Pareto Optimality in Coalition Formation

Haris Aziz, Felix Brandt, and Paul Harrenstein

Institut für Informatik, Technische Universität München, 80538 München, Germany
{aziz,brandtf,harrenst}@in.tum.de

Abstract. A minimal requirement on allocative efficiency in the social sciences is Pareto optimality. In this paper, we identify a far-reaching structural connection between Pareto optimal and perfect partitions that has various algorithmic consequences for coalition formation. In particular, we show that computing and verifying Pareto optimal partitions in general hedonic games and B-hedonic games is intractable while both problems are tractable for roommate games and W-hedonic games. The latter two positive results are obtained by reductions to maximum weight matching and clique packing, respectively.

1 Introduction

Topics concerning coalitions and coalition formation have come under increasing scrutiny of computer scientists. The reason for this may be obvious. For the proper operation of distributed and multiagent systems, cooperation may be required. At the same time, collaboration in very large groups may also lead to unnecessary overhead, which may even exceed the positive effects of cooperation. To model such situations formally, concepts from the social and economic sciences have proved to be very helpful and thus provide the mathematical basis for a better understanding of the issues involved.

Coalition formation games, which were first formalized by Drèze and Greenberg [9], model coalition formation in settings in which utility is *non-transferable*. In many such 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. Initiated by Banerjee et al. [3] and Bogomolnaia and Jackson [4], much of the work on coalition formation now concentrates on these so-called *hedonic games*. In this paper, we focus on Pareto optimality and individual rationality in this rich class of coalition formation games.

The main question in coalition formation games is which coalitions one may reasonably expect to form. To get a proper formal grasp of this issue, a number of stability concepts have been proposed for hedonic games—such as the core or Nash stability—and much research concentrates on conditions for existence, the structure, and computation of stable and efficient partitions. *Pareto optimality*—which holds if no coalition structure is strictly better for some player without being strictly worse for another—and *individual rationality*—which holds if every player is satisfied in the sense that no player would rather be on his own—are commonly considered minimal requirements for any reasonable partition.

Another reason to investigate Pareto optimal partitions algorithmically is that, in contrast to other stability concepts like the core, they are guaranteed to exist. This even holds if we additionally require individual rationality. Moreover, even though the *Gale-Shapley algorithm* returns a core stable matching for marriage games, it is already NP-hard to check whether the core is empty in various classes and representations of hedonic games, such as roommate games [12], general hedonic games [2], and games with $\mathscr{B}$- and $\mathscr{W}$-preferences [7, 6]. Interestingly, when the status-quo partition cannot be changed without the mutual consent of all players, Pareto optimality can be seen as a stability notion [11].

We investigate both the problem of finding a Pareto optimal and individually rational partition and the problem of deciding whether a partition is Pareto optimal. In particular, our results concern *general hedonic games*, *B-hedonic* and *W-hedonic games* (two classes of games in which each player's preferences over coalitions are based on his most preferred and least preferred player in his coalition, respectively), and *roommate* games.

Many of our results, both positive and negative, rely on the concept of *perfection* and how it relates to Pareto optimality. A *perfect* partition is one that is most desirable for every player. We find (*a*) that under extremely mild conditions, NP-hardness of finding a perfect partition implies NP-hardness of finding a Pareto optimal partition (Lemma 1), and (*b*) that under stronger but equally well-specified circumstances, feasibility of finding a perfect partition implies feasibility of finding a Pareto optimal partition (Lemma 2). The latter we show via a Turing reduction to the problem of computing a perfect partition. At the heart of this algorithm, which we refer to as the *Preference Refinement Algorithm (PRA)*, lies a fundamental insight of how perfection and Pareto optimality are related. It turns out that a partition is Pareto optimal for a particular preference profile if and only if the partition is perfect for another but related profile (Theorem 1). In this way PRA is also applicable to any other discrete allocation setting.

For general allocation problems, *serial dictatorship*—which chooses subsequently the most preferred allocation for a player given a fixed ranking of all players—is well-established as a procedure for finding Pareto optimal solutions (see, e.g., [1]). However, it is only guaranteed to do so, if the players' preferences over outcomes are strict, which is not feasible in many compact representations. Moreover, when applied to coalition formation games, there may be Pareto optimal partitions that serial dictatorship is unable to find, which may have serious repercussions if also other considerations, like fairness, are taken into account. By contrast, PRA handles weak preferences well and and is complete in the sense that it may return any Pareto optimal partition, provided that the subroutine that calculates perfect partitions can compute any perfect partition (Theorem 2).

2 Preliminaries

In this section, we review the terminology and notation used in this paper.

Hedonic games Let N be a set of n players. A *coalition* is any non-empty subset of N. By $\mathscr{N}_i$ we denote the set of coalitions player i belongs to, i.e., $\mathscr{N}_i = \{S \subseteq N : i \in S\}$. A *coalition structure*, or simply a *partition*, is a partition π of the players N into coalitions, where $\pi(i)$ is the coalition player i belongs to.

A *hedonic game* is a pair (N, R), where $R = (R_1, \dots, R_n)$ is a *preference profile* specifying the preferences of each player i as a binary, complete, reflexive, and transitive *preference relation* R_i over $\mathscr{N}_i$. If R_i is also anti-symmetric we say that i's preferences are *strict*. We adopt the conventions of social choice theory by writing $S\ P_i\ T$ if $S\ R_i\ T$ but not $T\ R_i\ S$—i.e., if *i strictly prefers* S to T—and $S\ I_i\ T$ if both $S\ R_i\ T$ and $T\ R_i\ S$—i.e., if i is *indifferent* between S and T.

For a player i, a coalition S in $\mathscr{N}_i$ is *acceptable* if for i being in S is at least preferable as being alone—i.e., if $S\ R_i\ \{i\}$—and *unacceptable* otherwise.

In a similar fashion, for X a subset of $\mathscr{N}_i$, a coalition S in X is said to be *most preferred in X by i* if $S\ R_i\ T$ for all $T \in X$ and *least preferred in X by i* if $T\ R_i\ S$ for all $T \in X$. In case $X = \mathscr{N}_i$ we generally omit the reference to X. The sets of most and least preferred coalitions in X by i, we denote by $\max_{R_i}(X)$ and $\min_{R_i}(X)$, respectively.

In hedonic games, players are only interested in the coalition they are in. Accordingly, preferences over coalitions naturally extend to preferences over partitions and we write $\pi\, R_i\, \pi'$ if $\pi(i)\, R_i\, \pi'(i)$. We also say that partition π is *acceptable* or *unacceptable* to a player i according to whether $\pi(i)$ is acceptable or unacceptable to i, respectively. Moreover, π is *individually rational* if π is acceptable to all players. A partition π is *Pareto optimal for R* if there is no partition π' with $\pi'\ R_j\, \pi$ for all players j and $\pi'\ P_i\, \pi$ for at least one player i. Partition π is, moreover, said to be *weakly Pareto optimal for* R_i if there is no π' with $\pi'\ P_i\, \pi$ for all players i.

Classes of hedonic games The number of potential coalitions grows exponentially in the number of players. In this sense, hedonic games are relatively large objects and for algorithmic purposes it is often useful to look at classes of games that allow for concise representations.

For *general hedonic games*, we will assume that each player expresses his preferences only over his acceptable coalitions. This representation is also known as *Representation by Individually Rational Lists of Coalitions* [2].

We now describe classes of hedonic games in which the players' preferences over coalitions are induced by their preferences over the other players. For R_i such preferences of player i over players, we say that a player j is *acceptable* to i if $j\, R_i\, i$ and *unacceptable* otherwise. Any coalition containing an unacceptable player is unacceptable to player i.

Roommate games. The class of *roommate games*, which are well-known from the literature on matching theory, can be defined as those hedonic games in which only coalitions of size one or two are acceptable and preferences R_i over other players are extended naturally over preferences over coalitions in the following way: $\{i, j\}\, R_i \{i, k\}$ if and only if $j\, R_i\, k$ for all $j, k \in N$.

B-hedonic and W-hedonic games. For a subset J of players, we denote by $\max_{R_i}(J)$ and $\min_{R_i}(J)$ the sets of the most and least preferred players in J by i, respectively. We will assume that $\max_{R_i}(\emptyset) = \min_{R_i}(\emptyset) = \{i\}$. In a *B-hedonic game* the preferences R_i of a player i over players extend to preferences over coalitions in such a way that, for all coalitions S and T in $\mathscr{N}_i$, we have $S\ R_i\ T$ if and only if $j\, R_i\, k$ for all $j \in \max_{R_i}(S \setminus \{i\})$ and $k \in \max_{R_i}(T \setminus \{i\})$ or some j in T is unacceptable to i. Analogously, in a *W-hedonic game*

(N, R), we have $S\ R_i\ T$ if and only if $j\, R_i\, k$ for all $j \in \min_{R_i}(S \setminus \{i\})$ and $k \in \min_{R_i}(T \setminus \{i\})$ or some j in T is unacceptable to i.[1]

3 Perfection and Pareto Optimality

Pareto optimality constitutes rather a minimal efficiency requirement on partitions. A much stronger condition is that of *perfection*. We say that a partition π is *perfect* if $\pi(i)$ is a most preferred coalition for all players i. Thus, every perfect partition is Pareto optimal but not necessarily the other way round. Perfect partitions are obviously very desirable, but, in contrast to Pareto optimal ones, they are unfortunately not guaranteed to exist. Nevertheless, there exists a strong structural connection between the two concepts, which we exploit in our algorithm for finding Pareto optimal partitions in Section 4.

The problem of finding a perfect partition (`PerfectPartition`) we formally specify as follows.

`PerfectPartition`
Instance: A preference profile R
Question: Find a perfect partition for R.
If no perfect partition exists, output $\emptyset$.

We will later see that the complexity of `PerfectPartition` depends on the specific class of hedonic games that is being considered. By contrast, the related problem of *checking* whether a partition is perfect is an almost trivial problem for virtually all reasonable classes of games. If perfect partitions exist, they clearly coincide with the Pareto optimal ones. Hence, an oracle to compute a Pareto optimal partition can be used to solve `PerfectPartition`. If this Pareto optimal partition is perfect we are done, if it is not, no perfect partitions exist. Thus, we obtain the following simple lemma, which we will invoke in our hardness proofs for computing Pareto optimal partitions.

Lemma 1. *For every class of hedonic games for which it can be checked in polynomial time whether a given partition is perfect, NP-hardness of* `PerfectPartition` *implies NP-hardness of computing a Pareto optimal partition.*

It might be less obvious that a procedure solving `PerfectPartition` can also be deployed as an oracle for an algorithm to compute Pareto optimal partitions. To do so, we first give a characterization of Pareto optimal partitions in terms of perfect partitions, which forms the mathematical heart of the Preference Refinement Algorithm to be presented in the next section.

The connection between perfection and Pareto optimality can intuitively be explained as follows. If all players are indifferent among all coalitions, all partitions are perfect. It follows that the players can always relax their preferences up to a point where

[1] W-hedonic games are equivalent to hedonic games with $\mathscr{W}$-preferences if individually rational outcomes are assumed. Unlike hedonic games with $\mathscr{B}$-preferences, B-hedonic games are defined in analogy to W-hedonic games and the preferences are not based on coalition sizes (cf. [7]).

perfect partitions become possible. We find that, if a partition is perfect for a minimally relaxed preference profile—in the sense that, if any one player relaxes his preferences only slightly less, no perfect partition is possible anymore—, this partition is Pareto optimal for the original unrelaxed preference profile. To see this, assume π is perfect in some minimally relaxed preference profile and that some player i reasserts some strict preferences he had previously relaxed, thus rendering π no longer perfect. Now, π does not rank among i's most preferred partitions anymore. By assumption, none of i's most preferred partitions is also most preferred by all other players. Hence, it is impossible to find a partition π' that is better for i than π, without some other player strictly preferring π to π'. It follows that π is Pareto optimal.

To make this argumentation precise, we introduce the concept of a coarsening of a preference profile and the lattices these coarsenings define. Let $R = (R_1, \dots, R_n)$ and $R' = (R'_1, \dots, R'_n)$ be preference profiles over a set X and let i be a player. We write $R_i \leq_i R'_i$ if

$$R_i|_{\{x,y\}} = R'_i|_{\{x,y\}} \text{ for all } x \in X \text{ and all } y \in X \setminus \max_{R_i}(X).$$

Accordingly, R_i is exactly like R'_i, except that in R'_i player i may have strict preferences among some of his most preferred coalitions in R_i. Thus, R'_i is finer than R_i. For instance, let R_i, R'_i, R''_i, and R'''_i be such that $\pi_1 \, P_i \, \pi_2 \, P_i \, \pi_3$, $\pi_1 \, I'_i \, \pi_2 \, P'_i \, \pi_3$, $\pi_1 \, P''_i \, \pi_2 \, I''_i \, \pi_3$, and $\pi_1 \, I'''_i \, \pi_2 \, I'''_i \, \pi_3$, then $R'''_i \leq_i R'_i \leq_i R_i$ and $R'''_i \leq_i R''_i$, but not $R''_i \leq_i R_i$. It can easily be established that $\leq_i$ is a linear order for each player i.

We say that a preference profile $R = (R_1, \dots, R_n)$ over X is a *coarsening of* or *coarsens* another preference profile $R' = (R'_1, \dots, R'_n)$ over X whenever $R_i \leq_i R'_i$ for every player i. In that case we also say that R' *refines* R and write $R \leq R'$. Moreover, we write $R{<}R'$ if $R{\leq}R'$ but not $R'{\leq}R$. Thus, if R' refines R, i.e., if $R \leq R'$, then for each i and all coalitions S and T we have that $S \, R'_i \, T$ implies $S \, R_i \, T$, but not necessarily the other way round. It is worth observing that, if a partition is perfect for some preference profile R, then it is also perfect for any coarsening of R. The same holds for Pareto optimal partitions.

For preference profiles R and R' with $R \leq R'$, let $[R, R']$ denote the set $\{R'' : R \leq R'' \leq R'\}$, i.e., the set of all coarsenings of R' that also refine R. $([R, R'], \leq)$ is a complete lattice with R and R' as bottom and top element, respectively. We say that R *covers* R' if R is a minimal refinement of R' with $R' \neq R$, i.e., if $R' < R$ and there is no R'' such that $R' < R'' < R$. Observe that, if R covers R', R and R' coincide for all but one player, say i, for whom R_i is the unique minimal refinement of R'_i such that $R'_i \neq R_i$. We also denote R_i by $\texttt{Cover}(R'_i)$. If no cover of R'_i exists, $\texttt{Cover}(R'_i)$ returns the empty set.

We are now in a position to prove the following theorem, which characterizes Pareto optimal partitions for a preference profile R as those that are perfect for particular coarsenings R' of R. These R' are such that no perfect partitions exist for any preference profile that covers R'.

Theorem 1. *Let $(N, R^\top)$ and $(N, R^\perp)$ be hedonic games such that $R^\perp \leq R^\top$ and π a perfect partition for $R^\perp$. Then, π is Pareto optimal for $R^\top$ if and only if there is some $R \in [R^\perp, R^\top]$ such that (i) π is a perfect partition for R and (ii) there is no perfect partition for any $R' \in [R^\perp, R^\top]$ that covers R.*

Proof. For the if-direction, assume there is some $R \in [R^{\perp}, R^{\top}]$ such that π is perfect for R and there is no perfect partition for any $R' \in [R^{\perp}, R^{\top}]$ that covers R. For contradiction, also assume π is not Pareto optimal for $R^{\top}$. Then, there is some π' such that $\pi' \, R^{\top}_j \, \pi$ for all j and $\pi' \, P^{\top}_i \, \pi$ for some i. By $R \leq R^{\top}$ and π being perfect for R, it follows that π' is a perfect partition for R as well. Hence, $\pi' \, I_i \, \pi$. It follows that there is some $R' = (R_1, \ldots, R_{i-1}, R'_i, R_{i+1}, \ldots, R_n)$ in $([R_{\perp}, R^{\top}], \leq)$ that covers R. Also observe that, because R'_i is the unique minimal refinement of R_i such that $R_i <_i R'_i$, and $\pi' \, P^{\top}_i \, \pi$ even if $\pi' \, I_i \, \pi$, π' is still perfect for R', a contradiction.

For the only-if direction assume that π is Pareto optimal for $R^{\top}$. Let R be the finest coarsening of $R^{\top}$ in $[R^{\top}, R^{\perp}]$ for which π is perfect. Observe that $R = (R_1, \ldots, R_n)$ can be defined such that $R_i = R^{\top}_i \cup \{(X, Y) : X \, R^{\top}_i \, \pi \text{ and } Y \, R^{\top}_i \, \pi\}$ for all i. Since π is perfect for $R^{\perp}$, we have $R^{\perp} \leq R$. If $R = R^{\top}$, we are done immediately. Otherwise, consider an arbitrary $R' \in [R^{\perp}, R^{\top}]$ that covers R and assume for contradiction that some perfect partition π' exists for R'. Then, in particular, $\pi' \, R'_k \, \pi$ for all k. Since R' covers R, there is exactly one i with $R'_i \neq R_i$, whereas $R'_j = R_j$ for all $j \neq i$. As π is perfect for R, we also have $\pi \, R'_j \, \pi'$ for all $j \neq i$. Since R' is a finer coarsening of $R^{\top}$ than R, π is not perfect for R' by assumption. It follows that $\pi' \, P'_i \, \pi$. Hence, π is not Pareto optimal for R'. As $R' \leq R^{\top}$, we may conclude that π is not Pareto optimal for $R^{\top}$, a contradiction. □

4 The Preference Refinement Algorithm

In this section, we present the *Preference Refinement Algorithm (PRA)*, a general algorithm to compute Pareto optimal and individually rational partitions. The algorithm invokes an oracle solving `PerfectPartition` and is based on the formal connection between Pareto optimality and perfection made explicit in Theorem 1. We define `PerfectPartition` to return $\emptyset$ if $R_i = \emptyset$ for some i.

The idea underlying the algorithm is as follows. To calculate a Pareto optimal and individually rational partition for a hedonic game (N, R), first find that coarsening R' of R in which each player is indifferent among all his acceptable coalitions and his preferences among unacceptable coalitions are as in R. In this coarsening, a perfect and individually rational partition is guaranteed to exist. Then, we search the lattice $([R', R], \leq)$ for a preference profile that allows for a perfect partition but none of the profiles covering it do. By virtue of Theorem 1, every perfect partition for such a preference profile will be a Pareto optimal partition for R. By only refining the preferences of one player at a time, we can use divide-and-conquer to conduct the search. A formal specification of PRA is given in Algorithm 1. $\texttt{Refine}(Q^{\perp}_i, Q^{\top}_i)$ returns a refinement $Q'_i \in (Q^{\perp}_i, Q^{\top}_i]$, i.e., Q'_i is a refinement of $Q^{\perp}_i$ but not a refinement of $Q^{\top}_i$. $\texttt{Refine}(Q^{\perp}_i, Q^{\top}_i)$ can be defined in at least three fundamental ways:

(*i*) $\texttt{Refine}(Q^{\perp}_i, Q^{\top}_i) = Q'_i$ such that the number of refinements from $Q^{\perp}_i$ to Q'_i is half of the number of refinements from $Q^{\perp}_i$ to $Q^{\top}_i$ (default divide-and-conquer setting);
(*ii*) $\texttt{Refine}(Q^{\perp}_i, Q^{\top}_i) = Q^{\top}_i$ (random dictatorship setting); and
(*iii*) $\texttt{Refine}(Q^{\perp}_i, Q^{\top}_i) = \texttt{Cover}(Q^{\perp}_i)$.

The following theorem shows the correctness and completeness of PRA.

Algorithm 1 Preference Refinement Algorithm (PRA)

Input: Hedonic game (N, R)
Output: Pareto optimal and individually rational partition

1 $Q_i^{\top} \leftarrow R_i$, for each $i \in N$
2 $Q_i^{\perp} \leftarrow R_i \cup \{(X, Y) : X\, R_i\, \{i\}$ and $Y\, R_i\, \{i\}\}$, for each $i \in N$
3 $J \leftarrow N$
4 **while** $J \neq \emptyset$ **do**
5     $i \in J$
6   **if** $\texttt{PerfectPartition}(N, (Q_1^{\perp}, \ldots, Q_{i-1}^{\perp}, \texttt{Cover}(Q_i^{\perp}), Q_{i+1}^{\perp}, \ldots, Q_n^{\perp})) = \emptyset$ **then**
7         $J \leftarrow J \setminus \{i\}$
8   **else**
9         $Q_i' \leftarrow \texttt{Refine}(Q_i^{\perp}, Q_i^{\top})$
10   **if** $\texttt{PerfectPartition}(N, (Q_1^{\perp}, \ldots, Q_{i-1}^{\perp}, Q_i', Q_{i+1}^{\perp}, \ldots, Q_n^{\perp})) \neq \emptyset$ **then**
11             $Q_i^{\perp} \leftarrow Q_i'$
12   **else**
13             $Q_i^{\top} \leftarrow Q_i''$ where $\texttt{Cover}(Q_i'') = Q_i'$
14   **end if**
15   **end if**
16 **end while**
17 **return** $\texttt{PerfectPartition}(N, Q^{\perp})$

Theorem 2. *For any hedonic game* (N, R)*,*

(i) *PRA returns an individually rational and Pareto optimal partition.*
(ii) *For every individually rational and Pareto optimal partition* π'*, there is an execution of PRA that returns a partition* π *such that* $\pi\, I_i\, \pi'$ *for all* i *in* N*.*

Proof. For (i), we prove that during an execution of PRA, for each assignment of $Q^{\perp}$, there exists a perfect partition π for that assignment. This claim certainly holds for the first assignment of $Q^{\perp}$, the coarsest acceptable coarsening of R. Furthermore, $Q^{\perp}$ is only refined (Step 9) if there exists a perfect partition for a refinement of $Q^{\perp}$. Let Q^* be the final assignment of $Q^{\perp}$. Then, we argue that the partition π returned by PRA is Pareto optimal and individually rational. By Theorem 1, if π were not Pareto optimal, there would exist a covering of Q^* for which a perfect partition still exists and Q^* would not be the final assignment of $Q^{\perp}$. Since, each player at least gets one of his acceptable coalitions, π is also individually rational.

For (ii), first observe that, by Theorem 1, for each Pareto optimal and individually rational partition π for a preference profile R there is some coarsening Q^* of R where π is perfect and no perfect partitions exist for any covering of Q^*. By individual rationality of π, it follows that Q^* is a refinement of the initial assignment of $Q^{\perp}$. An appropriate number of coverings of the initial assignment of $Q^{\perp}$ with respect to each player results in a final assignment Q^* of $Q^{\perp}$. The perfect partition for Q^* that is returned by PRA is then such that $\pi\, I_i\, \pi'$ for all i in N. □

We now specify the conditions under which PRA runs in polynomial time.

Lemma 2. *For any class of hedonic games for which any coarsening and* `PerfectPartition` *can be computed in polynomial time, PRA runs in polynomial time.*

Furthermore, if for a given preference profile R and partition π, a coarsening of R for which π is perfect can be computed in polynomial time, it can also be verified in polynomial time whether π is Pareto optimal.

Proof (Sketch). Under the given conditions, we prove that PRA runs in polynomial time. We first prove that the while-loop in PRA iterates a polynomial number of times. In each iteration of the while-loop, either a player i which cannot be further improved is removed from J (Step 7) or we enter the first else condition. In the first else, either $Q^{\perp}$ is set to Q_i' or $Q_i^{\top}$ is set to Q_i'' where $\texttt{Cover}(Q_i'') = Q_i'$. In either case, we discard from future consideration, half of the refinements of $Q_i^{\perp}$ due to the default divide-and-conquer definition of `Refine` in order to find a suitable refinement of the current $Q^{\perp}$ with respect to i. Therefore, even if the representation of (N, R) may be such that each player differentiates between an exponential number of coalitions, divide-and-conquer ensures that PRA iterates a polynomial number of times. As the crucial subroutine `PerfectPartition` takes polynomial time, PRA runs in polynomial time.

For the second part of the lemma, we run PRA to find a Pareto optimal partition that Pareto dominates π. We therefore modify Step 2 by setting $Q_i^{\perp}$ to a coarsening of R for which π is a perfect partition. Since such a coarsening can be computed in polynomial time as stated by the condition in the lemma, Step 2 takes polynomial time. Since an initial perfect partition exists for $Q_i^{\perp}$, we run PRA as usual after Step 2. □

PRA applies not only to general hedonic games but to many natural classes of hedonic games in which equivalence classes (of possibly exponentially many coalitions) for each player are implicitly defined.[2] In fact PRA runs in polynomial time even if there are an exponential number of equivalence classes. Note that the lattice $[Q^{\perp}, R]$ can be of exponential height and doubly-exponential width. PRA traverses though this lattice in an orderly way to compute a Pareto optimal partition.

Serial dictatorship is a well-studied mechanism in resource allocation, in which an arbitrary player is chosen as the 'dictator' who is then given his most favored allocation and the process is repeated until all players or resources have been dealt with. In the context of coalition formation, *serial dictatorship* is well-defined only if in every iteration, the dictator has a *unique* most preferred coalition.

Proposition 1. *For general hedonic games, W-hedonic games, and roommate games, a Pareto optimal partition can be computed in polynomial time when preferences are strict.*

Proposition 1 follows from the application of serial dictatorship to hedonic games with strict preferences over the coalitions. If the preferences over coalitions are not strict, then the decision to assign one of the favorite coalitions to the dictator may be

[2] For example, in W-hedonic games, $\max_{R_i}(N)$ specifies the set of favorite players of player i but can also implicitly represent all those coalitions S such that the least preferred player in S is also a favorite player for i.

sub-optimal. Even if players expresses strict preferences over other players, serial dictatorship may not work if the preferences induced over coalitions admit ties.

We see that if serial dictatorship works properly and efficiently in some setting, then so can PRA by simulating serial dictatorship. If in each iteration in Algorithm 1, the same player is chosen in Step 5 (until it is deleted from J) and $Q_i^{\top}$ is chosen in Step 9, then PRA can simulate serial dictatorship. Therefore PRA can also achieve the positive results of Proposition 1.

PRA has another advantage over serial dictatorship. Abdulkadiroğlu and Sönmez [1] showed that in the case of strict preferences and house allocation settings, every Pareto optimal allocation can be achieved by serial dictatorship. In the case of coalition formation, however, it is easy to construct a four-player hedonic game with strict preferences for which there is a Pareto optimal partition that serial dictatorship cannot return.

5 Computational results

In this section, we consider the problem of Verification (verifying whether a given partition is Pareto optimal) and Computation (computing a Pareto optimal partition) for the classes and representations of hedonic games mentioned in the preliminaries.

5.1 General hedonic games

As shown in Proposition 1, Pareto optimal partitions can be found efficiently for general hedonic games with strict preferences. If preferences are not strict, the problem turns out to be NP-hard. We prove this statement by utilizing Lemma 1 and showing that `PerfectPartition` is NP-hard by a reduction from ExactCoverBy3Sets (X3C).

Theorem 3. *For a general hedonic game, computing a Pareto optimal partition is NP-hard even when each player has a maximum of four acceptable coalitions and the maximum size of each coalition is three.*

Interestingly, *verifying* Pareto optimality is coNP-complete even for strict preferences.[3]

Theorem 4. *For a general hedonic game, verifying whether a partition π is Pareto optimal and whether π is weakly Pareto optimal is coNP-complete even when preferences are strict and π consists of the grand coalition of all players.*

5.2 Roommate games

For the class of roommate games, we obtain more positive results.

Theorem 5. *For roommate games, an individually rational and Pareto optimal coalition can be computed in polynomial time.*

[3] Theorem 4 contrasts with the general observation that "*in the area of matching theory usually ties are 'responsible' for NP-completeness*" [5].

Proof (Sketch). We utilize Lemma 1. It is sufficient to show that `PerfectPartition` can be solved in time $O(n^3)$.

We say that $j \in F(i)$ if and only if j is a favorite player in i's preference list. Construct an undirected graph $G = (V, E)$ where $V = N \cup (N \times \{0\})$, $E = \{\{i, j\} : i \neq j \wedge i \in F(j) \wedge j \in F(j)\} \cup \{\{i, (i, 0)\} : i \in F(i)\}$.

Then the claim is that there exists a perfect partition for (N, R) if and only if there exists a matching of size n in graph G. It is clear that in a matching of size n, each $v \in N$ is matched. If there exists a perfect partition, then each player in N is matched to a player $j \neq i$ such that $j \in F(i)$ or i is unmatched but $i \in F(i)$. In either case there exists a matching M in which i is matched. In the first case, i is matched to j in a matching M in G. In the second case, i is matched to $(i, 0)$.

Now assume that there exists a matching M of size n in G. Then, each $i \in N$ is matched to $j \neq i$ or $(i, 0)$. If i is matched to j, then we know $\{i, j\} \in E$ and therefore $j \in F(i)$. If i is matched to $(i, 0)$, then we know $\{i, (i, 0)\} \in E$ and therefore $i \in F(i)$. Thus, there exists a perfect partition. □

By utilizing the second part of Lemma 1, it can be seen that there exists an algorithm to compute a Pareto optimal improvement of a given roommate matching which takes time $O(n^3) \cdot O(n \log(n)) = O(n^4 \log(n))$. As a corollary we get the following.

Theorem 6. *For roommate games, it can be checked in polynomial time whether a partition is Pareto optimal.*

We can devise a tailor-made algorithm for roommate games which finds a Pareto optimal Pareto improvement of a given matching in $O(n^3)$—the same asymptotic complexity required by the algorithm of Morrill [11] for the restricted case of strict preferences.

5.3 W-hedonic games

We now turn to Pareto optimality in W-hedonic games.

Theorem 7. *For W-hedonic games, a partition that is both individually rational and Pareto optimal can be computed in polynomial time.*

Proof (sketch). The statement follows from Lemma 2 and the fact that `PerfectPartition` can be solved in polynomial time for W-hedonic games. The latter is proved by a polynomial-time reduction of `PerfectPartition` to a polynomial-time solvable problem called *clique packing*.

We first introduce the more general notion of graph packing. Let $\mathscr{F}$ be a set of undirected graphs. An $\mathscr{F}$-packing of a graph G is a subgraph H such that each component of H is (isomorphic to) a member of $\mathscr{F}$. The size of $\mathscr{F}$-packing H is $|V(H)|$. We will informally say that vertex i is *matched* by $\mathscr{F}$-packing H if i is in a connected component in H. Then, a maximum $\mathscr{F}$-packing of a graph G is one that matches the maximum number of vertices. It is easy to see that computing a maximum $\{K_2\}$-packing of a graph is equivalent to maximum cardinality matching. Hell and Kirkpatrick [10] and Cornuéjols et al. [8] independently proved that there is a polynomial-time algorithm to compute a maximum $\{K_2, \ldots, K_n\}$-packing of a graph. Cornuéjols et al. [8] note that finding a $\{K_2, \ldots, K_n\}$-packing can be reduced to finding a $\{K_2, K_3\}$-packing.

We are now in a position to reduce `PerfectPartition` for W-hedonic games to computing a maximum $\{K_2, K_3\}$-packing. For a W-hedonic game (N, R), construct a graph $G = (N \cup (N \times \{0, 1\}), E)$ such that $\{(i, 0), (i, 1)\} \in E$ for all $i \in N$; $\{i, j\} \in E$ if and only if $i \in \max_{R_j}(N)$ and $j \in \max_{R_i}(N)$ for $i, j \in N$ such that $i \neq j$; and $\{i, (i, 0)\}, \{i, (i, 1)\} \in E$ if and only if $i \in \max_{R_i}(N)$ for all $i \in N$. Let H be a maximum $\{K_2, K_3\}$-packing of G.

It can then be proved that there exists a perfect partition of N according to R if and only if $|V(H)| = 3|N|$. We omit the technical details due to space restrictions.

Since `PerfectPartition` for W-hedonic games reduces to checking whether graph G can be packed perfectly by elements in $\mathscr{F} = \{K_2, K_3\}$, we have a polynomial-time algorithm to solve `PerfectPartition` for W-hedonic games. Denote by $CC(H)$ the set of connected components of graph H. If $|V(H)| = 3|N|$ and a perfect partition does exist, then $\{V(S) \cap N : S \in CC(H)\} \setminus \{\emptyset\}$ is a perfect partition. □

Due to the second part of Lemma 2, the following is evident.

Theorem 8. *For W-hedonic games, it can be checked in polynomial time whether a given partition is Pareto optimal or weakly Pareto optimal.*

Our positive results for W-hedonic games also apply to hedonic games with $\mathscr{W}$-preferences.

5.4 B-hedonic games

We saw that for W-hedonic games, a Pareto optimal partition can be computed efficiently, even in the presence of unacceptable players. In the absence of unacceptable players, computing a Pareto optimal and individually rational partition is trivial in B-hedonic games, as the partition consisting of the grand coalition is a solution.

Interestingly, if preferences do allow for unacceptable players, the same problem becomes NP-hard. The statement is shown by a reduction from Sat.

Theorem 9. *For B-hedonic games, computing a Pareto optimal partition is NP-hard.*

By using similar techniques, the following can be proved.

Theorem 10. *For B-hedonic games, verifying whether a partition is weakly Pareto optimal is coNP-complete.*

We expect the previous result to also hold for Pareto optimality rather than weak Pareto optimality.

6 Conclusions

Pareto optimality and individual rationality are important requirements for desirable partitions in coalition formation. In this paper, we examined computational and structural issues related to Pareto optimality in various classes of hedonic games (see Table 1). We saw that unacceptability and ties are a major source of intractability when

Game	Verification	Computation
General	coNP-complete (Th. 4)	NP-hard (Th. 3)
General (strict)	coNP-complete (Th. 4)	in P (Prop. 1)
Roommate	in P (Th. 6)	in P (Th. 5)
B-hedonic	coNP-complete (Th. 10, weak PO)	NP-hard (Th. 9)
W-hedonic	in P (Th. 8)	in P (Th. 7)

Table 1. Complexity of Pareto optimality in hedonic games: positive results hold for both Pareto optimality and individual rationality.

computing Pareto optimal outcomes. In some cases, checking whether a given partition is Pareto optimal can be significantly harder than finding one.

It should be noted that most of our insights gained into Pareto optimality and the resulting algorithmic techniques—especially those presented in Section 3 and Section 4—do not only apply to coalition formation but to any discrete allocation setting.

References

1. A. Abdulkadiroğlu and T. Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–702, 1998.
2. C. Ballester. NP-completeness in hedonic games. *Games and Economic Behavior*, 49(1):1–30, 2004.
3. S. Banerjee, H. Konishi, and T. Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18:135–153, 2001.
4. A. Bogomolnaia and M. O. Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.
5. K. Cechlárová. On the complexity of exchange-stable roommates. *Discrete Applied Mathematics*, 116:279–287, 2002.
6. K. Cechlárová and J. Hajduková. Stable partitions with $\mathscr{W}$-preferences. *Discrete Applied Mathematics*, 138(3):333–347, 2004.
7. K. Cechlárová and J. Hajduková. Stability of partitions under $\mathscr{B}\mathscr{W}$-preferences and $\mathscr{W}\mathscr{B}$-preferences. *International Journal of Information Technology and Decision Making*, 3(4):605–618, 2004.
8. G. Cornuéjols, D. Hartvigsen, and W. Pulleyblank. Packing subgraphs in a graph. *Operations Research Letters*, 1(4):139–143, 1982.
9. J. H. Drèze and J. Greenberg. Hedonic coalitions: Optimality and stability. *Econometrica*, 48(4):987–1003, 1980.
10. P. Hell and D. G. Kirkpatrick. Packings by cliques and by finite families of graphs. *Discrete Mathematics*, 49(1):45–59, 1984.
11. T. Morrill. The roommates problem revisited. *Journal of Economic Theory*, 145 (5):1739–1756, 2010.
12. E. Ronn. NP-complete stable matching problems. *Journal of Algorithms*, 11:285–304, 1990.