

A Study of Sybil Manipulations in Hedonic Games

Thibaut Vallée, Grégory Bonnet, Bruno Zanuttini, François Bourdon
Normandie Univ, France
UNICAEN, GREYC, F-14032 Caen, France
CNRS, UMR 6072, F-14032 Caen, France
firstname.lastname@unicaen.fr

ABSTRACT

Hedonic games model agents that decide which other agents they will join, given some preferences on other agents. We study Sybil attacks on such games, by a malicious agent which introduces multiple false identities, so that the outcome of the game is more interesting for itself. First taking Nash stability as the solution concept, we consider two simple manipulations, and show that they are essentially the only possible Sybil manipulations. Moreover, small experiments show that they are seldom possible in random games. We exhibit another simple manipulation on the concepts of (contractual) individual stability afterwards. Then we show that such hedonic games are very sensitive to Sybil manipulations, which contrasts sharply with the Nash case.

Categories and Subject Descriptors

Computing Methodologies [Artificial Intelligence]: Distributed artificial intelligence

Keywords

Coalition Formation; Game Theory; Strategic Reasoning

1. INTRODUCTION

In decentralized multi-agent systems, a recurrent question is how, at a given instant, the agents decide together who they will join (for instance, for playing a game). Coalition formation models such problems. Canonical coalitional games are based on transferable utility, whereas in hedonic games utility is not transferable. In the latter, each agent expresses a preference relation telling whom it accepts to join. Then the problem consists in finding a partition in which all agents are satisfied. A Nash stable coalition structure is a partition in which no agent wants to change coalition individually. However, such partitions are not always the optimal outcome for all agents: a malicious agent may report another preference relation in order to affect the equilibria and get a better outcome. In this work, we consider a particular manipulation, called *Sybil* attack. It consists of an agent joining the system under multiple false identities, with honest agents believing them all to be distinct, unknown agents, to which they are assumed to be indifferent.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*
Copyright © 2014, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

The malicious agent reports preferences defined on purposed for itself and its false identities. This quite general attack encompasses false reports of preferences, as widely studied, for instance, in voting systems. To the best of our knowledge, Sybil attacks on hedonic games have not been studied so far. After presenting related work (Section 2), we describe our model for games and attacks (Section 3). Considering simple attacks (Section 4) for Nash stability as the solution concept, we show under which conditions they are successful, and that it is computationally hard to carry them out. Then we show that they are essentially the only possible attacks (Section 5), and we report on small experiments which suggest that hedonic games are robust to them on average. Finally, we extend our study to the concepts of (contractual) individual stability (Section 6). We exhibit another simple attack, and we show that such hedonic games are very sensitive to Sybil manipulations. This is in sharp contrast with the robustness of Nash stability.

2. RELATED WORK

The problem of partitioning a group of agents so that all of them are satisfied with their own coalition is widely studied in the literature. Several models propose how each agent decides which coalitions it wants to form [5, 9, 11, 13]. They consider diverse properties on the partition that guarantee an equilibrium, for instance, Pareto optimality or Nash stability. However, even deciding whether there is a Nash stable coalition structure at all is an NP-complete problem [3]. Since stability depends on the preferences of each agent, what happens if a malicious agent lies about its own preferences so as to manipulate the system? Manipulations have been studied on a large panel of systems, including P2P networks [17], voting systems [4], weighted voting games [2], combinatorial auctions [8], matching problems [18], social networks [7], and reputation systems [14]. For instance, in voting systems, constructive manipulations try to make a candidate win, and destructive ones try to make a candidate lose. Walsh [19] shows empirically that even if it is NP-hard to manipulate a vote in the worst case, it is easy in practice for the STV and veto voting rules. A *Sybil attack* [10], or *false-name manipulation* [20], consists of introducing false identities in the system. A system can be made robust to such attacks by using a central certified authority [10] or searching for suspect clusters in the graph that structures the system [6], but these proposals consider reputation systems or combinatorial auctions and, to the best of our knowledge, Sybil attacks on hedonic systems have not been studied so far. Indeed [16] investigated the strategy-

h_1	$12 \succ 13m \succ 13 \sim 12m \succ 123m \sim 1 \succ 1m \sim 123$
h_2	$12 \sim 23m \succ 123 \sim 2m \succ 12m \succ 23 \succ 123m \sim 2$
h_3	$13 \sim 23m \succ 3m \succ 123 \succ 23 \succ 123m \sim 3 \succ 13m$
m	$1m \succ 2m \succ 3m \succ m \succ 12m \succ 13m \sim 23m \succ 123m$
NS_G	$\Pi_1 = \{12, 3m\}, \Pi_2 = \{13, 2m\}$
UR_G	$\Pi_3 = \{1, 23m\}, \Pi_4 = \{12m, 3\}, \Pi_5 = \{123m\}$

Figure 1: Running example with four agents 1, 2, 3, m .

proofness of coalitions formations rules on hedonics games but they studied the case of malicious agents which lie on their own preferences only. Observe that the related problem of *strategic cloning* of candidates has been studied for voting systems [12]. Hedonic games can be seen as a voting system insofar the agents express preferences over partitions, and one partition is elected, but with the important difference that the set of candidates (partitions) depends on the set of voters (agents).

3. MODEL

3.1 Hedonic Games

Consider an online game platform, as the League of Legends¹ matchmaking platform, where players can join a game instance with several other players. Some players do not want to play with certain players but prefer to play with some others, depending on game skills, previous experiences, or social affinities. At a given instant, several game instances can be launched each with a subset of players, possibly all of them. Notice that each game instance is then independant from the others. How to decide who will join which game instance, given that once an instance has been launched, no new player can join it? Observe that a player can always play alone, but this is typically not satisfying for her. Such problems can be modelled by hedonic games. These are coalitional games where each agent (player) expresses preferences about the possible coalitions (online game instances), where a *coalition* is a subset of the players involved.

DEFINITION 1. A (hedonic) game is a pair $G = \langle N, \succeq \rangle$, with $N = (a_1, \dots, a_n)$ a set of agents and \succeq a preference profile $(\succeq_1, \dots, \succeq_n)$, which gives each agent a_i a preference relation \succeq_i on the subsets of N (coalitions) containing a_i .

The preference relation of an agent may come out from some notion of trust or reputation, or from the outcomes of previous games. As we only focus on the system at a given instant, we assume that the preference profile is given.

DEFINITION 2. Let N be a set of agents. A preference relation \succeq on N is a total preorder (i.e., a reflexive, transitive and total relation²) on the subsets of N . We write \succ (resp. \sim) for the strict (resp. symmetric) part of \succeq . For $C, C' \subseteq N$, $C \succ_i C'$ (resp. $C \sim_i C'$) means that a_i prefers C to C' (resp. is indifferent to C, C').

EXAMPLE 1. Throughout the paper, we use the example game of Fig. 1 (top), where for compactness we omit the

¹<http://euw.leagueoflegends.com/>

²Totality is assumed for simplicity, but all results carry over to partial preorders.

subscripts of preference relations and we write, e.g., $13m$ for the coalition $\{h_1, h_3, m\}$. Here, h stands for “honest”, m for “malicious” and s for “Sybil” (a false identity of m). According to h_1 , the coalition 12 is preferred to $13m$ which is preferred to both 13 and $12m$. It is indifferent to the two latter. It also prefers the singleton 1 to $1m$ and to 123 .

Solving a hedonic game means finding a set of coalitions which *satisfies* the preferences of all agents. As is common, we consider nonoverlapping coalitions. In our example application, this means that each agent participates in only one game instance. We write $\Pi = \{C_1, \dots, C_m\}$ for a partition of N , and C_i^Π for the unique coalition in Π with $a_i \in C_i^\Pi$. We first consider Nash stability as the solution concept. We relax this assumption in the final part of the paper (Section 6) by considering the concepts of (contractual) individual stability. The concept of core stability [5] is left for a future work. A partition is Nash stable if no agent wants to unilaterally change coalition in it.

DEFINITION 3. Let $G = \langle N, \succeq \rangle$ be a hedonic game. A partition Π of N is said to be Nash stable if the following holds: $\forall a_i \in N, \nexists C \in \Pi \cup \{\emptyset\}, C \cup \{a_i\} \succ_i C_i^\Pi$.

EXAMPLE 2. Fig. 1 (“ NS_G ” row) gives the two Nash stable partitions of the game.

In general, G may have zero, one, or several Nash stable partitions. This solution concept may seem restrictive, but unstability captures the case when agents do not reach an agreement. In our example application, it happens when some players want to play with some others who do not want to play with them. In the sequel, we refer to a Nash stable partition simply as a *stable partition*, and we write NS_G for the set of all stable partitions in G . There are many ways to chose the actual outcome of the game in NS_G , for instance by a negotiation protocol or by a random draw. To make things precise, we assume the following.

ASSUMPTION 1. The outcome of G is drawn uniformly at random from NS_G .

The important point is that the goal of a malicious agent is to increase the proportion of *satisfactory* outcomes among all possible outcomes of the game (Section 3.3). Our results hold whenever the actual outcome is chosen in such a way that such a goal makes sense. Finally, in a coalitional game, any agent may decide unilaterally to be in the singleton coalition, meaning not to join any other agent. Hence the preference relation of an agent a_i over the coalitions to which the singleton $\{a_i\}$ is preferred is irrelevant to the outcome of the game. Consequently, for ease of reading, we use a Representation by Individually Rational Lists of Coalitions (RIRLC) for \succeq_i , where the players give only the coalitions preferred or indifferent to the singleton [3].

3.2 Sybil Attacks

An agent performs a *Sybil attack* [10] if it appears in the system under multiple false identities. In our example application, a malicious player may join the online game platform using several accounts at the same time. The false accounts may be used for joining multiple game instances and continuing to play only with the weakest opponents (simulating, say, an unwanted disconnection from the network for other instances), or for joining a group which refuses to play with the malicious player (under its true identity). In this paper, we assume that only one malicious agent m tries to manipulate the game, by adding false identities.

DEFINITION 4. Let $G = \langle N, \succeq \rangle$ be a game, and $m \in N$ be an agent. A Sybil attack of G by m is a set of new agents $\{s_1, \dots, s_k\}$, called Sybil agents, a preference relation \succeq'_m for m , and a preference relation \succeq'_{s_i} for each Sybil s_i .

Let us notice that this definition of Sybil attacks is a generalization of the canonical manipulations where malicious agents falsely report their own preferences. Indeed, in such an attack the malicious agent manipulates the game, by reporting a false preference relation for itself, and introducing false identities with preference relations defined on purpose. This attack is quite general, since no assumption is made on the number of Sybil agents nor on the knowledge of the game by m . It may not even know the number of other agents, or at the other extreme it may know the full preference profile. There may be as many Sybil agents as needed by m , but observe that the case of no Sybil at all encompasses attacks which consist in simply lying about its own preferences. Since new agents are introduced by a manipulation, we need to determine the preference relation (written \succeq'_i) of each honest agent h_i over coalitions involving them. We introduce two assumptions. Independence to irrelevant alternatives [1] is a common requirement, *e.g.*, for voting systems. In our context, it imposes that if an agent prefers C_1 to C_2 , the arrival of a new agent does not change this preference. The second assumption models an *a priori* acceptance of unknown agents by honest agents.

ASSUMPTION 2 (IRRELEVANT ALTERNATIVES).
 $\forall C_1, C_2 \subseteq N, \forall a_i \in C_1 \cap C_2 : C_1 \succeq_i C_2 \Leftrightarrow C_1 \succeq'_i C_2$

ASSUMPTION 3 (BENEFIT OF THE DOUBT).
 $\forall C \subseteq N, \forall a_i \in C, \forall u \notin N : C \sim'_i C \cup \{u\}$

Assumption 2 is a commonsense assumption. Assumption 3 may seem beneficial to malicious agents, but precisely, if hedonic games are robust *even under favourable conditions for the manipulator*, they will be even more so under a weaker assumption. Moreover, Assumption 3 is in some sense necessary for an open system to allow new agents to cooperate with existing agents. Indeed, in the context of reputation systems, [15] state the following desirable property: *the new entrants should not be penalised by initially having low reputation values attributed to them*. In our example application, players are indifferent to one newcomer joining the online game platform, as they have no prior experience with her. Furthermore, we will relax this assumption in Section 5. Finally, the benefit of the doubt does not allow multiple unknown agents to join a single coalition. Indeed, this assumption is made only for $C \subseteq N$. As soon as an unknown agent u joins C , $C \cup \{u\} \subseteq N$ does not hold any more. Hence the benefit of the doubt is only granted to a

single unknown agent (with arbitrary preferences for more unknown agents joining a coalition).

EXAMPLE 3. If an unknown agent u enters the game of Fig. 1, then \succeq'_1 satisfies $12 \sim'_1 12u \succ'_1 13m \sim'_1 13mu \succ'_1 13 \sim'_1 13u \sim'_1 12m \dots 123 \sim'_1 123u$.

Finally, the new preferences of honest agents may be *any* preference profile as long as it satisfies Assumptions 2 and 3.

DEFINITION 5. Let $G = \langle N, \succeq \rangle$ be a game, with $N = \{h_1, \dots, h_n, m\}$. A game G' results from a Sybil attack $(\{s_1, \dots, s_k\}, \succeq'_m, (\succeq'_{s_1}, \dots, \succeq'_{s_k}))$ of G by m , if it is of the form $G' = \langle N \cup \{s_1, \dots, s_k\}, (\succeq_1, \dots, \succeq_n, \succeq'_m, \succeq'_{s_1}, \dots, \succeq'_{s_k}) \rangle$ where for $i = 1, \dots, n$, \succeq'_i satisfies Assumptions 2, 3.

3.3 Rationality of Malicious Agents

We are interested in *rational* malicious agents, in the sense that they perform an attack if and only if they prefer the outcome of the resulting game. We define the goal of a malicious agent in a quite general manner. An *effective* manipulation increases the proportion of *satisfactory* partitions, where satisfactory is defined relative to a *threshold coalition* C_θ , meaning the minimally preferred coalition in which m wants to be. We let C_θ be an input for m that models its goal. Hence, C_θ is chosen by m depending on its intentions. As particular cases, setting C_θ to the coalition maximally preferred by m means m wants to increase its chances to be precisely in this coalition, and setting it to one immediately preferred to the singleton $\{m\}$ means it simply wants to increase its chances not to be alone.

DEFINITION 6. Let $G = \langle N, \succeq \rangle$ be a game. A partition Π of N is satisfactory for m relative to a threshold coalition C_θ if $\Pi \in NS_G$ and $C_m^\Pi \succeq_m C_\theta$ hold. The set of all stable and satisfactory partitions is written NS_G^θ .

EXAMPLE 4. On Fig. 1, for $C_\theta = 1m$ no partition in the “ NS_G ” row is satisfactory for m , but for $C_\theta = 3m$ both are.

In the game G' resulting from a manipulation, m is involved both under its true identity m and false identities s_1, \dots, s_k . Intuitively, if m wants to join a coalition C , it will be equally happy if one of its false identities joins it instead. Hence, we redefine satisfactory partitions for G' as follows.

DEFINITION 7. Let $G' = \langle N \cup \{s_1, \dots, s_k\}, \succeq' \rangle$ result from a manipulation. A partition Π' is said to be satisfactory relative to C_θ if $\Pi' \in NS_{G'}$ holds and we have either $C_m^{\Pi'} \succeq_m C_\theta$ or $\exists s_i \in \{s_1, \dots, s_k\}, C_{s_i}^{\Pi'} \cup \{m\} \setminus \{s_i\} \succeq_m C_\theta$.

We insist that the satisfaction of m in the manipulated game G' is defined relative to its *initial preferences* \succeq_m (in G , and hence over N). In particular, a coalition containing several identities of m cannot make m satisfied. We formalize in this manner the fact that in the outcome of the game, m cannot concretely act under several identities in parallel, hence all but one of its identities must defect. We assume that such defections will not affect the rest of the game. Indeed, in games where coalitions are independent and act in parallel, an agent can quit a coalition by disconnecting from the network, simulating a failure, or simply without doing any costly action (such as folding in a poker game). For instance, our example application meets this assumption since a player can always quit a game, but the remaining players cannot join a game instance that has already started. The following definition is justified by Assumption 1.

DEFINITION 8. Let G be a game, m be an agent, and G' result from some manipulation of G by m . Let moreover C_θ be a coalition. We define r_θ^G to be the ratio $|NS_G^\theta|/|NS_G|$ (with $r_\theta^G = 0$ for $|NS_G| = 0$, by convention). The manipulation is effective relative to C_θ if $r_\theta^{G'} > r_\theta^G$ holds.

Observe that if C_θ is the singleton $\{m\}$, then all Nash stable partitions are satisfactory. Also observe that if all stable partitions in G are satisfactory and there is at least one, then r_θ^G is 1 and no manipulation can be (strictly) effective.

4. MANIPULATIONS

4.1 A Constructive Sybil Attack

We show a first attack. This manipulation is *constructive*: the malicious agent manipulates the game so that a desirable unstable partition becomes stable. In any unstable partition Π for a game G , we can split the agents into two groups: those which do not want to change coalition, and those which want to change. The latter are called *responsible* for the instability of Π .

DEFINITION 9. Let G be a game, a_i an agent, and Π an unstable partition. Then a_i is said to be responsible for the instability of Π if there is a coalition $C \in \Pi$ with $C \cup \{a_i\} \succ_i C_i^\Pi$. Such a coalition is said to be attractive (for a_i).

We write UR_G for the set of all partitions which are unstable with m as the unique responsible, and UR_G^θ for those partitions $\Pi \in UR_G$ which moreover contain a satisfactory attractive coalition C for m (i.e., $C \cup \{m\} \succ_m C_m^\Pi$ and $C \cup \{m\} \succeq_m C_\theta$).

EXAMPLE 5. Fig. 1 gives the set UR_G . For $C_\theta = 1m$ or $2m$, UR_G^θ is $\{\Pi_3\}$, and for $C_\theta = 3m$, UR_G^θ is $\{\Pi_3, \Pi_4\}$.

The constructive manipulation³ works when the malicious agent m is the unique responsible for the instability of a partition Π , and the attractive coalitions in Π are satisfactory for it. Roughly, m manipulates the game by becoming disinterested (accepting all coalitions), and introducing one false identity, which expresses its original preferences while benefiting from the doubt (Assumption 3). We define the *indifferent* profile for m , written \succeq_m^{indif} , in which m is indifferent to all coalitions ($C_1 \sim_m^{indif} C_2$ for all $C_1, C_2 \ni m$). We also write $\succeq_m [m/s]$ for the relation obtained from \succeq_m by replacing m with s in all coalitions.

DEFINITION 10. Let $G = \langle \{h_1, \dots, h_n, m\}, \succeq \rangle$ be a game. The constructive manipulation of G by m is the manipulation with one Sybil agent s , in which m reports the preference relation $\succeq'_m := \succeq_m^{indif}$ and s reports $\succeq'_s := \succeq_m [m/s]$.

Observe in particular that the Sybil agent reports that it does *not* want to join m (since m is replaced with s in \succeq'_s).

EXAMPLE 6. For the game on Fig. 1, the constructive manipulation introduces a Sybil s with $1s \succ'_s 2s \succ'_s 3s \succ'_s s$.

We now show under what conditions the constructive manipulation is effective. First, we examine under what conditions an agent wants to change coalition in some partition

³We abuse words by using “the”, as there may be other constructive manipulations, and similarly for destructive manipulations (Section 4.2). However, as we show in Section 5 the manipulations which we exhibit can be seen as canonical.

for the manipulated game. Obviously, m never wants to do so, since it reports indifference to all coalitions. Now fix a game $G = \langle N, \succeq \rangle$, a malicious agent $m \in N$, and a partition Π for G . Write G' for the game resulting from the constructive manipulation of G by m , $C_0 \in \Pi \cup \{\emptyset\}$ for a coalition, and $\Pi' = \Pi[s \rightarrow C_0]$ for the partition for G' obtained from Π when the Sybil agent s joins C_0 , i.e., $\Pi' = \Pi \setminus \{C_0\} \cup \{C_0 \cup \{s\}\}$.

LEMMA 1. An honest agent h wants to change coalition in Π' if and only if it wants to change coalition in Π .

PROOF. By definition, h wants to change in Π' if and only if $C' \cup \{h\} \succ'_h C_h^{\Pi'}$ (1) holds for some $C' \in \Pi'$. Now by Assumption 3 we have $C' \cup \{h\} \sim'_h C' \cup \{h\} \setminus \{s\}$ and $C_h^{\Pi'} \sim'_h C_h^{\Pi'} \setminus \{s\}$. Hence (1) is equivalent to $C' \cup \{h\} \setminus \{s\} \succ'_h C_h^{\Pi'} \setminus \{s\}$. But $C_h^{\Pi'} \setminus \{s\}$ is precisely C_h^Π and $C' \setminus \{s\}$ is in Π , hence h wants to change to C' in Π' if and only if it wants to change to $C' \setminus \{s\}$ in Π . \square

LEMMA 2. The Sybil agent s wants to change coalition in Π' if and only if $m \in C_0$ holds, or there is a coalition $C \in \Pi$ with $m \notin C$ and $C \cup \{m\} \succ_m C_0 \cup \{m\}$.

PROOF. First assume that s wants to change from $C_0 \cup \{s\}$ to $C' \in \Pi'$. We assume $m \notin C_0$ and define C as in the claim. Indeed, since s wants to change, we have $C' \cup \{s\} \succ'_s C_0 \cup \{s\}$. Hence, by definition of \succeq'_s we have $m \notin C'$ and $C' \cup \{m\} \succ_m C_0 \cup \{m\}$, i.e., C' is as in the claim. Conversely, if $m \in C_0$ holds, then s wants to change in Π' (at least to $\{s\}$). Finally, if there is C as in the claim, then by definition of \succeq'_s , s wants to change to $C \cup \{s\}$ in Π' . \square

From these two lemmas we easily obtain the following.

COROLLARY 1. A partition Π' is stable in G' if and only if, writing $\Pi' = \Pi[s \rightarrow C_0]$, m is not in C_0 , $C_0 \cup \{m\}$ is maximally preferred by m in Π , and either (1) Π is stable, or (2) m is the unique responsible of the instability of Π .

EXAMPLE 7. On Fig. 1, for $C_\theta = 1m$, $\Pi'_3 = \Pi_3[s \rightarrow 1] = \{1s, 23m\}$ is satisfactory and $\Pi'_1 = \Pi_1[s \rightarrow 12] = \{12s, 3m\}$ is stable but not satisfactory.

We can now give the exact conditions under which the constructive manipulation is effective on a game G . We only give the characterization for the case $NS_G^\theta = \emptyset$, since otherwise either $NS_G^\theta = NS_G$ and the agent is already fully satisfied, or $\emptyset \subsetneq NS_G^\theta \subsetneq NS_G$ and the destructive manipulation (Section 4.2) is also applicable, and fully effective (Proposition 3). However, a characterization can be given also for the case $NS_G^\theta \neq \emptyset$; for instance, when \succeq_m is strict, the constructive manipulation is effective if and only if $|UR_G^\theta|/|UR_G| > |NS_G^\theta|/|NS_G|$ holds.

PROPOSITION 1. Assume $NS_G^\theta = \emptyset$. The constructive manipulation is effective on G if and only if $UR_G^\theta \neq \emptyset$ holds.

PROOF. By definition of r_θ^G , if the manipulation is effective then G' has at least one satisfactory partition Π' . Write $\Pi' = \Pi[s \rightarrow C_0]$. From Corollary 1 it follows $\Pi \in NS_G$ or $\Pi \in UR_G$. Since Π' is satisfactory, either $C_0 \cup \{m\} \succeq_m C_\theta$ or $C_m^\Pi \succeq_m C_\theta$ holds. In both cases, from $NS_G^\theta = \emptyset$ we get $\Pi \notin NS_G$, hence $\Pi \in UR_G$ and finally, $\Pi \in UR_G^\theta$. The converse is shown similarly. \square

EXAMPLE 8. On Fig. 1, for $C_\theta = 1m$ the manipulation is effective ($NS_G^\theta = \emptyset$ and $UR_G^\theta = \{\Pi_3\}$). However, for $C_\theta = 2m$ it results in a strictly worse situation for m . Indeed, as $NS_G^\theta = \{\Pi_2\}$ and $UR_G^\theta = \{\Pi_3\}$, $r_\theta^G = 1/2$ and $r_\theta^{G'} = 2/5$.

PROPOSITION 2. The following problem is NP-hard: given a game G with an RIRLC representation, a player m , and a coalition C_θ , decide whether the constructive manipulation is effective on G for m relative to C_θ .

PROOF. We reduce from the problem of deciding whether a game in RIRLC, say G_0 , has at least one Nash stable partition. This problem is NP-complete [3]. Given G_0 , we build a game G with $NS_G^\theta = \emptyset$, but with $UR_G^\theta \neq \emptyset$ if and only if G_0 has a stable partition. From Proposition 1 it follows that the constructive manipulation is effective in G if and only if G_0 has a Nash stable partition. Write $G_0 = \langle N_0, \succeq_0 \rangle$ with $N_0 = \{h_1, \dots, h_n\}$. The game G is defined from G_0 by adjoining two new agents, h and m , and introducing the following preference relations: $\{h, m\} \succ_m \{m\}$, $\{h\} \succ_h C$ for all coalitions $C \neq \{h\}$, and \succeq_i as built from $(\succeq_0)_i$ with Assumptions 2 and 3. Intuitively, h wants to be alone and m wants to join h ; other agents are indifferent to them, and otherwise keep their preferences from G_0 . Clearly, G can be built in time polynomial in the size of G_0 . Finally, we let C_θ be the coalition $\{h, m\}$. No partition Π is stable in G , because if h is not in the singleton coalition $\{h\}$, then it wants to change to it, while if it is in $\{h\}$, then m wants to join it. Now assume that there is a stable partition Π_0 in G_0 , and consider the partition $\Pi = \Pi_0 \cup \{\{h\}, \{m\}\}$ for G . Then clearly m is the unique responsible for the unstability of Π . Moreover, in Π the attractive coalition for m is satisfactory for it. Hence $\Pi \in UR_G^\theta$ holds. Dually, if all partitions Π_0 for G_0 are unstable, then because h_1, \dots, h_n are indifferent to h, m , all partitions involving h, m must also be unstable. Finally, G has no stable partition, and $UR_G^\theta \neq \emptyset$ holds if and only if G_0 has a stable partition, as desired. \square

Observe that this manipulation is independent of the preferences of honest agents. However, deciding whether it is effective, beside being computationally hard, requires to know them. Moreover, such decision is in some sense necessary, since an ineffective constructive manipulation may (strictly) worsen the situation of m (Example 8).

4.2 A Destructive Sybil Attack

We now consider a *destructive* attack, in the sense that it results in undesirable stable partitions becoming unstable. With Nash stability, a single “veto” agent can refuse a coalition, and therefore make a given partition unstable. The destructive attack builds on this by using a single false identity, which vetoes any partition where m is not satisfied.

DEFINITION 11. Let $G = \langle N, \succeq \rangle$ be a game. The destructive manipulation G' by m uses one Sybil agent s , with $\succeq'_m := \succeq_m$, and \succeq'_s defined for all $C \subseteq N$ by $C \cup \{s\} \succ'_s \{s\}$ if $m \in C$ and $C \not\subseteq_m C_\theta$, or $\{s\} \succ'_s C \cup \{s\}$ otherwise.

In particular, we have $\{s\} \succ'_s C_\theta \cup \{s\}$, and the relative preferences between the coalitions in each case can be arbitrary. Informally, the Sybil agent wants to join all coalitions containing m and not preferred to C_θ . As m does not want to be with s , all unsatisfactory partitions become unstable.

EXAMPLE 9. On Fig. 1 with $C_\theta = 2m$, the preferences of s are given by $3ms, ms, 12ms, 13ms, 23ms, 123ms \succ'_s s$.

The destructive manipulation is effective when there is at least one satisfactory partition in the original game. Fix a game G , a malicious agent m , and a coalition C_θ . Write G' for the game resulting from the destructive manipulation.

LEMMA 3. There is a satisfactory partition in G' if and only if there is one in G . Moreover, all stable partitions in G' are satisfactory for m .

PROOF. For the first claim (“only if”), assume Π' is satisfactory in G' , and write $\Pi' = \Pi[s \rightarrow C_0]$. If m is in a satisfactory coalition in Π' , then Π is satisfactory in G . Otherwise only $C_0 \cup \{s\}$ is satisfactory in Π' , but then the definition of \succeq'_s implies that s wants to change to $C_m^{\Pi'}$, contradicting the stability of Π' . For the “if” direction, simply observe that if Π is satisfactory in G , then $\Pi \cup \{\{s\}\}$ is satisfactory in G' . For the second claim, let Π' be a stable but unsatisfactory partition in G' . Then by definition of \succeq'_s , s is in the same coalition as m . But then m prefers being in the singleton coalition, contradicting the stability of Π' . \square

Recall that if all stable partitions are satisfactory in G , no manipulation can be strictly effective. Interestingly, when the destructive manipulation is effective, it is *fully* so: all stable partitions are satisfactory in the manipulated game.

PROPOSITION 3. The destructive manipulation is effective on G iff G has at least one satisfactory partition, and at least one stable but unsatisfactory partition.

EXAMPLE 10. On Fig. 1, the manipulation is effective for $C_\theta = 2m$ (Π_2 is satisfactory, Π_1 is not, so only $\Pi_2[s \rightarrow \emptyset] = \{13, 2m, s\}$ remains), but it is not for $C_\theta = 3m$ (m is already fully satisfied in G) nor for $C_\theta = 1m$ (r_θ^G remains 0).

Like for the constructive manipulation, it is hard to decide whether this manipulation is effective, and this requires some knowledge about G . However, unlike the constructive case, the attack cannot strictly worsen the situation of m .

PROPOSITION 4. The following problem is NP-hard: given a game G with an RIRLC representation, a player m , and a coalition C_θ , decide whether the destructive manipulation is effective on G for m relative to C_θ .

PROOF. The construction is similar to the one in Proposition 2. Given $G_0 = \langle N_0, \succeq_0 \rangle$, we build a game G with both a satisfactory, and a stable but unsatisfactory partitions, if and only if G_0 has a stable partition. The game G is defined from G_0 by adding three agents, h, h' , and m , with the preference relations: $\{h, h', m\} \succ_a \{a\}$ for $a \in \{h, h', m\}$ and, for all agents h_i , \succeq_i as built from $(\succeq_0)_i$ with Assumptions 2 and 3. Intuitively, h, h' and m want to be all together or each alone, and other agents are indifferent to them. Finally, we let C_θ be the coalition $\{h, h', m\}$. Let Π be any partition in G . Then Π is not stable if at least one of h, h', m is with some agent h_i , since they prefer to be alone. It is not stable either if exactly two of them are together. In the two remaining cases, either each of them is in the singleton coalition or they are all together, and it is easily seen that Π is stable if and only if the partition $\Pi \setminus \{\{h\}, \{h'\}, \{m\}, \{h, h', m\}\}$ is stable for G_0 . Moreover, though both are stable, only the partition containing the coalition $\{h, h', m\}$ is satisfactory for m , as desired. \square

5. ROBUSTNESS FOR NASH STABILITY

5.1 Two Canonical Manipulations

We now show that the manipulations exhibited above are the only possible Sybil attacks on a hedonic game using at most one false identity, in the sense that games which are not manipulable (efficiently) by the constructive or the destructive attack defined above, are not manipulable by an attack at all under our assumptions.

PROPOSITION 5. *Let G be a hedonic game with the Nash stable solution concept, m an agent in G , and C_θ a threshold coalition for m . If neither the constructive nor the destructive manipulations are effective on G , then no Sybil attack using at most one false identity is effective on G .*

PROOF. We assume that there is an effective manipulation M but the destructive manipulation is not effective, and we show that the constructive one is effective. Since the destructive manipulation is not effective, by Proposition 3 either all stable partitions in G are satisfactory, or none is. In the former case M cannot be effective, contradicting the assumption. Hence G has no satisfactory partition. Write G' for the game resulting from the manipulation M , and s for the Sybil agent used by M . Since M is effective, in G' there is a satisfactory Π' . Write Π for the partition $\{C' \setminus \{s\} \mid C' \in \Pi'\}$. We show that either Π is satisfactory in G , yielding a contradiction, or $\Pi \in UR_G^\theta$ holds. First observe that no honest agent h_i wants to change coalition in Π ; otherwise, by the benefit of the doubt (Assumption 3) h_i desires the same change in Π' , contradicting the stability of Π' . As concerns m , we distinguish two cases. Assume that in Π' no coalition is preferred to that of m , precisely, that for all coalitions $C \in \Pi'$ it holds $C \setminus \{s\} \cup \{m\} \not\prec_m C_m^{\Pi'} \setminus \{s\}$. Then m does not want to change coalition in Π . So Π is stable. Moreover, because m is in its preferred coalition in Π' and Π' is satisfactory, Π is satisfactory as well, a contradiction. Hence there is $C \in \Pi'$ with $C \setminus \{s\} \cup \{m\} \succeq_m C_m^{\Pi'} \setminus \{s\}$, and m wants to change to such a C in Π . Moreover, since Π' is satisfactory there must be such a C which is satisfactory, and it follows $\Pi \in UR_G^\theta$. Because G has no satisfactory partition ($NS_G^\theta = \emptyset$), the constructive manipulation is effective (Proposition 1). \square

Proposition 5 hence completely characterizes the conditions under which a hedonic game is manipulable by a false report of preferences and/or the use of one Sybil agent. Moreover, it holds even for all imaginable manipulations, possibly with many Sybil agents, provided Assumption 3 is extended to any set of agents, that is, provided honest agents are indifferent to any number of unknown agents joining a coalition. While this may seem an unrealistic assumption (for instance, honest agents would be indifferent to play a game with some friends, or with the same friends plus a thousand of unknowns), this extended result is unexpected as it shows that, under the extended assumption, using many Sybil agents instead of just one does not help the malicious agent. Another interesting relaxation of Assumption 3 is the following form of *subadditivity*: honest agents prefer unknown agents not to join a coalition, but otherwise maintain their preferences when unknown agents are disregarded.

ASSUMPTION 4 (WEAK SUBADDITIVITY). $\forall C_1, C_2 \subseteq N$, $\forall a_i \in N$ with $C_1 \succeq_i C_2$, $\forall u \notin N : C_1 \succeq_i C_1 \cup \{u\} \succeq_i C_2$

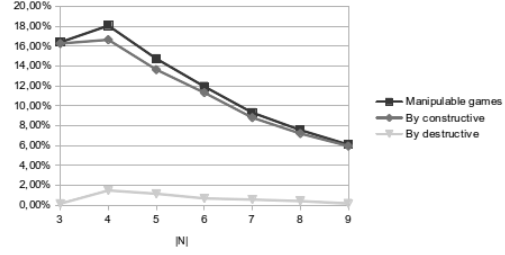


Figure 2: Ratio of manipulable games

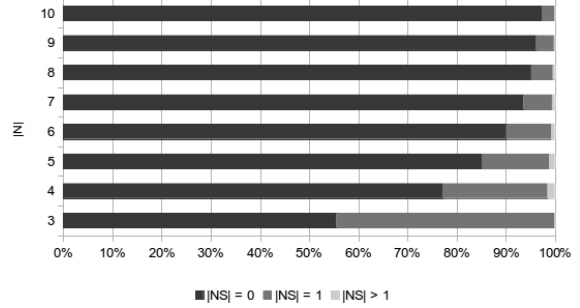


Figure 3: Ratio of stable partitions

Lemma 1 for the constructive manipulation and the characterization of Proposition 5 are based on the fact that an honest agent h wants to change to C' in Π' if and only if it wants to change to $C' \setminus \{s\}$ in Π . Since this is still true under the relaxed Assumption 4 and the other results do not use Assumption 3, all our results for Nash stability hold under the relaxed assumption. Hence, they also hold under the dual assumption of *weak superadditivity*, including the characterization of Proposition 5, despite the fact that superadditivity may seem more beneficial to malicious agents.

5.2 Empirical Study

We present here small experiments which suggest that, even if some games are manipulable, this seldom occurs in practice. In order to give a rough estimate of the probability for a game to be manipulable, we ran a set of experiments with 3 to 10 agents. For each experiment, we ran 10,000 simulations, each of which consists of generating a hedonic game G with preference profiles drawn uniformly at random. Then we measured the proportion of those games which are manipulable (Fig. 2), and the proportion of games with given numbers of stable partitions (Fig. 3). Fig. 2 suggests that as the number of agents increases, a random game has a decreasing probability to be manipulable. This is quite intuitive since the malicious agent has less and less chances to be the unique responsible for the unstability of a partition. For instance, with 6 agents, only 67 of the 10,000 generated games were manipulable by the destructive attack, and beyond 7 agents, less than 10% of the games were manipulable at all. Similarly, Fig. 3 suggests that as the number of agents increases, the number of stable partitions in the original game decreases. Again, this typical observation for Nash stability can be explained by the fact that more and more agents are candidate for changing coalition.

6. OTHERS SOLUTION CONCEPTS

As we saw, the robustness of hedonic games with Nash stability as a solution concept is due to the fact that the set of Nash stable partitions is small, and may even be empty. Consequently, other less restrictive solution concepts were proposed, such as individual or contractual individual stability [5]. We now study these solution concepts.

6.1 Contractual individual stability

This solution concept provides a nonempty set of stable partitions. Informally, contractual individual stability means that no agent can change its coalition without the acceptance of both the coalitions which it joins and leaves.

DEFINITION 12. Let $G = \langle N, \succeq \rangle$ be a hedonic game. A partition Π of N is said to be contractually individually stable (C.I. stable for short) if the following holds: $\forall a_i \in N, \nexists C \in \Pi \cup \{\emptyset\}$ such that (1) $C \cup \{a_i\} \succ_i C_i^\Pi$, (2) $\forall a_j \in C : C \cup \{a_j\} \succeq_j C$, and (3) $\forall a_k \in C_i^\Pi : C_i^\Pi \setminus \{a_i\} \succeq_k C_i^\Pi$.

In the sequel, we write CIS_G for the set of C.I. stable partitions of a game G . It is known that $NS_G \subseteq CIS_G$ and $CIS_G \neq \emptyset$ hold for any game G [5]. We now propose a constructive manipulation of such games⁴. For a game $G = \langle N, \succeq \rangle$ and a coalition $C \subseteq N$, we write $G \setminus C$ for the game $\langle N \setminus C, \succeq_{|N \setminus C} \rangle$, where $\succeq_{|N \setminus C}$ is the restriction of \succeq to the preferences of the agents in $N \setminus C$.

DEFINITION 13. Let $G = \langle N, \succeq \rangle$ be a game with the solution concept of C.I. stability. The constructive manipulation of G by m uses one Sybil agent s , with $\succeq'_m := \succeq_m$, and $\succeq'_s := C_\theta \setminus \{m\} \cup \{s\} \succeq_s \{s\}$.

PROPOSITION 6. A partition Π' is C.I. stable in G' if and only if, writing $\Pi' = \Pi[s \rightarrow C_0]$, either (1) $C_s^{\Pi'} = \{s\}$, $C_\theta \setminus \{m\} \notin \Pi'$, and Π is C.I. stable in G , or (2) $C_s^{\Pi'} = C_\theta \setminus \{m\} \cup \{s\}$, and $\Pi' \setminus \{C_s^{\Pi'}\}$ is C.I. stable in $G \setminus (C_\theta \setminus \{m\})$.

PROOF. For the “only if” direction, let Π' be a C.I. stable partition of G' . If $C_s^{\Pi'} \neq \{s\}$ and $C_s^{\Pi'} \neq C_\theta \setminus \{m\} \cup \{s\}$ then $\{s\}$ wants to change to $\{s\}$, and no agent can veto this because of Assumption 3. Hence $\Pi' \notin CIS_{G'}$, a contradiction. Now for $C_s^{\Pi'} = \{s\}$, $C_\theta \setminus \{m\} \notin \Pi'$ holds because otherwise s would want to join it (other agents cannot veto this because of Assumption 3); moreover, since Π' is C.I. stable no other agent can change coalition in it, and because of Assumptions 2 and 3, the same holds in Π , which is thus C.I. stable. Finally, for $C_s^{\Pi'} = C_\theta \setminus \{m\} \cup \{s\}$, the same reasoning shows that $\Pi' \setminus \{C_s^{\Pi'}\}$ is stable in $G \setminus (C_\theta \setminus \{m\})$.

We now show the “if” direction. For $C_s^{\Pi'} = \{s\}$, s does not want to change coalition because of the assumption $C_\theta \setminus \{m\} \notin \Pi'$, and the other agents cannot change because otherwise they could perform the same change in Π , contradicting its C.I. stability. Hence Π' is stable. Finally, for $C_s^{\Pi'} = C_\theta \setminus \{m\} \cup \{s\}$, clearly s does not want to change coalition. As for honest agents, none of them can quit or join $C_\theta \setminus \{m\} \cup \{s\}$ because of the veto of s . Hence possibilities of changing coalitions can be only among other coalition, but there cannot be any of them since $\Pi' \setminus \{C_s^{\Pi'}\}$ is C.I. stable in $G \setminus (C_\theta \setminus \{m\})$. This concludes the proof. \square

⁴Due to space constraints, we let the case of the destructive manipulation in individual and contractual individual stability solution concepts for future work.

We now show that this manipulation is always effective, which contrasts with the robustness of the Nash case.

PROPOSITION 7. Let G be a hedonic game with the solution concept of C.I. stability. Then the constructive manipulation is effective on G as soon as the malicious agent is not fully satisfied in G (i.e., as soon as $r_\theta^{G'}$ is not 1).

PROOF. The manipulation is effective if $r_\theta^{G'} > r_\theta^G$, by definition. From Proposition 6, the number of C.I. stable partitions in G' is $|CIS_G \cup CIS_{G \setminus (C_\theta \setminus \{m\})}|$. As $C_s^{\Pi'} = C_\theta \setminus \{m\} \cup \{s\}$ for $\Pi \in CIS_{G \setminus (C_\theta \setminus \{m\})}$, every C.I. stable partition built from $G \setminus (C_\theta \setminus \{m\})$ (Case (1) of Proposition 6) is satisfactory. Moreover, a partition Π' built from a partition Π for G (Case (2) of Proposition 6) is satisfactory if and only if so is Π . Finally, both cases yield distinct partitions Π' . It follows $r_\theta^{G'} = \frac{|CIS_G^\theta| + |CIS_{G \setminus (C_\theta \setminus \{m\})}|}{|CIS_G \cup CIS_{G \setminus (C_\theta \setminus \{m\})}|} > \frac{|CIS_G^\theta|}{|CIS_G|} = r_\theta^G$ (using $\frac{|CIS_G^\theta|}{|CIS_G|} = r_\theta^G \neq 1$ and $CIS_{G \setminus (C_\theta \setminus \{m\})} \neq \emptyset$ [5]). \square

6.2 Individual stability

Contractual individual stability may seem to be constraining, since agents cannot decide alone to leave their coalitions. Hence we now consider individual stability. This solution concept means that no agent can change its coalition without the acceptance of the coalition it joins.

DEFINITION 14. Let $G = \langle N, \succeq \rangle$ be a hedonic game. A partition Π of N is said to be individually stable if the following holds: $\forall a_i \in N, \nexists C \in \Pi \cup \{\emptyset\}$ such that (1) $C \cup \{a_i\} \succ_i C_i^\Pi$ and (2) $\forall a_j \in C : C \cup \{a_j\} \succeq_j C$.

We write IS_G for the set of all individually stable partitions of G . It is known that $IS_G \subseteq CIS_G$ always holds, but IS_G may be empty [5]. We now investigate the effectiveness of the constructive manipulation defined for the concept of contractual individual stability, when used with individual stability. For simplicity, in this section we assume that C_θ is the coalition maximally preferred by the malicious agent (we assume there is only one maximally preferred coalition) and that C_θ is not the singleton $\{m\}$. Nevertheless, the construction and results can easily be extended to the general case.

PROPOSITION 8. Let Π be an individually stable partition for G . Then (1) if $C_\theta \setminus \{m\}$ is not in Π , then $\Pi \cup \{\{s\}\} \in IS_{G'}$ holds, and (2) if $C_\theta \setminus \{m\}$ is in Π , then $\Pi[s \rightarrow C_\theta \setminus \{m\}] \in IS_{G'}$ holds. In both cases, no other partition of the form $\Pi[s \rightarrow C_0]$ is individually stable in G .

PROOF. The proof is similar to the one for Proposition 6. In the first case, write $\Pi' = \Pi \cup \{\{s\}\}$. Clearly s does not want to change coalition (since $C_\theta \setminus \{m\}$ is not in Π'), and no other agent can change coalition since otherwise it could perform the same change in Π , contradicting its stability. Now in the second case, write $\Pi' = \Pi[s \rightarrow C_\theta \setminus \{m\}]$. Clearly again, s does not want to change coalition in Π' , and any change available to another agent would be available in Π as well (using Assumption 3), contradicting its stability. We now prove the last claim. In the first case, if s is not in the coalition $\{s\}$, then it prefers to be alone, and no agent can veto this by definition of individual stability. In the second case, if s is not in $C_\theta \setminus \{m\} \cup \{s\}$, then it wants to join it, and no agent can veto this by Assumption 3. \square

As previously, we can now give the conditions under which the constructive manipulation is effective on a game G .

PROPOSITION 9. *Let G be a hedonic game with individual stability as the solution concept. Then the constructive manipulation is effective on G if and only if there is an individually stable partition Π for G which contains $C_\theta \setminus \{m\} \in \Pi$.*

PROOF. From Proposition 8 it follows that each partition $\Pi \in IS_G$ gives rise to exactly one partition $\Pi' \in IS_{G'}$, so the number of stable partitions in G' is the same as in G . Now consider an individually stable partition Π for G , and the corresponding partition Π' for G' (built as in Proposition 8). If Π does not contain $C_\theta \setminus \{m\}$, then we distinguish two cases. If Π is satisfactory, *i.e.*, contains the unique satisfactory coalition C_θ , then so is $\Pi' = \Pi \cup \{\{s\}\}$. Dually, if Π is not satisfactory, *i.e.*, does not contain C_θ , then neither does Π' . Hence the number of satisfactory partitions is preserved from G to G' for the case $C_\theta \setminus \{m\} \notin \Pi$. Now consider the case $C_\theta \setminus \{m\} \in \Pi$. Then clearly Π is not satisfactory for m while $\Pi' = \Pi[s \rightarrow C_\theta \setminus \{m\}]$ is, hence the ratio $r_\theta^{G'}$ is greater than r_θ^G if and only if there is such a partition Π , which concludes the proof. \square

7. CONCLUSION AND FUTURE WORK

We studied the robustness of hedonic games to a quite general type of manipulations, called Sybil attacks, with Nash stability as a solution concept. We showed that they are manipulable only under particular conditions, and we exhibited two manipulations which cover all these conditions. These manipulations involve only one false identity and no knowledge of the game (not even the number of honest agents). We showed that it is computationally hard for a malicious agent to decide whether one or the other is effective, and that these conditions are seldom met by random games. From all these results we conclude that hedonic games with Nash stability as a solution concept are very robust to Sybil attacks. Observe that our results do *not* imply that there are no more effective constructive manipulations. There may well be one which increases the satisfaction of the malicious agent by more than the one which we exhibited, or which cannot worsen its situation, or whose effectivity can be decided efficiently. However, our results and experiments *do* show that it would be seldom effective. Consequently, we investigate other solution concepts than Nash stability, such as individual or contractual individual stability. As these solution concepts are less restrictive, the conditions for a rational manipulation to exist are less restrictive as well. In particular, in sharp contrast with the Nash case, we showed that when contractual individual stability is the solution concept, then every game is manipulable (using only one Sybil agent). Our results rely on two assumptions about the attitude of honest agents in presence of new agents, namely, irrelevance of independent alternatives and benefit of the doubt. These assumptions may seem overly beneficial to malicious agents, but we showed that (slightly) weakening the benefit of the doubt still allows manipulations in the Nash case, which reinforces our conclusions. Nevertheless, it would be interesting to consider relaxing the assumptions on solution concepts other than Nash stability, as the core stability. Intuitively, manipulations should be easier to achieve when honest agents prefer new agents to join, and harder when they avoid them. Finally, we defined our setting by introducing a threshold coalition and an uniform draw to choose

the stable partitions. It would be interesting to do without such input and assumption by extending the players' preferences over partitions to preferences over sets of partitions and using well-known social choice functions.

8. REFERENCES

- [1] K. J. Arrow. *Social Choice and Individual Values*. Yale University Press, 1970.
- [2] H. Aziz, Y. Bachrach, E. Elkind, and M. Paterson. False-name manipulations in weighted voting games. *JAIR*, 40:57–93, 2011.
- [3] C. Ballester. NP-completeness in hedonic games. *GEB*, 49(1):1–30, 2004.
- [4] J. J. Bartholdi, C. A. Tovey, and M. A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.
- [5] A. Bogomolnaia and M. O. Jackson. The stability of hedonic coalition structures. *Games and Econ. Beh.*, 38:201–230, 2002.
- [6] G. Bonnet. A protocol based on a game-theoretic dilemma to prevent malicious coalitions in reputation systems. In *20th ECAI*, pages 187–191, 2012.
- [7] V. Conitzer, N. Immorlica, J. Letchford, K. Munagala, and L. Wagman. False-name-proofness in social networks. *LNCS*, 6484:209–221, 2010.
- [8] V. Conitzer and M. Yokoo. Using mechanism design to prevent false-name manipulations. *Artificial Intelligence Magazine*, 31(4):65–78, 2010.
- [9] D. Dimitrov and S. C. Sung. Enemies and friends in hedonic games: individual deviations, stability and manipulation. Technical report, Tilburg Univ., 2004.
- [10] J. R. Douceur. The Sybil attack. In *1st IPTPS*, pages 251–260, 2002.
- [11] E. Elkind and M. Wooldridge. Hedonic coalition nets. In *8th AAMAS*, pages 417–424, 2009.
- [12] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. Cloning in elections: Finding the possible winners. *JAIR*, 42(1):529–573, 2011.
- [13] T. Génin and S. Aknine. Coalition formation strategies for self-interested agents in hedonic games. In *19th ECAI*, pages 1015–1016, 2010.
- [14] K. Hoffman, D. Zage, and C. Nita-Rotaru. A survey of attack and defense techniques for reputation systems. *ACM Computer Survey*, 42(1):1–31, 2009.
- [15] S.D. Ramchurn, T.D. Huynh, and N. R. Jennings. Trust in multiagent systems. *The Knowledge Engineering Review*, 19:1–25, 2004.
- [16] Carmelo Rodríguez-Álvarez. Strategy-proof coalition formation. *International Journal of Game Theory*, 38(3):431–452, 2009.
- [17] H. Rowaihy, W. Enck, P. D. McDaniel, and T. La Porta. Limiting Sybil attacks in structured P2P networks. In *26th INFOCOM*, pages 2596–2600, 2007.
- [18] T. Todo and V. Conitzer. False-name-proof matching. In *12th AAMAS*, 2013.
- [19] T. Walsh. Where are the hard manipulation problems? *JAIR*, 42:1–29, 2011.
- [20] M. Yokoo, Y. Sakurai, and S. Matsubara. The effect of false-name bids in combinatorial auctions: new fraud in Internet auctions. *GEB*, Vol. 46:174–188, 2004.