then without loss of generality, let $T_i \neq 1$. This means that $P_i(t+1)$ (or $R_i(t+1)$) is a weighted average of $P_i(t)$ (or $R_i(t)$) and f, so $f'(t+1)$ is a weighted average of f and $f'(t)$. This means that f' approaches f.

Using the same techniques as Lemmas 2, 3, and 4 with f instead of c, all P_i and R_i will eventually converge to a window of size $2 \cdot \inf\{s : s \in S\} \cdot \text{diam}_{IN}(G)$ or $\inf\{s : s \in S\} \cdot \text{diam}_{IN}(H)$ around f.

Finally, for case 3: if $\exists \, P_i(0) = R_i(0) \neq P_j(0) = R_j(0)$ for $T_i = T_j = 1$, then for any $t, P_i(t) = R_i(t) \neq P_j(t) = R_j(t)$, so trivially no consensus is possible. $\diamond$

In summary, if content sensitivity can be arbitrarily small, unless there is trivially no way to establish consensus, then all players will converge to a consensus based on their stubbornness values. The rate of convergence will actually depend on the topology of the network, and on how homogeneous users' comfort values and stubbornness levels are. We provide some insights on these dimensions in the next section.

5 Empirical Results

Our convergence results guide understanding of behavioral trajectories in a social network. However, some interesting and more practical issues are unaccounted for in our analysis, especially with respect to the effects of scale. That is, large social networks may have multiple stubborn users, users who interact very often, or those who interact very infrequently. Informed by our theoretical findings, we can further our understanding of these effects through controlled experiments, varying specific parameters of the game that we anticipate may play a significant role in real-world networks.

Through simulation, we explore the effects of specific personal and structural characteristics (e.g., stubbornness, degree) at the node level as they relate to short- and long-term evolution of privacy preferences for jointly managed content throughout the system. We study:

1. The role of the stubborn users, and their evolution in the network (e.g. how does an extremely opinionated user affect others? Does his/her behavior change over time?);
2. The role of high-degree users, and their relative rates of successful or failed interactions (i.e. are popular users more likely to experience successful interactions?); and
3. The short- vs. long-term nature of observed effects (i.e. is convergence to a fair value possible in the short term? if so, under which conditions)?

5.1 Local influencers: stubbornness and connectivity

Two types of users are likely to affect the dynamics of our system. These are highly stubborn users (i.e. $T_i \simeq 1$), and highly popular users $deg(i) > avg(deg(n))$. Users with high stubbornness (who are slower to concede their preferences) are influential in their neighborhood. Recall that, per Theorem 1 in the extreme case of a fully-connected graph with exactly one perfectly stubborn user ($T_i = 1$), given the conditions of our model described, all users' proposal and response values will converge to his comfort level.

In the case that a network has multiple stubborn users, each becomes a *local influencer*, with the speed and diameter of influence dependent on local connectivity patterns and proximity to "competing" stubborn users. In this way, stubborn users typically serve as centers of "communities", closely aligned with community structure detected by classic community structure detection algorithms.

We study this case through a simulation through the benchmark Karate Club network ($N = 34$) [41]. The Karate Club network is used as a first network topology as it is well understood and its small size allows for explicit tracking and visualization at the individual node level. In addition, the close-knit peer group represents a micro-scale view of a larger social system.

Over this network structure, we start from baseline assumptions that: 1) Users' inherent privacy preferences; 2) Users' stubbornness scores T_i; 3) Influence scores over pairwise links $IN(i, j)$; and 4) Content sensitivity scores are all uniformly distributed in $[0, 1]$. We let zero represent least inclined to share, least stubborn, least influenced and least public (most sensitive), respectively.

Consider the following representative example (Figure 4). User comfort levels, i.e., initial proposal and response values for all users are taken from a uniform random distribution in $[0, 1]$, with the exception that user 1 is seeded with a comfort of 0.1 (strict sharing) and user 34 with a comfort of 0.9 (public sharing). In addition, users 1 and 34 are seeded as perfectly stubborn, i.e., $T_1 = T_{34} = 1$. The left hand image visualizes initial comforts, equivalently initial proposal values, for all individuals in the network. Nodes are colored on a temperature scale, with blue representing 0 and red representing 1. The game is run to convergence (10K iterations of play), and the resultant final proposal values are reflected on the right; note that proposal and response values are equivalent in the limit. Consistent with our theoretical findings, we see convergence around stubborn users. In this case wherein multiple perfectly stubborn users are present within a single connected component of the graph, convergence is localized around each and specific diffusion of influence depends on the local connectivity patterns.

As such, high centrality (degree, betweenness, closeness) users may play an interesting and notable role in system-wide behaviors. Specifically, we suggest that users embedded in their local communities may have more rapid influence; in time-limited real-world scenarios of evolving social graphs, more rapid influence likely means wider influence as well. In addition, high-degree users "play" more often (are involved in more shared content and subsequent negotiations); in the framework we have described, highly connected users may be selected as responder any time a friend is selected as proposer. In sum, location, degree, number and extent of stubborn users are interrelated determinants of system-wide preferences, i.e., proposal and response values, at convergence.

5.2 Evolution of the Ultimatum Game at the slow time scale

We have theoretically (Sect. 4) and experimentally described the behavior of the ultimatum game at the *long time horizon*. These analyses allowed for a formal understanding of limit behavior, but were not necessarily realistic in real-world, time-constrained scenarios. Here, we consider the implications of our findings at the shorter-term horizon, or for more sparse interactions.

Estimates indicate [5] that Facebook users share on average 0.35 photos per day, or 1 photo every 3 days (350 million photos per day, divided by 1 billion active users per day). In our small network of 34 users, we estimate 12 instances of sharing/interactions per day. Provided a static network structure, over the course of one month the game is played for 360 iterations. Figure 5 illustrates the influence of one stubborn user (user 3) after 500 iterations of play. We seeded

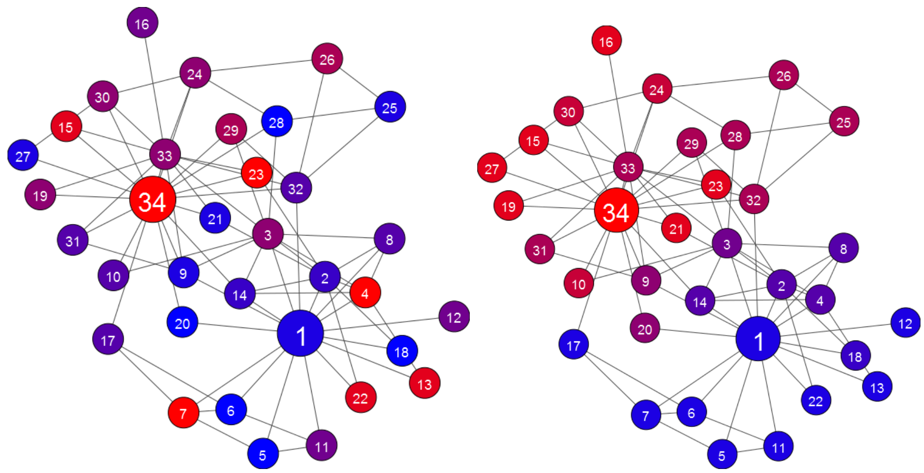

Fig. 4: Initial (left) and converged (right) proposal values on the karate club network.

user 3 with a sharing comfort of 0.9 and all other users with comfort 0.1. User 3 was seeded as perfectly stubborn, $T_3 = 1$, while other users' stubbornness values were taking from a uniform distribution on $[0, 1]$. We have proven that in this contrived but important extreme case, all users will eventually converge to proposal and response values at 0.9. However at the shorter horizon, notice the local influence of user 3, where variants in neighbors' final values are attributable to connectivity patters, number of interactions and inherent stubbornness.

Stubborn users, then, seem to be playing a winning strategy. In the long term, they pull other users toward their own preferences and exhibit greater influence regionally over time. However, we note one consequence for stubborn users, namely a greater expectation of failed interactions. As their peers move more quickly toward compromise and bring their own preferences in line with their neighbors, stubborn users are slower to narrow this gap. Accordingly, as pairs of less stubborn users reach preferences within the window of consensus and begin to increase their rate of successful interactions, all else (connectivity, preferences) being equal, stubborn users continue to fail further into the game. In addition to less stubborn users, it can also be said that users whose comfort level is nearer to the mean (in our case, the mean is fixed at 0.5) experience more successful interactions with their peers. That is, the expected value of the difference between their own proposal/response value and that of their neighbor is lesser than the expected value of that difference for a user with a preference nearer to either end of interval.

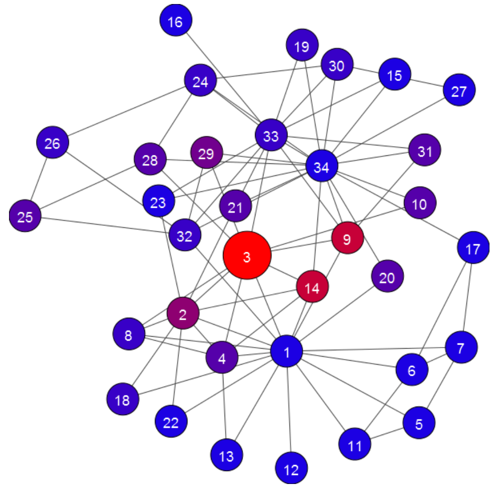

Fig. 5: Proposal values after 500 iterations of the ultimatum game. Observe node 3 and its local influence in its neighborhood.

5.3 The importance of homophily

Our last observation brings us to an important consideration. The examples we have provided thus far have involved fixing personal preferences and stubbornness near extreme values in order to demonstrate effects. However, consider expected scenarios where connected users have generally similar preferences and are in general moderately stubborn. The social science literature on homophily provides evidence that real world social systems are well-modeled using 'birds of a feather' assumptions [27, 23].

Consider the same network of Figure 5, wherein initial preferences and stubborness are distributed in a uniform (and random) fashion from the interval $[0.3, 0.7]$. Figure 6 represents the preferences of all 34 users over 5000 rounds of play. Notice, in this framework, all players' values tend to converge to a common small range of values, with less than 0.2 separating the preferences of any two users in the network.

This tendency toward agreement in more homogeneous communities holds implications for prototypical real-world social networks wherein densely linked groups of users tend to be more 'similar' by some measure. It is an open question whether documented instances of homophily in social systems extends to privacy preferences as well, but our model would suggest that it may. Furthermore, anomalous cases for which this tendency is not observed may be indicative of areas for deeper investigation.

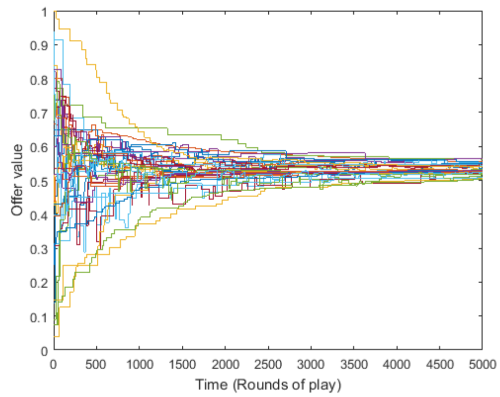

Fig. 6: Proposal values plotted for 34 players over 5000 iterations of the ultimatum game for multi-party content.

6 Related Work

Our work lies at the crossroad of game theory for modeling social interactions and multi-user access control.

There is a long history using the ultimatum game to model pairwise interactions amongst individuals seeking to rectify opposing forces of cooperation and selfishness[8, 37]. In particular, in the Ultimatum Game, one player proposes a division of a sum of money between herself and a second player, who either accepts or rejects. Based on rational self-interest, responders should accept any nonzero offer and proposers should offer the smallest possible amount. Traditional, deterministic models of evolutionary game theory agree: in the one-shot anonymous Ultimatum Game, natural selection favors low offers and demands. However, experiments in real populations reveal a preference for fairness. When carried out between members of a shared social group (e.g., a village, a tribe, a nation, humanity) people offer "fair" splits close to 50-50, and offers of less than 30% are often rejected [28, 14]. There are several theories as to why this difference between theoretically optimal and practical behaviors may exist, including reputation and memory effects [6], natural selection [26], empathy and perspective taking [24]. In [42], we study this phenomenon using a similar model to that presented in this paper, but in a general setting unrelated to privacy and access control. Accordingly, the formulation explored in [42] involves a more general rule set, leaning on notions of greed and charity, rather than consensus-formation.

With respect to privacy and related decision making processes, researchers from many communities have noted the trade-off between privacy and utility (e.g., [21, 31, 4, 7, 32, 29]). The majority of this prior work tends to view the privacy/utility trade-off as mutually exclusive: an increase in privacy (resp. utility) results in an immediate decrease in utility (resp. privacy). We note that the interplay of multiple entities in any access control/privacy decision where privacy and utility are unevenly distributed among the players and context-dependency results in a complex relationship between these concepts [1, 19]. A growing body of recent work has focused on multi-party access control mechanisms, some of which have used game-theoretical concepts. Chen et al. 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 [9]. Hu et al. tackle the problem of multi-party access control in [17], proposing a logic-based approach for identification and resolution of privacy conflicts. In [18] these authors extend this work, this time proposing adopting a game-theoretic framework, specifically a multi-party control game to model the behavior of users in collaborative data sharing in OSNs. Another game-theoretic model is given in [35], in which automated agents negotiate on behalf of users access control settings in a multi-user scenario. Other very recent approaches to multi-party access control mechanism use a mediator [33] or a recommendation system [12] to suggest the optimal decision in one-shot multiparty access control scenarios. The primary difference between our work and previous ones on multi-party access control (whether game-theoretic or not) is our unique focus on the effects of one-time interactions to a given network, and the related consequences for users in the network over a number of interactions.

7 Conclusions and Future Work

In this paper, we presented a macro-model to describe how individuals' sharing decisions change over time, who are the most influential users and how they benefit from it, along with privacy gains and losses from a collective perspective. Through a carefully designed ultimatum game, informed by the body of work on multi-party access control, we were able to capture the most important dynamics underlying privacy decision making in online social networks. Our results show users' overall tendency to converge toward a self-adjusted environment, wherein successes and failures commensurate with users' stubbornness and underlying network dynamics.

This work is the first step toward a more systematic analysis of how people's privacy attitudes evolve over time, and change their personal information sharing patterns as a result. As such, we anticipate several extensions and possible avenues for research.

Further theoretical work may look into the system's convergence properties for nonvanishing content sensitivity and study time to convergence (within some bounds) in network topologies that reflect real-world social structure. Related to this, convergence in a practical sense will reflect agreement on a *discrete* privacy setting and accounting for this will impact these findings.

With respect to discretization of privacy settings, our model is thus far agnostic to the actual privacy settings or access control paradigm used by the online social network provider. We plan to define a mapping that converts available settings into values in $[0, 1]$ and vice versa. For instance, default Facebook settings go from private, to friends, to friends of friends, to public. Also, users may choose particular users or groups. In that case, the comfort value would be the distance between a user's desired privacy policy and the one she may finally accept, in a similar way to [35], in which the euclidean distance is used to compare the distance between two privacy policies to quantify the actual concession being made during an access control negotiation

Further empirical, simulated studies may look at larger network graphs and regimes of influence. That is, we have shown that stubborn users have disproportionate influence in their local neighborhood, but their global influence is dependent on their place in the network topology. We envision that these considerations may support a full taxonomy of users categorized in multiple dimensions including centrality, stubbornness and inherent privacy preferences. Ultimately, categorizing users in this way and developing a common language with which to discuss different user privacy behaviors will be very useful to further understand the interplay between local one-shot decisions and overall sharing dynamics at the social network in multi-party access control.

Finally, as more detailed data becomes available on instances of multi-party access control negotiations in the wild, especially longitudinal data about repeated negotiations over time, either through collected data from popular networking sites or through smaller and more targeted user studies, this data may be used to verify and parameterize the proposed model. We believe that the ultimatum game framework is a reasonable starting point, given its fundamental role in modeling social cooperation broadly and existing evidence on one-shot multi-party access control decisions. The update rules we have chosen are motivated by the psychology literature on in-group/out-group behaviors, peer pressure, and one-shot multi-party access control decisions. However, these rules and parameters thereof should be further researched in the specific context of repeated decisions on multi-party access control settings.

Acknowledgements

Portions of Dr. Squicciarini's and Mr. Semonsen's work were funded by Army Research Office under the auspices of Grant W911NF-13-1-0271. Portions of Dr. Squicciarini's work were additionally supported by a National Science Foundation grant 1453080. Dr. Such was supported by the EPSRC under grant EP/M027805/2.

References

1. A. Acquisti. Privacy in Electronic Commerce and the economics of immediate gratification. In ACM Press, editor, *Proceedings of the 5th ACM Electronic Commerce Conference*, pages 21–29, 2004.
2. Dustin L Arendt and Leslie M Blaha. Opinions, influence, and zealotry: a computational study on stubbornness. *Computational and Mathematical Organization Theory*, 21(2):184–209, 2015.

3. Andrew Besmer and Heather Richter Lipford. Moving beyond untagging: photo privacy in a tagged world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1563–1572. ACM, 2010.
4. Bhume Bhumiratana and Matt Bishop. Privacy aware data sharing: balancing the usability and privacy of datasets. In *Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments*, page 73. ACM, 2009.
5. Brandwatch.com. Facebook statistics from Brandwatch, 2016. https://www.brandwatch.com/blog/47-facebook-statistics-2016/.
6. Giangiacomo Bravo, Marco Castellani, Flaminio Squazzoni, and Riccardo Boero. Reputation and judgment effects in repeated trust games. 2008.
7. AJ Brush, John Krumm, and James Scott. Exploring end user preferences for location obfuscation, location-based services, and the value of location. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 95–104. ACM, 2010.
8. Colin Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2003.
9. Jundong Chen, Matthias R. Brust, Ankunda R. Kiremire, and Vir V. Phoha. Modeling privacy settings of an online social network from a game-theoretical perspective. In *Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 2013 9th International Conference Conference on*, pages 213–220. IEEE, October 2013.
10. Dan Cosley, Daniel P Huttenlocher, Jon M Kleinberg, Xiangyang Lan, and Siddharth Suri. Sequential influence models in social networks. *ICWSM*, 10:26, 2010.
11. David Crandall, Dan Cosley, Daniel Huttenlocher, Jon Kleinberg, and Siddharth Suri. Feedback effects between similarity and social influence in online communities. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 160–168. ACM, 2008.
12. Ricard L Fogues, Pradeep K Murukannaiah, Jose M Such, and Munindar P Singh. Sharing policies in multiuser privacy scenarios: Incorporating context, preferences, and arguments in decision making. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(1):5, 2017.
13. Noah E Friedkin. *A structural theory of social influence*, volume 13. Cambridge University Press, 2006.
14. Joseph Patrick Henrich. *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford University Press on Demand, 2004.
15. Hongxin Hu and Gail-Joon Ahn. Multiparty authorization framework for data sharing in online social networks. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 29–43. Springer, 2011.
16. Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. Detecting and resolving privacy conflicts for collaborative data sharing in online social networks. In *Proceedings of the 27th Annual Computer Security Applications Conference*, pages 103–112. ACM, 2011.
17. Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. Multiparty access control for online social networks: Model and mechanisms. *IEEE Transactions on Knowledge and Data Engineering*, 25(7):1614–1627, July 2013.
18. Hongxin Hu, Gail-Joon Ahn, Ziming Zhao, and Dejun Yang. Game theoretic analysis of multiparty access control in online social networks. In *Proceedings of the 19th ACM Symposium on Access Control Models and Technologies (SACMAT)*, pages 93–102, London, Ontario, 2014. ACM.

19. Leslie K John, Alessandro Acquisti, and George Loewenstein. Strangers on a plane: context-dependent willingness to divulge sensitive information. *Journal of Consumer Research*, 37(5):858–873, 2011.
20. Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. We're in it together: interpersonal management of disclosure in social network services. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3217–3226. ACM, 2011.
21. Kun Liu and Evimaria Terzi. A framework for computing the privacy scores of users in online social networks. *ACM Trans. Knowl. Discov. Data*, 5(1):6:1–6:30, December 2010.
22. Alice E Marwick and Danah Boyd. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society*, 13(1):114–133, 2011.
23. Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.
24. Karen M Page and Martin A Nowak. Empathy leads to fairness. *Bulletin of mathematical biology*, 64(6):1101–1116, 2002.
25. Sarah Rajtmajer, Anna Squicciarini, Christopher Griffin, Sushama Karumanchi, and Alpana Tyagi. Constrained social-energy minimization for multi-party sharing in online social networks. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 680–688. International Foundation for Autonomous Agents and Multiagent Systems, 2016.
26. David G Rand, Corina E Tarnita, Hisashi Ohtsuki, and Martin A Nowak. Evolution of fairness in the one-shot anonymous ultimatum game. *Proceedings of the National Academy of Sciences*, 110(7):2581–2586, 2013.
27. B. Ryan and N. Gross. The diffusion of hybrid seed corn in two iowa communities. *Rural Sociology*, 8(1):15–24, 1943.
28. Alan G Sanfey, James K Rilling, Jessica A Aronson, Leigh E Nystrom, and Jonathan D Cohen. The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626):1755–1758, 2003.
29. Roman Schlegel, Apu Kapadia, and Adam J Lee. Eyeing your exposure: quantifying and controlling information sharing for improved privacy. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 14. ACM, 2011.
30. Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. The post that wasn't: exploring self-censorship on facebook. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 793–802. ACM, 2013.
31. Daniel J Solove. *Nothing to hide: The false tradeoff between privacy and security.* Yale University Press, 2011.
32. Frederic Stutzman and Woodrow Hartzog. Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 769–778. ACM, 2012.
33. Jose M Such and Natalia Criado. Resolving multi-party privacy conflicts in social media. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1851–1863, 2016.
34. Jose M Such, Joel Porter, Sören Preibusch, and Adam Joinson. Photo privacy conflicts in social media: A large-scale empirical study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3821–3832. ACM, 2017.
35. Jose M. Such and Michael Rovatsos. Privacy policy negotiation in social media. *ACM Transactions on Autonomous and Adaptive Systems*, 11(4):1–29, 2016.

36. R. van Baaren, L. Janssen, T. L. Chartrand, and A. Dijksterhuis. Where is the love? the social aspects of mimicry. *Philosophical Trans. of the Royal Society B*, 364(1528):2381–2389, 2009.
37. Björn Wallace, David Cesarini, Paul Lichtenstein, and Magnus Johannesson. Heritability of ultimatum game responder behavior. *Proceedings of the National Academy of Sciences*, 104(40):15631–15634, 2007.
38. Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 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*, page 10. ACM, 2011.
39. Pamela Wisniewski, Heather Lipford, and David Wilson. Fighting for my space: Coping mechanisms for sns boundary regulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 609–618. ACM, 2012.
40. Ercan Yildiz, Daron Acemoglu, Asuman E Ozdaglar, Amin Saberi, and Anna Scaglione. Discrete opinion dynamics with stubborn agents. 2011.
41. W.W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33:452–473, 1977.
42. Qianru Zhu, Sarah Rajtmajer, and Andrew Belmonte. The emergence of fairness in an agent-based ultimatum game. *In preparation*, 2016.