Boundaries to Single-Agent Stability in Additively Separable Hedonic Games

Martin Bullinger
Technical University of Munich, Germany

Abstract

Coalition formation considers the question of how to partition a set of agents into coalitions with respect to their preferences. Additively separable hedonic games (ASHGs) are a dominant model where cardinal single-agent values are aggregated into preferences by taking sums. Output partitions are typically measured by means of stability, and we follow this approach by considering stability based on single-agent movements (to join other coalitions), where a coalition is defined as stable if there exists no beneficial single-agent deviation. Permissible deviations should always lead to an improvement for the deviator, but they may also be constrained by demanding the consent of agents involved in the deviations, i.e., by agents in the abandoned or welcoming coalition. Most of the existing research focuses on the unanimous consent of one or both of these coalitions, but more recent research relaxes this to majority-based consent. Our contribution is twofold. First, we settle the computational complexity of the existence of contractually Nash-stable partitions, where deviations are constrained by the unanimous consent of the abandoned coalition. This resolves the complexity of the last classical stability notion for ASHGs. Second, we identify clear boundaries to the tractability of stable partitions under majority-based stability concepts by proving elaborate hardness results for restricted classes of ASHGs. Slight further restrictions lead to positive results.

2012 ACM Subject Classification Computing methodologies → Multi-agent systems; Theory of computation → Design and analysis of algorithms

Keywords and phrases Coalition Formation, Hedonic Games, Stability

Digital Object Identifier 10.4230/LIPIcs.MFCS.2022.26

Funding *Martin Bullinger*: Deutsche Forschungsgemeinschaft, grants BR 2312/11-2, BR 2312/12-1.

Acknowledgements I would like to thank Felix Brandt and Leo Tappe for the helpful discussions.

1 Introduction

Coalition formation is a vibrant topic in multi-agent systems at the intersection of theoretical computer science and economic theory. Given a set of agents, e.g., humans or machines, the central concern is to determine a coalition structure, or partition, of the agents into subsets, or so-called coalitions. Agents have preferences over coalition structures, and therefore coalition formation naturally generalizes the matching problem under preferences [22]. As in the special case of matchings, a common assumption is that externalities outside one's own coalition play no role, i.e., agents are only concerned about the coalition they are part of. This assumption leads to the popular framework of hedonic games [18].

In contrast to matchings, the number of coalitions an agent can be part of is not polynomially bounded in coalition formation, and therefore, a lot of effort has been put into identifying reasonable and succinct classes of hedonic games (see, e.g., [2, 5, 8, 20]). In many such classes, agents extract cardinal preferences from a weighted and possibly directed graph by some aggregation method. Probably the most natural and thoroughly researched way to aggregate preferences is by taking the sum of the weights of edges towards agents in one's own coalition. This leads to the concept of additively separable hedonic games (ASHGs) [8]. This paper continues to investigate this class of hedonic games.

47th International Symposium on Mathematical Foundations of Computer Science (MFCS 2022).
Editors: Stefan Szeider, Robert Ganian, and Alexandra Silva; Article No. 26; pp. 26:1–26:15
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The desirability of an output, i.e., of a coalition structure, is frequently measured with respect to stability, which captures the prospect of agents maintaining their coalitions. A coalition structure is stable if no single agent or group of agents has an incentive to deviate by leaving their coalitions and joining other coalitions or forming new coalitions. Depending on the requirements that deviators need to meet, one can define various specific stability notions. In this paper, we focus on stability based on single-agent deviations. This means that a deviation consists of a single agent that abandons her current coalition to join another existing coalition or to form a new coalition of her own.

In this case, a reasonable minimum requirement is that a deviating agent should improve her coalition. If no such deviation is possible, then a coalition structure is said to be Nash-stable. However, this leads to an immensely strong stability concept because the deviation is only constrained weakly. As a consequence, Nash-stable outcomes hardly ever exist. For instance, consider a game with two agents x and y where x prefers to form a coalition with y over staying alone, whereas y prefers to stay alone. Then, x always has an incentive to join y whenever she is in a coalition of her own, whereas y would always leave x. Such run-and-chase situations occur in most classes of hedonic games.[1]

Therefore, various weakenings of Nash stability have been proposed. These restrict the possible deviations by adding further requirements on other agents involved in the deviation. Typically, two types of constraints are considered, namely the demanding of some kind of consent from the abandoned or the welcoming coalition. Most of the research has focused on the unanimous consent of these coalitions. This leads to the concepts of contractual Nash stability and individual stability where all agents in the abandoned or welcoming coalition have to approve the deviation. Still, unanimous consent of involved coalitions is a strong requirement. Hence, a reasonable compromise is to merely demand partial consent. Therefore, we also study stability where deviations are constrained by the approval of a majority vote of the abandoned or welcoming coalition.

1.1 Contribution

Our contribution is twofold. First, we settle the complexity of the existence problem of contractually Nash-stable coalition structures. Despite knowing for quite long that No-instances, i.e., additively separable hedonic games which do not admit a contractually Nash-stable coalition structure, exist [28], detailed computational investigations of single-agent stability during the last decade have left this problem open [10, 29]. Hence, we complete the picture of the complexity of unanimity-based single-agent stability concepts in ASHGs.

Second, we investigate majority-based stability concepts. We will show that, even under significant weight restrictions, stable coalition structures need not exist and we can leverage No-instances to obtain computational intractabilities. This complements very recent results by Brandt et al. [10] and resolves problems left open by this work. In particular, we completely pinpoint the complexity of majority-based stability notions in friends-and-enemies games and appreciation-of-friends games.

These results are in line with the repeatedly observed theme in hedonic games research that the existence of counterexamples is the key to computational intractabilities (see, e.g., [3, 10, 11, 16, 29]).[2] On the other hand, we demonstrate that the observed intractabilities lie at the computational boundary by carving out further weak restrictions that lead to the existence and efficient computability of stable states.

[1] Notably, Nash-stable coalition structures always exist in ASHGs if the input graph is symmetric [8], and in a generalization of this class of games called subset-neutral hedonic games [27].

[2] A notable exception is provided by Bullinger and Kober [13] who identify a class of hedonic games where partitions in the core always exist, but are still hard to compute.

1.2 Related Work

The study of hedonic games was initiated by Drèze and Greenberg [18] but was only popularized two decades later by Banerjee et al. [6], Cechlárová and Romero-Medina [15], and Bogomolnaia and Jackson [8]. Aziz and Savani [4] review many important concepts in their survey. Two important research questions concern the design of reasonable computationally manageable subclasses of hedonic games and the detailed investigation of their computational properties. The former has led to a broad landscape of game representations. Some of these representations [5, 20] are ordinal and fully expressive, i.e., they can, in principle, express every preference relation over coalitions. Still, representing certain preference relations requires exponential space. These representations are contrasted by cardinal representations based on weighted graphs [2, 8, 26], which are not fully expressive but only require polynomial space (except when weights are artificially large). Apart from the already discussed additively separable hedonic games, important aggregation methods consider the average of weights leading to the classes fractional hedonic games [2] and modified fractional hedonic games [26]. Additively separable hedonic games have important subclasses where the focus lies in distinguishing friends and enemies, and therefore only two different weights are present in the underlying graph [16].

The computational properties of hedonic games have been extensively studied and we focus on literature related to additively separable hedonic games. Various versions of stability have been investigated [1, 3, 10, 16, 29, 21]. The closest to our work are the detailed studies of single-agent stability by Sung and Dimitrov [29] and Brandt et al. [10]. Gairing and Savani [21] settle the complexity of single-agent stability for symmetric input graphs. Majority-based stability has only received little attention thus far [10, 21]. Apart from stability, other desirable axioms concern efficiency and fairness. Aziz et al. [3] cover a wide range of axioms, whereas Elkind et al. [19] and Bullinger [12] focus on Pareto optimality, and Brandt and Bullinger [9] investigate popularity, an axiom combining ideas from stability and efficiency which is also related to certain majority-based stability notions [10]. Finally, a recent trend in the research on coalition formation is to complement the static view of existence problems by considering dynamics based on stability concepts (see, e.g., [7, 10, 11, 14, 23]).

2 Preliminaries

In this section, we formally introduce hedonic games and our considered stability concepts.

2.1 Hedonic Games

Let $N = [n]$ be a set of $n \in \mathbb{N}$ agents, where we define $[n] = \{1, \dots, n\}$. The output of a coalition formation problem is a coalition structure, that is, a partition of the agents into different disjoint coalitions according to their preferences. A *partition* of N is a subset $\pi \subseteq 2^N$ such that $\bigcup_{C\in\pi} C = N$, and for every pair $C, D \in \pi$, it holds that $C = D$ or $C \cap D = \emptyset$. An element of a partition is called a *coalition* and, given a partition π, the unique coalition containing agent i is denoted by $\pi(i)$. We refer to the partition π given by $\pi(i) = \{i\}$ for every agent $i \in N$ as the *singleton partition*, and to $\pi = \{N\}$ as the *grand coalition*.

Let $\mathcal{N}_i$ denote all possible coalitions containing agent i, i.e., $\mathcal{N}_i = \{C \subseteq N : i \in C\}$. A *hedonic game* is a tuple $(N, \succsim)$, where N is an agent set and $\succsim = (\succsim_i)_{i\in N}$ is a tuple of weak orders $\succsim_i$ over $\mathcal{N}_i$ representing the preferences of the respective agent i. Hence, as mentioned before, agents express preferences only over the coalitions of which they are part without considering externalities. The strict part of an order $\succsim_i$ is denoted by $\succ_i$, i.e., $C \succ_i D$ if and only if $C \succsim_i D$ and not $D \succsim_i C$.

Additively separable hedonic games assume that every agent is equipped with a cardinal utility function that is aggregated by taking the sum of single-agent values. Formally, following [8], an *additively separable hedonic game* (ASHG) (N, v) consists of an agent set N and a tuple $v = (v_i)_{i \in N}$ of utility functions $v_i \colon N \to \mathbb{R}$ such that $\pi \succsim_i \pi'$ if and only if $\sum_{j \in \pi(i)} v_i(j) \geq \sum_{j \in \pi'(i)} v_i(j)$. Clearly, ASHGs are a subclass of hedonic games. When we specify ASHG utilities, we neglect, without loss of generality, $v_i(i)$ because the preferences do not depend on it and we implicitly assume that it is set to an appropriate constant if an ASHG has to fit into a certain subclass of games.

Every ASHG can be naturally represented by a complete directed graph $G = (N, E)$ with weight $v_i(j)$ on arc (i, j). There are various subclasses of ASHGs that allow a natural interpretation in terms of friends and enemies. An agent $j \in N$ is called a *friend* (or *enemy*) of agent $i \in N$ if $v_i(j) > 0$ (or $v_i(j) < 0$). An ASHG is called a *friends-and-enemies game* (FEG) if $v_i(j) \in \{-1, 1\}$ for every pair of agents $i, j \in N$ [10]. Further, following [16], an ASHG is called an *appreciation-of-friends game* (AFG) (or an *aversion-to-enemies game* (AEG)) if $v_i(j) \in \{-1, n\}$ (or $v_i(j) \in \{-n, 1\}$). In such games, agents seek to maximize their number of friends while minimizing their number of enemies, where these goals have a different priority in each case. Based on the friendship of agents, we define the *friendship relation* (or *enemy relation*) as the subset $R \subseteq N \times N$ where $(i, j) \in R$ if and only if $v_i(j) > 0$ (or $v_i(j) < 0$).

2.2 Single-Agent Stability

We want to study stability under single agents' incentives to perform deviations. A *single-agent deviation* performed by agent i transforms a partition π into a partition π' where $\pi(i) \neq \pi'(i)$ and, for all agents $j \neq i$,

$$\pi'(j) = \begin{cases} \pi(j) \setminus \{i\} & \text{if } j \in \pi(i), \\ \pi(j) \cup \{i\} & \text{if } j \in \pi'(i), \text{ and} \\ \pi(j) & \text{otherwise.} \end{cases}$$

We write $\pi \xrightarrow{i} \pi'$ to denote a single-agent deviation performed by agent i transforming partition π to partition π'.

We consider myopic agents whose rationale is to only engage in a deviation if it immediately makes them better off. A *Nash deviation* is a single-agent deviation performed by agent i making her better off, i.e., $\pi'(i) \succ_i \pi(i)$. Any partition in which no Nash deviation is possible is said to be *Nash-stable* (NS).

Following [10], we introduce consent-based stability concepts via favor sets. Let $C \subseteq N$ be a coalition and $i \in N$ an agent. The *favor-in set* of C with respect to i is the set of agents in C (excluding i) that strictly favor having i inside C rather than outside, i.e., $F_{\text{in}}(C, i) = \{j \in C \setminus \{i\} \colon C \cup \{i\} \succ_j C \setminus \{i\}\}$. The *favor-out set* of C with respect to i is the set of agents in C (excluding i) that strictly favor having i outside C rather than inside, i.e., $F_{\text{out}}(C, i) = \{j \in C \setminus \{i\} \colon C \setminus \{i\} \succ_j C \cup \{i\}\}$.

An *individual deviation* (or *contractual deviation*) is a Nash deviation $\pi \xrightarrow{i} \pi'$ such that $F_{\text{out}}(\pi'(i), i) = \emptyset$ (or $F_{\text{in}}(\pi(i), i) = \emptyset$). Then, a partition is said to be *individually stable* (IS) or *contractually Nash-stable* (CNS) if it allows for no individual or contractual deviation, respectively. A related weakening of both stability concepts is contractual individual stability (CIS), based on deviations that are both individual and contractual deviations [8, 17].

Finally, we define hybrid stability concepts according to [10] where the consent of the abandoned or welcoming coalition is decided by a majority vote. A Nash deviation $\pi \xrightarrow{i} \pi'$ is

$$
\begin{tikzcd}
 & \text{Majority-Out Stability} \arrow[r] & \text{Contractual Nash Stability} \arrow[dr] & \\
\text{Nash Stability} \arrow[ur] \arrow[dr] & & & \text{Contractual Individual Stability} \\
 & \text{Majority-In Stability} \arrow[r] & \text{Individual Stability} \arrow[ur] &
\end{tikzcd}
$$

Figure 1 Logical relationships between stability notions. An arrow from concept S to concept S' indicates that if a partition satisfies S, it also satisfies S'. Conversely, this means that every S' deviation is also an S deviation.

called a *majority-in deviation* (or *majority-out deviation*) if $|F_{\text{in}}(\pi'(i), i)| \geq |F_{\text{out}}(\pi'(i), i)|$ (or $|F_{\text{out}}(\pi(i), i)| \geq |F_{\text{in}}(\pi(i), i)|$). Similar to before, a partition is said to be *majority-in stable* (MIS) or *majority-out stable* (MOS) if it allows for no majority-in or majority-out deviation, respectively. The concepts MIS and MOS are special cases of the voting-based stability notions by Gairing and Savani [21] for a threshold of 1/2. Brandt et al. [10] also consider stability concepts that require voting-based consent by both the abandoned and welcoming coalition, similar to CIS.

For a stability concept $S \in \{\text{NS}, \text{IS}, \text{CNS}, \text{MIS}, \text{MOS}\}$, we denote the deviation corresponding to S as S *deviation*, e.g., CNS deviation for a contractual deviation. A taxonomy of our related solution concepts is provided in Figure 1.

3 Contractual Nash Stability

Our first result settles the computational complexity of contractual Nash stability in ASHGs. All of our reductions in this and the subsequent sections are from the NP-complete problem Exact3Cover (E3C) [25]. An instance of E3C consists of a tuple (R, S), where R is a ground set together with a set S of 3-element subsets of R. A Yes-instance is an instance such that there exists a subset $S' \subseteq S$ that partitions R.

Before giving the complete proof, we briefly describe the key ideas. Given an instance (R, S) of E3C, the reduced instance consists of three types of gadgets. First, every element in R is represented by a subgame that does not contain a CNS partition. In principle, any such game can be used for a reduction, and we use the game identified by Sung and Dimitrov [28]. Moreover, we have further auxiliary gadgets that also consist of the same No-instance. The number of these auxiliary gadgets is equal to the number of sets in S that would remain after removing an exact cover of R, i.e., there are $|S| - |R|/3$ such gadgets. By design, the agents in the subgames corresponding to No-instances have to form coalitions with agents outside of their subgame in every CNS partition. The only agents that can achieve this are agents in gadgets corresponding to elements in S. A gadget corresponding to an element $s \in S$ can either prevent non-stability caused by exactly one auxiliary gadget, or by the three gadgets corresponding to the elements $r \in R$ with $r \in s$. Hence, the only possibility to deal with all No-instances simultaneously is if there exists an exact cover of R by sets in S. Then, the gadgets corresponding to elements in R can be dealt with by the cover and there are just enough elements in S to additionally deal with the other auxiliary gadgets.

▶ **Theorem 1.** *Deciding whether an ASHG contains a CNS partition is NP-complete.*

Proof. We provide a reduction from E3C. Let (R, S) be an instance of E3C and set $a = |S| - |R|/3$ (this is the number of additional sets in S if removing some exact cover). Without

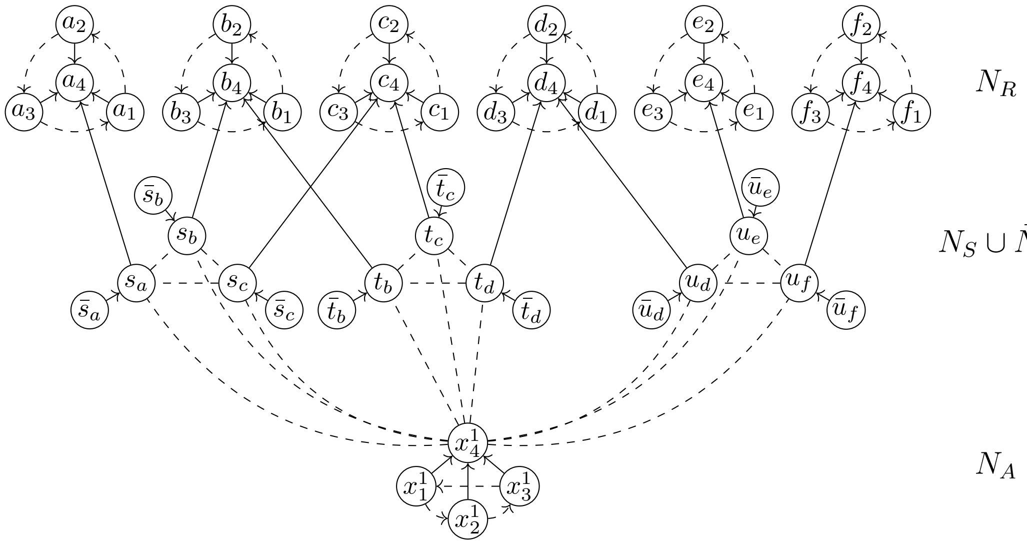

Figure 2 Schematic of the reduction from the proof of Theorem 1. We depict the reduced instance for the instance (R, S) of E3C where $R = \{a, b, c, d, e, f\}$, and $S = \{s, t, u\}$, with $s = \{a, b, c\}$, $t = \{b, c, d\}$, and $u = \{d, e, f\}$. Fully drawn edges mean a positive utility, which is usually 1 except between agents of the types $\bar{s}_r$ and s_r, where $v_{\bar{s}_r}(s_r) = 3$. Dashed edges represent a utility of 0. For agents in $\bar{N}_S$, only the single positive utility is displayed. Other omitted edges represent a negative utility of -4.

loss of generality, $a \geq 0$. We define an ASHG (N, v) as follows. Let $N = N_R \cup N_S \cup \bar{N}_S \cup N_A$ where

- $N_R = \cup_{r \in R} N_r$ with $N_r = \{r_i : i \in [4]\}$ for $r \in R$,
- $N_S = \cup_{s \in S} N_s$ with $N_s = \{s_r : r \in s\}$ for $s \in S$,
- $\bar{N}_S = \cup_{s \in S} \bar{N}_s$ with $\bar{N}_s = \{\bar{s}_r : r \in s\}$ for $s \in S$, and
- $N_A = \cup_{1 \leq j \leq a} N^j$ with $N^j = \{x_i^j : i \in [4]\}$ for $1 \leq j \leq a$.

We define valuations v as follows:

- For each $r \in R$, $i \in [3]$: $v_{r_i}(r_4) = 1$.
- For each $r \in R$, $(i, j) \in (1, 2), (2, 3), (3, 1)$: $v_{r_i}(r_j) = 0$.
- For each $1 \leq j \leq a$, $i \in [3]$: $v_{x_i^j}(x_4^j) = 1$.
- For each $1 \leq j \leq a$, $(i, k) \in (1, 2), (2, 3), (3, 1)$: $v_{x_i^j}(x_k^j) = 0$.
- For each $s \in S$, $r \in s$: $v_{s_r}(r_4) = 1$.
- For each $s \in S$, $r \in s$, $1 \leq j \leq a$: $v_{s_r}(x_4^j) = v_{x_4^j}(s_r) = 0$.
- For each $s \in S$, $r, r' \in s$: $v_{s_r}(s_{r'}) = 0$.
- For each $s \in S$, $r, r' \in s$, $r \neq r'$, $z \in (N_S \cup N_A) \setminus N_s$: $v_{\bar{s}_r}(s_r) = 3$, $v_{\bar{s}_r}(s_{r'}) = -2$, and $v_{\bar{s}_r}(z) = 0$.
- All other valuations are -4.

An illustration of the game is given in Figure 2. The agents in N_R in the reduced instance form gadgets consisting of a subgame without CNS partition for every element in R. The agents in N_A constitute further such gadgets. The agents in N_S consist of triangles for every set in S and are the only agents who can bind agents in the gadgets in any CNS partition. Finally, agents in $\bar{N}_S$ avoid having agents in N_S in separate coalitions to bind agents in N_A.

We claim that (R, S) is a Yes-instance if and only if (N, v) contains a CNS partition. Suppose first that $S' \subseteq S$ partitions R. Consider any bijection $\phi\colon S \setminus S' \to [a]$. Define a partition π by taking the union of the following coalitions:

- For every $r \in R, i \in [3]$, form $\{r_i\}$.
- For $s \in S', r \in s$, form $\{s_r, r_4\}$.
- For $s \in S \setminus S'$, form $\{s_r : r \in s\} \cup \{x_4^{\phi(s)}\}$.
- For $s \in S, r \in s$, form $\{\bar{s}_r\}$.
- For $1 \leq j \leq a, i \in [3]$, form $\{x_i^j\}$.

We claim that π is CNS. We will show that no agent can perform a deviation.

- For $r \in R$, $i \in [3]$, it holds that $v_{r_i}(\pi) = 0$ and joining any other coalition results in a negative utility. In particular, $v_{r_i}(\pi(r_4) \cup \{r_i\}) = -3$.
- For $r \in R$, r_4 is not allowed to leave her coalition.
- For $s \in S'$, $r \in s$, it holds that $v_{s_r}(\pi) = 1$ and joining any other coalition results in a negative utility. The agent s_r is in a most preferred coalition.
- For $s \in S \setminus S'$, $r \in s$, it holds that $v_{s_r}(\pi) = 0$ and joining any other coalition results in a negative utility. In particular, $v_{s_r}(\pi(r_4) \cup \{s_r\}) = -3$.
- For $s \in S'$, $r \in s$, the agent $\bar{s}_r$ obtains a non-positive utility by joining any other coalition. In particular, $v_{\bar{s}_r}(\pi(s_r) \cup \{\bar{s}_r\}) = -1$.
- For $s \in S \setminus S'$, $r \in s$, the agent $\bar{s}_r$ obtains a non-positive utility by joining any other coalition. In particular, $v_{\bar{s}_r}(\pi(s_r) \cup \{\bar{s}_r\}) = -1$.
- For $1 \leq j \leq a$, $i \in [3]$, it holds that $v_{x_i^j}(\pi) = 0$ and joining any other coalition results in a negative utility. In particular, $v_{x_i^j}(\pi(x_4^j) \cup \{x_i^j\}) = -11$.
- For $1 \leq j \leq a$, x_4^j is in a best possible coalition (achieving utility 0).

Conversely, assume that (N, v) contains a CNS partition π. Define $S' = \{s \in S\colon \pi(s_r) \cap N_R \neq \emptyset \text{ for some } r \in s\}$. We will show first that S' covers all elements in R and then show that $|S'| = |R|/3$.

Let $r \in R$. Then, for all $i \in [3]$, $\pi(r_i) \subseteq N_r$. This follows because there is no agent who favors r_i in her coalition. Therefore, she would leave any coalition with an agent outside N_r to receive non-negative utility in a singleton coalition. Further, if there is no $s \in S$ with $r \in s$ such that $r_4 \in \pi(r_s)$, then $\pi(r_4) \subseteq N_r$. Indeed, if r_4 forms any coalition except a singleton coalition, she will receive negative utility, and then there must exist an agent who favors her in the coalition. Consequently, if $r_4 \notin \pi(r_s)$ for all $s \in S$ with $r \in s$, then r_4 is in a singleton coalition, or there exists $i \in [3]$ with $r_4 \in \pi(r_i)$, for which we already know that $\pi(r_i) \subseteq N_r$.

Assume now that $\pi(r_4) \subseteq N_r$. For $i, i' \in [3]$, $r_i \notin \pi(r_{i'})$ because then one of them would receive a negative utility and could perform a CNS deviation to form a singleton coalition. If $\{r_4\} \in \pi$, then r_1 would deviate to join her. Hence, there exists exactly one $i \in [3]$ with $\{r_i, r_4\} \in \pi$. Suppose without loss of generality that $\{r_1, r_4\} \in \pi$. But then, r_3 would perform a CNS deviation to join them, a contradiction. We can conclude that there exists $s \in S$ with $r \in s$ such that $r_4 \in \pi(r_s)$. Hence, $s \in S'$ and we have shown that S' covers R.

To bound the cardinality of S', we will show that, for every $1 \leq j \leq a$, there exists $s \in S \setminus S'$ with $N_s \subseteq \pi(x_4^j)$. Let therefore $1 \leq j \leq a$ and let $C = \pi(x_4^j)$. Similar to the considerations about agents in N_r, we know that $\pi(x_i^j) \subseteq X^j$ for $i \in [3]$, and that it cannot happen that $C \subseteq X^j$, and therefore $C \cap X^j = \{x_4^j\}$. In particular, there must be an agent $y \in N \setminus X^j$ with $y \in C$. Since no agent in C favors x_4^j to be in her coalition, we know that $v_{x_4^j}(\pi) \geq 0$ and therefore $C \subseteq \{x_4^j\} \cup N_S$. Let $s \in S$ and $r \in s$ with $s_r \in C$. As we already know that $\bar{s}_r \notin C$, it must hold that $N_s \subseteq C$ to prevent her from joining. It follows that $s \notin S'$. Since $\pi(x_4^j) \cap \pi(x_4^{j'}) = \emptyset$ for $1 \leq j' \leq a$ with $j' \neq j$, we find an injective mapping $\phi\colon [a] \to S \setminus S'$ such that, for every $1 \leq j \leq a$, $N_{\phi(j)} \subseteq \pi(x_4^j)$. Consequently, $|S'| \leq |S| - |\phi([a])| \leq |S| - a = |R|/3$. Hence, S' covers all elements from R with (at most) $|R|/3$ sets and therefore is an exact cover. ◀

The reduction in the previous proof only uses a very limited number of different weights, namely the weights in the set $\{1, 0, -2, -4\}$, where the weight -4 may be replaced by an arbitrary smaller weight. By contrast, CNS partitions always exist if the utility functions of an ASHG assume at most one nonpositive value, and can be computed efficiently in this case [10, Theorem 4]. This encompasses for instance FEGs, AFGs, and AEGs. Hence, the hardness result is close to the boundary of computational feasibility.

4 Appreciation-of-Friends Games

In this section, we consider appreciation-of-friends games. Typically, these games behave well with respect to stability. In particular, IS, CNS, and MIS partitions always exist and can be computed efficiently, while it is only known that NS leads to non-existence and computational hardness among single-agent stability concepts [10, 16]. By contrast, we show in our next result that MOS partitions need not exist in AFGs. In other words, despite their conceptual complementarity, the stability concepts MOS and MIS lead to very different behavior in a natural class of ASHGs. The constructed game has a sparse friendship relation in the sense that almost all agents only have a single friend. After discussing the counterexample, we show how requiring slightly more sparsity yields a positive result. Due to space restrictions, some proofs are omitted or sketched.

▶ **Proposition 2.** *There exists an AFG without an MOS partition.*

Proof. We define the game formally. An illustration is given in Figure 3. Let $N = \{z\} \cup \bigcup_{x\in\{a,b,c\}} N_x$, where $N_x = \{x_i : i \in [5]\}$ for $x \in \{a, b, c\}$. In the whole proof, we read indices modulo 5, mapping to the respective representative in [5]. The utilities are given as:

- For all $i \in [5], x \in \{a, b, c\}$: $v_{x_i}(x_{i+1}) = n$.
- For all $x \in \{a, b, c\}$: $v_{x_1}(z) = n$.
- All other valuations are -1.

The AFG consists of 3 cycles with 5 agents each, together with a special agent that is liked by a fixed agent of each cycle and has no friends herself. The key insight to understanding why there exists no MOS partition is that agents of type x_1 where $x \in \{a, b, c\}$ have conflicting candidate coalitions in a potential MOS partition. Either, they want to be with z (a coalition that has to be small because z prefers to stay alone) or they want to be with x_2 which requires a rather large coalition containing their cycle.

Before going through the proof that this game has no MOS partition, it is instructional to verify that, for cycles of 5 agents, the unique MOS partition is the grand coalition, i.e., the unique MOS partition of the game restricted to N_x is $\{N_x\}$, where $x \in \{a, b, c\}$. This is a key idea of the construction and is implicitly shown in Case 2 of the proof for $x = b$.

Assume for contradiction that the defined AFG admits an MOS partition π. To derive a contradiction, we perform a case distinction over the coalition sizes of z.

Case 1. $|\pi(z)| = 1$.

In this case, it holds that $\pi(z) = \{z\}$. Then, $\pi(a_1) \in \{\{a_1, a_2\}, \{a_1, a_5\}\}$. Indeed, if $\pi(a_1) \neq \{a_1, a_2\}$, then a_1 has an NS deviation to join z, and is allowed to perform it unless $\pi(a_1) = \{a_1, a_5\}$. We may therefore assume that $\{a_i, a_{i+1}\} \in \pi$ for some $i \in \{1, 5\}$.

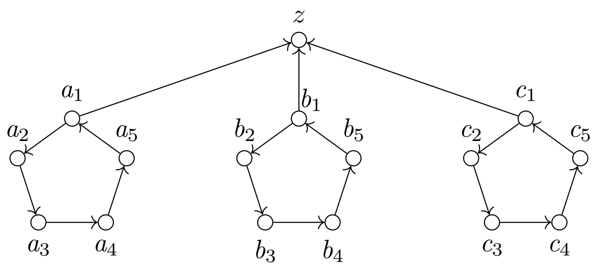

Figure 3 AFG without an MOS partition. The depicted (directed) edges represent friends, i.e., a utility of n, whereas missing edges represent a utility of -1.

Then, $\pi(a_{i-1}) = \{a_{i-1}, a_{i-2}\} =: C$. Otherwise, a_{i-1} can perform an MOS deviation to join $\{a_i, a_{i+1}\}$. But then a_{i+2} can perform an MOS deviation to join C. This is a contradiction and concludes the case that $|\pi(z)| = 1$.

Case 2. $|\pi(z)| > 1$.
Let $F := \{a_1, b_1, c_1\}$, i.e., the set of agents that have z as a friend. Note that z can perform an NS deviation to be a singleton. Hence, as π is MOS, $|F \cap \pi(v)| \geq |\pi(z)|/2$. In particular, there exists an $x \in \{a, b, c\}$ with $\pi(z) \cap N_x = \{x_1\}$. We may assume without loss of generality that $\pi(z) \cap N_a = \{a_1\}$. Then, $\pi(a_5) = \{a_4, a_5\}$. Otherwise, a_5 has an MOS deviation to join $\pi(z)$. Similarly, $\pi(a_3) = \{a_2, a_3\}$ (because of the potential deviation of a_3 who would like to join $\{a_4, a_5\}$). Now, note that $v_{a_1}(\{a_1, a_2, a_3\}) = n - 1$. We can conclude that $|\pi(z)| \leq 3$ as a_1 would join $\{a_2, a_3\}$ by an MOS deviation, otherwise. Hence, we find $x \in \{b, c\}$ with $N_x \cap \pi(z) = \emptyset$. Assume without loss of generality that $x = b$ has this property.

Assume first that $\pi(b_1) = \{b_1, b_5\}$. Then, $\pi(b_4) = \{b_3, b_4\}$. Otherwise, b_4 has an MOS deviation to join $\{b_1, b_5\}$. But then b_2 has an MOS deviation to join $\{b_3, b_4\}$, a contradiction. Hence, $\pi(b_1) \neq \{b_1, b_5\}$. Note that we have now excluded the only case where b_1 is not allowed to perform an NS deviation. In all other cases, no majority of agents prefers her to stay in the coalition. We can conclude that $b_2 \in \pi(b_1)$ because otherwise, b_1 can perform an MOS deviation to join $\pi(z)$. If $b_5 \notin \pi(b_1)$, then $\pi(b_5) = \{b_4, b_5\}$ (to prevent a potential deviation by b_5). But then b_3 has an MOS deviation to join them. Hence, $b_5 \in \pi(b_1)$. Similarly, if $b_4 \notin \pi(b_1)$, then $\pi(b_4) = \{b_3, b_4\}$ and b_2 has an MOS deviation to join $\{b_3, b_4\}$ (which is permissible because $b_5 \in \pi(b_1)$). Hence $\{b_1, b_2, b_4, b_5\} \subseteq \pi(b_1)$, and therefore even $N_b \subseteq \pi(b_1)$. Hence, b_1 has an MOS deviation to join $\pi(v)$ (recall that $|\pi(v)| \leq 3$). This is the final contradiction, and we can conclude that π is not MOS. ◀

Note that most agents in the previous example have at most 1 friend (only three agents have 2 friends). By contrast, if every agent has at most one friend, MOS partitions are guaranteed to exist. This is interesting because it covers in particular directed cycles, which cause problems for Nash stability. The constructive proof of the following proposition can be directly converted into a polynomial-time algorithm.

▶ **Proposition 3.** *Every AFG where every agent has at most one friend admits an MOS partition.*

Proof. We prove the statement by induction over n. Clearly, the grand coalition is MOS for $n = 1$. Now, assume that (N, v) is an AFG with $n \geq 2$ such that every agent has at most one friend. Consider the underlying directed graph $G = (N, A)$ where $(x, y) \in A$ if and only if $v_x(y) > 0$, i.e., y is a friend of x. By assumption, G has a maximum out-degree of 1, hence it can be decomposed into directed cycles and a directed acyclic graph.

Assume first that there exists $C \subseteq N$ such that C induces a directed cycle in G. We call an agent y *reachable* by agent x if there exists a directed path in G from x to y. Let $c \in C$ and define $R = \{x \in N : c \text{ reachable by } x\}$. Note that $C \subseteq R$ and that R is identical to the set of agents that can reach *any* agent in C. By induction, there exists an MOS partition π' of the subgame of (N, v) induced by $N \setminus R$ that is MOS. Define $\pi = \pi' \cup \{R\}$. We claim that π is MOS. Let $x \in N \setminus R$. By our assumptions on π', there exists no MOS deviation of x to join $\pi(y)$ for $y \in N \setminus R$. In particular, if x is allowed to perform a deviation, then x must have a non-negative utility (otherwise, she can form a singleton coalition contradicting that π' is MOS). So her only potential deviations are to a coalition where she has a friend. Note that x has no friend in R. Indeed, if y was a friend of x in R, then c is reachable for x in G through the concatenation of (x, y) and the path from y to c. Hence, x has no MOS deviation. Now, let $x \in R$. Then, $v_x(\pi) > 0$ because she forms a coalition with her unique friend. By assumption, x has no friend in any other coalition. Therefore, x has no MOS deviation either.

We may therefore assume that G is a directed acyclic graph. Hence, there exists an agent $x \in N$ with in-degree 0. If x has no friend, let $T = \{x\}$. If x has a friend y, we claim that there exists an agent w such that (i) w is the friend of at least one agent and (ii) every agent that has w as a friend has in-degree 0, i.e., such agents are not the friend of any agent. We provide a simple linear-time algorithm that finds such an agent. We will maintain a tentative agent w that will continuously fulfill (i) and update w until this agent also fulfills (ii). Start with $w = y$. Note that this agent w fulfills (i) because y is a friend of x. If w is the friend of some agent z that is herself the friend of some other agent, update $w = z$. For the finiteness (and efficient computability) of this procedure, consider a topological order σ of the agents N in the directed acyclic graph G [24], i.e., a function $\sigma : N \to [n]$ such that $\sigma(a) < \sigma(b)$ whenever $(a, b) \in A$. Note that if w is replaced by the agent z in the procedure, then $\sigma(z) < \sigma(w)$. Hence, w is replaced at most n times, and our procedure finds the desired agent w after a linear number of steps. Now, define $T = \{a \in N : w \text{ reachable by } a\}$, i.e., T contains precisely w and all agents that have w as a friend.

We are ready to find the MOS partition. By induction, we find a partition π' that is MOS for the subgame induced by $N \setminus T$. Consider $\pi = \pi' \cup \{T\}$. Then, $a \in T \setminus \{w\}$ has no incentive to deviate, because she has no friend in any other coalition and has w as a friend. Also, w is not allowed to perform a deviation, because the non-empty set of agents $T \setminus \{w\}$ unanimously prevents that. Possible deviations by agents in $N \setminus T$ can be excluded as in the first part of the proof because these agents have no friend in T. Together, we have completed the induction step and found an MOS partition. ◀

On the other hand, it is NP-complete to decide whether an AFG contains an MOS partition. For a proof, we use the game constructed in Proposition 2 as a gadget in a greater game. The difficulty is to preserve bad properties about the existence of MOS partitions because the larger game might allow for new possibilities to create coalitions with the agents in the counterexample.

▶ **Theorem 4.** *Deciding whether an AFG contains an MOS partition is NP-complete.*

5 Friends-and-Enemies Games

Friends-and-enemies games always contain efficiently computable stable coalition structures with respect to the unanimity-based stability concepts IS and CNS [10]. In this section, we will see that the transition to majority-based consent crosses the boundary of tractability.

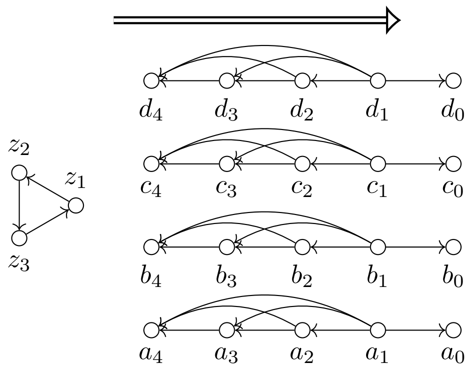

Figure 4 FEG without an MOS partition. The depicted (directed) edges represent friends. The double arrow means that every agent to the left of the tail of the arrow has every agent below the arrow as a friend.

The closeness to this boundary is also emphasized by the fact that it is surprisingly difficult to even construct No-instances for MOS and MIS, i.e., FEGs which do not contain an MOS or MIS partition, respectively. Indeed, the smallest such games that we can construct are games with 23 and 183 agents, respectively. We will start by considering MOS.

▶ **Proposition 5.** *There exists an FEG without an MOS partition.*

Proof sketch. We only give a brief overview of the instance by means of the illustration in Figure 4. The FEG consists of a triangle of agents together with 4 sets of agents whose friendship relation is complete and transitive, together with one additional agent each that gives a temptation for the agent of the transitive substructures with the most friends.

An important reason for the non-existence of MOS partitions is that there is a high incentive for the transitive structures to form coalitions. This gives incentive to agents z_i to join them. If z_1, z_2, and z_3 are in disjoint coalitions, then they would chase each other according to their cyclic structure. If they are all in the same coalition, then agents x_0 for $x \in \{a, b, c, d\}$ prevent the complete transitive structures to be part of this coalition and other transitive structures are more attractive. ◀

In the previous proof, it is particularly useful to establish disjoint coalitions of groups of agents who dislike each other. On the other hand, if we make the further assumption that one agent from every pair of agents likes the other agent, then this does not work anymore and the grand coalition is MOS. This condition essentially means completeness of the friendship relation.[3] Note that this proposition is not true for other stability concepts such as NS or even IS.

▶ **Proposition 6.** *The grand coalition is MOS in every FEG with complete friendship relation.*

Proof. Let (N, v) be an FEG with complete friendship relation, and let π be the grand coalition. We claim that π is MOS. Suppose that there is an agent $x \in N$ who can perform an NS deviation to form a singleton.

[3] Technically, the friendship relation may not be reflexive, but we can set $v_i(i) = 1$ for all $i \in N$ in an FEG to formally achieve completeness.

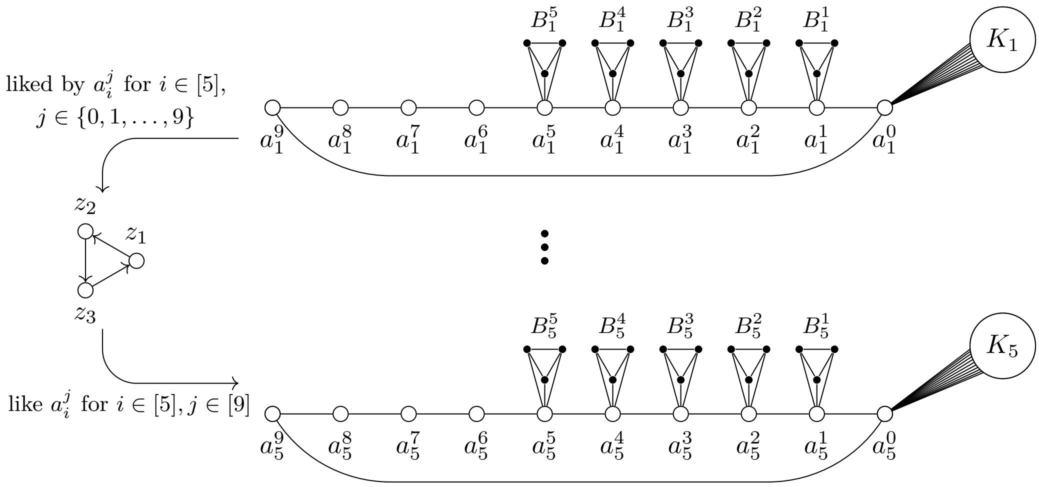

Figure 5 FEG without an MIS partition. The depicted edges represent friends. Undirected edges represent mutual friendship. For $i \in [5]$, some of the edges of agents in A_i are omitted. In fact, these agents form cliques. Also, each K_i represents a clique of 11 agents.

Then, $v_x(N) < 0$ and therefore $|\{y \in N \setminus \{x\}: v_x(y) = -1\}| > \{y \in N \setminus \{x\}: v_x(y) = 1\}|$. Hence,

$$\begin{aligned}|F_{\text{in}}(N, x)| &\geq |\{y \in N \setminus \{x\}: v_x(y) = -1\}| \\ &> |\{y \in N \setminus \{x\}: v_x(y) = 1\}| \\ &\geq |F_{\text{out}}(N, x)|.\end{aligned}$$

In the first inequality, we use that x is a friend of all of her enemies. In the final inequality, we use that x can only be an enemy of her friends. Hence, x is not allowed to perform an MOS deviation. ◀

Still, the non-existence of MOS partitions in FEGs shown in Proposition 5 can be leveraged to prove an intractability result. Interestingly, in contrast to the proofs of Theorem 1 and Theorem 4, the next theorem merely uses the existence of an FEG without an MOS partition to design a gadget and does not exploit the specific structure of a known counterexample.

▶ **Theorem 7.** *Deciding whether an FEG contains an MOS partition is NP-complete.*

In our next result, we construct an FEG without an MIS partition. Despite a lot of structure, the game is quite large encompassing 183 agents.

▶ **Proposition 8.** *There exists an FEG without an MIS partition.*

Proof sketch. We illustrate the example with the aid of Figure 5 and briefly discuss some key features. Again, the central element is a directed cycle of three agents. These agents are connected to five copies of the same gadget. This gadget consists of a main clique $\{a_i^0, \ldots, a_i^9\}$ of 10 mutual friends and further cliques that cause certain temptations for agents in the main clique. Cliques are linked by agents that have an incentive to be part of two cliques, which are part of disjoint coalitions. Since it is possible to balance all diametric temptations, the instance does not admit an MIS partition. ◀

Similar to Proposition 6, it is easy to see that the singleton partition is MIS in every FEG with complete enemy relation. Indeed, then an agent either has no incentive to join another agent, or the other agent will deny her consent. Hence, MIS can also prevent typical run-and-chase games which do not admit NS partitions. We are ready to prove hardness of deciding on the existence of MIS partitions in FEGs.

▶ **Theorem 9.** *Deciding whether an FEG contains an MIS partition is NP-complete.*

6 Discussion and Conclusion

We have investigated single-agent stability in additively separable hedonic games. Our main results determine strong boundaries to the efficient computability of stable partitions. Table 1 provides a complete picture of the computational complexity of all considered stability notions and subclasses of ASHGs, where our results close all remaining open problems. First, we resolve the computational complexity of computing CNS partitions, which considers the last open unanimity-based stability notion in unrestricted ASHGs. The derived hardness result stands in contrast to positive results when considering appropriate subclasses such as FEGs, AEGs, or AFGs [10]. Second, our intractability concerning AFGs stands in contrast to known positive results for all other consent-based stability notions, and can also be circumvented by considering AFGs with a sparse friendship relation. Finally, we provide sophisticated hardness proofs for majority-based stability concepts in FEGs. These turn into computational feasibilities when transitioning to unanimity-based stability, or under further assumptions to the structure of the friendship graph.

A key step of all hardness results in restricted classes of ASHGs was to construct the first No-instances, that is, games that do not admit stable partitions for the respective stability notion. This is no trivial task as can be seen from the complexity of the constructed games. Once No-instances are found, we can leverage them as gadgets of hardness reductions, which is a typical approach for complexity results about hedonic games. We have provided both reductions where the explicit structure of the determined No-instances is used as well as reductions where the mere existence of No-instances is sufficient and used as a black box.

Our results complete the picture of the computational complexity for all considered stability notions and game classes. Still, majority-based stability notions deserve further attention because they offer a natural degree of consent to perform deviations. Their thorough investigation in other classes of hedonic games might lead to intriguing discoveries.

Table 1 Overview of the computational complexity of single-agent stability concepts in different classes of ASHGs. The NP-completeness results concern deciding on the existence of a stable partition. Membership in Function-P means that the search problem of constructing a stable partition can be solved in polynomial time.

ASHG	Unrestricted	Friends-and-enemies games	Appreciation-of-friends games
NS	NP-complete [29]	NP-complete [10]	NP-complete [10]
IS	NP-complete [29]	Function-P [10]	Function-P [16]
CNS	NP-complete (Th. 1)	Function-P [10]	Function-P [10]
MIS	NP-complete [10]	NP-complete (Th. 9)	Function-P [10]
MOS	NP-complete [10]	NP-complete (Th. 7)	NP-complete (Th. 4)

References

1 H. Aziz and F. Brandl. Existence of stability in hedonic coalition formation games. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 763–770, 2012.

2 H. Aziz, F. Brandl, F. Brandt, P. Harrenstein, M. Olsen, and D. Peters. Fractional hedonic games. *ACM Transactions on Economics and Computation*, 7(2):1–29, 2019.

3 H. Aziz, F. Brandt, and H. G. Seedig. Computing desirable partitions in additively separable hedonic games. *Artificial Intelligence*, 195:316–334, 2013.

4 H. Aziz and R. Savani. Hedonic games. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 15. Cambridge University Press, 2016.

5 C. Ballester. NP-completeness in hedonic games. *Games and Economic Behavior*, 49(1):1–30, 2004.

6 S. Banerjee, H. Konishi, and T. Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18:135–153, 2001.

7 V. Bilò, A. Fanelli, M. Flammini, G. Monaco, and L. Moscardelli. Nash stable outcomes in fractional hedonic games: Existence, efficiency and computation. *Journal of Artificial Intelligence Research*, 62:315–371, 2018.

8 A. Bogomolnaia and M. O. Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.

9 F. Brandt and M. Bullinger. Finding and recognizing popular coalition structures. *Journal of Artificial Intelligence Research*, 74:569–626, 2022.

10 F. Brandt, M. Bullinger, and L. Tappe. Single-agent dynamics in additively separable hedonic games. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, 2022. Forthcoming.

11 F. Brandt, M. Bullinger, and A. Wilczynski. Reaching individually stable coalition structures in hedonic games. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, pages 5211–5218, 2021.

12 M. Bullinger. Pareto-optimality in cardinal hedonic games. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 213–221, 2020.

13 M. Bullinger and S. Kober. Loyalty in cardinal hedonic games. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 66–72, 2021.

14 R. Carosi, G. Monaco, and L. Moscardelli. Local core stability in simple symmetric fractional hedonic games. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 574–582, 2019.

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

16 D. Dimitrov, P. Borm, R. Hendrickx, and S. C. Sung. Simple priorities and core stability in hedonic games. *Social Choice and Welfare*, 26(2):421–433, 2006.

17 D. Dimitrov and S. C. Sung. On top responsiveness and strict core stability. *Journal of Mathematical Economics*, 43(2):130–134, 2007.

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

19 E. Elkind, A. Fanelli, and M. Flammini. Price of pareto optimality in hedonic games. *Artificial Intelligence*, 288:103357, 2020.

20 E. Elkind and M. Wooldridge. Hedonic coalition nets. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 417–424, 2009.

21 M. Gairing and R. Savani. Computing stable outcomes in symmetric additively separable hedonic games. *Mathematics of Operations Research*, 44(3):1101–1121, 2019.

22 D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

23 M. Hoefer, D. Vaz, and L. Wagner. Dynamics in matching and coalition formation games with structural constraints. *Artificial Intelligence*, 262:222–247, 2018.

24 A. B. Kahn. Topological sorting of large networks. *Communications of the ACM*, 5(11):558–562, 1962.

25 R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

26 M. Olsen. On defining and computing communities. In *Proceedings of the 18th Computing: Australasian Theory Symposium (CATS)*, volume 128 of *Conferences in Research and Practice in Information Technology (CRPIT)*, pages 97–102, 2012.

27 W. Suksompong. Individual and group stability in neutral restrictions of hedonic games. *Mathematical Social Sciences*, 78:1–5, 2015.

28 S. C. Sung and D. Dimitrov. On myopic stability concepts for hedonic games. *Theory and Decision*, 62(1):31–45, 2007.

29 S. C. Sung and D. Dimitrov. Computational complexity in additive hedonic games. *European Journal of Operational Research*, 203(3):635–639, 2010.