Stability in Online Coalition Formation

Martin Bullinger^{1*}, René Romen²

¹Department of Computer Science, University of Oxford ²School of Computation, Information and Technology, Technical University of Munich martin.bullinger@cs.ox.ac.uk, rene.romen@tum.de

Abstract

Coalition formation is concerned with the question of how to partition a set of agents into disjoint coalitions according to their preferences. Deviating from most of the previous work, we consider an online variant of the problem, where agents arrive in sequence and whenever an agent arrives, they have to be assigned to a coalition immediately and irrevocably. The scarce existing literature on online coalition formation has focused on the objective of maximizing social welfare, a demanding requirement, even in the offline setting. Instead, we seek to achieve *stable* coalition structures in an online setting, and focus on stability concepts based on deviations by single agents. We present a comprehensive picture in additively separable hedonic games, leading to dichotomies, where positive results are obtained by deterministic algorithms and negative results even hold for randomized algorithms.

Introduction

The formation of stable coalition structures is an important concern in multi-agent systems. The key question is how to partition a set of agents into reasonable coalitions. A standard framework for this is the consideration of hedonic games (Drèze and Greenberg 1980). In these games, a set of agents expresses their preferences over subsets of agents containing themselves, i.e., their potential coalitions. The output is then a coalition structure, where all agents are assigned to a unique coalition. In our work, we consider additively separable hedonic games, one of the most prominent classes of hedonic games, where cardinal utilities for single agents encode the preferences, and a sum-based aggregation obtains the utility for a coalition.

Hedonic games have been used to model various aspects of group interaction, such as the formation of research teams (Alcalde and Revilla 2004) or the detection of communities (Aziz et al. 2019). A commonality of most research is that the focus is on a single game, which is fully specified and for which a desirable outcome is searched. However, this misses an important feature of many real-life scenarios: Agents might arrive over time and have to be assigned to existing coalitions. For instance, in a company, most workers are assigned to a department or team, and new hires join existing teams.

Based on such considerations, Flammini et al. (2021b) have introduced an online variant of hedonic games that adds the arrival of agents over time. Their goal is to achieve high social welfare, and they find that a greedy algorithm performs best in a competitive analysis. However, this algorithm has an unbounded competitive ratio even if the number of agents or the utility range is unbounded. In subsequent work, Bullinger and Romen (2023a) show that it is possible to get rid of utility dependencies of the competitive ratio under certain model assumptions, e.g., a uniformly random arrival of agents. They achieve a competitive ratio of $\Theta(n)$, which is essentially the best approximation guarantee that we can hope for by efficient algorithms because, for every $\epsilon > 0$, it is NP-complete to approximate social welfare by a factor of $n^{1-\epsilon}$ (Flammini, Kodric, and Varricchio 2022, Theorem 17).¹

By contrast, the scarce existing literature on online coalition formation omits other common objectives in coalition formation. Stability probably is the most studied solution concept for hedonic games in general and additively separable hedonic games in particular (see, e.g., Bogomolnaia and Jackson 2002; Sung and Dimitrov 2010; Aziz, Brandt, and Seedig 2013; Woeginger 2013; Gairing and Savani 2019; Brandt, Bullinger, and Tappe 2022; Brandt, Bullinger, and Wilczynski 2023; Bullinger 2022). In our work, we close this research gap and consider the question of whether notions of stability can be achieved in an online manner.

We focus on stability concepts based on deviations by single agents, namely Nash stability, individual stability, contractual Nash stability, and contractual individual stability. In addition, we study Pareto optimality, which is particularly interesting because it is a natural weakening of the demanding objective of maximizing social welfare while it can still be interpreted as a notion of stability (Morrill 2010). We present a comprehensive picture of the capabilities and impossibilities of online algorithms aiming to compute stable partitions. For this, we consider natural utility restrictions, such as symmetry or the distinction of friends and en-

^{*}Most of this research was done when Martin Bullinger was a PhD student at the Technical University of Munich.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹This result even holds for the class of aversion-to-enemies games that we will introduce and investigate later.

	strict	FENG	FEG	AFG	AEG
NS	X (Th. 5)				
IS	X (Th. 5)				
CNS	X (Th. 2)	X (Th. 4)	✔(Th. 1)	X (Th. 2)	✔(Th. 1)
CIS	✔(Co. 1)	X (Th. 4)	✔(Co. 1)	✔(Co. 1)	✔(Co. 1)
РО	✔(Th. 3)	X (Th. 4)	✔(Th. 3)	√ (Th. 3)	✔(Th. 3)

Table 1: Computability of stable partitions by online algorithms; for definitions of stability concepts and utility restrictions, see ?? . A checkmark (\checkmark) means that there exists a deterministic online algorithm that can compute the desired partition. A cross (\varkappa) means that there exists no randomized online algorithm that outputs the desired partition with probability bounded away from 0. All negative results hold even for the case of symmetric games. Of the positive results, only the results for CNS need symmetry.

emies, which lead to subclasses of additively separable hedonic games like appreciation-of-friends games or aversionto-enemies games (Dimitrov et al. 2006). Our findings are summarized in Table 1.

A large part of our results is negative and proves the nonexistence of randomized algorithms capable of computing stable coalition structures under strong utility restrictions. This is even the case for solution concepts like Pareto optimality or contractual Nash stability, for which solutions are guaranteed to exist in any hedonic game (Aziz and Savani 2016). By contrast, we obtain deterministic online algorithms capable of computing contractually Nash-stable and Pareto-optimal coalition structures in restricted classes of games. While such positive results seem rare, they entail very strong stability guarantees. The associated algorithms do not only output a final stable coalition structure but they maintain stability throughout the entire arrival process of agents. Otherwise, they would fail their promised guarantee on a partial instance. Hence, these algorithms are suitable for every application with an indefinite time horizon, where new agents can arrive continuously.

Related Work

Our work contributes to two streams of work: the consideration of coalition formation models in economic theory and, more recently, the AI literature as well as the investigation of online algorithms in related settings, mostly in theoretical computer science. Here, we give an account of both.

Coalition formation in the framework of hedonic games was first studied by Drèze and Greenberg (1980) and popularized two decades later (Banerjee, Konishi, and Sönmez 2001; Cechlárová and Romero-Medina 2001; Bogomolnaia and Jackson 2002). Bogomolnaia and Jackson (2002) introduced additively separable hedonic games (ASHGs), which are since then an ongoing subject of study. Aziz and Savani (2016) present a survey of this stream of work. The majority of the research on ASHGs considers the offline setting and focuses on the computational complexity of stability concepts (Dimitrov et al. 2006; Olsen 2009; Sung and Dimitrov 2010; Aziz, Brandt, and Seedig 2013; Gairing and Savani 2019; Flammini et al. 2021b; Brandt, Bullinger, and Tappe 2022; Bullinger 2022), but some more recent studies also consider economic efficiency in the sense of Pareto optimality (Elkind, Fanelli, and Flammini 2020; Bullinger 2020), or strategyproofness (Flammini et al. 2021a). Most important to our work, Dimitrov et al. (2006) and Brandt, Bullinger, and Tappe (2022) consider stability in succinct classes of hedonic games based on the distinction of friends and enemies, and the previously cited work settles the complexity of many single-agent stability notions (including all of the notions we consider here) in the offline setting. Moreover, Bullinger (2020) presents a polynomial-time algorithm to compute Pareto-optimal partitions for symmetric ASHGs.

Like we mentioned in the introduction, online ASHGs have been introduced by Flammini et al. (2021b) and subsequently been studied by Bullinger and Romen (2023a). Moreover, Pavone et al. (2022) study online hypergraph matching. Their model can be interpreted as coalition formation with bounded coalition sizes. In contrast to Flammini et al. (2021b) and Bullinger and Romen (2023a), Pavone et al. (2022) do not assume additively separable utilities and agents do not have to be matched immediately at arrival, but they depart from the market unmatched after a fixed time. All three works solely consider the maximization of social welfare or the minimization of total cost.

In addition, a recent series of work considers deviation dynamics for hedonic games, which constitute another time-dependent model of coalition formation (see, e.g., Bilò et al. 2018; Brandt, Bullinger, and Wilczynski 2023; Carosi, Monaco, and Moscardelli 2019). In particular, ASHGs and close variants are studied in depth (Bilò, Monaco, and Moscardelli 2022; Boehmer, Bullinger, and Kerkmann 2023; Brandt, Bullinger, and Tappe 2022; Bullinger and Suksompong 2023).

From the literature on online algorithms, most related to our setting is online matching, which originates from the seminal paper by Karp, Vazirani, and Vazirani (1990). Huang and Tröbst (2023) very recently survey this line of work. Matchings can be seen as a variant of hedonic games, where coalitions are restricted to be of size at most 2. Different to our work, the input instances are bipartite and only one side of the agents appears online. The objective in online matching is usually to find a matching of maximum cardinality or weight. Karp, Vazirani, and Vazirani (1990) introduce the famous ranking algorithm, which achieves a competitive ratio of $1 - \frac{1}{e}$. Subsequent work considers related models with edge weights, all agents arriving online, or nonbipartite matching (Feldman et al. 2009; Wang and Wong 2015; Huang et al. 2018). While it is possible to achieve the competitive ratio of $1 - \frac{1}{e}$ in the weighted setting (Feldman et al. 2009), this is usually impossible if all agents arrive online (Wang and Wong 2015; Huang et al. 2018).

Additionally, stability has been considered for online matching to some extent. Doval (2022) extends stability according to Gale and Shapley (1962) to an online setting, and shows that her extension can always be satisfied. Still, her model has several conceptual differences to the online models discussed thus far. Most notably, agents do not have to be matched immediately (but suffer from a discount in utility if

matched later) and may arrive in batches. Moreover, Gajulapalli et al. (2020) study a two-stage process for school choice with the goal of preserving stability. While this is in general not possible, they present efficient algorithms that maximize the number of additionally matched agents or minimize the number of reallocations compared to the matching of the first stage. More loosely related, Benadè and Sahoo (2023) touch upon an online model for recommender systems, where they investigate stability mostly experimentally.

Preliminaries

In this section, we present preliminaries. For an integer $i \in \mathbb{N}$, we define $[i] := \{1, \ldots, i\}$.

Additively Separable Hedonic Games

Let N be a finite set of n agents. Any subset of N is called a *coalition*. We denote the set of all possible coalitions containing agent $i \in N$ by $\mathcal{N}_i := \{C \subseteq N : i \in C\}$. A *coalition structure* (or *partition*) is a partition of the agents. Given an agent $i \in N$ and a partition π , let $\pi(i)$ denote the coalition of i, i.e., the unique coalition $C \in \pi$ with $i \in C$.

A hedonic game is a pair (N, \succeq) consisting of a set Nof agents and a preference profile $\succeq = (\succeq_i)_{i \in N}$ where \succeq_i is a weak order over \mathcal{N}_i which represents the preferences of agent *i*. The semantics is that agent *i* strictly prefers coalition C to coalition C' if $C \succ_i C'$ and is indifferent between coalitions C and C' if $C \sim_i C'$.

An additively separable hedonic game (ASHG) consists of a set N of agents and a tuple $u = (u_i)_{i \in N}$ of *utility* functions $u_i: N \to \mathbb{Q}$, such that, for every pair of coali-tions $C, C' \in \mathcal{N}_i$, it holds that $C \succeq_i C'$ if and only if $\sum_{j \in C} u_i(j) \geq \sum_{j \in C'} u_i(j)$ (Bogomolnaia and Jackson 2002). We usually represent an ASHG by the pair (N, u). Clearly, an ASHG is a hedonic game. We abuse notation and extend the definition of u to coalitions $C \in \mathcal{N}_i$ and partitions π by $u_i(C) := \sum_{i \in C} u_i(j)$ and $u_i(\pi) := u_i(\pi(i))$, respectively. Also, an ASHG can be represented equivalently by a complete directed graph G = (N, E) with weight $u_i(j)$ on arc (i, j). An ASHG is said to be *symmetric* if, for every pair of agents $i, j \in N$, it holds that $u_i(j) = u_i(i)$. In this case, we also write $u(i, j) := u_i(j)$. A symmetric ASHG can be naturally represented by a complete undirected graph. Following Aziz, Brandt, and Seedig (2013), an ASHG is said to be *strict* if, for every pair of agents $i, j \in N$, it holds that $u_i(j) \neq 0.$

There are various important subclasses of ASHGs with restricted utility values. Given a subset $U \subseteq \mathbb{Q}$, an ASHG is called a *U-ASHG* if, for every pair of agents $i, j \in N$, it holds that $u_i(j) \in U$. In particular, there are ASHGs that allow a natural interpretation in terms of friends and enemies. A *U*-ASHG is called an *appreciation-of-friends* game (AFG), aversion-to-enemies game (AEG), friendsand-enemies game (FEG), or friends-enemies-and-neutrals game (FENG) if $U = \{n, -1\}, U = \{1, -n\}, U = \{1, -1\}$, or $U = \{1, 0, -1\}$, respectively (Dimitrov et al. 2006; Brandt, Bullinger, and Tappe 2022). In all of these games, the utility for a coalition depends on the distinction of friends and enemies, i.e., players that yield positive and



Figure 1: Logical relationships between our solutions concepts (see, e.g., Aziz and Savani 2016). An arrow from concept α to concept β indicates that if a partition satisfies α , then it also satisfies β . For reference, we also depict welfare optimality.

negative utility, respectively. In FEGs and FENGs, friends and enemies have equal importance, whereas in AFGs, a single friend outweighs an arbitrary number of enemies, and in AEGs, a single enemy annihilates any number of friends.

Solution Concepts

In this section, we define the solution concepts considered in this paper. An overview of their logical relationships is depicted in Figure 1. We assume that we are given a fixed ASHG (N, u).

Notions of stability capture agents' incentives to perform deviations (Bogomolnaia and Jackson 2002; Dimitrov and Sung 2007). 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 $j \neq i$, it holds that $\pi(j) \setminus \{i\} =$ $\pi'(j) \setminus \{i\}$. The basic idea of deviations is that the deviating agent should have an immediate benefit from a deviation. A Nash deviation is a single-agent deviation performed by agent *i* such that $u_i(\pi') > u_i(\pi)$. Any partition in which no Nash deviation is possible is said to be Nash-stable (NS).

The drawback of Nash stability is that only the preferences of the deviating agent are considered, which might seem too demanding in a context of cooperation. Therefore, various refinements have been proposed, which additionally require the consent of the abandoned or the welcoming coalition. An individual deviation (or contractual deviation) is a Nash deviation by agent *i* transforming π into π' such that, for all agents $j \in \pi'(i) \setminus \{i\}$ (or $j \in \pi(i) \setminus \{i\}$), it holds that $u_i(\pi') \ge u_i(\pi)$. Then, a partition is said to be *individ*ually stable (IS) or contractually Nash-stable (CNS) if it allows for no individual or contractual deviation, respectively. A single-agent deviation is called a contractual individual deviation if it is both a contractual deviation and an individual deviation. A partition is said to be contractually individually stable (CIS) if it allows for no contractual individual deviation. For a more concise notation, we refer to devia-CIS as α deviations, e.g., IS deviations for α = IS. Similarly, we refer to a partition satisfying stability concept α as α partition.

Finally, we also consider Pareto optimality, which can be seen as a stability guarantee, where the whole set of agents cannot perform a group deviation. A partition π' is said to

Pareto-dominate a partition π if, for every agent $i \in N$, it holds that $u_i(\pi') \ge u_i(\pi)$, and there exists an agent $j \in N$ with $u_j(\pi') > u_j(\pi)$. A partition π is said to be *Paretooptimal* (PO) if it is not Pareto-dominated by another partition. Pareto optimality is a classical concept of economic efficiency, and already was a primal objective during the birth of hedonic games (Drèze and Greenberg 1980). Note that Pareto optimality is a weakening of welfare optimality, which was the objective in the literature on online AS-GHs thus far (Flammini et al. 2021b; Bullinger and Romen 2023a).

Online Coalition Formation

In this section, we introduce our model of online coalition formation following the notation of Bullinger and Romen (2023a). The online setting is not restricted to ASHGs, so we define it for a general hedonic game $G = (N, \succeq)$. Let $\Sigma(G) := \{\sigma : [|N|] \to N \text{ bijective}\}$ be the set of all orders of the agent set N. Given a subset of agents $M \subseteq N$, let G[M] be the hedonic game restricted to agent set M. Moreover given a partition π of a set N and a subset of agents $M \subseteq N$, we define $\pi[M]$ as the partition restricted to M as $\pi[M] := \{C \cap M : C \in \pi, C \cap M \neq \emptyset\}$. Specifically, if M consists of all agents except a single agent $i \in N$, then we write $\pi - i := \pi[N \setminus \{i\}]$.

An instance of an *online coalition formation* problem is a pair (G, σ) , where $G = (N, \succeq)$ is a hedonic game and $\sigma \in \Sigma(G)$. An *online coalition formation algorithm* for instance (G, σ) gets as input the sequence G_1, \ldots, G_n , where, for every $i \in [n]$, $G_i := G[\{\sigma(j): 1 \leq j \leq i\}]$. Then, for every $i \in [n]$, the algorithm has to produce a partition π_i of $\{\sigma(j): 1 \leq j \leq i\}$ such that

- the algorithm has only access to G_i and
- $\pi_i \sigma(i) = \pi_{i-1}$.

The output of the algorithm is the partition π_n . Given an online coalition formation algorithm ALG, let $ALG(G, \sigma)$ be its output for instance (G, σ) . If σ is clear from the context, we omit it from this notation and simply write ALG(G).

More informally, the algorithm iteratively creates a partition such that it only has access to the utilities of the currently present agents when irrevocably adding a new agent to an existing or new coalition. In addition to deterministic algorithms, we also consider randomized algorithms. This means that the decisions as to which coalition an agent is added to can be random.

Unlike welfare optimality, stability concepts do not naturally yield a quantitative maximization objective, and we cannot directly perform the usual competitive analysis. Instead, we have qualitative objectives that are either satisfied or not by an output. Therefore, we desire algorithms that output stable partitions with high probability if agents arrive online, which once again is a quantitative objective.

Consider a solution concept $\alpha \in \{NS, IS, CNS, CIS, PO\}$ and an algorithm *ALG*. We define the α guarantee of *ALG* as

$$W_{\alpha}(ALG) := \inf_{G} \min_{\sigma \in \Sigma(G)} \mathbb{P}(ALG(G, \sigma) \text{ is } \alpha).$$

Here, the probability is taken according to the randomized decisions of ALG. Hence, $W_{\alpha}(ALG)$ is the worst-case probability that ALG outputs an α partition.

Note that the α guarantee of deterministic algorithms is either one—if the algorithm always outputs an α partition or zero—if the algorithm does not output an α partition for some input instance.

Results

In this section, we present our results. For the consideration of stability in online coalition formation, only instances that allow for a desired partition are relevant. Otherwise, no algorithm, and therefore especially no online algorithm, can make any guarantee.²

In the literature on stability in ASHGs, two restrictions have turned out to be vital for the existence of stable partitions, namely symmetry and severe utility restrictions (Bogomolnaia and Jackson 2002; Brandt, Bullinger, and Tappe 2022). In particular, in symmetric ASHGs, Nash-stable partitions (and therefore partitions satisfying all weaker stability notions) are guaranteed to exist (Bogomolnaia and Jackson 2002). Moreover, in (possibly asymmetric) FEGs, AEGs, and AFGs, it is guaranteed that individually stable and contractually Nash-stable partitions exist (Brandt, Bullinger, and Tappe 2022). In this section, we will see (the conjunction of) which of these assumptions are sufficient to allow for the computation of stable outcomes in an online manner, and which of the results in the offline setting cause problems online.

Contractual Nash Stability

We start with the consideration of CNS, where every agent in the abandoned coalition can veto a single-agent deviation. As a warm-up, we start with a simple proposition that gives a first insight, why computing stable partitions in an online manner is a nontrivial task. Even the conjunction of symmetry and utility restrictions is not sufficient for computing CNS partitions.

Proposition 1. There exists no deterministic online algorithm, which always outputs a CNS partition for symmetric AFGs.

Proof. Assume for contradiction that ALG always outputs a CNS partition for symmetric AFGs. Consider a game, with agent set $\{a, b, c\}$ and symmetric utilities u(a, b) = -1 and u(a, c) = u(b, c) = 3 with the arrival order a, then b, then c. When b arrives, ALG has to create a new coalition as otherwise the adversary stops with a coalition that is not CNS. When c arrives, ALG cannot form a new singleton coalition as otherwise a (or b) has a CNS deviation to join c. Assume without loss of generality that ALG forms $\{a, c\}$. Then, b has a CNS deviation to join them.

 $^{^{2}}$ We can also take the viewpoint of comparing the capabilities of online algorithms with offline possibilities: if no stable partition exists in an instance, then any online algorithm is as good as an optimal offline algorithm.

Algorithm 1: Contractually Nash-stable partition of online symmetric $\{-y, x\}$ -ASHGs for $y \ge x > 0$.

Input: Symmetric $\{-y, x\}$ -ASHG **Output:** Contractually Nash-stable partition π

$$\begin{aligned} \pi \leftarrow \emptyset \\ \text{for } i &= 1, \dots, n \text{ do} \\ N_i \leftarrow \{j \in [i-1] \colon u(a_i, a_j) > 0\} \\ \text{if } \exists j \in N_i \text{ with } |\pi(a_j)| &= 1 \text{ then} \\ \pi \leftarrow \pi \setminus \{\{a_j\}\} \cup \{\{a_i, a_j\}\} \\ \text{else if } \exists j \in N_i \text{ with } |\pi(a_j)| > 1 \text{ then} \\ \pi \leftarrow \pi \setminus \{\pi(a_j)\} \cup \{\pi(a_j) \cup \{a_i\}\} \\ \text{else} \\ \pi \leftarrow \pi \cup \{\{a_i\}\} \\ \text{return } \pi \end{aligned}$$

However, the previous result seems to rely on small negative utilities. In fact, for other restricted classes, we can compute CNS partitions with an online algorithm. The basic idea of our algorithm is to establish that agents in a coalition of size at least 2 are not allowed to leave their coalition because some other agent would veto this. Moreover, we use our assumption on the utility values to achieve that agents in singleton coalitions can never gain positive utility by joining a constructed coalition.

Theorem 1. Let $y \ge x > 0$. Then, there exists a deterministic online algorithm, which always outputs a CNS partition for symmetric $\{-y, x\}$ -ASHGs.

Proof. Let $y \ge x > 0$ and consider a symmetric $\{-y, x\}$ -ASHG with agent set $N = \{a_i : 1 \le i \le n\}$ and arrival order a_1, \ldots, a_n . We apply Algorithm 1 to compute a partition π . This algorithm proceeds as follows. Whenever a new agent arrives, we check if there are agents with positive utility for this agent. Assume that this is the case. First, we check if some of them is in a singleton coalition, and if yes, then the new agent joins such a singleton coalition. Otherwise, we add a_i to any coalition of an agent with positive utility. If no such agent exists, we form a new singleton coalition.

We claim that the partition π is contractually Nash-stable. We show the claim by induction. Recall that, for every $i \in [n], \pi_i$ is the partition created by the algorithm after agent a_i has been assigned to a coalition.

Claim 1. For every $i \in [n]$, the following statements are true:

- 1. The partition π_i is CNS.
- 2. For every coalition $C \in \pi_i$ with $|C| \ge 2$ and $a_k \in C$, there exists an agent $a_\ell \in C$ with $u(a_k, a_\ell) = x$.
- 3. For every $k \in [i]$ with $|\pi_i(a_k)| = 1$, it holds that, if $\ell \in [i]$ with $u(a_k, a_\ell) = x$, then $|\pi_i(a_\ell)| \ge 2$ and $u(a_k, b) = -y$ for all $b \in \pi_i(a_\ell) \setminus \{a_\ell\}$

Proof. All three statements are true for i = 1. Assume now that the statements are true for some $1 \le i < n$. Let $N_{i+1} := \{j \in [i] : u(a_{i+1}, a_j) > 0\}$. We start by proving

the second and third assertion by a case distinction according to the different cases in the algorithm.

Assume first that there exists $j \in N_{i+1}$ with $|\pi_i(a_j)| = 1$ and that we have $\pi_{i+1} = \pi_i \setminus \{\{a_j\}\} \cup \{\{a_j, a_{i+1}\}\}$. Clearly, the second assertion for i + 1 follows by induction and $u(a_{i+1}, a_j) = x$. For the third assertion, let $k \in [i+1]$ with $|\pi_{i+1}(a_k)| = 1$ and consider $\ell \in [i+1]$ with $u(a_k, a_\ell) = x$. Since $|\pi_{i+1}(a_{i+1})| = 2$, it holds that $k \neq i + 1$. By the induction hypothesis for the third assertion, the third assertion is true unless $\ell \in \{i+1, j\}$. Moreover, again by the induction hypothesis for the third assertion, $u(a_k, a_j) = -y$, and therefore the assertion is true if $\ell = i + 1$.

Next, assume that there exists no $j \in N_{i+1}$ with $|\pi_i(a_j)| = 1$, but $N_{i+1} \neq \emptyset$ and that $\pi_{i+1} = \pi_i \setminus {\pi_i(a_j)} \cup {\pi_i(a_j)} \cup {\pi_i(a_j) \cup {a_{i+1}}}$. Note that the second assertion is true for agent a_{i+1} because $a_j \in \pi_{i+1}(a_{i+1})$. For all other agents, the second assertion follows by induction. As there exists no $j \in N_{i+1}$ with $|\pi(a_j)| = 1$, the third assertion follows by induction.

Finally, if $N_{i+1} = \emptyset$, then $\pi_{i+1} = \pi_i \cup \{\{a_{i+1}\}\}$. Hence, the second assertion follows by induction and the third assertion follows by induction for all agents except for a_{i+1} . For a_{i+1} , it is true because $N_{i+1} = \emptyset$.

It remains to prove the first assertion. We show how it follows from the second and third assertion. By the second assertion, no agent in a coalition of size at least 2 is allowed to leave their coalition and can therefore not perform a CNS deviation. On the other hand, by the third assertion, no agent in a singleton coalition can improve their utility by joining any other coalition. Hence, π_{i+1} is a CNS partition. This completes the proof of the claim.

The assertion of Theorem 1 follows from Claim 1 for the case i = n.

In particular, Theorem 1 applies to symmetric FEGs and AEGs. For the latter, we need to deal with variable utility values that depend on the number of players. However, Theorem 1 applies to individual games, and each AEG satisfies the conditions of the theorem. By contrast, Theorem 1 breaks down if we additionally allow for the utility value of 0, even if the positive and negative utilities are restricted further. We defer this result to the consideration of CIS, where we get it as a byproduct.

To conclude this section, we show how to strengthen Proposition 1 for randomized algorithms. The idea is to construct a random instance where every deterministic algorithm performs badly. We can then apply Yao's principle (Yao 1977) to bound the performance of any randomized algorithm. The idea is to create a random version of the game in Proposition 1. In this game, every deterministic algorithm succeeds to compute a CNS partition with probability at most 1/2. Modifying this instance by concatenating k copies of this instance, every deterministic algorithm succeeds with probability at most 2^{-k} .

Theorem 2. Let ALG be any randomized online algorithm for symmetric AFGs. Then, it holds that $W_{\text{CNS}}(ALG) = 0$.

Proof. Let $k \in \mathbb{N}$ be a positive integer. We define the random AFG G = (N, u) where $N = \bigcup_{i \in [k]} N_i$ for $N_i =$ $\{a_i, b_i, c_i\}$. The random utilities are given by $u(a_i, b_i) = -1$ and, with probability 1/2 each, it holds that $u(a_i, c_i) =$ $u(b_i, c_i) = 3k$ or $u(a_i, c_i) = u(b_i, c_i) = -1$. Note that 3k is the number of agents. All other utilities are set to -1. The randomizations for the utilities within N_i and N_j for $1 \le i < j \le k$ are performed independently. The agents arrive in the order N_1, \ldots, N_k and within a set N_i , first a_i , then b_i , then c_i . In other words, G is a uniformly random choice from a set of 2^k instances, each of which is a composition of k gadgets drawn independently from the same distribution. Each of the gadgets is similar to the game considered in Proposition 1.

Now, let ALG be an arbitrary deterministic algorithm for AFGs and define $\pi := ALG(G)$. By the proof of Proposition 1, for every $i \in [k]$, ALG fails with probability 1/2 on $G[N_i]$.

Moreover, by design of the random instance, if ALG computes a CNS partition on G, then, for all $i \in [k], \pi[N_i]$ is CNS for $G[N_i]$. Indeed, if $u(a_i, c_i) = u(b_i, c_i) = -1$, then π is CNS only if all agents in N_i are in singleton coalitions, and hence $\pi[N_i]$ is CNS for $G[N_i]$. If $u(a_i, c_i) = u(b_i, c_i) = 3k$, then $a_i \in \pi(c_i)$ and $b_i \in \pi(c_i)$ as these agents would perform a CNS deviation to join c_i , otherwise. Hence, $\pi[N_i] = \{N_i\}$, which is CNS for $G[N_i]$.

Now, observe that, by the arrival sequence of the agents, the performance of ALG on N_i is at most as good as the performance of the best algorithm for $G[N_i]$. Therefore, using independence of the random selection of the utilities, the probability that ALG computes a CNS partition on G is bounded by the product of the probabilities that $\pi[N_i]$ is CNS for $G[N_i]$. Hence, π is a CNS partition with probability at most 2^{-k} .

By Yao's principle, no randomized algorithm can compute a CNS partition with probability more than 2^{-k} for every (deterministic) symmetric AFG. Since k is chosen arbitrarily, this proves the assertion.

Contractual Individual Stability and Pareto Optimality

Next, we consider CIS, which is the weakening of CNS, where an agent in the welcoming coalition can also veto a single-agent deviation. Algorithm 1 can be used to compute CIS partitions, even if we allow for strict and symmetric ASHGs as input. However, our next result achieves even more and shows the existence of an online algorithm for computing PO partitions in strict (and possibly nonsymmetric) ASHGs. Recall that PO is a stronger notion than CIS. The presented algorithm is an online adaptation of serial dictatorships, an algorithmic approach that is known to be successful for offline ASHGs (Aziz, Brandt, and Seedig 2013; Bullinger 2020) and online fair division (Aleksandrov and Walsh 2019).³ The idea is to assign a dictator with every created coalition and these are asked in the order of their arrival whether they want newly arriving agents to be part of their coalition.

Algorithm 2: Pareto-optimal partition of online strict ASHG

Input: Strict ASHG

Output: Pareto-optimal partition π

```
\begin{aligned} \pi \leftarrow \emptyset, k \leftarrow 0 \\ \text{for } i = 1, \dots, n \text{ do} \\ \text{if } \{j \in [k] \colon u_{\ell_j}(a_i) > 0\} \neq \emptyset \text{ then} \\ j^* \leftarrow \min_{j \in [k]} \{u_{\ell_j}(a_i) > 0\} \\ \pi \leftarrow \pi \setminus \{C_{j^*}\} \cup \{C_{j^*} \cup \{a_i\}\} \\ \text{else} \\ k \leftarrow k + 1 \\ C_k \leftarrow \{a_i\}, \ell_k \leftarrow a_i \\ \pi \leftarrow \pi \cup \{C_k\} \end{aligned}return \pi
```

Theorem 3. There exists a deterministic online algorithm, which always outputs a PO partition for strict ASHGs.

Proof. Consider a strict ASHG with agent set $N = \{a_i: 1 \le i \le n\}$ and arrival order a_1, \ldots, a_n . Apply Algorithm 2 to compute a partition π . This algorithm proceeds as follows: Whenever a new agent arrives, we ask for the existing coalitions whether the first agent in that coalition has a positive utility for the new agent. If there exists such a coalition, the new agent joins the coalition, which was created first. Otherwise, the algorithm starts a new coalition with the new agent.

For the proof, we use the notation from the algorithm and assume that $\pi = \{C_i : 1 \le i \le m\}$ for some m > 0. Assume further that these coalitions were formed in the order C_1, \ldots, C_m and that agent ℓ_j was the first agent in coalition C_j for all $j \in [m]$. Clearly, the algorithm fulfills the property that, for all $j \in [m]$, and agents $x \in C_j \setminus \{\ell_j\}$, it holds that $u_{\ell_j}(x) > 0$. We refer to this property as observation (*).

We are ready to prove that π is Pareto-optimal. Assume that π' is any partition such that, for all agents $x \in N$, it holds that $u_x(\pi') \ge u_x(\pi)$. We claim that $\pi' = \pi$.

By observation (*) and the design of the algorithm, agent ℓ_1 is in their best coalition in π and, by strictness of the utilities, their best coalition is unique, so $\pi'(\ell_1) = \pi(\ell_1) = C_1$. We call this fact observation (**). We now prove our claim that $\pi' = \pi$ by induction over m, i.e., the number of coalitions in the partition π .

First, consider the case m = 1. Then, by observation (**), it holds that $\pi' = \{\pi'(\ell_1)\} = \{\pi(\ell_1)\} = \pi$.

Now, assume that m > 1. By observation (**), it suffices to show that $\pi' \setminus \{C_1\} = \pi \setminus \{C_1\}$. Consider the ASHG restricted to the agent set $N' = \bigcup_{2 \le j \le m} C_j$. Then, $\pi \setminus \{C_1\}$ is the output of Algorithm 2 of the restricted ASHG if the arrival order is the subsequence of the original arrival order. Moreover, by assumption, for all $x \in N$, it holds that $u_x(\pi' \setminus \{C_1\}) \ge u_x(\pi \setminus \{C_1\})$. Hence, by induction, it holds that $\pi' \setminus \{C_1\} = \pi \setminus \{C_1\}$, as desired. This shows that $\pi' = \pi$. As a consequence, there exists no partition that Pareto-

dominates π and therefore π is Pareto-optimal.

Since, every Pareto-optimal partition is also a CIS partition, we obtain the following corollary.

³Similar to the online fair division literature, our online serial dictatorship algorithm can be shown to have the additional desirable property of strategyproofness.

Corollary 1. There exists a deterministic online algorithm, which always outputs a CIS partition for strict ASHGs.

Of course, Theorem 3 and Corollary 1 work for subclasses of ASHGs like FEGs, AFGs, and AEGs. Interestingly, however, once we allow for a utility of 0, it is impossible to compute CIS partitions, and therefore PO partitions, in an online manner and this already holds for FENGs. This is a clear contrast to offline hedonic games, where PO (and CIS) partitions are guaranteed to exist without any restriction on the game.

Proposition 2. There exists no deterministic online algorithm, which always outputs a CIS partition for symmetric FENGs.

Proof. Assume for contradiction that ALG always outputs a CIS partition for symmetric FENGs. Assume that the first two agents to arrive are a and b with u(a, b) = 0.

If ALG forms $\{a, b\}$, then an agent c arrives with u(a, c) = -1 and u(b, c) = 1. Then, $\{\{a, b\}, \{c\}\}$ is no CIS partition because b has a CIS deviation to join agent c. However, forming $\{a, b, c\}$ does not lead to a CIS partition either, because then agent a has a CIS deviation to form a singleton coalition.

On the other hand, if ALG forms two singleton coalitions for a and b, then consider an agent c with u(a, c) = u(b, c) = 1. If ALG forms $\{a, c\}$ or $\{c\}$, then b has a CIS deviation to join this coalition. If, however, ALG forms $\{b, c\}$, then a has a CIS deviation to join.

Similar to the previous section, we can extend this result to randomized algorithms. We defer the details to the full version (Bullinger and Romen 2023b).

Theorem 4. Let ALG be any randomized online algorithm for symmetric FENGs. Then, it holds that $W_{CIS}(ALG) = 0$.

Individual Stability

Finally, we consider individual stability, which is a strengthening of CIS and the complementary (but logically incomparable) notion of CNS, where each agent in the welcoming (instead of abandoned) partition has the power to veto a single-agent deviation. Even for the combination of symmetry and utilities restricted to any positive and negative value, online algorithms fail to be able to compute IS partitions.

Compared to the proofs of Theorems 2 and 4, simply concatenating identical games with negative utilities in between can be problematic for some utility values. For instance, if the positive utility is sufficiently large compared to the negative utility (e.g., in AFGs), then the grand coalition is IS if each agent has a positive utility for some other agent. Instead, we prove the statement by considering one large random adversarial instance for deterministic algorithms.

In this instance, depicted in Figure 2, we first have k^2 agents arriving, which all have a mutual utility of -y. Then, an agent b arrives that has a positive utility for a random subset B of k of the first k^2 agents. Finally, an agent c arrive that has a positive utility for b and one random agent in B. It can be shown that every deterministic algorithm computes an IS partition with probability at most $\frac{1}{k}$ and we can apply Yao's principle once again to obtain our theorem about



Figure 2: Adversarial instance for achieving individual stability in $\{-y, x\}$ -ASHGs for x, y > 0. We only depict the positive utilities of x. All remaining utilities are -y.

randomized algorithms. The proof can be found in the full version.

Theorem 5. Let x, y > 0 and let ALG be any randomized online algorithm for symmetric $\{-y, x\}$ -ASHGs. Then, it holds that $W_{IS}(ALG) = 0$.

Conclusion

In this paper, we have studied stability in online coalition formation. We have considered stability notions based on single-agent deviations and Pareto optimality. An overview of our results is displayed in Table 1 in the introduction. Our positive results follow from two deterministic algorithms. The first one outputs CNS partitions for symmetric games with utility restrictions that include FEGs and AEGs. The second one applies to strict ASHGs and outputs PO partitions. The latter is interesting because PO has the flavor of both stability and optimality.⁴

By contrast, we obtain negative results in the sense that there exists no randomized algorithm that can guarantee any fixed probability to output a stable partition. Surprisingly, our negative results even encompass concepts like CIS and PO, for which solutions are guaranteed to exist in every hedonic game. Hence, the online capabilities of algorithms can be severely weaker than offline possibilities. Negative results naturally extend to stronger solution concepts. For instance, a consequence of our negative results for IS and CNS is the NS guarantee of 0 for all considered game restrictions.

We believe that it is an important step to depart from the mere consideration of welfare maximality in online coalition formation. There is plenty of space for future research in this direction. Possible directions include to consider other solution concepts, such as fairness notions. We want to remark, however, that notions of group stability might not yield many positive results. For instance, since the strict core is a strengthening of CIS, Theorem 4 also applies for the strict core. Moreover, it would be interesting to consider game classes different from ASHGs.

⁴Notably, however, both of the two algorithms may output partitions of negative social welfare for instances where the maximum social welfare is positive, and therefore yield no approximation guarantee for maximizing social welfare.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft under grants BR 2312/11-2 and BR 2312/12-1, and by the AI Programme of The Alan Turing Institute. We would like to thank Saar Cohen and the anonymous reviewers for their comments.

References

Alcalde, J.; and Revilla, P. 2004. Researching with whom? Stability and manipulation. *Journal of Mathematical Economics*, 40(8): 869–887.

Aleksandrov, M.; and Walsh, T. 2019. Strategy-proofness, envy-freeness and Pareto efficiency in online fair division with additive utilities. In *Proceedings of the 16th Pacific Rim International Conference on Artificial Intelligence* (*PRICAI*), 527–541. Springer.

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

Aziz, H.; Brandt, F.; and Seedig, H. G. 2013. Computing Desirable Partitions in Additively Separable Hedonic Games. *Artificial Intelligence*, 195: 316–334.

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

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

Benadè, G.; and Sahoo, N. 2023. Stability, Fairness and the Pursuit of Happiness in Recommender Systems. Technical Report 4241170, SSRN.

Bilò, V.; Fanelli, A.; Flammini, M.; Monaco, G.; and Moscardelli, L. 2018. Nash Stable Outcomes in Fractional Hedonic Games: Existence, Efficiency and Computation. *Journal of Artificial Intelligence Research*, 62: 315–371.

Bilò, V.; Monaco, G.; and Moscardelli, L. 2022. Hedonic Games with Fixed-Size Coalitions. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, 9287–9295.

Boehmer, N.; Bullinger, M.; and Kerkmann, A. M. 2023. Causes of Stability in Dynamic Coalition Formation. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, 5499–5506.

Bogomolnaia, A.; and Jackson, M. O. 2002. The Stability of Hedonic Coalition Structures. *Games and Economic Behavior*, 38(2): 201–230.

Brandt, F.; Bullinger, M.; and Tappe, L. 2022. Single-Agent Dynamics in Additively Separable Hedonic Games. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, 4867–4874.

Brandt, F.; Bullinger, M.; and Wilczynski, A. 2023. Reaching Individually Stable Coalition Structures. *ACM Transactions on Economics and Computation*, 11(1–2): 4:1–65.

Bullinger, M. 2020. Pareto-Optimality in Cardinal Hedonic Games. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 213–221.

Bullinger, M. 2022. Boundaries to Single-Agent Stability in Additively Separable Hedonic Games. In *Proceedings of the 47th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, 26:1–26:15.

Bullinger, M.; and Romen, R. 2023a. Online Coalition Formation under Random Arrival or Coalition Dissolution. In *Proceedings of the 31st Annual European Symposium on Algorithms (ESA)*, 27:1–27:18.

Bullinger, M.; and Romen, R. 2023b. Stability in Online Coalition Formation. Technical report, https://arxiv.org/abs/2312.09119.

Bullinger, M.; and Suksompong, W. 2023. Topological Distance Games. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, 5549–5556.

Carosi, R.; Monaco, G.; and Moscardelli, L. 2019. Local Core Stability in Simple Symmetric Fractional Hedonic Games. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 574–582.

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

Dimitrov, D.; Borm, P.; Hendrickx, R.; and Sung, S. C. 2006. Simple Priorities and Core Stability in Hedonic Games. *Social Choice and Welfare*, 26(2): 421–433.

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

Doval, L. 2022. Dynamically stable matching. *Theoretical Economics*, 17(2): 687–724.

Drèze, J. H.; and Greenberg, J. 1980. Hedonic Coalitions: Optimality and Stability. *Econometrica*, 48(4): 987–1003.

Elkind, E.; Fanelli, A.; and Flammini, M. 2020. Price of Pareto Optimality in hedonic games. *Artificial Intelligence*, 288: 103357.

Feldman, J.; Korula, N.; Mirrokni, V.; Muthukrishnan, S.; and Pál, M. 2009. Online Ad Assignment with Free Disposal. In *Proceedings of the 5th International Conference on Web and Internet Economics (WINE)*, 374–385.

Flammini, M.; Kodric, B.; Monaco, G.; and Zhang, Q. 2021a. Strategyproof Mechanisms for Additively Separable and Fractional Hedonic Games. *Journal of Artificial Intelligence Research*, 70: 1253–1279.

Flammini, M.; Kodric, B.; and Varricchio, G. 2022. Strategyproof mechanisms for friends and enemies games. *Artificial Intelligence*, 302: 103610.

Flammini, M.; Monaco, G.; Moscardelli, L.; Shalom, M.; and Zaks, S. 2021b. On the Online Coalition Structure Generation Problem. *Journal of Artificial Intelligence Research*, 72: 1215–1250.

Gairing, M.; and Savani, R. 2019. Computing Stable Outcomes in Symmetric Additively Separable Hedonic Games. *Mathematics of Operations Research*, 44(3): 1101–1121.

Gajulapalli, K.; Liu, J.; Mai, T.; and Vazirani, V. V. 2020. Stability-preserving, time-efficient mechanisms for school choice in two rounds. In *Proceedings of the 40th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS).*

Gale, D.; and Shapley, L. S. 1962. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1): 9–15.

Huang, Z.; Kang, N.; Tang, Z. G.; Wu, X.; Zhang, Y.; and Zhu, X. 2018. How to Match When All Vertices Arrive Online. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC)*, 17–29.

Huang, Z.; and Tröbst, T. 2023. Online Matching. In Echenique, F.; Immorlica, N.; and Vazirani, V. V., eds., *Online and Matching-Based Market Design*. Cambridge University Press.

Karp, R. M.; Vazirani, U. V.; and Vazirani, V. V. 1990. An optimal algorithm for on-line bipartite matching. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (STOC)*, 352–358.

Morrill, T. 2010. The roommates problem revisited. *Journal* of Economic Theory, 145(5): 1739–1756.

Olsen, M. 2009. Nash Stability in Additively Separable Hedonic Games and Community Structures. *Theory of Computing Systems*, 45: 917–925.

Pavone, M.; Saberi, A.; Schiffe, M.; and Tsao, M. W. 2022. Online hypergraph matching with delay. *Operations Research*, 70(4): 2194 – 2212.

Sung, S. C.; and Dimitrov, D. 2010. Computational Complexity in Additive Hedonic Games. *European Journal of Operational Research*, 203(3): 635–639.

Wang, Y.; and Wong, S. C. 2015. Two-sided online bipartite matching and vertex cover: Beating the greedy algorithm. In *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming (ICALP)*, 1070–1081.

Woeginger, G. J. 2013. A hardness result for core stability in additive hedonic games. *Mathematical Social Sciences*, 65(2): 101–104.

Yao, A. C.-C. 1977. Probabilistic computations: Toward a unified measure of complexity. In *Proceedings of the 18th Symposium on Foundations of Computer Science (FOCS)*, 222–227. IEEE Computer Society Press.