Individual-based Stability in Hedonic Games depending on the Best or Worst Players

(Extended Abstract)

Haris Aziz Paul Harrenstein Evangelia Pyrga
Institut für Informatik, TU München
{aziz,harrenst,pyrga}@in.tum.de

ABSTRACT

We consider classes of hedonic games in which each player's preferences over coalition structures are induced by the best player ($\mathcal{B}$- and B-hedonic games) or the worst player ($\mathcal{W}$- and W-hedonic games) in his coalition. For these classes, which allow for concise representation, we analyze the computational complexity of deciding the existence of and computing individually stable, Nash stable, and individually rational and contractually individually stable coalition partitions. We identify a key source of intractability in compact coalition formation games in which preferences over players are extended to preferences over coalitions.

Categories and Subject Descriptors

F.2 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity; J.4 [**Computer Applications**]: Social and Behavioral Sciences - Economics

General Terms

Economics, Theory and Algorithms

Keywords

Game theory (cooperative and non-cooperative)

1. HEDONIC GAMES

Coalition formation games, as introduced by Drèze and Greenberg [5], provide a simple but versatile formal model for modeling and analyzing how agents join in groups. In many situations it is natural to assume that a player's appreciation of a coalition structure only depends on the coalition he is a member of and not on how the remaining players are grouped. Much of the work on coalition formation concentrates on these so-called *hedonic games*.

Formally, a *hedonic game* is a pair $(N, \succsim)$, where N is a set of players and $\succsim = (\succsim_1, \dots, \succsim_{|N|})$ a profile specifying the preferences of each player i as a transitive and complete relation $\succsim_i$ over the set $\mathcal{N}_i = \{S \subseteq N \mid i \in S\}$ of coalitions i may belong to. If $\succsim_i$ is also anti-symmetric we say that i's preferences are *strict*. A coalition $S \in \mathcal{N}_i$ is *acceptable to i* if i prefers S to being alone, i.e., $S \succsim_i \{i\}$ and *unacceptable*, otherwise.

As the set of coalitions a player may be member of grows exponentially in the number of players, for hedonic games concise representations do not exist in general. However, concise representations are possible if we assume the players to have preferences over the *players* in N and that their appreciation of a coalition S systematically depends on their most or least preferred players in S. We distinguish four such classes of hedonic games: $\mathcal{B}$-hedonic games [2, 4], B-hedonic games, $\mathcal{W}$-hedonic games [3, 4], and W-hedonic games.

As no confusion is likely, we also use $\succsim_i$ to denote player i's preferences over N. For J a subset of players, we denote by $\max_i(J)$ and $\min_i(J)$ the sets of players that are *most*, respectively, *least* preferred by i in J, on the understanding that $\max_i(\emptyset) = \min_i(\emptyset) = \{i\}$. With a slight abuse of notation we write $\max_i(S) \succsim_i \max_i(T)$ ($\min_i(S) \succsim_i \min_i(T)$) if $s \succsim_i t$ for all $s \in \max_i(S)$ and all $t \in \max_i(T)$ ($s \in \min_i(S)$ and all $t \in \min_i(T)$, respectively). Moreover, player j is said to be *acceptable* to i if $j \succsim_i i$, and *unacceptable* otherwise.

In a *B-hedonic* game, the preferences $\succsim_i$ of a player i over players extend to preferences over coalitions in such a way that $S \succsim_i T$ if and only if either (a) some j in T is unacceptable to i or (b) all players in S and T are acceptable to i and $\max_i(S \setminus \{i\}) \succsim_i \max_i(T \setminus \{i\})$. Analogously, in a *W-hedonic* game we have that $S \succsim_i T$ if and only if either (a) some j in T is unacceptable to i or (b) all players in S and T are acceptable to i and $\min_i(S \setminus \{i\}) \succsim_i \min_i(T \setminus \{i\})$. For *hedonic games with $\mathcal{W}$-preferences* (or *$\mathcal{W}$-hedonic games*) are such that $S \succsim_i T$ if and only if $\min_i(S \setminus \{i\}) \succsim_i \min_i(T \setminus \{i\})$. Finally, *hedonic games with $\mathcal{B}$-preferences* (or *$\mathcal{B}$-hedonic games*) are defined such that $S \succ_i T$ if and only if (a) $\max_i(S \setminus \{i\}) \succ_i \max_i(T \setminus \{i\})$ or (b) both $\max_i(S \setminus \{i\}) \sim_i \max_i(T \setminus \{i\})$ and $|S| < |T|$.[1]

A solution of a hedonic game is a partition of the players in coalitions. In this respect, the main focus has been on solutions that capture a notion of stability. Thus, a partition π is said to be *Nash stable* (*NS*) if no player can benefit from moving to another (possibly empty) coalition in π. Partition π is *individually stable* (*IS*) if no player can benefit from moving to another (possibly empty) coalition T in π without making the members of T worse off. Finally, π is *contractually individually stable* (*CIS*) if no player would strictly prefer to move from his coalition S to another existing (pos-

[1] $\mathcal{W}$- and $\mathcal{B}$-hedonic games were originally introduced by Cechlárová and Romero-Medina [4]. For $\mathcal{B}$-hedonic games the dependence on coalition size prevents the grand coalition N to be trivially the most preferred one by all players.

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

sibly empty) coalition T in π without making neither the members of S nor the members of T worse off. It is easily seen that Nash stability implies individual stability and that individual stability implies contractual individual stability. Another, minimal, requirement, automatically satisfied by NS and IS partitions, is that a partition is *individually rational* (*IR*), i.e., that it assigns each player to a coalition that is acceptable to him.

We analyze the computational complexity of deciding the existence of and computing IS, NS, and CIS & IR partitions in B-, $\mathcal{B}$-, W-, and $\mathcal{W}$-hedonic games.

2. RESULTS

We first note that W-hedonic games are equivalent to hedonic games with $\mathcal{W}$-preferences if only individually rational outcomes are considered. For both $\mathcal{W}$- and B-hedonic games, if preferences do not allow unacceptable players, then the partition consisting of the grand coalition is Nash stable and therefore individually stable. However, if unacceptability of players is expressed, we obtain relatively more negative results. Our hardness results are by reductions from SAT and rely on the idea of a so-called *stalker game*. The simplest example is the hedonic game $(N, \succsim)$ where $N = \{1,2\}$ and $\{1\} \succ_1 \{1,2\}$ and $\{1,2\} \succ_2 \{2\}$. Then, player 2 will stalk player 1 and the game has no NS partition.

THEOREM 1. *For $\mathcal{W}$-hedonic and B-hedonic games, deciding whether a NS partition exists is NP-complete.*

PROOF (SKETCH). By a reduction from SAT. Let $\varphi = X_1 \wedge \cdots \wedge X_k$ be a Boolean formula in conjunctive normal form in which all and only the Boolean variables $p_1, \ldots, p_m$ occur. Now define the B-hedonic game $(N, \succsim)$, where $N = \{X_1, \ldots, X_k\} \cup \{p_1, \neg p_1, \ldots, p_m, \neg p_m\} \cup \{0, 1\}$.

Define the preferences $\succsim$ such that for each literal p or $\neg p$, and each clause $X = (x_1 \vee \cdots \vee x_\ell)$,

$$\begin{array}{rl} p\colon & (0, 1, \text{———}, \; p \,\|\, \neg p, X_1, \ldots, X_k) \\ \neg p\colon & (0, 1, \text{———}, \neg p \,\| \;\; p, X_1, \ldots, X_k) \\ X\colon & (1, \text{———} \mid X_1, \ldots, X_k \,\|\, 0, x_1, \ldots, x_\ell) \\ 0\colon & (\text{———}, 0 \,\|\, 1, X_1, \ldots, X_k) \\ 1\colon & (\text{———}, 1 \,\|\, 0, X_1, \ldots, X_k), \end{array}$$

where " ——— " stands for the players not explicitly mentioned in the list, " $|$ " for $\succ_i$, commas for $\sim_i$, and the players to the right of " $\|$ " are unacceptable.

To prove that φ is satisfiable if and only if an NS partition for $(N, \succsim)$ exists, first assume that there exists a valuation v that satisfies φ. Define the partition $\pi = \{\{1, x'_1, \ldots, x'_{\ell'}\}, \{0, x''_1, \ldots, x''_{\ell''}\}, \{X_1, \ldots, X_k\}\}$ where $x'_1, \ldots, x'_{\ell'}$ are the literals rendered true by v and $x''_1, \ldots, x''_{\ell''}$ are those that are rendered false. It can easily be verified that π is NS-stable.

For the opposite direction, assume that there is a NS partition π. Then, for each clause $X = (x_1 \vee \cdots \vee x_\ell)$ there is some literal $x \in \{x_1 \vee \cdots \vee x_\ell)\}$ that is in π in the same coalition as 1; if not, X would become the stalker of 1. One can show that setting to true all the literals that are in the same coalition as 1 results in an assignment that satisfies φ. $\square$

The reduction in the proof of Theorem 1 is the prototype for more complicated reductions used to establish the results on NP-completeness in Table 1. These results involve an extended concept of a stalker game.

class	preferences	NS	IS	CIS & IR
$\mathcal{B}$	general	?	in P*	in P*
$\mathcal{B}$	strict	in P	in P*	in P*
B	general	NPC	NPC	in P*
B	strict	NPC	NPC	in P*
$\mathcal{W}$/W	general	NPC	NPC	in P*
$\mathcal{W}$/W	strict	NPC	?	in P*

Table 1: Complexity of individual-based stability. The positive results even hold for computing stable partitions whereas the NP-completeness results even hold for checking the existence of a stable partition. An asterisk indicates that a stable partition is guaranteed to exist.

EXAMPLE 1 (EXTENDED STALKER GAME). *Let $N = \{0, \ldots, 4\}$ and, assuming arithmetic modulo 5, the preferences over N of each player i be given by:*

$$i+1 \succ_i i-1 \succ_i i \succ_i \cdots$$

Then, in the B-, W- and $\mathcal{W}$-hedonic games induced by these preferences each player i stalks player $i+1$, joining him in any coalition whenever $i+1$ is alone. Consequently, no IS partition exists.

We also obtain some positive results. Firstly, a CIS and IR partition can be computed in polynomial time for all classes of games considered by starting with the individually rational partition of singletons and allowing arbitrary CIS deviations. For $\mathcal{B}$-hedonic games, in which a coalition is unacceptable only if all other players are unacceptable, positive results are even easier to obtain. In particular, we show that for $\mathcal{B}$-hedonic games, an IS partition is guaranteed to exist and can be computed in polynomial time.

Our results are summarized in Table 1 (for details and proofs, please see [1]). We obtain a general insight that in hedonic games based on extensions of preferences over players to preferences over coalitions, the following property can lead to intractability: the presence of an unacceptable player rendering a coalition unacceptable.[2]

REFERENCES

[1] H. Aziz, P. Harrenstein, and E. Pyrga. Individual-based stability in hedonic games depending on the best or worst players. Technical report, http://arxiv.org/abs/1105.1824, 2011.

[2] K. Cechlárová and J. Hajduková. Computational complexity of stable partitions with B-preferences. *International Journal of Game Theory*, 31(3):353–354, 2002.

[3] K. Cechlárová and J. Hajduková. Stable partitions with $\mathcal{W}$-preferences. *Discrete Applied Mathematics*, 138(3): 333–347, 2004.

[4] K. Cechlárová and A. Romero-Medina. Stability in coalition formation games. *International Journal of Game Theory*, 29:487–494, 2001.

[5] J. H. Drèze and J. Greenberg. Hedonic coalitions: Optimality and stability. *Econometrica*, 48(4):987–1003, 1980.

[2]This work is supported by the Deutsche Forschungsgemeinschaft under grants BR 2312/10-1, BR 2312/3-3, and BR 2312/9-1.