The Thirty-Third International
FLAIRS Conference (FLAIRS-33)

Internal Stability in Hedonic Games

Jacob Schlueter, Judy Goldsmith
University of Kentucky
Lexington, Kentucky
tjacobschlueter@gmail.com, goldsmit@cs.uky.edu

Abstract

We investigate internal stability, a stability notion that has applications in distributed hedonic coalition formation games. We prove that internal stability and Nash stability are equivalent in some classes of hedonic games and different in others. We show that Price of Stability for internal stability is equal to 1 in some cases, but unbounded in others.

Introduction

Many situations exist wherein individuals will choose to act as a group, or coalition. Examples include social clubs, political parties, marriage partner selection, and legislative voting (Woeginger 2013; Bogomolnaia and Jackson 2002; Gale and Shapley 1962). Coalition formation games are a class of cooperative game where the goal is to partition a set of agents into coalitions, according to some criteria. We are interested in a subclass of coalition formation games, hedonic games, which were first proposed by Drèze and Greenberg (1980) and later formalized by Banerjee, Konishi, and Sönmez (2001) and Bogomolnaia and Jackson (2002). Hedonic games are distinguished from general coalition formation games by the requirement that each agent's utility is wholly derived from the members of their own coalition (Drèze and Greenberg 1980; Banerjee, Konishi, and Sönmez 2001; Bogomolnaia and Jackson 2002).

A central problem in hedonic games research, and for coalition formation games in general, is deciding whether or not a proposed set of coalitions, or partition, is *stable* (Woeginger 2013). Several concepts have been introduced to characterize the ways in which a partition is or is not stable; Woeginger's (2013) survey and book chapters by Aziz and Savani (2016) and by Elkind and Rothe (2016) provide overviews of these notions. One stability concept that is of perennial interest in coalition formation games is core stability (Guesnerie and Oddou 1979; Greenberg and Weber 1986; 1993; Demange 1994; Banerjee, Konishi, and Sönmez 2001; Bogomolnaia and Jackson 2002; Dimitrov et al. 2006; Aziz, Brandt, and Harrenstein 2014; Nguyen et al. 2016). Core stability is defined as the lack of agents who are incentivized to leave their assigned coalition(s) and form a *blocking coalition*. This assumes that agents can easily identify others to form a blocking coalition; however, this is not always realistic. Consider the task of partitioning first-semester students into project groups. Some of the assigned groups may find that a subgroup would rather work together, abandoning the rest of their group. Since the students were not acquainted beforehand, they are less likely to coordinate across separate groups. To address this possibility, we focus on the situation where any new groups formed after the semester starts will consist solely of subgroups of previously-existing groups.

Cases where blocking coalitions were confined to subcoalitions that split away from an existing coalition were first considered in the context of hedonic games by Dimitrov et al. (2006) and Alcalde and Romero-Medina (2006) who referred to it as *internal stability*. We extend the notion of an internally stable coalition to partitions, by defining that internal stability holds for a partition of agents into coalitions when no subgroup of any assigned coalition is incentivized to break away. This stability notion has appeared in other hedonic games works (e.g., in Taywade, Goldsmith, and Harrison's (2018) work on decentralized hedonic games, coalitions are only blocked by subcoalitions).

A natural question is what the relationship is between internal stability and Nash stability. We show that Nash stability implies internal stability in some classes of hedonic games, but not in others. While communication channels such as Slack have increased workers' abilities to interact with distant co-workers, the workers may still defect from the larger group and form private sub-channels. In such myopic situations, internal stability seems a more realistic measure of the sustainability of a work-group assignment.

Stability notions focus on outcomes that are likely to occur due to agents' selfish behavior. Stable outcomes may provide the best individual outcomes for all agents, but this is not always the case. The price of stability (PoS) provides a metric to gauge the utility lost in order to achieve stability.

Related Work

Works published in 2006 by Dimitrov et al. (2006) and Alcalde and Romero-Medina (2006) introduce internal stability for coalitions, observing that a singleton coalition is al-

ways internally stable. Alcalde and Romero-Medina (2006) use internally stable coalitions as a tool to investigate conditions that guarantee the existence of core stable partitions in hedonic games. Dimitrov et al. (2006) use internal stability to define another stability notion, deviation stability, which they use to prove the existence of core stable partitions in friend and enemy-oriented hedonic games. Since its introduction, the only other paper that discusses internal stability is a paper by Liu, Tang, and Fang (2014) which adapts it to matching and exchange contexts. One can view internal stability as relevant when agents are *myopic,* meaning that their awareness of the desirability of other agents is limited to those agents that are nearby, namely in the same coalition. Another work that considered agents with limited preference knowledge is Sliwinski and Zick's (2017) notion of PAC-stability, which could be understood as resistance to random attempts by somewhat myopic agents to form blocking coalitions.

Contributions

We introduce here a natural and important extension of Dimitrov et al. (2006) and Alcalde and Romero-Medina's (2006) internal stability. We investigate the relationships of internal stability to other, more common notions, and show that it is distinct from core and Nash stability, for multiple types of hedonic games. We investigate the prices of anarchy and stability for internal stability with respect to several hedonic games. This work gives us insight into a key stability notion when agents are myopic.

Preliminaries

Hedonic coalition formation games are driven by agents' preferences over coalitions. As there are exponentially more coalitions than agents, these preferences need to be expressed succinctly, restricting the expressivity of representations. We review some ways preferences can be constrained, which are relevant to our work.

Types of Hedonic Games

Below, we outline several classes of hedonic games. In each class, a game G consists of a finite set of n agents N, with preference set $P = \{P_i : i \in N\}$, where P_i is the preference of each agent i over partitions of N into coalitions. P_i may exhaustively list the preferences of agent i or provide a succinct representation from which preferences are derived. When preferences are given as utilities, we assume that $u_i(\{i\}) = 0$: for each i, the utility of agent i for being in a coalition of size 1 is 0.

Definition 1. ***Hedonic games*** *(Banerjee, Konishi, and Sönmez 2001; Bogomolnaia and Jackson 2002) are coalition formation games with nontransferable utility wherein players' preferences are concerned only with their own coalition. This inherently self-interested means of determining utility makes such games hedonic in nature.*

Let $\mathcal{N}_i$ be the set of possible coalitions containing agent $i \in N$. A preference ordering of $\mathcal{N}_i$ is derived from the preference set $P_i \in P$. A solution for a game is a partition π, which is contained in the set of all distinct partitions Γ. Each player $i \in N$ has preferences over all partitions $\pi \in \Gamma$ based on their assigned coalition in each π.

Additively Separable Hedonic Games (ASHG) (Banerjee, Konishi, and Sönmez 2001) are a class of hedonic games where each agent $i \in N$ assigns values to each agent $j \in N$, expressed as $v_i(j)$; $v_i(i)$ is always set to 0. The utility an agent derives from each $S \in \mathcal{N}_i$ is defined as $u_i(S) = \sum_{j \in S} v_i(j)$.

Friend-oriented Hedonic Games (FOHGs) (Dimitrov et al. 2006) are a subclass of ASHGs games where each agent regards all other agents as either a friend or an enemy. FOHGs are often represented by graphs where an edge from some agent $i \in N$ to another agent $j \in N$ indicates that i regards j as a friend. Lack of an edge from i to j indicates that i regards j as an enemy. Utility for each agent is the sum of values they assign to other agents, friends being assigned a value of n while enemies are valued at -1.

An **Enemy-oriented Hedonic Game (EOHG)** follows the same basic principles of FOHGs, but gives friends a value of 1 and enemies a value of $-n$.

We use "general ASHG" when we're not restricting attention to special cases such as FOHGs or EOHGs.

Fractional Hedonic Games (FHGs) (Aziz, Brandt, and Harrenstein 2014) are a class of Hedonic Games where agents assign values to each other agent. In contrast to ASHGs, Fractional Hedonic Games define utility as an average rather than a sum: $u_i(S) = (\sum_{j \in S} v_i(j))/|S|$.

An **Altruistic Hedonic Game (AHG)** (Nguyen et al. 2016) is a hedonic game in which agents derive utility from both their own basic preferences and those of any friends in the same coalition.

Let each agent $i \in N$ have utility u_i, and let i partition other agents into friends and enemies, given by F_i, E_i. Three levels of altruism are considered in AHGs: *selfish-first*, *equal treatment*, and *altruistic first*. The function used to determine an agent's utility depends on their altruism level and on pre-utility preference values calculated as the utility agents would have from some coalition C in a FOHG based on the same graph ($n|C \cap F_i| - |C \cap E_i|$). Two of these functions utilize a weight parameter of $M = n^5$ to ensure that one of the terms in the equation dominates the other. This weight value is the smallest whole number exponent of n which guarantees this for both equations that make use of M.

Selfish-First: agents prioritize their own preferences, but use the preferences of others to break ties. $u_i = M(n|C \cap F_i| - |C \cap E_i|) + \sum_{a \in C \cap F_i} \frac{n|C \cap F_a| - |C \cap E_a|}{|C \cap F_i|}$

Equal Treatment: all preferences are treated equally. $u_i = \sum_{a \in C \cap (F_i \cup \{i\})} \frac{n|C \cap F_a| - |C \cap E_a|}{|C \cap (F_i \cup \{i\})|}$

Altruistic First: agents prioritize the preferences of others, but use their own preferences to break ties. $u_i = n|C \cap F_i| - |C \cap E_i| + M \cdot \sum_{a \in C \cap F_i} \frac{n|C \cap F_a| - |C \cap E_a|}{|C \cap F_i|}$

Super Altruistic Hedonic Games (SAHGs) (Schlueter and Goldsmith 2020) extend the central principle of AHGs so agents consider the preferences of all agents in their coalition. Agents weight their consideration of each other's preferences according to some polynomially computable value.

Let parameters (a, g, M, L) be non-negative weights

where a and g represent the weights associated with friends and enemies, respectively, while M and L represent the weights associated with personal preference and the weighted average of others' preferences. Next, let $D(i,j)$ be a polynomial-time computable function that is non-increasing with the graph distance between i and j. Let the number of other agents in coalition C_i be $h_i = |C_i \setminus \{i\}|$. For each agent $i \in N$, let that agent's base preference be $b_i = a|C_i \cap F_i| - g|C_i \cap E_i|$, and let their utility be $u_i = Mb_i + L\sum_{j \in C_i \setminus \{i\}} \frac{D(i,j) \cdot b_j}{h_i}$. If $C_i = \{i\}$ then the sum is set to 0. The default definition of D is the inverse graph distance function: for any pair of agents $i, j \in N : i \neq j$, let d_{ij} be the shortest path distance between them, then let $D(i,j) = 1/d_{ij}$. The **total utility** of a partition π is given by $U_T = \sum_{i \in N} u_i$.

A **Role Based Hedonic Game (RBHG)** (Spradling and Goldsmith 2015) instance consists of a population of agents P, a set of roles A, a set of available team compositions C, where a composition $c \in C$ is a multiset (bag) of roles from R, and $U = P \times R \times C \rightarrow Z$ define the utility function $u_i(r, c)$ for each player p_i. We assume that for all $p_i \in P$ and for all $r \in R$, $u_i(r, \{r\}) = 0$.

An instance of the **Group Activity Selection Problem (GASP)** (Darmann et al. 2012) is given by a set of agents $N = \{1, ..., n\}$, a set of activities $A = A^* \cup \{a_\emptyset\}$, where $A^* = \{a_1, ..., a_p\}$, and a *profile* P, which consists of n *votes* (one for each agent): $P = (V_1, ..., V_n)$. The vote of agent i describes their preferences over the set of *alternatives* $X = X^* \cup \{a_\emptyset\}$, where $X^* = A^* \times \{1, ..., n\}$; alternative (a, k), $a \in A^*$, is interpreted as "activity a with k participants," and $a_\emptyset$ is the *void* activity.

The vote V_i of an agent $i \in N$ (also denoted by $\succeq_i$) is a weak order over X^*; its induced strict preference and indifference relations are denoted by $\succ_i$ and $\sim_i$ respectively. We set $S_i = \{(a, k) \in X^* | (a, k) \succ_i a_\emptyset\}$; we say that voter i *approves* of all alternatives in S_i, and refer to the set S_i as the *induced approval vote* of voter i.

In some cases, we consider both *symmetric* and *asymmetric* games, where the (a)symmetry refers to preferences. Note that we exclusively consider symmetric FOHGs and EOHGs. In some hedonic games, including GASPs and RBHGs, players do not express preferences over other individuals, so the distinction does not apply.

Stability

One of the major topics of hedonic games is *stability*, the idea that a partition will not be disrupted by individuals deviating from their assigned coalitions. There are many sets of constraints placed on such disruptions such as the number of agents that can move simultaneously, whether all moving agents must see an increase in utility, whether agents left behind must see their utility increase, or whether agents being joined must not see their utility decrease.

We next define stability notions used here. In these definitions π is a partition composed of a set of k disjoint coalitions $\{C_1, C_2, ...C_k\}$. The term $\pi(i)$ refers to the coalition $C \in \pi$ such that $i \in C$.

Definition 2. *A partition is **individually rational** if no individual agent can improve their utility by leaving their current coalition to become a singleton.*

Individual rationality is a precondition for many other stability notions, such as Nash stability.

Definition 3. *(Bogomolnaia and Jackson 2002) A partition is **Nash stable** if no individual agent can improve their utility by deviating from their current coalition to join another coalition or to become a singleton.*

Nash stability focuses only on the selfish behavior of an individual. Bogomolnaia and Jackson (2002) build on work by Greenberg and Weber (1993), and by Demange (1994), proposing blocking coalitions and core stability to examine whether groups can benefit from cooperative deviation.

Definition 4. *(Bogomolnaia and Jackson 2002) A coalition **blocks** a partition if all agents in the coalition prefer it over their current coalitions; formally, given coalition C and blocking coalition C_b, $\forall i \in C_b \subset C : C_b \succ_i C$.*

*A partition is **core stable** if no coalition blocks it.*

Internal stability considers group deviations where all deviating agents must come from the same coalition.

Definition 5. *A coalition C is **internally stable** if there is not subset $D \subset C$ such that all of the agents of D are better off leaving C and forming a new coalition (Dimitrov et al. 2006; Alcalde and Romero-Medina 2006).*

*A partition π is **internally stable** if all coalitions $C \in \pi$ are internally stable.*

Price of Stability

Notions of stability are useful tools for predicting outcomes. However, there may be costs that result from the imposition of a given stability notion. Further, a given stability notion may admit good outcomes, but may also permit particularly subpar outcomes as well. The notions of price of stability and price of anarchy formally define these ideas.

Definition 6. *For stability notion X, the **price of stability (PoS_X)** is the ratio between the overall utility-maximizing partition and the utility-maximizing X-stable partition. For internal stability, we use the representation PoS_{IS}.*

Relationship to Core and Nash Stability

It might seem, at first, that internal stability is very similar to Nash stability, or perhaps to core stability. We investigate the relationship of internal stability to Nash stability and core stability for a variety of hedonic games.

Observation 1 immediately follows from Dimitrov et al.'s (2006) observation that all coalitions in a core stable partition are internally stable.

Observation 1. *Core stability implies internal stability for all classes of hedonic games.*

Note that all singleton partitions (i.e., each agent in a coalition of size 1) are trivially internally stable. However, in most instances of most hedonic games, a singleton partition will neither be core stable nor Nash stable. Thus, internal stability does not, generally, imply core stability or Nash stability. Theorem 1 shows our first step in analyzing the relationship between Nash stability and internal stability.

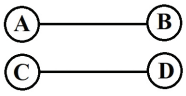

Figure 1: Nash stable, not internally stable

Theorem 1. *Nash stability does not guarantee internal stability in FOHGs.*

Proof. Consider a graph $G = (V, E)$ with $V = \{A, B, C, D\}$ and $E = \{(A, B), (C, D)\}$. (See Figure 1.) Consider a FOHG based on this graph such that the set of agents $N = V$ and each edge $(i, j) \in E$ defines a mutual friendship. All pairs $(i, j) \notin E$ indicate mutual enmity.

Now consider the grand coalition for this FOHG. Each agent $i \in N$ gains 4 utility from the presence of their mutual friend, but also loses 2 utility due to the 2 enemies present. This gives each agent i a net utility of 2, making the grand coalition individually rational. The only deviation a single agent can make to the grand coalition is to leave the coalition and become a singleton, thereby earning 0 utility. Thus, the grand coalition is Nash stable. If we view the grand coalition from the viewpoint of internal stability, however, we see that for pairs of agents (A, B) and (C, D), each agent pair can leave the grand coalition and increase the utility of both agents from 2 to 4. Thus, the grand coalition is not internally stable and we conclude that not all Nash stable partitions are internally stable. □

Since FOHGs are a subclass of ASHGs, Corollary 1 immediately follows from Theorem 1.

Corollary 1. *Nash stability does not imply internal stability for ASHGs.*

While internal stability is not implied by Nash stability in general case ASHGs, there are subclasses where Nash stability does imply internal stability. We show that enemy-oriented hedonic games are one such subclass; in particular, we show that not only does Nash stability imply internal stability, but individual rationality is sufficient to imply internal stability.

Lemma 1. *Individual rationality guarantees internal stability in EOHGs.*

Proof. By the definition of EOHGs the only individually rational coalitions for some agent $i \in N$ are those which contain none of i's enemies. Equivalently, all individually rational coalitions can be described by cliques in a graph of friendship relations.

Since all agents in an individually rational coalition are friends with each other, no subset of agents in such a coalition can increase their utility by leaving. Therefore, all individually rational coalitions are internally stable. A partition is only individually rational if all its coalitions are individually rational and, therefore, internally stable. Thus, all individually rational partitions are internally stable. □

Theorem 2. *Nash stability guarantees internal stability for EOHGs.*

Proof. Because all Nash stable partitions are individually rational, it follows from Lemma 1 that Nash stability implies internal stability. □

Theorems 1 and 2 show that the relationship between Nash and internal stability varies between subclasses of ASHGs. We expand our understanding of the relationship between Nash and internal stability beyond the scope of ASHGs in Theorem 3 by analyzing fractional hedonic games.

Theorem 3. *Nash stability does not guarantee internal stability in fractional hedonic games.*

Proof. Construct a fractional hedonic game, G, with agents $\{A, B, C, D\}$ with $u_A(B) = u_B(A) = 4$ and $u_X(Y) = 1$ for all other $X \neq Y$. Consider the grand coalition, S. Then $u_A(S) = u_B(S) = \frac{4+1+1}{4} = \frac{6}{4}$ and $u_C(S) = u_D(S) = \frac{1+1+1}{4} = \frac{3}{4}$. This coalition is individually rational, since each agent has positive utility. It is also Nash Stable, since an agent's only defection option is to leave S and form a singleton coalition, with utility 0. However, the coalition $\{A, B\}$ provides utility 2 for each of A and B, so S is not internally stable. Therefore, for FHGs, Nash stability does not imply internal stability. □

In Theorem 4 we clarify the relationship between Nash and internal stability in altruistic hedonic games.

Theorem 4. *Nash stability does not guarantee internal stability for any of the three altruism levels of altruistic hedonic games (AHGs).*

Proof. Consider an AHG based with agents $\{A, B, C, D\}$. Let $\{A, B\}$ and $\{C, D\}$ be mutual friends that regard all other agents as enemies. Since there are 4 agents in this game, we set $M = n^5 = 4^5 = 1024$.

In the grand coalition all agents achieve a utility of 2050 in the *Selfish-First* and *Altruistic First* paradigms and 2 in *Equal Treatment*. Individual agents can only deviate by becoming a singleton, so the grand coalition is Nash stable for all three paradigms.

Now consider if $\{A, B\}$ or $\{C, D\}$ broke away; the agents breaking away derive a utility of 5000 in the *Selfish-First* and *Altruistic First* paradigms and 4 in the *Equal Treatment* paradigm. The grand coalition is not internally stable for any of the three paradigms, because both $\{A, B\}$ and $\{C, D\}$ have incentive to break away. □

Because Schlueter and Goldsmith (2020) showed that selfish-first AHGs are special cases of Super Altruistic Hedonic Games, Corollary 2 follows from Theorem 4.

Corollary 2. *Nash stability does not imply internal stability in Super Altruistic Hedonic Games.*

The proofs for Theorems 1–4 also give Observation 2.

Observation 2. *If there exists an instance of a coalition formation game such that the grand coalition is individually rational, but not internally stable, then Nash stability does not imply internal stability for that class of games.*

The games we have examined thus far assume that agents' utility is based directly on the other agents in their coalition. In Role Based Hedonic Games, utility is based instead on an agent's assigned role and the team's role composition.

Theorem 5. *Nash stability does not imply internal stability in RBHGs.*

Proof. Consider an RBHG instance with the following setup: $P = \{p_1, p_2, p_3, p_4\}$, $R = \{r_1, r_2\}$, $C = \{\{r_1, r_2\}, \{r_1, r_1, r_1, r_1\}, \{r_1\}, \{r_2\}\}$. Now $\forall i \in P$ let $u_i(r_1, \{r_1, r_1, r_1, r_1\}) = 1$ and $\forall (r, c) \in R \times C$ let $u_{p_1}(r, c) = u_{p_3}(r, c)$ and let $u_{p_2}(r, c) = u_{p_4}(r, c)$. Let $u_{p_1}(r_1, \{r_1, r_2\}) = 2$ and $u_{p_1}(r_2, \{r_1, r_2\}) = -1$. Let $u_{p_2}(r_1, \{r_1, r_2\}) = -1$ and $u_{p_2}(r_2, \{r_1, r_2\}) = 2$. As usual for singletons, $\forall i \in P$ let $u_i(x, \{x\}) = 0$ for $x = r_1, r_2$.

Now let a partition π form where all four agents are put in the grand coalition with the composition $\{r_1, r_1, r_1, r_1\}$. Since every role in the composition is r_1, all agents derive a utility of 1 from this partition. The only way an individual can deviate from this partition is to leave to become a singleton of either role r_1 or r_2; in either case, agents derive zero utility as a singleton. Since the only deviations available to individuals will reduce their utility from 1 to 0, the partition of all agents in the grand coalition with composition $\{r_1, r_1, r_1, r_1\}$ is Nash stable.

While the grand coalition with composition $\{r_1, r_1, r_1, r_1\}$ is Nash stable, it is not internally stable. Either agent p_1 or p_3 could join agent p_2 or p_4 and split away to form a new coalition with composition $\{r_1, r_2\}$ with p_1 or p_3 taking role r_1 and p_2 or p_4 taking role r_2; doing this would increase the utility of both deviating agents from 1 to 2. Thus, Nash stability does not imply internal stability in RBHGs. □

Roles and Teams Hedonic Games (Spradling et al. 2013) are a subclass of RBHGs that impose a strict team size rule that makes it impossible for a subset of a valid existing team to break away to form a new, valid team.

Observation 3. *All valid partitions are internally stable in RTHGs.*

In instances of the Group Activity Selection Problem, agents derive utility from the selected activity and the size of their coalition.

Theorem 6. *Nash stability does not imply internal stability in GASPs.*

Proof. Consider a GASP instance where there are n agents, and activity set $A = \{a_1, a_\emptyset\}$ and $\forall i \in N$ $(a_1, n-2) \succ_i (a_1, n) \succ_i a_\emptyset$; all alternatives not included in the preference profile are seen as worse than the void activity.

Now consider the case where all n agents form a single coalition to participate in activity a_1. We can see from the preference profile shared by all agents that this outcome is preferable to the void activity and to participating in activity a_1 alone. Since the only way for an individual to deviate from this outcome is to leave and do nothing, or participate in a_1 by themselves, no individual agent has incentive to deviate from this outcome. Thus, the outcome is Nash stable.

From the perspective of internal stability, we see that any subset of $n-2$ agents could break away and achieve a more favorable outcome. Thus, the grand coalition participating in a_1 is not a internally stable outcome. □

Price of Stability (PoS)

Theorem 7. *$PoS_{IS}(G)$ is unbounded in hedonic games where agents assign asymmetric values to each other. This includes general case ASHGs and FHGs.*

Proof. Consider a hedonic game with the following setup: $N = \{1, 2, 3\}$, $v_1(2) = 10$, $v_1(3) = -1$, $v_2(1) = -1$, $v_2(3) = -1$, $v_3(1) = -1$, $v_3(2) = 10$. In both ASHGs and FHGs, the utility is maximized by the grand coalition, but the only stable partition is the partition of singletons. Since the partition of singletons has a sum utility of zero, PoS$= \infty$. This scenario can be adapted to any hedonic game where agents derive utility from potentially asymmetric values assigned to each other. □

Theorem 8. *PoS_{IS} for symmetric ASHGs is 1.*

Proof. Consider a symmetric ASHG. Consider an optimal coalition A and suppose that it contains a blocking subcoalition $B \subset A$. Note that the total utility for A is the sum of utilities of agents within B, agents in $A \setminus B$, and utilities (in each direction) between B and $A \setminus B$: $\sum_{x \in B} u_x(B) + \sum_{y \in A \setminus B} u_y(A \setminus B) + 2 \cdot \sum_{x \in B, y \in A \setminus B} u_x(y)$. Since B is a blocking subcoalition $\sum_{x \in B} u_x(B) > \sum_{x \in B} u_x(A) = \sum_{x \in B} u_x(B) + \sum_{x \in B, y \in A \setminus B} u_x(y)$. Therefore, $\sum_{x \in B, y \in A \setminus B} u_x(y) < 0$.

Thus, the total utility for the partition $B, A \setminus B$ has utility $\sum_{x \in B} u_x(B) + \sum_{y \in A \setminus B} u_y(A \setminus B) > \sum_{x \in B} u_x(B) + \sum_{y \in A \setminus B} u_y(A \setminus B) + 2 \cdot \sum_{x \in B, y \in A \setminus B} u_x(y)$, contradicting the optimality of A.

Thus, any optimal coalition has maximum utility over sub-partitions, and is internally stable. Therefore, $\text{PoS}_{\text{IS}}(G) = 1$ for all symmetric ASHGs. □

Theorem 9. *PoS_{IS} for symmetric FHGs is bounded by 2.*

Proof. We define an FHG, G, which maximizes the ratio between the utility-maximizing partition and the utility-maximizing internally stable partition. If $\text{PoS}_{\text{IS}}(G) > 1$, then the utility-maximizing partition is not internally stable. Without loss of generality, we consider a single coalition. (The maximum PoS will occur when each coalition is split; since we take the average utility over all agents, it suffices to consider a single coalition A to find the maximum PoS as if that were the grand coalition.) Let $B \subset A$ be a blocking coalition. We also assign $a = |A|$, $b = |B|$, and $r = a - b$. Keep in mind that, for an agent in B, its average utility in B is higher than its average utility in A. Because, for FHGs, agents' utilities for other agents in their coalition are averaged, and agent's utilities for others need only distinguish three cases. We define v_b as the value agents in B assign to each other; $\forall i, j \in B, v_i(j) = v_j(i) = v_b$. We define v_r as the value agents in $A \setminus B$ assign to each other; $\forall i, j \in A \setminus B, v_i(j) = v_j(i) = v_r$. We define v_a as the

value agents in B assign to agents in $A \setminus B$ and vice-versa; $\forall i \in B, j \in A \setminus B, v_i(j) = v_j(i) = v_a$.

In order for agents in B to be incentivized to break away from the rest of A, the value each agent in B receives in B (which approaches v_b as $|B|$ grows) is greater than the value it receives in A, which is $\frac{(b-1)v_b + rv_r}{b+r}$. This implies that $v_b > v_r$. We see that the grand coalition maximizes the sum of agents' utilities when $v_b - \frac{1}{b} + v_r - \frac{1}{r} < 2v_a$. We are able to show that the ratio between $v_b : 2v_a$ is a hard upper bound on the ratio between the sum utilities of the grand coalition and $\{A \setminus B, B\}$, allowing us to construct examples G where the the $\text{PoS}_{\text{IS}}(G)$ is arbitrarily close to 2. □

Related to PoS is the price of anarchy (PoA), which gauges the potential loss of utility caused by agents' selfish behavior; PoA is defined as the ratio of the highest-utility partition over the lowest-utility stable partition. Since the partition into singletons is vacuously internally stable, we have $\text{PoA}_{\text{IS}}(G) = \infty$ for any game where singletons derive zero utility.

Conclusions

We have extended the notion of internal stability for coalitions (Dimitrov et al. 2006; Alcalde and Romero-Medina 2006) to a partition stability notion. Internal stability is important whenever agents have a local view of their preferences. The relationship of internal stability to other, more commonly investigated notions turns out to depend on the particularity of preference representation used in the hedonic game. For instance, any individually rational hedonic game has an internally stable partition of singletons; finding others is expected to be more difficult computationally.

We showed that internal stability is, in some hedonic game types, not equivalent to Nash or core stability. We have seen instances where the price of stability for internal stability is bounded, and where it is unbounded. We predict that this notion will continue to generate interesting insight into locally-aware coalition formation games.

References

Alcalde, J., and Romero-Medina, A. 2006. Coalition formation and stability. *Social Choice and Welfare* 27(2):365–375.

Aziz, H., and Savani, R. 2016. *Handbook of computational social choice*. Cambridge University Press. chapter 15, 356–376.

Aziz, H.; Brandt, F.; and Harrenstein, P. 2014. Fractional hedonic games. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 5–12. International Foundation for Autonomous Agents and Multiagent Systems.

Banerjee, S.; Konishi, H.; and Sönmez, T. 2001. Core in a simple coalition formation game. *Social Choice and Welfare* 18(1):135–153.

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

Darmann, A.; Elkind, E.; Kurz, S.; Lang, J.; Schauer, J.; and Woeginger, G. 2012. Group activity selection problem. In *International Workshop on Internet and Network Economics*, 156–169. Springer.

Demange, G. 1994. Intermediate preferences and stable coalition structures. *Journal of Mathematical Economics* 23(1):45–58.

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.

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

Elkind, E., and Rothe, J. 2016. *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*. Springer Texts in Business and Economics. chapter 3.

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

Greenberg, J., and Weber, S. 1986. Strong tiebout equilibrium under restricted preferences domain. *Journal of Economic Theory* 38(1):101–117.

Greenberg, J., and Weber, S. 1993. Stable coalition structures with a unidimensional set of alternatives. *Journal of Economic Theory* 60(1):62–82.

Guesnerie, R., and Oddou, C. 1979. *Second Best Taxation as a Game*. Centre d'études prospectives d'économie mathématique appliquées à la planification.

Liu, Y.; Tang, P.; and Fang, W. 2014. Internally stable matchings and exchanges. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Nguyen, N.-T.; Rey, A.; Rey, L.; Rothe, J.; and Schend, L. 2016. Altruistic hedonic games. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 251–259. International Foundation for Autonomous Agents and Multiagent Systems.

Schlueter, J., and Goldsmith, J. 2020. Super altruistic hedonic games. In *The Thirty-Third International Florida Artificial Intelligence Research Society Conference*.

Sliwinski, J., and Zick, Y. 2017. Learning hedonic games. In *Proceedings of the 2017 International Joint Conferences on Artificial Intelligence*.

Spradling, M., and Goldsmith, J. 2015. Stability in role based hedonic games. In *The Twenty-Eighth International Florida Artificial Intelligence Research Society Conference*.

Spradling, M.; Goldsmith, J.; Liu, X.; Dadi, C.; and Li, Z. 2013. Roles and teams hedonic game. In *International Conference on Algorithmic Decision Theory*, 351–362. Springer.

Taywade, K.; Goldsmith, J.; and Harrison, B. 2018. Decentralized multiagent approach for hedonic games. In *16th European Conference on Multi-Agent Systems*.

Woeginger, G. J. 2013. Core stability in hedonic coalition formation. In *Software Seminar*, 33–50. Springer.