

A cooperative game-theoretic approach to quantify the value of personal data in networks*

Michela Chessa,[†] and Patrick Loiseau[‡]

April 19, 2016

Abstract

The explosion of online services created a surge of interest on how to quantify the value of personal data. Recent studies have focused on its private cost, i.e., the amount at which users are willing to give it away. An appropriate quantification should, however, also take into account the benefit (or profit) users can get by releasing it. In this paper, we argue that this makes the question much more complex because profit is not extracted directly from personal data but from the information derived from it, which has a very different economic nature: it is a public good and it has strong local externalities driven by the social network in which users are naturally embedded.

We propose a cooperative game-theoretic approach to quantify the value of personal data in networks. We model the system as a local public good game and we propose a natural extension to a cooperative game. We apply classical allocation solutions (the core and the Shapley value) to quantify the value of each user's personal data. We prove that the game and the proposed allocations have good mathematical properties. Finally, we analyze the impact of the social network structure.

Keywords: Personal Data, Social Network, Cooperative Game Theory, Core, Shapley Value

1 Introduction

1.1 Motivations

The Internet has become an essential part of the citizens' life and of the economy in many countries. In this online ecosystem, service providers collect large

*This work was partially done while Michela Chessa was at EURECOM. This work was partially funded by the French Government (National Research Agency, ANR) through the "Investments for the Future" Program reference # ANR-11-LABX-0031-01; and by the Institut Mines-Télécom through the "Futur & Rupture" program.

[†]I.S.E.M., University of Nice Sophia-Antipolis / GREDED, CNRS; email: michela.chessa@unice.fr

[‡]EURECOM, email: patrick.loiseau@eurecom.fr

amounts of personal data about individuals (e.g., their shopping behaviors collected through tracking or login mechanisms) and use it to offer services (e.g., product recommendations or targeted advertising) from which they derive high profits. Personal data therefore has clear intrinsic economic value, which is extensively exploited by online services. At the same time, users also benefit from this ecosystem by being granted free access to services but it is unclear whether this appropriately compensates them for the release of their personal data. As the amount of data collected by online services is exploding, users and organizations increasingly ask for more fair, transparent, and personalized compensations. This naturally raises the question of *how much is personal data worth?*, which recently gained very high interest [38].

Quantifying the value of personal data is a very difficult task. Firstly, most research studies have focused on quantifying the private cost of releasing personal data, that is the amount at which users are willing to give it away [3]. An appropriate quantification, however, should also take into account the profit that can be extracted from the data by online service providers. Secondly, profit is not extracted directly from the personal data itself but from the information the provider can derive from it (e.g., users movie interests or sports activities), which has a very different economic nature. Indeed, while personal data is a private good that users can voluntarily decide to disclose at a given private cost, the information derived from it has a strong public component: information derived from personal data of a user may benefit other users (e.g., by improving the quality of the recommendation algorithm). Taking into account such externalities is key for an appropriate quantification. Lastly, as many online services collect data not only about individual users but also about the social interactions among them, the strongest externalities are local externalities: information can be derived about a user from the personal data released by her neighbors in the social graph. The position of the users in the social graph therefore plays a crucial role in the quantification of the value of their personal data.

1.2 Contribution

In this paper, we propose a game-theoretic approach to quantify the value of personal data that addresses the key difficulties mentioned above. Following the economic literature [4, 43], we model the public component of the provider’s information by treating it as a *public good*.¹ More specifically, to take into account the local externalities, we propose to model users interactions as a local public good game: users contribute by releasing personal data (at a private cost) and benefit from the local public good (the information derived about them that allows the provider to give them good service). The information about the users depends on the data released by themselves and by their neighbors in the social graph. Then, we propose to use solutions of a cooperative extension

¹A public good is a good that is non-rivalrous (i.e., such that consumption of the good by a user does not reduce availability to others) and non-excludable (i.e., it is not possible to exclude users from benefiting from the good).

of this game, which measure the contribution of a user to the value extracted from the information derived from personal data (minus private cost), to quantify the value of the personal data released by a user. A cooperative solution can provide fair quantification of the value of personal data taking into account both the private cost of users and the profit extracted by the provider. Such a quantification can be used for monetization, i.e., to price the data. It could also be used by an online service to determine the users with most valuable information, e.g., to offer discounts, or to target them as participants of an online survey, when involving all the players can be costly. Cooperation amongst users could be implemented by allowing them to negotiate effectively, or through an intermediary between the provider and the users; the intermediary would implement our solutions to compute the monetary transfers on each side. Since, as is typical with public goods, the strategic equilibrium of our game (i.e., when users are rewarded solely by being granted access to the service) is inefficient, the cooperative ecosystem would be likely sustained by all actors because cooperation yields a strictly higher welfare which can be shared in the interest of all parties.

Motivated by the application to quantifying the value of personal data as explained above, our main technical contribution is the proposition of a cooperative extension of a classic local public good game [8, 9] and its analysis. More specifically:

1. We propose a sensible extension of the strategic local public good game introduced by Bramoullé and Kranton [8, 9]. Our extension follows in spirit other extensions previously adopted in studies of economies with public goods (see Section 1.3), but extends those to include the network. We prove that the resulting cooperative game is equivalent to the minimax extension proposed by von Neumann and Morgenstern [48].
2. We analyze the properties of the cooperative game. In particular, we show that the game is monotonic and superadditive.
3. We propose to use the core and the Shapley value to quantify the value of personal data. We show that the core is non-empty and that it is therefore possible to find a stable allocation of welfare, which is an allocation such that no coalition will have incentives to stop cooperating with the other players. The Shapley value, on the other hand, is not guaranteed to be stable but has other desirable properties such as symmetry/fairness. We show examples in which the Shapley value belongs to the core.
4. Finally, we analyze how the allocations depend on the graph. We use *network games*, an extension of cooperative games proposed by Jackson and Wolinsky [28] which permit to model the dependence of the value function on the graph, to carry the analysis. In particular, we show that it is beneficial for the players to create new links and that the only stable network is given by the complete graph, in which each player is connected to every other player.

While games on graphs and local public good games more specifically are well-known and widely used in economics and computer science, our work is, to the best of our knowledge, the first study on cooperative extensions of such games in the field of the information economics and, more specifically, of the Internet economics. Besides the main motivation of quantification of the value of personal data, we expect that it will find applications in other areas that are well modeled by local public goods.

The remainder of the paper is organized as follows. We review related works in Section 1.3. In Section 2 we present the local public good game. Section 3 describes the cooperative extension and the solution concepts. In Section 4, we analyze some properties of the cooperative game and solutions. We study the graph influence in Section 5. We conclude in Section 6. Due to space constraints, proofs can be found in appendix.

1.3 Related Works

Economics of personal data. Academic concerns about the economic value of personal data date back to the 90’s with the informal discussions of Laudon [31] and Varian [47]. More recently, some experimental studies have been conducted that aim to quantify the value that users assign to their personal data in different scenarios, see e.g., Huberman et al. [25], Acquisti et al. [2] or Acquisti and Grossklags [1]. Pu and Grossklags also conducted a recent experimental study to measure how much users value their friend’s personal data [39].

The game-theoretic analysis of the private data monetization was pioneered by Kleinberg et al. [30], who proposed fair compensation mechanisms for personal data based on cooperative game theory. Their solutions are based on the core and on the Shapley value like ours, but their simple model does not take into account the public good nature of information which is the main focus of our paper. Recently, a significant thread of research started on selling personal data using auctions: Ghosh and Roth [22], Dandekar et al. [16], Roth and Schoenebeck [42] and Ligett and Roth [33]. In all those works, the loss of privacy by releasing data is quantified using differential privacy [18]. Riederer et al. [41] propose a mechanism called “transactional privacy” where users can sell access to their data through an unlimited supply auction.

All the aforementioned works only consider the private cost of revealing personal data, i.e., the cost incurred by an individual on account of his loss of privacy independently of the data revealed by others. In a recent work, Ioannidis and Loiseau [26] (see also extensions of the initial model by Chessa, Grossklags and Loiseau [14, 13]) propose a different model which takes into account the public good nature of information. They analyze the game as a non-cooperative game and provide results on Nash equilibrium and price of anarchy, but they do not investigate cooperative solutions to reach efficiency. Their model also does not take into account the local interactions that happen within a social network. An extension to a graph setting was proposed in [40], but also only in a non-cooperative setting.

Network games. In this paper, we consider users embedded in a social

network and use a graph to model the local public good nature of personal information. Games on graphs have been studied in different contexts and with different utility functions of the players, see e.g., the books of Jackson [27] and Easley and Kleinberg [19] and the recent survey by Jackson and Zenou [29]. In particular games on graphs have recently been studied in two situations: to model pricing problems with network externalities [10, 6, 20, 21, 15] and to model strategic propagation representing for instance product adoption [24, 32]. But the literature on the strategic interaction of the agents in networks is even larger [46, 5] and, among many others, we mention Bramoullé and Kranton [8], who described the model we adopt in this paper. These papers differ in the utility model and interactions that they consider. All of these works, however, consider a non-cooperative setting at the exception of Jackson and Wolinsky [28] who partly use a cooperative approach to model the strategic interaction of the agents on the link formation. Moreover, to the best of our knowledge, no prior work has used games on graphs to model questions related to personal information and quantification of its value.

Cooperative games for public goods. In this paper, we use a cooperative game-theory approach: we transform our local public good model in a cooperative game and use classical allocation solutions to quantify the value of personal information. The use of cooperative game theory methods to analyze public good problems is standard in the economics literature. We mention in particular the works of Champsaur [11], Moulin [35] and Chander [12]. To the best of our knowledge, however, no prior work has analyzed the impact of possible variations of the network in local public good problems using cooperative game-theory methods.

2 The Model

In this section, we describe our basic model, in which users are embedded in a social network and choose strategically the level of personal data to reveal.

2.1 The Social Network as a Graph

Let $N = \{1, \dots, n\}$ be a *set of players*, representing the users. In our model, the players are embedded in a social network (e.g., the Facebook network), that we represent by an undirected graph, where the nodes identify the players and the links their pairwise relations. Formally, an *undirected graph* g on N is a set of links, i.e., a set of unordered pairs of players $\{i, j\}$, that for simplicity we denote by ij . For any pair of players i and j , $ij \in g$ indicates that players i and j are linked in the graph g (in our example, the users are “Facebook friends”). Let g^N be the *complete graph*, i.e., the set of all the unordered pairs of players in N and $\mathcal{G}(N) = \{g \subseteq g^N\}$ be the set of all the graphs on N . Given $g \in \mathcal{G}(N)$, let $N_i(g) = \{j \in N \setminus \{i\} : ij \in g\}$ be the set of *neighbors* of i in g , i.e., the set of players which are linked to player i by the graph g , and $n_i(g) = |N_i(g)|$ its cardinality, called *degree* of player i . Player i 's *neighborhood* in g is defined as

herself and her set of neighbors, i.e., as $\bar{N}_i(g) = \{i\} \cup N_i(g)$.

2.2 The utility functions

Given a set of players N and an undirected graph g , we adopt a model from [9] for public goods in networks to describe their strategic interaction. Throughout the paper, at the exception of Section 5, the graph g is supposed to be a fixed parameter (in Section 5, we analyze the case where the graph is assumed to be a variable of the model). Each player in N chooses a *disclosure level* $d_i \in [0, +\infty)$ of her personal data, which corresponds to her contribution to the public good, the information. Depending on the situation at stake and its interpretation, this disclosure level can represent the precision in revealing a single piece of data or the total amount of data revealed.² Let $\mathbf{d} = (d_1, \dots, d_n)$ denote a disclosure profile of all players. In our model, data disclosed by a user benefits herself and her neighbors. This situation typically occurs with personalized applications: assuming that neighbors have similar tastes, data revealed by a user leads to better personalization for herself and for all her “friends”. We consider a model with positive externalities between neighbors, in which a neighbor’s disclosure is a substitute with one’s own proportionally to a factor $\delta \in [0, 1]$, but a player does not derive benefits from the disclosure of players with whom she does not share a link. Intuitively, δ is the measure of how much i ’s neighbors directly affect i . Player i ’s utility from profile \mathbf{d} in graph g is then given by

$$U_i(\mathbf{d}, g) = f \left(d_i + \delta \sum_{j \in N_i(g)} d_j \right) - kd_i, \quad (1)$$

where f is a twice differentiable function, with $f(0) = 0$, $f' > 0$ and $f'' < 0$ and $k \in \mathbb{R}_+$ is the cost to player i for one unit of disclosure level. The first component f captures the *benefit for player i* , i.e., the benefit that player i receives when the provider infers some information from her data and her neighbors’ (in terms, for example, of good ad hoc recommendation); in particular, it is an increasing function of the level of disclosure of player i and of the total level of disclosure of her neighbors. The second component is the *privacy cost*; it represents the cost that player i incurs (for instance due to loss of privacy) by choosing a given disclosure level and then it depends only on d_i .

Note that the public good model we consider includes only positive externalities, i.e., benefits obtained by a user when her neighbors disclose data. This assumption is well justified in our scenario. The utility defined in (1) has a positive component, the benefit, which is a function of the inferred information (the public good), and a negative component, the privacy cost, which is proportional to the amount of personal data a user decides to disclose (the private good). There could be negative externalities due to inference of information

² Note that we assume for simplicity that there is no bound on the disclosure level ($d_i \in [0, +\infty)$). This assumption is very mild since, if the disclosure level is bounded by \bar{d} sufficiently large so that the disclosed levels at equilibrium or at the optimum are interior, all our results are unchanged.

about player i from his neighbors data. However, such negative externalities are affecting only the public good component and can be captured by our model in the benefit component f , for instance through a smaller δ . Nevertheless, by assuming $\delta \geq 0$, we implicitly model situations in which the benefits from neighbors' disclosure (positive externalities) is higher than the negative externalities.

In our model, both the function f and the cost k are independent of the player i . When collecting real data on a social network, an online service may easily obtain information about the graph structure of its users, but not as easily about their privacy concerns. Many works have studied the problem of estimating how much individuals value their personal data (see Section 1.3), but implementing these mechanisms can be costly and time consuming for the online service. In this paper, we take an orthogonal approach and choose to propose solutions to quantify the value of personal data when the users differ only due to their position in the social network, but are otherwise totally symmetric.

2.3 The Local Public Good Game

We describe the strategic interaction of the players through the following local public good game

$$\Gamma = \langle N, [0, +\infty)^n, (U_i)_{i \in N} \rangle, \quad (2)$$

where N is the set of players, each player has strategy space given by $[0, +\infty)$ and her utility function is defined by (1). A *Nash equilibrium* (in pure strategy) of Γ is defined as a disclosure profile $\mathbf{d}^* \in [0, +\infty)^n$ s.t.

$$d_i^* \in \arg \max_{d_i \in [0, +\infty)} U_i(d_i, \mathbf{d}_{-i}^*), \quad \forall i \in N, \quad (3)$$

where we use the standard notation \mathbf{d}_{-i} to denote the collection of actions of every player but i .

A complete analysis of the local public good game Γ is provided in [9]. In particular, the authors show that Γ always has a Nash equilibrium, which may not be unique. If we denote by \bar{d} the disclosure level at which $f'(\bar{d}) = k$ (i.e., the marginal benefit to an individual equals its marginal cost), then any Nash equilibrium \mathbf{d}^* is s.t., for every agent i either $\delta \sum_{j \in N_i(g)} d_j^* \geq \bar{d}$ and $d_i^* = 0$ or $\delta \sum_{j \in N_i(g)} d_j^* < \bar{d}$ and $d_i^* = \bar{d} - \delta \sum_{j \in N_i(g)} d_j^*$.

We define the aggregate utility of the players in N , given a disclosure profile \mathbf{d} and the graph g , as the sum of their utilities:

$$W(\mathbf{d}, g) = \sum_{i \in N} U_i(\mathbf{d}, g). \quad (4)$$

The aggregate utility represents the *social welfare* of the population. We say that a profile \mathbf{d} is *efficient* for a given graph g if it maximizes the social welfare, i.e., if and only if there exists no other profile \mathbf{d}' such that $W(\mathbf{d}', g) > W(\mathbf{d}, g)$; and we denote by \mathbf{d}^e an efficient profile. A Nash equilibrium profile of Γ is in

general non efficient. This outcome is typical of public good models, which often yield non-efficient equilibria.

A powerful way to achieve efficient outcomes is to rely on cooperation. Agreeing to act cooperatively, the players may increase the aggregate utility to be shared, choosing a strategy profile which maximizes the social welfare. To provide a complete analysis of the cooperative setting, which will be at the basis of our quantification of the personal data of the users, in the next section we propose a cooperative extension of the local public good model in which the players can decide to make binding agreement and to coordinate to choose an efficient profile.

Before moving to the cooperative analysis of the game, we introduce a last notation. Given a disclosure profile \mathbf{d} and the graph g , we define the total service value

$$V(\mathbf{d}, g) = \sum_{i \in N} f \left(d_i + \delta \sum_{j \in N_i(g)} d_j \right), \quad (5)$$

which corresponds to the sum of the value of the service for all users, and can also be interpreted as the turnover of the provider (a fraction of which will become its net revenue).

3 The Cooperative Extension

Public good models often yield non-efficient equilibria. An efficient outcome can arise when regulatory mechanisms force the players to reach an efficient solution (e.g., by establishing the amount of taxes citizens have to pay by law, instead of letting them to freely contribute). Alternatively, in some different real-world situations, the purpose of achieving efficiency may lead the users to spontaneously create binding agreement between them and to commit to play an efficient strategy. In this paper, we suppose that the players are moved by this latter goal. We assume that they may decide to coordinate and strategically centralize their choices about the quantity of personal data to disclose to the provider, but that, at the same time, they maintain full control, as they can choose whether or not to cooperate with the rest of the group.

3.1 The Cooperative Extension of the Local Public Good Game

We extend the model proposed in Section 2 assuming that the players can choose to centralize their choices. Situations in which players can cooperate by establishing binding agreements among themselves (possibly through an intermediary) and forming coalitions in order to coordinate their strategies and share their joint payoff, are studied by cooperative game theory. We refer to [36] for

a complete account on the subject. Throughout the paper, in particular, we model our situation as a TU (transferable utility) cooperative game.³

Given a *coalition* of players $S \subseteq N$ who decided to sign a binding agreement between them, and given the graph g , we consider the *subgraph* of g reduced to S , $g|_S = \{ij | ij \in g, i, j \in S\} \in \mathcal{G}(S)$. The idea is that the players in S , while making an agreement to cooperate, accept to reveal their connections, while the group is not informed about the rest of the graph. In particular, information about the links between two players in $N \setminus S$, or between one player in S and one player in $N \setminus S$ is not available. For each player in S , a disclosure level of personal data is selected. We denote by \mathbf{d}_S a disclosure profile for the players in S . Given a disclosure profile \mathbf{d}_S , the utility of a player in S is still defined as in (1), where the set of neighbors is now restricted to the set of neighbors who are in S . Formally, when restricting to the coalition S , for each player $i \in S$ her utility is defined as

$$U_i(\mathbf{d}_S, g|_S) = f \left(d_i + \delta \sum_{j \in N_i(g|_S)} d_j \right) - kd_i. \quad (6)$$

We define the aggregate utility of the coalition S given a disclosure profile \mathbf{d}_S as the sum of the utilities of the players in S on $g|_S$, i.e., as

$$W(\mathbf{d}_S, g|_S) = \sum_{i \in S} U_i(\mathbf{d}_S, g|_S). \quad (7)$$

We denote an efficient profile by \mathbf{d}_S^e . Note again that the aggregate utility of coalition S depends only on the pairwise relations between its players, and not on the other links outside S .

A cooperative game is defined by a couple (N, v) , where N is the set of players and $v : 2^N \rightarrow \mathbb{R}$, with $v(\emptyset) = 0$, is the *characteristic function*, which assigns to every coalition S a *worth* $v(S)$. In our model, we define the worth of a coalition as the maximal aggregate utility the coalition can reach, i.e., for each $S \subseteq N$, $S \neq \emptyset$,

$$v(S) = W(\mathbf{d}_S^e, g|_S) = \max_{\mathbf{d}_S \in [0, +\infty)^s} W(\mathbf{d}_S, g|_S). \quad (8)$$

Intuitively, the worth of a coalition is given by the maximum social welfare that its players can reach by cooperating. Note that it is possible that there exist multiple efficient disclosure profiles. However, as they all provide the same aggregate utility, function v is univocally defined.

Note that our definition (8) of the characteristic function assumes that the worth of a coalition does not depend on the links which are external to the coalition (and on the data revealed by players outside the coalition, even if they have links with players inside). This is a standard assumption when extending a

³In such a game, we assume that players have the option to freely transfer among themselves a commodity, usually money, to increase or decrease their payoffs.

non-cooperative game with externalities to a cooperative game and is the most basic way to extend the local public good game, but it is also well justified in our model. Indeed, it is reasonable to assume either that links from players outside the coalition would not be revealed to the provider or, even if they were known, that it would be illegal for an online service provider to make use of links regarding a user who decided to refuse participating in the coalition. However, the definition of the cooperative game takes into account all the possible coalition structures, and the solutions we will propose in the following section are a weighted average of all the contributions a user can bring to the different groups of players. As a consequence, these solutions implicitly take into account all the links which are present in our original network structure and strongly depend on the local externalities.

If a coalition $S \subseteq N$ of players decide to make an agreement for the disclosure of their personal data, an efficient disclosure profile \mathbf{d}_S^e is selected. The aggregate utility of the players in S will be given by $W(\mathbf{d}_S^e, g|_S)$, while the total service value for players in S will be given by $V(\mathbf{d}_S^e, g|_S)$.

3.2 The Allocation of Utilities

Given a set of players, we quantify the value of their personal data by a sharing of the worth of the coalition between these players, as a reward for their contribution.

Suppose for simplicity that all the players in N decide to cooperate, and that the worth $v(N) = W(\mathbf{d}_N^e, g)$ has to be shared between them (everything still holds for a coalition $S \subseteq N$). A utility share is given by an *efficient payoff allocation*, i.e., by a vector $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ s.t., $\sum_{i \in N} x_i = v(N)$ (*efficiency* property). Each component x_i is interpreted as the utility payoff to player i . A *solution* for the game (N, v) is a function which associates to every characteristic function v a (possibly empty) set of payoff allocations. In the following, we propose to use the two most common concepts of solution for cooperative game.

The first solution we propose to use is the *core* [23], which provides the following set of efficient payoff allocations

$$C(v) = \left\{ \mathbf{x} \mid \sum_{i \in N} x_i = v(N), \sum_{i \in S} x_i \geq v(S), \forall S \subset N \right\}.$$

The fundamental idea of the core is that an agreement among the players in N can only be binding if every coalition $S \subset N$ receives collectively at least its worth, i.e., the value that the players in S could obtain without collaborating with the players in $N \setminus S$. In this way, single players and coalitions have no incentive to deviate. The core is a very appealing solution due to this stability property, but it may be empty.

The second solution we propose to use is the *Shapley value* [44]. Differently from the core, this solution proposes a unique efficient payoff allocation and it

is always defined. The Shapley value, that we denote by ϕ , is defined by the following equation: given the characteristic function v , for every $i \in N$

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{s!(n-s-1)!}{n!} (v(S \cup \{i\}) - v(S)).$$

It is a weighted sum of the contributions that player i brings while joining all possible coalitions $S \subseteq N \setminus \{i\}$.

The Shapley value can also be defined axiomatically as the unique solution satisfying efficiency, and the following three properties. (a) *Symmetry*: if two players i and j contribute to any coalition in exactly the same way, they must be granted the same payoff. (b) *Null player*: if a player contributes nothing to every coalition, he should receive zero. (c) *Additivity*: the Shapley value of the sum of two games must be equal to the sum of the Shapley values of the two games.

In the remaining of this paper, we present the properties of our model, by analyzing the cooperative game and the two solutions we proposed to share the aggregated utility between the players and to quantify the value of their personal data.

4 Properties of the Cooperative Extension

In this section, we first describe the coalition formation, in which we show that it is convenient for all the players to cooperate together. Then, we show that it is possible to implement some stable solutions for sharing the aggregate utility. Finally, we provide some examples, to show that the Shapley value may also be stable.

4.1 The Properties of the Cooperative Game

Deriving a characteristic function from an n -person strategic-form game is a standard issue in game theory [37]. When analyzing cooperative games associated to economies with public goods and externalities, defining a worth for the coalition S requires to also take into account the strategic choices made by players who are not members of S . In particular, the characteristic function should specify explicitly the actions of both the players who are and who are not members of the coalition. A standard way to get around this problem is to assume that the players outside the coalition choose the strategies which are the least favorable to the coalition, as in the well-known way of deriving a characteristic function from an n -person strategic-game form proposed by Von Neumann and Morgenstern [48]. Following their model, the characteristic function starting from our game Γ is defined for all $S \subseteq N$ as

$$v'(S) = \min_{\sigma_{N \setminus S} \in \Delta(D_{N \setminus S})} \max_{\sigma_S \in \Delta(D_S)} \sum_{i \in S} \tilde{U}_i(\sigma_S, \sigma_{N \setminus S}, g),$$

where $\Delta(D_S)$ (resp. $\Delta(D_{N \setminus S})$) is the set of *correlated strategies* available to coalition S (resp. $N \setminus S$). Here, $\tilde{U}_i(\sigma_S, \sigma_{N \setminus S}, g)$ denotes player i 's expected payoff when the correlated strategies σ_S and $\sigma_{N \setminus S}$ are implemented:

$$\tilde{U}_i(\sigma_S, \sigma_{N \setminus S}, g) = \sum_{\mathbf{d}_S} \sum_{\mathbf{d}_{N \setminus S}} \sigma_S(\mathbf{d}_S) \sigma_{N \setminus S}(\mathbf{d}_{N \setminus S}) U_i(\mathbf{d}_N, g).$$

The idea behind the definition of von Neumann and Morgenstern is to assert that the value of coalition S , $v'(S)$, is the maximum sum of utilities that the players of coalition S can guarantee themselves against the best offensive threat by the complementary coalition $N \setminus S$. The game (N, v') is called the *minimax representation* in cooperative form of the strategic-form game Γ with transferable utility.

The minimax representation is a very convenient way to derive a characteristic function from Γ thanks to its properties. In particular, if we plan to apply the core as solution concept, the inequality $\sum_{i \in S} x_i < v'(S)$ is necessary and sufficient for the players in S to be able to guarantee themselves payoffs that are strictly better than in x , no matter what the other players in $N \setminus S$ do. However, the definition of the minimax representation is not trivial, both from an analytic and a computational point of view, as it is defined over all the correlated strategies. Also, in our setting, the minimax representation does not have a clear interpretation as the characteristic function we defined in (8) has. Yet, our first result shows that the definition of our characteristic function v in (8) and of the minimax representation v' coincide for our game.

Proposition 1 *The cooperative game (N, v) is equivalent to the minimax representation of the game Γ , i.e.,*

$$v(S) = v'(S), \quad \forall S \subseteq N.$$

In particular, this result implies that our game (N, v) enjoys all the nice properties of (N, v') , but through a simpler definition of the characteristic function, both from the analytical and from the computational point of view, with a clearer interpretation.

We now define two important properties of cooperative games. A cooperative game (N, v) is said to be *monotonic* if

$$S \subseteq T \subseteq N \Rightarrow v(S) \leq v(T),$$

meaning that a larger coalition always has a larger worth. A cooperative game (N, v) is said to be *superadditive* if

$$S \cap T = \emptyset \Rightarrow v(S) + v(T) \leq v(S \cup T).$$

Superadditivity implies that the worth of the coalition N of all players is at least as large as the sum of the worths of the members of any partition of N . The following corollary states that, as a consequence of Proposition 1, our game satisfies these two properties.

Corollary 1 *The game (N, v) is monotonic and superadditive, i.e., every player has incentives to cooperate with the others and the grand coalition will form.*

The monotonicity ensures that, under cooperation, the aggregate utility increases when a new player decides to make an agreement with the others. Most importantly, superadditivity ensures that it is optimal to form the grand coalition. In particular, whenever the worth is shared in an advantageous way for everyone, it is not convenient for a player to abandon the coalition and it is not convenient to decentralize the cooperation between two different groups of players. Because of these important properties, in the following we can assume that the grand coalition forms and that the total utility to be shared is equal to $v(N)$.

Now, we observe that also the total service value increases, when the players decided to cooperate to adopt an efficient disclosure level.

Theorem 1 *For any efficient disclosure profile \mathbf{d}^e and any Nash equilibrium profile \mathbf{d}^* , we have $V(\mathbf{d}^e, g) \geq V(\mathbf{d}^*, g)$, i.e., the total service value is higher under cooperation. In particular, $V(\mathbf{d}^e, g) > V(\mathbf{d}^*, g)$ if g is not totally disconnected.*

Corollary 1 and Theorem 1 ensure that, as the cooperation can improve the utility of both the players (the users) and the provider, it is likely to occur and to be sustained by all the actors. In the following, we assume that the players act cooperatively and we analyze the properties of the payoff allocations we have proposed in the previous section.

4.2 The Properties of the Allocations

The first solution we proposed, the core, is very appealing, in view of the assumption that a payoff allocation in the core ensures that any coalition can negotiate effectively and that such a solution is stable against deviation by selfish coalitions of players. Unfortunately, for some games, the core may be empty, and if this happens, no matter what allocation may occur, there is always a coalition that would gain by deviating. The following result ensures that this is not the case for our game (N, v) , and that there always exists at least a stable payoff allocation.

Theorem 2 *The game (N, v) has a nonempty core, i.e., there exists an efficient and stable allocation such that no subset of players has incentives to deviate.*

Theorem 2 shows that the core of our game (N, v) is non-empty, but it does not provide a concrete construction to find a vector in the core. Regardless of its importance, when the core is large, this solution can be difficult to apply as a predictive theory, due to the high number of stable allocations. To provide a specific vector in the core, it is possible to compute the *nucleolus*, a powerful pointwise solution that has been shown to be stable whenever the core is nonempty [17, 36]. The nucleolus is defined in terms of objections and

counterobjections, or, equivalently, as the unique vector minimizing the dissatisfaction of the members of any possible coalition, in a lexicographic order. Unfortunately, computing the nucleolus is, in general, costly.

The second solution we proposed, the Shapley value, is a very powerful tool because of its fairness properties. However, it may be unstable. Even if also the Shapley value is in general really costly, when it belongs to the core, it ensures both a fair and stable outcome for the cooperating players, and it becomes a very appealing solution concept, in particular for our purpose of quantifying the value of personal data of the users. A sufficient condition for the Shapley value to belong to the core is that our game satisfies the convexity property.⁴ Unfortunately, the convexity of our game remains an open question. Below, we present two examples in which we can show that the Shapley value belongs to the core. Whether this holds in general remains a conjecture whose investigation is left as future work.

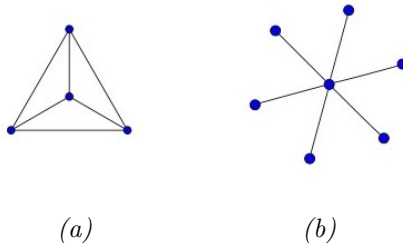


Figure 1: Symmetric and star graphs.

Example 1 Let g be a symmetric graph. Then each player $i \in N$ has the same Shapley value equal to

$$\phi_i(v) = \frac{1}{n}v(N).$$

This Shapley value allocation belongs to the core.

In particular, we observe that circles and complete graphs belong to the class of symmetric graphs.

Example 2 Let g be a star graph. For simplicity, we denote by S_h a star of cardinality h . Then the center player has the highest Shapley value, given by

$$\phi_C(v) = \frac{1}{n}v(S_n) + \frac{1}{n} \sum_{k=2}^{n-2} [v(S_k) - kv(S_1)],$$

while each player at the periphery has a Shapley value given by

$$\phi_P(v) = \frac{1}{n}v(S_n) - \frac{1}{n(n-1)} \sum_{k=2}^{n-2} [v(S_k) - kv(S_1)].$$

This Shapley value allocation belongs to the core.

⁴A cooperative game (N, v) is *convex* if $\forall S, T \subseteq N, v(S) + v(T) \leq v(S \cap T) + v(S \cup T)$.

5 The Impact of the Network

So far, we have assumed that the network structure was a fixed parameter. In this section, we analyze the impact of this network, by considering the graph as a variable of the model. In particular, we want to understand how the two solutions we proposed in the previous section change when we modify the graph and if they can induce the players to create or to remove some links strategically in order to get a higher payoff. To this end, we use the formalism introduced by Jackson and Wolinsky [28], who propose to define the notions of *network games* and *allocation rules for network games* as natural extensions of cooperative games and of allocation rules for cooperative games, respectively. These notions allow keeping track of the overall value generated by a particular network as well as how it is allocated across players, and of the way they vary depending on the graph. We first introduce some additional notations about graphs complementing the ones in Section 2.1, which are necessary for the analysis of this section.

5.1 Background on Graphs

In this section, we assume that a graph $g \in \mathcal{G}(N)$ may contain loops, where a *loop* is a link ii which connects a player with herself.⁵ Given the set of players N and a graph $g \in \mathcal{G}(N)$, we denote by $g + ij = g \cup \{ij\}$ the graph obtained by adding the link ij to the existing graph g and by $g - ij = g \setminus \{ij\}$ the graph obtained by removing the link. Let $N(g) = \{i | \exists j \text{ s.t. } ij \in g\}$ be the set of players who have at least one link in the graph g . A *path* in g connecting i_1 and i_t is a set of distinct nodes $\{i_1, \dots, i_t\} \subseteq N(g)$ such that $\{i_1i_2, \dots, i_{t-1}i_t\} \subseteq g$. The subgraph $g' \subseteq g$ is a *component* of g if either (a) $N(g') = \{i\}$ and for all $j \in N(g)$, $ij \in g$ implies that $i = j$ (i.e., g' contains a single node which is connected only to itself in g); or (b) for all $i \in N(g')$ and $j \in N(g')$, $i \neq j$, there exists a path in g' connecting i and j , and for any $i \in N(g')$ and $j \in N(g)$, $ij \in g$ implies that $ij \in g'$ (i.e., any two nodes in $N(g')$ are connected by a path and there is no path from a node in $N(g')$ to any other node in $N(g) \setminus N(g')$). The set of components of g is denoted by $C(g)$.

5.2 The Network Game

The cooperative extension of the public good game Γ can be naturally embedded into the definition of network games. Given the cooperative game (N, v) with characteristic function defined by (8), we now underline the dependence on the graph structure g , denoting it by (N, v_g) . This game is equivalent to a *network game* (N, w) , where N is the set of players, and $w : \mathcal{G}(N) \rightarrow \mathbb{R}$, with $w(\emptyset) = 0$, is the *value function*, which assigns to every graph g the *worth* $w(g)$. Formally, our cooperative game is equivalent to the network game defined for each $g \in \mathcal{G}(N)$

⁵Adding the presence of loops does not influence the model and the analysis of the previous sections.

by

$$w(g) = v_g(N(g)). \tag{9}$$

Given the corresponding network game, it is possible to go back to the original cooperative game in (8), as for each $g \in \mathcal{G}(N)$ and for each $S \subseteq N$, $v_g(S) = w(g|_S)$.

A network game is an extension of a cooperative game, in which the particular structure of the network, and not only the subsets of the players, matters. These two representations, focusing on different aspects, the graph and the coalitions, permit to analyze different properties of the model. In particular, the network game representation allows us to analyze how the social network structure influences the efficient disclosure profile selected under cooperation and the value of the personal data of the users depending on their position.

Differently from the original model in [28], we allow the graph to contain loops. This is necessary to represent our cooperative game as a network game because we need to be able to distinguish between two different cases. In the first case, when an isolated player does not belong to the coalition S , she does not contribute to the aggregate utility; in the network game, we model the situation by assuming that she does not have links. In the second case, the isolated player belongs to the coalition, but she does not share links with any other player and then she maximizes her utility independently from the others; then, we assume that this player is connected only to herself by a loop. In the original model [28], it was not necessary to distinguish between these two cases, as the utility of an isolated player was assumed to be zero.

5.3 The Properties of the Network Game

We now define some important properties for network games and we analyze if they are satisfied by our model. Given the network game (N, w) , we say that the value function is *component additive* if $\sum_{h \in C(g)} w(h) = w(g)$. This condition rules out externalities across components, but still allows them within components. Given a permutation of players π (a bijection from N to N), and any $g \in \mathcal{G}(N)$, let $g^\pi = \{\pi(i)\pi(j) | ij \in g\}$. Thus, g^π is a network that shares the same architecture as g but with the players relabeled according to π . A value function is *anonymous* if for any permutation of the set of players π , $w(g^\pi) = w(g)$. Anonymity says that the value of a network is derived from the structure of the network and not the labels of the players who occupy various positions. The following proposition states that our game, represented as a network game with value function (9), satisfies both properties.

Proposition 2 *The value function w defined in (9) is component additive and anonymous.*

The property of component additivity is a consequence of the fact that our model allows externalities only between connected players, in particular only between neighbors. The property of anonymity is a consequence of the fact that

f and k are independent from i . It guarantees that our model really focuses on the network structure and that it captures the utility that a set of players can generate because of their pairwise relations, independently from their labels.

A graph $g \in \mathcal{G}(N)$ is said to be *strongly efficient* if $w(g) \geq w(g')$ for all $g' \in \mathcal{G}(N)$, i.e., if it guarantees maximal total value. The following theorem shows that the complete graph satisfies this important property for our model.

Theorem 3 *The complete graph g^N is strongly efficient.*

Theorem 3 states that the complete graph produces a maximal aggregate utility for the grand coalition N . We can generalize this result noticing that for each possible coalition S , the aggregate utility of S is maximal when all the links are present. In the following result, we show that there may exist a strongly efficient graph which is not complete, but that in this case, missing links are only between non contributing players.

Corollary 2 *Any strongly efficient graph $g \subset g^N$ is s.t. $N(g) = N$ and if $ij \notin g$, then there exists an efficient disclosure profile \mathbf{d}^e such that $d_i^e = d_j^e = 0$.*

To complete our analysis of the network game, we first notice that a player that discloses zero at an efficient profile cannot be isolated. Then, that in order to be strongly efficient, a graph has to be fully connected. Moreover we observe that, typically, players contributing zero at the efficient profiles are players such that their neighborhood is included in the neighborhood of another player, hence, intuitively, players whose links are not necessary for reaching any other player.

5.4 The Properties of the Allocations for Network Games

After analyzing the property of the game as a function of the graph, we now finally analyze the properties of the proposed solutions as functions of the graph. We start by extending them to network games. In particular, the definition of Shapley value for our game (N, v) corresponds to the definition of the Shapley value as allocation rule for network games in [28], while also any vector in the core of the cooperative game (N, v) can be interpreted as an allocation for the corresponding network game (N, w) . Formally, an *allocation rule* for network games is a function $Y : \mathcal{G}(N) \times \mathcal{W} \rightarrow \mathbb{R}^n$ that associates, to a couple of a graph and a value function, a payoff allocation (x_1, \dots, x_n) .

A notion of stability for the graph with respect to a given allocation is given by the following definition. A network g is *pairwise stable* with respect to allocation rule Y and value function w if

- (i) for all $ij \in g$, $Y_i(g, w) \geq Y_i(g - ij, w)$ and $Y_j(g, w) \geq Y_j(g - ij, w)$, and
- (ii) for all $ij \notin g$, if $Y_i(g + ij, w) > Y_i(g, w)$ then $Y_j(g + ij, w) < Y_j(g, w)$.

In this definition, we assume that the formation of a link requires the consent of both players involved, while severance can be done by one player unilaterally.

Then, a network is pairwise stable if (i) no player wishes to sever a link that she is involved in; and (ii) if one link is not in the graph and one of the involved players would benefit from adding it, then it must be that the other player would suffer from the addition of the link.

Our last two results provide stability results for the core and the Shapley value as allocation rules of the network game (N, w) .

Proposition 3 *Given a non-complete graph g , for each payoff allocation in the core $C(v_g)$ there exists a payoff allocation in the core $C(v_{g+ij})$ which assigns at least the same payoff to players i and j .*

In particular, Proposition 3 implies that there exists a vector in the core such that, if the graph is complete, no player has an incentive to break a link. Such a payoff vector is then a stable allocation. Finally, the last corollary provides an important stability result for the Shapley value as well.

Proposition 4 *The complete graph g^N is the only pairwise stable graph with respect to the Shapley value allocation. In particular, every player has incentives to create a new link to augment her own payoff.*

Proposition 4 shows that in our model, if the Shapley value is implemented as solution to share the aggregate utility between the players, every player gains by forming a new link. In this setting, the only social network which is stable is the one in which each player is connected to every other player. When the complete graph is the only strongly efficient graph, this result is intuitive because the players may share a higher aggregate utility when creating links. Interestingly though, it still holds when there exists a strongly efficient graph $g \subset g^N$ as the following example illustrates.

Example 3 *Suppose that g is a graph as in Figure 2-(a), with parameter $\delta = 1$ (i.e., in a situation of perfect substitutability). Such a graph is strongly efficient. However, while implementing the Shapley value to share their payoff, the two extreme agents have incentives to create a new link as in Figure 2-(b), at the expense of the central player.*

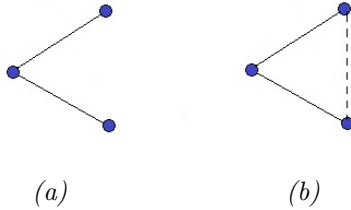


Figure 2: Strategic creation of links.

At first sight, this last result may suggest that, in a situation in which players from a social network are rewarded for their data, they have incentives to create artificial links, in order to get a higher final payoff. This, however, would be an

over interpretation. Indeed, our results suppose that all created links have the same positive externality and affirm that players do have incentives to have a higher number of effective real relations. On the contrary, to fully understand the creation of artificial links that could occur, we would need to introduce in the model a cost of link creation and potential negative externalities that would arise (a fake link could bring bad recommendation and consequently to a decreasing of the utility of the users). We leave this extension of the model as future work. More simply, in the context of personal data, our results should be interpreted as showing that when the graph is dense, this is beneficial for the users and it augments the value of their personal contribution.

6 Concluding Remarks

In this paper, we propose a cooperative game-theoretic approach to quantify the value of personal data in social networks with positive externalities. We use a local public good game to model the private cost of revealing data and the local public good component of the information derived from it. We propose a cooperative extension and analyze two solutions: the core and the Shapley value. In particular, we show that the core is non-empty and that the game is super-additive, which indicates that these allocations are efficient and that the grand coalition will form and is stable. To the best of our knowledge, our study is the first cooperative extension of the proposed local public good game. As local public good games have become a classic model and the cooperative approach is often sensible in public good situations, we hope that our results can find applications and extensions outside the area of personal data.

Our cooperative game extension assumes that players can form a coalition and maximize the welfare of their coalition when all links to players outside the coalition are removed and players outside the coalition do not take part in the game. This is a sensible assumption since it is reasonable that players who decide to join a platform within which information is shared will cooperate towards a higher welfare (as long as the benefits are appropriately shared). Other cooperative extensions should be studied to refine our model. Cooperative approaches are widely adopted in many different fields (from political science to inventory management) to predict the coalition structure and the way players will negotiate to increase their utility. We hope that our work will enhance cooperative game theory as a valuable tool in the Internet economics.

Our model proposes a new approach to the quantification of personal data which is focused on the network structure. Differently from existing works, this model allows analyzing how users can strategically manipulate the graph (and not only the information revealed). Our results in Section 5 show how the links between users may increase the value of personal data, due to the externalities. In particular, users belonging to dense networks have very high incentives to reveal their data; and users always have incentives to reveal existing links. These results do not suggest, however, that users have incentives to create artificial links in order to increase their own payoff. Indeed, our model is too simple

to discuss creation of fake links, which may not be associated with the same positive externality as true links and may instead have negative externality. An interesting extension of our work will be to include this effect and analyze which pairs of users (if any) have an incentive to create fake links. This will guide us towards a fully incentive compatible model where users reveal links truthfully.

In our study, we have made a number of simplifying assumptions. In particular, in order to focus on the effect of the local public good aspect of personal information derived from the graph on its value, we have assumed that the private cost is the same for all players. As all previous researches have focused only on the private cost of personal information, we believe that it was necessary to clearly separate the effect of the position of a user in the graph from its private cost and that our paper provides useful results in this direction. As future work, we plan, however, to extend our analysis to the case of heterogeneous private costs in order to quantify the value of personal data in more realistic scenarios that combine heterogeneous private costs and local public good components.

A typical problem suffered by cooperative game-theory solutions is their computational complexity. In this paper, we provide results about the core. In the case of a convex game, its extreme points can be easily computed with a greedy algorithm, but the computation becomes less efficient for a generic game. Also the Shapley value cannot be computed efficiently. There exist, however, various approximations of the Shapley value that can be computed efficiently (such as some variations of Monte Carlo sampling methods [34]). We plan to extend our work in two directions: *(i)* analyze how the use of approximations affect our results and *(ii)* study the limit regime where the number of players is very large and how this can provide more computationally efficient or more intuitive results.

References

- [1] Alessandro Acquisti and Jens Grossklags. An online survey experiment on ambiguity and privacy. *Communications & Strategies*, 88(4):19–39, 2012.
- [2] Alessandro Acquisti, Leslie K. John, and George Loewenstein. What is privacy worth? *The Journal of Legal Studies*, 42(2):249–274, 2013.
- [3] Alessandro Acquisti, Curtis Taylor, and Curtis Wagman. The economics of privacy. *Journal of Economic Literature*, 2015. forthcoming.
- [4] Kenneth J. Arrow. Economic welfare and the allocation of resources for invention. In Universities-National Bureau, editor, *The Rate and Direction of Economic Activity: Economic and Social Factors*, pages 609–626. Princeton University Press, 1962.
- [5] C. Ballester and A. Calvó-Armengol. Moderate interactions in games with induced complementarities. Mimeo, 2007.

- [6] Francis Bloch and Nicolas Quérou. Pricing in social networks. *Games and Economic Behavior*, 80(0):243–261, 2013.
- [7] Olga N. Bondareva. Some applications of linear programming methods to the theory of cooperative games (in russian). *Problemy Kybernetiki*, 10:119–139, 1963.
- [8] Yann Bramoullé and Rachel Kranton. Public goods in networks. *Journal of Economic Theory*, 135(1):478–494, 2007.
- [9] Yann Bramoullé, Rachel Kranton, and Martin D’Amours. Strategic interaction and networks. *The American Economic Review*, 104(3):898–930, 2014.
- [10] Ozan Candogan, Kostas Bimpikis, and Asuman Ozdaglar. Optimal pricing in networks with externalities. *Operations Research*, 60(4):883–905, 2012.
- [11] P. Champsaur. How to share the cost of a public good. *International Journal of Game Theory*, 4(3):113–129, 1975.
- [12] P. Chander. Dynamic procedures and incentives in public good economies. *Econometrica*, 61(6):1341–1354, 1993.
- [13] Michela Chessa, Jens Grossklags, and Patrick Loiseau. A game-theoretic study on non-monetary incentives in data analytics projects with privacy implications. In *Proceedings of CSF*, 2015.
- [14] Michela Chessa, Jens Grossklags, and Patrick Loiseau. A short paper on the incentives to share private information for population estimates. In *Proceedings of FC*, 2015.
- [15] Maxime C. Cohen and Pavithra Harsha. Designing price incentives in a network with social interactions. Preprint available at <http://ssrn.com/abstract=2376668>, 2013.
- [16] Pranav Dandekar, Nadia Fawaz, and Stratis Ioannidis. Privacy auctions for recommender systems. In *Proceedings of WINE*, pages 309–322, 2012.
- [17] Theo S. H. Driessen. *Cooperative Games, Solutions and Applications*. Kluwer Academic Publishers, 1988.
- [18] Cynthia Dwork. Differential privacy. In *In Proceedings of ICALP*, pages 1–12, 2006.
- [19] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.
- [20] Matthew Elliott and Benjamin Golub. A network approach to public goods. In *Proceedings of EC*, pages 377–378, 2013.

- [21] Michal Feldman, David Kempe, Brendan Lucier, and Renato Paes Leme. Pricing public goods for private sale. In *Proceedings of EC*, pages 417–434, 2013.
- [22] Arpita Ghosh and Aaron Roth. Selling privacy at auction. In *Proceedings of EC*, pages 199–208, 2011.
- [23] D. B. Gillies. Solutions to general non-zero-sum games. In A. W. Tucker and R. D. Luce, editors, *Contributions to the Theory of Games IV*, pages 47–85. Princeton University Press, 1959.
- [24] Sanjeev Goyal and Michael Kearns. Competitive contagion in networks. In *Proceedings of STOC*, pages 759–774, 2012.
- [25] Bernardo A. Huberman, Eytan Adar, and Leslie R. Fine. Valuating privacy. *IEEE Security & Privacy*, 3(5):22–25, 2005.
- [26] Stratis Ioannidis and Patrick Loiseau. Linear regression as a non-cooperative game. In *Proceedings of WINE*, 2013.
- [27] Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, 2010.
- [28] Matthew O. Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71:44–74, 1996.
- [29] Matthew O. Jackson and Yves Zenou. Games on networks. In H. Peyton Young and Shmuel Zamir, editors, *Handbook of Game Theory with Economic Applications*, volume 4, pages 95–163. Elsevier, 2015.
- [30] Jon Kleinberg, Christos H. Papadimitriou, and Prabhakar Raghavan. On the value of private information. In *Proceedings of TARK*, pages 249–257, 2001.
- [31] Kenneth C. Laudon. Markets and privacy. *Communications of the ACM*, 39(9):92–104, 1996.
- [32] Marc Lelarge. Diffusion and cascading behavior in random networks. *Games and Economic Behavior*, 75(2):752–775, 2012.
- [33] Katrina Ligett and Aaron Roth. Take it or leave it: Running a survey when privacy comes at a cost. In *Proceedings of WINE*, pages 378–391, 2012.
- [34] Irwin Mann and Lloyd S. Shapley. Value of large games, iv: Evaluating the electoral college by monte carlo techniques. In *RM-2651*. The Rand Corporation, Santa Monica, California, 1960.
- [35] H. Moulin. Egalitarian-equivalent cost sharing of a public good. *Econometrica*, 55:963–976, 1987.

- [36] Roger B. Myerson. *Game theory: analysis of conflict*. Harvard University Press, 1991.
- [37] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1997.
- [38] OECD. *Data-Driven Innovation: Big Data for Growth and Well-Being*. OECD Publishing, 2015.
- [39] Yu Pu and Jens Grossklags. Using conjoint analysis to investigate the value of interdependent privacy in social app adoption scenarios. In *Proceedings of ICIS*, 2015.
- [40] Arthi Ramachandran and Augustin Chaintreau. The network effect of privacy choices. In *Proceedings of NetEcon*, pages 59–62, 2015.
- [41] Christopher Riederer, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, and Pablo Rodriguez. For sale : your data: by : you. In *Proceedings of HotNets*, pages 13:1–13:6, 2011.
- [42] Aaron Roth and Grant Schoenebeck. Conducting truthful surveys, cheaply. In *Proceedings of EC*, pages 826–843, 2012.
- [43] Paul A. Samuelson. The pure theory of public expenditures. *Review of Economics and Statistics*, 40(4):387–389, 1954.
- [44] Lloyd S. Shapley. A value for n-person games. In A. W. Tucker and R. D. Luce, editors, *Contributions to the Theory of Games II*, pages 307–317. Princeton University Press, 1953.
- [45] Lloyd S. Shapley. On balanced sets and cores. *Naval Research Logistics Quarterly*, 14:453–460, 1967.
- [46] S. Sharma and D Teneketzis. Local public good provisioning in networks: A nash implementation mechanism. *IEEE Journal on Selected Areas in Communications (JSAC)*, 30(11):2105–2116, 2012.
- [47] Hal R. Varian. Economic aspects of personal privacy. In Larry Irving, editor, *Privacy and Self-Regulation in the Information Age*. National Telecommunications and Information Administration, 1997. http://www.ntia.doc.gov/reports/privacy/privacy_rpt.htm.
- [48] John von Neumann and Oscar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1947.

APPENDIX

A Proof of Proposition 4.1

To prove the first inequality, we can observe that

$$\begin{aligned}
v'(S) &= \min_{\sigma_{N \setminus S} \in \Delta(D_{N \setminus S})} \max_{\sigma_S \in \Delta(D_S)} \sum_{i \in S} \tilde{U}_i(\sigma_S, \sigma_{N \setminus S}, g) \\
&\leq \max_{\sigma_S \in \Delta(D_S)} \sum_{i \in S} \left[\sum_{\mathbf{d}_S} \sigma_S(\mathbf{d}_S) \left(f \left(d_i + \delta \sum_{j \in N_i(g|_S)} e_j \right) - kd_i \right) \right] \\
&= \max_{\sigma_S \in \Delta(D_S)} \sum_{\mathbf{d}_S} \sigma_S(\mathbf{d}_S) \left[\sum_{i \in S} \left(f \left(d_i + \delta \sum_{j \in N_i(g|_S)} d_j \right) - kd_i \right) \right] \\
&= \max_{\mathbf{d}_S \in D_S} \sum_{i \in S} \left[f \left(d_i + \delta \sum_{j \in N_i(g|_S)} d_j \right) - kd_i \right] = v(S),
\end{aligned}$$

where the inequality is because we are minimizing on a smaller set (in particular, we are fixing the space to one point), and then the minimum value is necessarily higher. Moreover, to prove the opposite inequality, we can observe that

$$\begin{aligned}
v'(S) &= \min_{\sigma_{N \setminus S} \in \Delta(D_{N \setminus S})} \max_{\sigma_S \in \Delta(D_S)} \sum_{i \in S} \tilde{U}_i(\sigma_S, \sigma_{N \setminus S}, g) \\
&\geq \max_{\sigma_S \in \Delta(D_S)} \min_{\sigma_{N \setminus S} \in \Delta(D_{N \setminus S})} \sum_{i \in S} \tilde{U}_i(\sigma_S, \sigma_{N \setminus S}, g) \\
&= \max_{\sigma_S \in \Delta(D_S)} \sum_{\mathbf{e}_S} \sigma_S(\mathbf{e}_S) \left[\sum_{i \in S} \left(f \left(d_i + \delta \sum_{j \in N_i(g|_S)} e_j \right) - kd_i \right) \right] \\
&= \max_{\mathbf{d}_S \in D_S} \sum_{i \in S} \left[f \left(d_i + \delta \sum_{j \in N_i(g|_S)} d_j \right) - kd_i \right] = v(S),
\end{aligned}$$

where the inequality follows from the standard inequality $\min \max \geq \max \min$, often referred to as the *max-min inequality*.

B Proof of Corollary 4.2

Superadditivity follows from Proposition 1, because the minimax representation of a game in strategic form is always superadditive. Moreover, superadditivity implies monotonicity.

C Proof of Theorem 4.3

For each $i \in N$, the equilibrium condition is given by

$$f'(d_i^* + \delta \sum_{j \in N_i(g|_S)} d_j^*) = k,$$

and the efficiency condition is instead given by

$$f'(d_i^e + \delta \sum_{j \in N_i(g|_S)} d_j^e) + \delta \sum_{j \in N_i(g|_S)} f'(d_j^e + \delta \sum_{h \in N_j(g|_S)} d_h^e) = k.$$

It follows that for each $i \in N$, $f'(d_i^e + \delta \sum_{j \in N_i(g|_S)} d_j^e) \leq f'(d_i^* + \delta \sum_{j \in N_i(g|_S)} d_j^*)$ and, as f is concave, that for each $i \in N$, $d_i^e + \delta \sum_{j \in N_i(g|_S)} d_j^e \geq d_i^* + \delta \sum_{j \in N_i(g|_S)} d_j^*$. These two inequalities are strict for at least one $h \in N$ if the graph is not totally disconnected. As f is increasing, then for each $i \in N$, $f(d_i^e + \delta \sum_{j \in N_i(g|_S)} d_j^e) \geq f(d_i^* + \delta \sum_{j \in N_i(g|_S)} d_j^*)$, and the inequality is strict for h . It follows that their sum is s.t. $V(\mathbf{d}^e, g) \geq V(\mathbf{d}^*, g)$ for each graph g and the inequality is strict if g is not totally disconnected.

D Proof of Theorem 4.4

To prove the nonemptiness of the core, we introduce the notion of balancedness. A family \mathcal{B} of coalitions is called a *balanced collection* provided that for each $S \in \mathcal{B}$, there exists $\lambda_S > 0$ such that, for all $i \in N$, $\sum_{S \in \mathcal{B}, i \in S} \lambda_S = 1$. A very important result, known as Bondareva-Shapley Theorem ([7] and [45]), states that the cooperative game (N, v) has a nonempty core if and only if for every balanced collection \mathcal{B} ,

$$\sum_{S \in \mathcal{B}} \lambda_S v(S) \leq v(N). \quad (10)$$

To prove the non-emptiness of the core, we show that (N, v) verifies (10). At first, we write the characteristic function v as

$$v(S) = \max_{\mathbf{f}} \left\{ \sum_{i \in S} \bar{U}_i(\mathbf{f}_i) \mid \mathbf{f} = (\mathbf{f}_1, \dots, \mathbf{f}_n) \in [0, +\infty)^{2n}, \right. \\ \left. f_{ij} = 0 \ \forall ij \notin g|_S, f_{ij} = f_{ji} \ \forall i, j \in S \right\},$$

where $\bar{U}_i(\mathbf{f}_i) = f(f_{ii} + \delta \sum_{j \in N \setminus \{i\}} f_{ij}) - kf_{ii}$. This definition is equivalent to the definition in (8). The equivalence can be shown by writing the first order conditions and by seeing they are the same than the first order conditions of our original maximization problem defining (8). Then let \mathcal{B} be any balanced collection, and for each $S \subseteq N$, let $\mathbf{f}^S = (\mathbf{f}_i^S)_{i \in S}$ be an arg max allocation which provides $v(S)$. Define the allocation \mathbf{f} , where for each $i \in N$, \mathbf{f}_i is the following convex combination of the \mathbf{f}_i^S ,

$$\mathbf{f}_i = \sum_{S \in \mathcal{B}, i \in S} \lambda_S \mathbf{f}_i^S.$$

The allocation \mathbf{f} is feasible for $v(N)$, since for each $i, j \in N$

$$\begin{aligned} f_{ij} &= \sum_{S \in \mathcal{B}, \mathcal{B} \ni i} \lambda_S \mathbf{f}_{ij}^S = \sum_{S \in \mathcal{B}, \mathcal{B} \ni i, j} \lambda_S \mathbf{f}_{ij}^S \\ &= \sum_{S \in \mathcal{B}, \mathcal{B} \ni i} \lambda_S \mathbf{f}_{ji}^S = f_{ji}. \end{aligned}$$

Then, by the concavity of the U_i , we have that

$$\begin{aligned} &\sum_{S \in \mathcal{B}} \lambda_S v(S) \\ &= \sum_{S \in \mathcal{B}} \lambda_S \sum_{i \in S} U_i(\mathbf{f}_i^S) = \sum_{i \in N} \sum_{S \in \mathcal{B}, S \ni i} \lambda_S U_i(\mathbf{f}_i^S) \\ &\leq \sum_{i \in N} U_i \left(\sum_{S \in \mathcal{B}, S \ni i} \lambda_S \mathbf{f}_i^S \right) = \sum_{i \in N} U_i(\mathbf{f}_i) \leq v(N). \end{aligned}$$

This proves (10) and, consequently, the nonemptiness of the core.

E Example 4.5

Given a symmetric graph g , the corresponding game (N, v) is such that each player is symmetric and then she will have Shapley value equal to $v(N)/n$. Moreover, given the symmetry of the game, the core is a symmetric convex set, which contains the symmetric vector given by the Shapley value.

F Example 4.6

Because of the symmetry property of the Shapley value, every player at the periphery has the same value. Such a value is given by

$$\begin{aligned}
\phi_P(v) &= \\
& \sum_{t=0}^{n-2} \binom{n-2}{t} \frac{t!(n-t-1)!}{n!} v(S_1) \\
& + \sum_{t=0}^{n-2} \binom{n-2}{t} \frac{(t+1)!(n-t-2)!}{n!} [v(S_{t+2}) - v(S_{t+1})] \\
& = \sum_{t=0}^{n-2} \frac{(n-2)!}{t!(n-t-2)!} \frac{t!(n-t-1)!}{n!} v(S_1) \\
& + \sum_{t=0}^{n-2} \frac{(n-2)!}{t!(n-t-2)!} \frac{(t+1)!(n-t-2)!}{n!} [v(S_{t+2}) - v(S_{t+1})] \\
& = \sum_{t=0}^{n-2} \frac{n-t-1}{n(n-1)} v(S_1) + \sum_{k=0}^{n-2} \frac{k+1}{n(n-1)} [v(S_{t+2}) - v(S_{t+1})] \\
& = \left(\frac{n(n-1)}{2} - 1 \right) \frac{1}{n(n-1)} v(S_1) \\
& + \sum_{t=2}^{n-1} \left(-\frac{1}{n(n-1)} \right) v(S_t) + \frac{1}{n} v(S_n) \\
& = \frac{1}{n} v(S_n) - \frac{1}{n(n-1)} \sum_{t=2}^{n-1} [v(S_t) - tv(S_1)]
\end{aligned}$$

where we observe that the terms $v(S_t) - tv(S_1)$ are all bigger than 0 because of the superadditivity of the game.

Because of the efficiency property, the center node has Shapley value given by

$$\begin{aligned}
\phi_C(v) &= v(S_n) - (n-1)\phi_P(v) \\
&= \frac{1}{n} v(S_n) + \frac{1}{n} \sum_{t=2}^{n-1} [v(S_t) - tv(S_1)].
\end{aligned}$$

To prove that the Shapley value belongs to the core, we need to prove that for each $S \subseteq N$, $\sum_{i \in S} \phi_i \geq v(S)$. At first, we observe that this is true for each coalition of cardinality 1, as a consequence of the superadditivity of the game⁶. Second, we consider a coalition R of $r \geq 2$ peripheric players. We need to prove that $r\phi_P(v) \geq v(R)$, and this is true as

$$v(R) = rv(S_1) \leq r\phi_P(v)$$

⁶Efficient payoff allocations which satisfy this inequality for each coalition of cardinality 1 are called *imputations*. The Shapley value of a superadditive game is always an imputation [17].

and the equality holds whenever we have disconnected players. Finally, we consider a coalition R of $r - 1$ peripheric players and 1 central player. We need to prove that

$$(r - 1)\phi_P(v) + \phi_C(v) \geq v(S_r),$$

which is equivalent to proving that

$$\frac{r}{n}v(S_n) + \frac{n - r}{n(n - 1)} \sum_{t=2}^{n-1} [v(S_t) - tv(S_1)] \geq v(S_r).$$

To show that this last inequality holds, it is sufficient to prove that

$$\frac{v(S_n)}{n} \geq \frac{v(S_r)}{r} \quad (11)$$

for each $r \leq n$. At first we observe that, as the neighborhood of a peripheric player is contained in the neighborhood of the central player, there always exists an efficient profile such that the central player is the only one to disclose a nonzero amount of personal data. Denote d_t^e the disclosure level of such an efficient profile for a star S_T with t players. The inequality holds as

$$\begin{aligned} \frac{v(S_n)}{n} \geq \frac{v(S_r)}{r} &\Leftrightarrow \\ \frac{nf(d_n^e) - kd_n^e}{n} \geq \frac{rf(d_r^e) - kd_r^e}{r} &\Leftrightarrow \\ f(d_n^e) - \frac{k}{n}d_n^e \geq f(d_r^e) - \frac{k}{r}d_r^e \end{aligned}$$

and d_r^e is the argmax of a function $f(x) - \frac{k}{r}x$ which is increasing in the parameter r , and then it will have a higher value at maximum for a higher parameter.

G Proof of Proposition 5.1

Both the properties follows from the definition of w . In fact, for each $g \in \mathcal{G}(N)$,

$$w(g) = v_g(N(g)) = \max_{\mathbf{d}_{N(g)} \in D_{N(g)}} W(\mathbf{d}_{N(g)}, g),$$

where

$$W(\mathbf{d}_{N(g)}, g) = \sum_{i \in N(g)} \left[f \left(d_i + \delta \sum_{j \in N_i(g)} d_j \right) - kd_i \right].$$

The aggregate utility W as a function of g is component additive, as for each $i \in N(g)$, her utility depends only on her neighborhood, which is a subset of the component to which player i belongs to. Then, also the maximum value of W as a function of g is component additive, as the players of each component maximize independently their aggregate utility, and this gives the component additivity of w . Moreover, as f and k do not depend on i , the utility of each player depends only on her position on the graph and not on her label, and this gives the anonymity property for w .

H Proof of Theorem 5.2

At first, observe that, given a graph $g \in \mathcal{G}(N)$, the monotonicity of the game (N, v_g) in Theorem 1 implies that $w(g + ii) \geq w(g)$ for each $i \in N$. From this it immediately follows that, given g , the graph g' which contains all the links of g and in addition all the loops, will provide a higher aggregate utility than g . Such a graph is s.t. every player in N has at least one link, i.e., $N(g') = N$. Moreover, for each $i \in N(g')$, we observe that

$$f\left(d_i + \delta \sum_{j \in N_i(g')} d_j\right) \leq f\left(d_i + \delta \sum_{j \in N_i(g^N)} d_j\right), \quad (12)$$

because f is an increasing function and because $g' \subseteq g^N$ implies that $N_i(g') \subseteq N_i(g^N)$. Given the graph g' with $N(g') = N(g^N) = N$, (12) implies that $W(\mathbf{d}_N, g') \leq W(\mathbf{d}_N, g^N)$ for each $\mathbf{d}_N \in D_N$, from which it follows

$$\begin{aligned} w(g') &= v_{g'}(N) = \max_{\mathbf{d}_N} W(\mathbf{d}_N, g') \\ &\leq \max_{\mathbf{d}_N} W(\mathbf{d}_N, g^N) = v_{g^N}(N) = w(g^N) \end{aligned}$$

and this proves that g^N is strongly efficient.

I Proof of Corollary 5.3

Let $g \subset g^N$ be a strongly efficient graph. Then, $N(g) = N$ follows from the proof of Theorem 3. Now, let \mathbf{d}^e be an efficient disclosure profile such that $W(\mathbf{d}^e, g) = w(g)$. Adding a link ij when player i is s.t. $d_i^e \neq 0$, the aggregate utility becomes $W(\mathbf{d}^e, g + ij) > W(\mathbf{d}^e, g)$ (meaning that the maximum value for the graph $g + ij$ will be also greater). As a consequence, in order for g to be efficient we need to have that if $ij \notin g$, then $d_i^e = d_j^e = 0$.

J Proof of Proposition 5.4

At first, we observe that $w(g|_S) \leq w(g + ij|_S)$ for each $S \subseteq N$, i.e., $v_g(S) \leq v_{g+ij}(S)$, and, in particular $v_g(S) = v_{g+ij}(S)$ for each $S \subseteq N \setminus \{i, j\}$. By writing the core conditions, we observe that they become more restrictive for each player different from i and j , but eventually less restrictive for i and j .

K Proof of Proposition 5.5

To show this result, it is sufficient to show that $\phi_i(g + ij, w) > \phi_i(g, w)$ for at least one $i \in N$. By the definition of Shapley value, in particular the result is true if all the marginal contributions are s.t., $w(g + ij|_{S \cup \{i\}}) - w(g + ij|_S) \geq w(g|_{S \cup \{i\}}) - w(g|_S)$, i.e., if the marginal contributions of player i are always

bigger when the link ij is present, and at least one inequality is strict. We notice that $w(g + ij|_S) = w(g|_S)$, and this implies that it sufficient to show that $w(g + ij|_{S \cup \{i\}}) \geq w(g|_{S \cup \{i\}})$ for each $S \subseteq N \setminus \{i\}$ and the inequality is strict for at least one coalition S . We have that the weak inequality is true, because of what we showed in the last part of the proof of Theorem 3.

L Example 5.6

We may observe at first that an efficient disclosure profile for both the graph is given by the player on the left giving \tilde{d} s.t., $f'(\tilde{d}) = k/3$. Then, both the graph are providing a value $v(N) = 3f(\tilde{d}) - k\tilde{d}$. The first graph in Figure 3-(a) is a star S_3 , s.t. each peripheric player has a Shapley value equal to

$$\phi_P = \frac{v(N)}{3} - \frac{1}{6}[v(S_2) - 2v(S_1)] < \frac{v(N)}{3}.$$

While adding the link as in Figure 3-(b), the players become symmetric and the ex-peripheric players now get a utility equal to $v(N)/3$, then strictly bigger than before. Obviously, as the aggregate utility remains the same, the Shapley value of the ex-central player decreases.