Robustness against Agent Failure in Hedonic Games

Ayumi Igarashi[1] , **Kazunori Ota**[2] , **Yuko Sakurai**[3]**,** and **Makoto Yokoo**[2]
[1]University of Tokyo, Tokyo, Japan,
[2]Kyusu University, Fukuoka, Japan,
[3]National Institute of Advanced Industrial Science and Technology, Tokyo, Japan
igarashi@mist.i.u-tokyo.ac.jp, {ota,yokoo}@inf.kyushu-u.ac.jp, yuko.sakurai@aist.go.jp

Abstract

We study how stability can be maintained even after any set of at most k players leave their groups, in the context of hedonic games. While stability properties ensure an outcome to be robust against players' deviations, it has not been considered how an unexpected change caused by a sudden deletion of players affects stable outcomes. In this paper, we propose a novel criterion that reshapes stability form robustness aspect. We observe that some stability properties can be no longer preserved even when a single agent is removed. However, we obtain positive results by focusing on symmetric friend-oriented hedonic games. We prove that we can efficiently decide the existence of robust outcomes with respect to Nash stability under deletion of any number of players or contractual individual stability under deletion of a single player. We also prove that symmetric additively separable games always admit an individual stable outcome that is robust with respect to individual rationality.

1 Introduction

Coalition formation is everywhere in human activities. Companies group their employers into project teams. Countries form coalitions to promote international trade among them. Individuals interact with each other and form groups in order to achieve objectives they cannot seek for on their own.

Hedonic coalition formation games (for short, hedonic games), introduced by Bogomolnaia and Jackson [2002] and Banerjee *et al.* [2001], provide an elegant framework to formulate coalition formation. In these games, each player has a preference over the coalitions to which she belongs, and desirable outcomes often correspond to *stable* partitions. The basic intuition behind stable partitioning is that group structures need to be robust under certain changes *within* the system; that is, outcomes must be immune to players' coalitional or individual deviations to other coalitions.

In many real-world scenarios, however, groups may encounter unexpected changes and challenges, imposed from the *outside* of the system. For instance, a certain country can go bankrupt and be enforced to leave a political alliance. In this respect, a group structure that satisfies a standard stability requirement can become immediately unstable due to unexpected circumstances. A case in point is a political coalition of three countries with one intermediate country connecting two other countries who are enemies to each other: if the intermediate player happens to disappear from the coalition, one cannot maintain the stability of the whole system.

In this paper, we propose a novel criterion that redefines stability from robustness aspect. We define an outcome to be *robust* with respect to a certain stability requirement α if removing any set of at most k players still preserves α. Besides the preceding example of a political alliance, there are several applications of hedonic games, such as project team formation [Okimoto *et al.*, 2015], research team formation [Alcalde and Revilla, 2004], and group activity selection [Darmann *et al.*, 2017], in which unexpected players' non-participation may severely affect stability of the system. To the best of our knowledge, however, no attempt has been ever made to connect two important considerations, robustness and stability. Our goal is to make the first step filling this gap.

We focus on friend-oriented games, introduced by Dimitrov *et al.* [2006], where players' preferences are succinctly encoded via the binary friendship relations. While it is known that such games always guarantee the existence of stable outcomes, we observe that a simple example of one player connecting two enemies shows impossibility in maintaining most of the stability properties, such as *core stability*, *Nash stability*, *individual stability*, and *contractual individual stability*. Not surprisingly, this negative result holds even under a very small change of the system, i.e., only a single player can disappear.

Given these non-existence results, we investigate the computational complexity of deciding the existence of a robust outcome in a symmetric friend-oriented game. Specifically, we show that we can efficiently decide the existence of an outcome that is robust with respect to Nash stability, irrespective of the number of players leaving the game. We then prove that any symmetric friend-oriented game admits a polynomial time algorithm that finds a robust outcome with respect to contractual individual stability in case of removing a single player. To this end, we obtain a non-trivial characterization of games whose corresponding robust outcomes are non-empty. Moreover, we complement this result by showing that the problem becomes NP-hard when $k = 2$. We also show that the positive results do not extend to individual stability:

NS-robustness	polytime (Cor. 1)
IS-robustness	NP-complete ($k = 1$) (Th. 6)
CIS-robustness	polytime ($k = 1$) (Th. 4)
	NP-complete ($k = 2$) (Th. 5)
IS & IR-robustness	exists and polytime (Th. 8)

Table 1: Overview of our complexity results in a symmetric friend-oriented game, where k is the maximum number of players who can leave the entire game.

we prove that the associated problem for individual stability is NP-hard even when only a single player is allowed to leave.

Finally, we consider the question of whether a minimum stability requirement, *individual rationality*, can be maintained while ensuring that an outcome of a game itself satisfies stronger stability desiderata. It turns out that when players have symmetric additively separable preferences, an individually stable partition which is robust with respect to individual rationality always exist. Our complexity results are summarized in Table 1.

1.1 Related Work

A similar notion of robustness appears in a variety of contexts ranging from multi-agent systems to graph theory. In particular, the robustness concept adapted in this paper is close to the notion of *fault tolerance* in the theory of distributed systems. We refer the reader to the work of Fedoruk and Deters [2002] for an overview on fault tolerance. Recent works of Kouvaros and Lomuscio [2017] and Kouvaros *et al.* [2018] also considered a fault-tolerance problem in the context of temporal epistemic specifications.

In cooperative games with transferable utility, several papers studied robustness against agent failures. Bachrach *et al.* [2011] proposed the reliability extension of cooperative games, where each agent has an independent failure probability. Okimoto *et al.* [2015] introduced a similar concept to ours, so-called, *k-robustness* for team formation problems; under their definition, each team needs to accomplish their task even after k agents fail. Further, we note that our definition of robustness resembles some graph connectivity concepts, such as the k-vertex-connectivity, capturing the robustness of a given network (see, e.g., Schrijver [2003]).

Our work is related to the rich body of the literature on the study of hedonic games [Bogomolnaia and Jackson, 2002; Dimitrov *et al.*, 2006; Aziz *et al.*, 2014; 2017]. The relation between stability and the networks capturing agents' preferences has been fairly well-explored, in which the nodes of a graph represent players and edges correspond to the degree of preference [Bilò *et al.*, 2014; Peters, 2016; Igarashi and Elkind, 2016].

Full version. The full version of the paper is available on arXiv [Igarashi *et al.*, 2019]. It contains full proofs of Lemma 1, Corollary 1, and Theorems 2, 3, 4, 6, 7, 8.

2 Preliminaries

For a natural number $s \in \mathbb{N}$, we write $[s] = \{1, 2, \dots, s\}$. A hedonic game is defined as a pair $(N, (\succeq_i)_{i \in N})$ where $N = [n]$ is a finite set of *players* and each $\succeq_i$ is a preference over the subsets of N (also referred to as *coalitions*); specifically for every $i \in N$, we let $\mathcal{N}_i$ denote the collection of all coalitions containing i; each $\succeq_i$ describes a complete and transitive preference over the sets in $\mathcal{N}_i$. Let $\succ_i$ denote the strict preference derived from $\succeq_i$, i.e., $S \succ_i T$ if $S \succeq_i T$, but $T \not\succeq_i S$. For $i \in N$ and $S, T \in \mathcal{N}_i$, we say that player i *strictly prefers* a coalition S to another coalition T if $S \succ_i T$; player i *weakly prefers* S to T if $S \succeq_i T$. We call a coalition $S \subseteq N$ *individually rational* if every player $i \in S$ weakly prefers S to $\{i\}$.

A preference profile $(\succeq_i)_{i \in N}$ is said to be *additively separable* if there exists a *weight function* $w : N \times N \to \mathbb{R}$ such that for each $i \in N$ and each $S, T \in \mathcal{N}_i$ we have $S \succeq_i T$ if and only if $\sum_{j \in S} w(i, j) \geq \sum_{j \in T} w(i, j)$ [Bogomolnaia and Jackson, 2002]; we will assume that $w(i, i) = 0$ for each $i \in N$. An additively separable preference is said to be *symmetric* if the weight function $w : N \times N \to \mathbb{R}$ is symmetric, i.e., $w(i, j) = w(j, i)$ for all $i, j \in N$. We use the notation (N, w) to denote an additively separable game with weight function $w : N \times N \to \mathbb{R}$. For additively separable games, each player can consider every other player to be either a friend, a neutral player, or an enemy; specifically, for each pair of distinct players $i, j \in N$, we say that j is a *friend* of i if $w(i, j) > 0$, and j is an *enemy* of i if $w(i, j) < 0$.

Dimitrov *et al.* [2006] introduced a subclass of additively separable preferences, which they called *friend-oriented preferences*. Under friend-oriented preferences, each player has strong favour towards her friends: $w(i, j) \in \{n, -1\}$ for each $i, j \in N$ with $i \neq j$. For a symmetric additively separable game (N, w), let G_w denote the *friendship graph* where the set of vertices is given by the set of players and two players i, j are adjacent if and only if they are friends; each coalition S is said to have *minimum degree t* if each player in S has at least t other friends in S.

An *outcome* of a hedonic game is a partition of players into disjoint coalitions. Given a partition π of N and a player $i \in N$, let $\pi(i)$ denote the unique coalition in π that contains i. Much of the existing literature is concerned with outcomes that satisfy certain stability requirements. A minimum stability property we require is *individual rationality*. A partition π of N is said to be *individually rational* (IR) if each player prefers their coalition to staying alone, i.e., all coalitions in π are individually rational. If we extend this to a group deviation, we obtain the definition of the *core*. Specifically, a coalition $S \subseteq N$ *strongly blocks* a partition π of N if every player $i \in S$ strictly prefers S to her own coalition $\pi(i)$. A partition π of N is said to be *core stable* (CR) if no coalition $S \subseteq N$ strongly blocks π. We also consider deviations based on individual movements. Specifically, consider a player $i \in N$ and a pair of coalitions $S \notin \mathcal{N}_i$, $T \in \mathcal{N}_i$. A player $j \in S$ *accepts* a deviation of i to S if j weakly prefers $S \cup \{i\}$ to S; a player $j \in T$ *accepts* a deviation of i from T if j weakly prefers $T \setminus \{i\}$ to T. A deviation of i from T to S is an *NS-deviation* if i strictly prefers $S \cup \{i\}$ to T, an *IS-deviation* if it is an NS-deviation and all players in S accept it, and a *CIS-deviation* if it is an IS-deviation and all players in T accept it. A partition π is called *Nash stable* (NS) (respectively, *individually stable* (IS) and *contractually individually stable* (CIS)) if no player $i \in N$ has an NS-deviation (respectively, an IS-deviation and a CIS-deviation) from $\pi(i)$ to another coalition $S \in \pi$ or to $\emptyset$.

Trivially, Nash stability implies individual stability, which also implies contractually individual stability. Usually, core stability does not imply the stability based on individual deviations. However, we note that for a symmetric friend-oriented game, core stability implies individual stability. Also, contractual individual stability implies individual rationality with symmetric friend-oriented preferences.

Lemma 1 *Let (N, w) be a symmetric friend-oriented game (N, w). Then core stability implies individual stability. Further, contractually individual stability implies individually rationality.*

3 Agent Failure in Hedonic Games

Earlier we defined the robustness informally: A sudden deletion of players upon an outcome should preserve the property it has achieved before. We are now in a position to make the definition more formal. For each $S \subseteq N$ and $i \in N$, we denote by $\succeq_i |_S$ the preference relation restricted to $\mathcal{N}_i \cap 2^S$.

Definition 1 *Given $\alpha \in \{CR, NS, IS, CIS, IR\}$ and a natural number $k > 0$, a partition π is said to be α-robust under deletion of at most k players if π satisfies the property α, and for any $S \subseteq N$ with $|S| \leq k$, the partition $\pi_{-S} := \{S' \setminus S \mid S' \in \pi\}$ still satisfies the property α in the subgame $(N \setminus S, (\succeq_i |_{N\setminus S})_{i \in N\setminus S})$. When k is clear from the context, we will simply call such partition α-robust.*

By definition, if an outcome is α-robust under deletion of $k+1$ players, then it is α-robust under deletion of any $\ell \leq k$ players. Also, fixing parameter k, the relations between the above robustness concepts are the same as those among the corresponding stability concepts. Namely, we have the following containment relation among the classes of outcomes: NS-robust $\subseteq$ IS-robust $\subseteq$ CIS-robust and CR-robust $\subseteq$ IR-robust. Also, by Lemma 1, CR-robustness implies IS-robustness and CIS-robustness implies IR-robustness for a symmetric friend-oriented game.

It is known that a stable outcome of a symmetric friend-oriented game is guaranteed to exist and can be found in polynomial time: a partition that divides the players into the connected components satisfies the preceding stability requirements. However, the example below illustrates that even when players have symmetric friend-oriented preferences, an α-robust partition under deletion of a single player may not exist for any $\alpha \in \{CR, NS, IS, CIS\}$.

Example 1 Consider a symmetric friend-oriented game (N, w) with three players a, b, and c. The friendship graph G_w forms a star with the center being b (Figure 1).

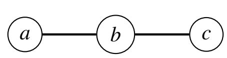


Figure 1: Non-existence of a CR-robust partition for symmetric friend-oriented games.

Suppose that π is CIS-robust under deletion of a single player. First, suppose $\pi = \{\{a, b, c\}\}$. Then, without b, the coalition is not individually rational, a contradiction. Second, suppose $\pi = \{\{a\}, \{b, c\}\}$ or $\pi = \{\{a, b\}, \{c\}\}$. Then, if the player who belongs to the same coalition as b disappears, player b would have a CIS-deviation to the other coalition, a contradiction. Third, if $\pi = \{\{a\}, \{b\}, \{c\}\}$, it would not satisfy contractually individual stability, a contradiction. Finally, if $\pi = \{\{a, c\}, \{b\}\}$, then π would not be individually rational, a contradiction. We have exhausted all possible cases and hence the game admits no CIS-robust partition. This means that the game does not have an α-robust outcome for any $\alpha \in \{CR, NS, IS, CIS\}$. □

4 NS-robustness

We saw that a symmetric friend-oriented game may not admit an NS-robust outcome. In this section, we show that deciding the existence of an NS-robust outcome remains easy for a symmetric friend-oriented game. We warm up by observing that in order to preserve individual rationality, each coalition must be a clique or have minimum degree at least $k + 1$.

Lemma 2 *For any symmetric friend-oriented game, $k > 0$, and any IR-robust partition π, each $S \in \pi$ is either a clique or has minimum degree at least $k + 1$.*

Proof: Let π be an IR-robust partition. Suppose towards a contradiction that there is a coalition $S \in \pi$ such that S does not form a clique and there is a player $i \in S$ who has at most k friends in S. This means that by IR-robustness, S has size at most $k + 1$; otherwise, removing all i's friends in S would violate individual rationality for i. Now since S is not a clique, there is a player j who has an enemy in S. Observe that j has at most $k - 1$ friends in S, since S has size at most $k + 1$ and at least one of the players is an enemy of j. Hence if all the friends of j in S disappear, this would cause the deviation of j to staying alone, contradicting IR-robustness. □

Observe that if there is a coalition of size at most $k + 1$ and some player has a friend in other coalitions, the player would have an NS-deviation to the other coalition after removal of k players. Hence, any NS-robust outcome cannot contain such coalitions, which leads to the following characterization of the classes of friend-oriented games whose NS-robust outcomes are non-empty.

Theorem 2 *The following conditions are equivalent for any symmetric friend-oriented game (N, w) and any natural number $k > 0$:*

(*i*) *There exists an NS-robust partition.*

(*ii*) *Each connected component of G_w is either a clique or has minimum degree at least $k + 1$.*

Corollary 1 *For a symmetric friend-oriented game (N, w), deciding the existence of NS-robust outcomes can be done in polynomial time.*

5 CIS-robustness and IS-robustness

We now turn our attention to a weaker stability concept, *contractually individual stability*. Usually, such stability requirement is not difficult to achieve: Gairing and Savani [2010] observed that a CIS partition is guaranteed to exist for any symmetric additively separable game and can be efficiently computed. As we have seen before, the presence of a star with two leaves complicates the existence of CIS-robust outcomes. In what follows, we will show that by decomposing the friendship graph appropriately, one can determine the existence of CIS-robust outcomes under deletion of a single player. We start by showing that a leaf player and its unique neighbor playing a role of *pseudo-center* form a pair in a CIS-robust outcome. For a graph $G = (V, E)$ and a subset $X \subseteq V$, we denote by $G \setminus X$ the subgraph of G induced by $V \setminus X$. We say that a vertex j is a *pseudo-center* in a graph G if at most one neighbor of j is a non-leaf vertex.

Lemma 3 *For a symmetric friend-oriented game* (N, w) *and* $k = 1$*, let* π *be an arbitrary CIS-robust partition. If* j *is the unique friend of* i*, and* j *is a pseudo-center in* G_w*, then* $\pi(i) = \{i, j\}$*.*

Proof: By Lemma 2, i's coalition is either the singleton $\{i\}$ or the pair $\{i, j\}$. Assume towards a contradiction that $\pi(i) = \{i\}$. If $|\pi(j)| = 1$, then i has a CIS-deviation to j's coalition, a contradiction. If $|\pi(j)| = 2$, then removing the player $h \neq j$ in $\pi(j)$ would cause the CIS-deviation of i to j's coalition, a contradiction. If $|\pi(j)| \geq 3$, then this means that j has at least two friends a, b in $\pi(j)$ by Lemma 2. However, this means that at least one of the players a, b has only one friend j but has at least one enemy in $\pi(j)$, contradicting Lemma 2. In either case, we obtain a contradiction. □

The above lemma can recursively apply to all such pairs in the following way: as long as there is an edge $\{i, j\}$ satisfying the property in Lemma 3, we need to put the players into a pair and examine whether such an edge still exists in the remaining instance. This allows us to partially determine the structure of a CIS-robust outcome. Figure 2 illustrates the sequence of pairs of players that need to be formed in a robust outcome. We now formalize the above idea as follows. For a friendship graph G_w, a sequence of edges $\{i_t, j_t\}$ for $t = 1, 2, \ldots, t^*$ is called an *outer elimination sequence* if the following two hold:

(E1) j_t is the unique friend of i_t in G_t, and j_t is a pseudo-center in G_t; or

(E2) j_t is the unique friend of i_t in G_t, and i_t is a friend of some player in $\bigcup_{h=1}^{t-1}\{i_h, j_h\}$.

Here $G_t = G_w \setminus \bigcup_{h=1}^{t-1}\{i_h, j_h\}$ for each $t = 1, 2, \ldots, t^*$. An outer elimination sequence $(\{i_t, j_t\})_{t=1,2,\ldots,t^*}$ is said to be *maximal* if it cannot be made any longer, i.e., there is no outer elimination sequence $(\{i_t, j_t\})_{t=1,2,\ldots,t^*+1}$.

Lemma 4 *For a symmetric friend-oriented game* (N, w) *and* $k = 1$*, let* π *be an arbitrary CIS-robust partition. If there is an outer elimination sequence* $\{i_t, j_t\}$ *for* $t = 1, 2, \ldots, t^*$*, then we have* $\pi(i_t) = \{i_t, j_t\}$ *for each* $t = 1, 2, \ldots, t^*$*.*

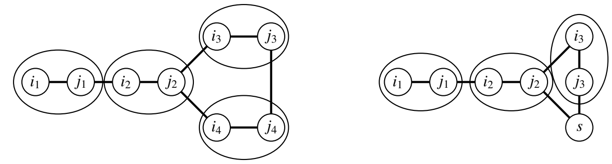


Figure 2: Examples of maximal outer elimination sequences. Observe that the left friendship graph admits a unique CIS-robust partition that consists of elimination pairs, whereas the right friendship graph admits no CIS-robust partition since deleting i_3 would cause the CIS-deviation of j_3 to s.

Proof: We prove the statement by induction on t. When $t = 1$, the claim holds due to Lemma 3. Suppose that the claim holds for $t \leq d - 1$ and we prove it for $t = d$. Now by the induction hypothesis, $\pi(i_h) = \{i_h, j_h\}$ for each $h = 1, 2, \ldots, t - 1$, and thus players i_t and j_t form a coalition within G_t. Now, we have either $\pi(i_t) = \{i_t\}$ or $\pi(i_t) = \{i_t, j_t\}$ by Lemma 2. Assume towards a contradiction that $\pi(i_t) = \{i_t\}$. First suppose that j_t is a pseudo-center in G_t. Again, if $|\pi(j_t)| \geq 3$, then $\pi(j_t)$ contains at least two friends a, b of j_t by Lemma 2 where at least one of the players has only one friend j_t in $\pi(j_t)$, contradicting Lemma 2. Thus, j_t either stays alone at π or forms a coalition with his another friend in G_t; however, in the former case, i_t would have a CIS-deviation to j_t; and in the latter case, deleting the other friend of j_t would cause the CIS-deviation of i_t to $\pi(j_t)$, a contradiction. Second, suppose that i_t is a friend of some player $i \in \bigcup_{h=1}^{t-1}\{i_h, j_h\}$. Then, by removing $j \in \pi(i)$ with $j \neq i$, player i would have a CIS-deviation to $\pi(i_t)$, a contradiction. We thus conclude that $\pi(i_t) = \{i_t, j_t\}$. □

Given a symmetric friend oriented game (N, w), we say that a pair of players is an *elimination pair* if it appears in some outer elimination sequence. We denote by P_w the set of players who belong to some elimination pair, by S_w the set of players who belong to $N \setminus P_w$ and have exactly one friend in $N \setminus P_w$, by B_w the set of players who belong to $N \setminus P_w$ and have at least two friends in $N \setminus (P_w \cup S_w)$, and by R_w the set of remaining players, i.e., $R_w = N \setminus (P_w \cup S_w \cup B_w)$. Before we proceed, we observe the following.

Lemma 5 *For a symmetric friend oriented game* (N, w)*, each player in* S_w *has no friend in* P_w*.*

Proof: Suppose that there is a player $i \in S_w$ who is a friend of some player in P_w. Let j be the unique friend of i in $N \setminus P_w$. Then, the pair of players i and j is an elimination pair satisfying the condition (E2), a contradiction. □

Lemma 6 *For a symmetric friend oriented game* (N, w)*, let* $j \in R_w$*. Then,* j *has no friend in* $N \setminus P_w$*. Further if there is a CIS-robust outcome,* j *has no friend in* P_w*.*

Proof: Consider $j \in R_w$. Assume towards a contradiction that j has some friend in $N \setminus P_w$. If j has a friend $i \in S_w$, then i together with j is an elimination pair satisfying (E1) and must be included in P_w, a contradiction. If j has exactly one friend in $N \setminus (P_w \cup S_w)$, then this means that $j \in S_w$, a contradiction. If j has at least two friends in $N \setminus (P_w \cup S_w)$, then this means that $j \in B_w$, a contradiction. Hence j has no friend in $N \setminus P_w$.

Further assume otherwise that there is a CIS-robust outcome π but there is a player $i \in P_w$ who is a friend of j. By Lemma 4, i forms a pair, say with player h at π. By Lemma 4 and Lemma 2, j stays alone at π. Thus deleting h would cause the CIS-deviation of i to j, a contradiction. □

Figure 3 illustrates the partition of the player set into P_w, S_w, B_w, and R_w. Now, a CIS-robust outcome must include all the elimination pairs, and hence if such outcome exists, there is at most one maximal outer elimination sequence. We can thus completely characterise the class of symmetric friend-oriented games that admit a CIS-robust outcome under deletion of a single player.

Theorem 3 *For a symmetric friend-oriented game and $k = 1$, a CIS-robust outcome exists if and only if the following holds:*

(i) *the set of elimination pairs that appear in each maximal elimination sequence is the same; and*

(ii) *there are no elimination pairs $\{i, j\}$ and $\{u, v\}$ where i is a friend of both u and v; and*

(iii) *for each player $i \in P_w$ and each player $j \in R_w$, i and j are enemies to each other; and*

(iv) *if there is a player $i \in P_w$ who is a friend of every player in B_w, then every player $j \in B_w$ is a friend of exactly one player in S_w and j is an enemy of at least one player in each elimination pair.*

Proof Sketch: Suppose that there is a CIS-robust outcome π. To show (i), take any maximal elimination sequences $(\{i_t, j_t\})_{t=1,2,\ldots,t^*}$ and $(\{a_h, b_h\})_{h=1,2,\ldots,s^*}$. If there are two elimination pairs $\{i_t, j_t\}$ and $\{a_h, b_h\}$ with $i_t = a_h$ and $j_t \neq b_h$, this would imply that $\pi(i_t) = \{i_t, j_t\} = \{i_t, b_h\}$ by Lemma 4, a contradiction. If the two sequences are disjoint, then one can create a longer elimination sequence by adding edge $\{a_1, b_1\}$ to the last position of the other sequence, contradicting maximality. To see (ii), if there are two pairs $\{i, j\}$ and $\{u, v\}$ where i is a friend of both u and v, then π has to include both pairs, which however implies that i would have a CIS-deviation to the coalition $\{u, v\}$ after the removal of player j, a contradiction. The statement (iii) holds due to Lemma 6. To see (iv), assume that some player $i \in P_w$ is a friend of every player in B_w. Consider any player $j \in B_w$. If $\pi(j) \subseteq B_w$, then by deleting the other friend $h \neq i$ with $h \in \pi(i)$, i would have a CIS-deviation to $\pi(j)$, a contradiction. Thus we have $\pi(j) \not\subseteq B_w$. Further, by Lemma 6, each player in R_w has no friend and hence stays alone at π by Lemma 2. This means $\pi(j) \cap R_w = \emptyset$, and thus $\pi(j) \cap S_w \neq \emptyset$. However, if j has no friend in S_w or multiple friends in S_w, player in S_w who belongs to j's coalition has no friend in $\pi(j)$ or has at most one friend and at least one enemy in $\pi(j)$, contradicting Lemma 2. Hence j has exactly one friend s in S_w; by Lemma 2 and by the fact that $\pi(j) \cap S_w \neq \emptyset$, we have $\pi(j) = \{j, s\}$. If there is an elimination pair $\{u, v\}$ where both of them are adjacent to j, then j would have a CIS-deviation to the coalition $\{u, v\}$ after removal of s. Hence, j is an enemy of at least one player in each elimination pair.

Conversely, suppose that all the properties (i) − −(iv) hold. Let $P_w = \bigcup_{t=1}^{t^*}\{i_t, j_t\}$ where $\{i_t, j_t\}_{t=1,2,\ldots,t^*}$ is a maximal outer elimination sequence. We note that P_w is empty if there is no elimination pair. We define the partition π as follows: First, for each $t = 1, 2, \ldots, t^*$, we set $\pi(i_t) = \{i_t, j_t\}$. Second, for each player $i \in R_w$, we set $\pi(i) = \{i\}$. Finally, we partition the players in B_w and S_w as follows.

- If there is a player $i \in P_w$ who is a friend of every player in B_w, we put each $j \in B_w$ and the unique friend of j in S_w into a pair.
- Otherwise, all players in B_w form a coalition and put each player in S_w into a singleton.

It can be verified that π is CIS-robust when $k = 1$. □

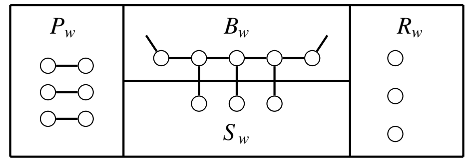


Figure 3: A partition of the player set into P_w, S_w, B_w, R_w.

Building on the above characterization, it is easy to see that we can decide in polynomial time whether a symmetric friend-oriented game admits a CIS-robut outcome under deletion of a single player. The proof employs a simple procedure, which iteratively expands an outer elimination sequence and eventually decompose the player set into P_w, S_w, B_w, and R_w.

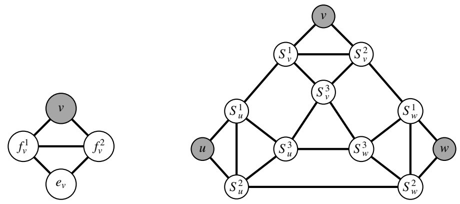
Theorem 4 *For a symmetric friend-oriented game and $k = 1$, deciding the existence of a CIS-robust outcome can be done in polynomial time.*

The above result turns out to be tight in several aspects. We first show that for $k = 2$, finding a CIS-robust outcome of a symmetric friend-oriented game is NP-hard.

Theorem 5 *For a symmetric friend-oriented game (N, w), it is NP-complete to decide the existence of a CIS-robust outcome even for $k = 2$.*

Proof Sketch: CIS-robustness can be verified in polynomial time: for each set $X \subseteq N$ of size at most two, one can check in polynomial time whether π_{-X} is contractually individually stable. So our problem is in NP. To show hardness, we give a reduction from Exact-3-Cover (X3C). Recall that an instance of X3C is given by a set of elements $V = \{v_1, v_2, \ldots, v_{3r}\}$ and a family $\mathcal{S}$ of three-element subsets of V; it is a 'yes'-instance if and only if there is an *exact cover* $\mathcal{S}' \subseteq \mathcal{S}$ with $|\mathcal{S}'| = r$ and $\bigcup_{S \in \mathcal{S}'} S = V$.

Construction: Given an instance $(V, \mathcal{S})$ of X3C, we construct an instance of a friend-oriented game as follows. For each $v \in V$, we create a *vertex player* v. For each vertex $v \in V$, we create a *vertex gadget* G_v, which enforces the corresponding vertex player v to have at least two friends in his robust coalition. Specifically, G_v consists of vertex player v, two friends f_v^1 and f_v^2 of v, and one enemy e_v of v. All the three players f_v^1, f_v^2, and e_v are friends to each other, f_v^1 and f_v^2 are enemies of all the vertex players except for v, and the player e_v is an enemy of all the vertex players. Figure 4(a)



(a) Vertex gadget G_v (b) Set gadget G_S for $S = \{u, v, w\}$

Figure 4: Gadgets constructed in the proofs of Theorem 5. The grey nodes correspond to vertex players.

illustrates G_v. For each $S = \{u, v, w\} \in \mathcal{S}$, we create a *set gadget G_S* which consists of its vertex players u, v, w, and cliques $\{S^1_v, S^2_v, S^3_v\}$ for $v \in S$. Specifically, S^1_v and S^2_v are a friend of v for each $v \in S$; S^3_u, S^3_v, S^3_w form a clique; and the pairs of S^1_u and S^1_v, S^2_u and S^2_w, and S^1_w and S^2_v are friends to each other. See Figure 4(b) for an illustration. Unless specified otherwise, players are enemies to each other. Finally, we set $k = 2$.

Correctness: Suppose that there is an exact cover $\mathcal{S}' \subseteq \mathcal{S}$. Then, we define π as follows. For each set $S \in \mathcal{S}'$ and each $v \in S$, we set $\pi(v) = \{v, S^1_v, S^2_v\}$; the remaining players of the set gadget forms a coalition, i.e., $\pi(S^3_u) = \{S^3_u, S^3_v, S^3_w\}$. For each $S \notin \mathcal{S}'$, the non-vertex players in the set gadget G_s form a coalition, i.e., $\pi(S^1_u) = \{ S^i_v \mid i = 1, 2, 3 \wedge v \in S \}$. For each $v \in V$, we set $\pi(f^1_v) = \{f^1_v, f^2_v, e_v\}$. The resulting partition π can be easily verified to be IR-robust. Also, as each player has at least two friends in his coalition and one enemy in the other coalitions, no player has a CIS-deviation to other coalitions even after an arbitrary player disappears.

Conversely, let π be a CIS-robust partition. To maintain robustness within each vertex gadget G_v, for $v \in V$, it can be verified that the three players f^1_v, f^2_v, e_v must form a coalition, which means that each vertex player v must have at least two friends in his coalition. Thus for each $v \in V$, we have $\{v, S^1_v, S^2_v\}$ for some $S \in \mathcal{S}$ with $v \in S$. Further, to maintain robustness within each set gadget, if $\pi(v) = \{v, S^1_v, S^2_v\}$, then each vertex player $u \in S$ also *selects* S, i.e., $\pi(u) = \{u, S^1_u, S^2_u\}$. Now let $\mathcal{S}' = \bigcup_{v \in V}\{S \in \mathcal{S} \mid \pi(v) = \{v, S^1_v, S^2_v\}\}$. Then it can be easily verified that $\mathcal{S}'$ is an exact cover. □

A similar proof of Theorem 5 shows that finding an IS-robust outcome is NP-hard even if only a single player is allowed to disappear.

Theorem 6 *For a symmetric friend-oriented game (N, w), deciding the existence of an IS-robust outcome is NP-complete even for $k = 1$.*

We have not been able to identify whether the problem of computing a CR-robust outcome is polynomial-time solvable; we leave this question for future work.

6 IR-robustness

In Example 1, we have seen that an outcome of a hedonic game can fail to preserve some stability properties, even when players' preferences are symmetric friend-oriented. Our next question is the following: is it still possible to guarantee a minimum stability requirement, i.e., individual rationality, under deletion of players, while ensuring desirable property of the original partition? The answer is positive for individual stability when players have symmetric additively separable preferences. In these games, one can guarantee the existence of an individually stable partition that is IR-robust.

Theorem 7 *For any symmetric additively separable game (N, w) and any natural number $k > 0$, there exists an individually stable and IR-robust partition.*

In general, finding an individually stable outcome of a symmetric additively separable game is known to be computationally intractable [Gairing and Savani, 2011]. In contrast, we can efficiently construct an individually stable partition that is IR-robust in symmetric friend-oriented games, in which each weight only takes two values.

Theorem 8 *For any symmetric friend-oriented (N, w) and any natural number $k > 0$, one can compute an individually stable and IR-robust partition in polynomial time.*

We note that without symmetry, the set of outcomes that are both individually stable and IR-robust can be empty.

7 Conclusion

We believe that this paper has made a first important step towards a future stream of research, sparked by the chemistry of two concepts, robustness and stability. Below, we list several interesting questions for future work.

Most obviously, while our main focus was on robustness against agents' non-participation, studying other types of robustness would be an important topic of research. For instance, one might want to consider sudden failure of agents' friendship relations, due to individual conflicts (see, e.g., Varma and Yoshida [2019]). We expect to see a different impact on stable outcomes as communication failure has usually less affect on the the structure of the underlying network.

Also, our work presents the worst-case analysis for agent failure. That is, our definition requires an outcome to be immune to *any* possibility of agent failure. However, one might want to consider specific coalitional failure, rather than all of them. For example, our model can be extended to the probabilistic setting where each agent/friendship link may have different probability of failure.

Finally, it would be interesting to extend this line of work to other settings where stability plays an important role. Examples include fractional hedonic games [Aziz *et al.*, 2014; 2017], stable marriage problem [Gale and Shapley, 1962], and group activity selection problem [Darmann *et al.*, 2017; Igarashi *et al.*, 2017].

Acknowledgements

We thank reviewers at AAMAS-19 for helpful feedback. We are also grateful to Ilan Nehama and Nathanaël Barrot for useful discussions. This work was supported by JSPS KAKENHI (Grant-in-Aid for JSPS Fellows, No. 18J00997; JP17H00761; JP18H03299; JP17KK0008) and partially supported by JST Strategic International Collaborative Research Program, SICORP.

References

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

[Aziz *et al.*, 2014] H. Aziz, F. Brandt, and P. Harrenstein. Fractional hedonic games. In *Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 5–12, 2014.

[Aziz *et al.*, 2017] H. Aziz, F. Brandl, F. Brandt, P. Harrenstein, M. Olsen, and D. Peters. Fractional hedonic games. *CoRR*, abs/1705.10116, 2017.

[Bachrach *et al.*, 2011] Y. Bachrach, R. Meir, M. Feldman, and M. Tennenholtz. Solving cooperative reliability games. In *Proceedings of the 27th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 27–34, 2011.

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

[Bilò *et al.*, 2014] V. Bilò, A. Fanelli, M. Flammini, G. Monaco, and L. Moscardelli. Nash stability in fractional hedonic games. In *Proceedings of the 10th Conference on Web and Internet Economics (WINE)*, pages 486–491, 2014.

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

[Darmann *et al.*, 2017] A. Darmann, E. Elkind, S. Kurz, J. Lang, J. Schauer, and G. Woeginger. Group activity selection problem. *International Journal of Game Theory*, 2017.

[Dimitrov *et al.*, 2006] 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.

[Fedoruk and Deters, 2002] A. Fedoruk and R. Deters. Improving fault-tolerance by replicating agents. In *Proceedings of the 1st International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 737–744, 2002.

[Gairing and Savani, 2010] M. Gairing and R. Savani. Computing stable outcomes in hedonic games. In *Proceedings of the 3rd International Symposium on Algorithmic Game Theory (SAGT)*, pages 174–185, 2010.

[Gairing and Savani, 2011] M. Gairing and R. Savani. Computing stable outcomes in hedonic games with voting-based deviations. In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 559–566, 2011.

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

[Igarashi and Elkind, 2016] A. Igarashi and E. Elkind. Hedonic games with graph-restricted communication. In *Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 242–250, 2016.

[Igarashi *et al.*, 2017] A. Igarashi, D. Peters, and E. Elkind. Group activity selection on social networks. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pages 565–571, 2017. Extended version available as arXiv:1611.04524 [cs.GT].

[Igarashi *et al.*, 2019] A. Igarashi, K. Ota, Y. Sakurai, and M. Yokoo. Robustness agains agent failure in hedonic games. *CoRR*, abs/1903.05534, 2019.

[Kouvaros and Lomuscio, 2017] Panagiotis Kouvaros and Alessio Lomuscio. Verifying fault-tolerance in parameterised multi-agent systems. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 288–294, 2017.

[Kouvaros *et al.*, 2018] P. Kouvaros, A. Lomuscio, and E. Pirovano. Symbolic synthesis of fault-tolerance ratios in parameterised multi-agent systems. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 324–330, 2018.

[Okimoto *et al.*, 2015] T. Okimoto, N. Schwind, M. Clement, T. Ribeiro, K. Inoue, and P. Marquis. How to form a task-oriented robust team. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 395–403, 2015.

[Peters, 2016] D. Peters. Graphical hedonic games of bounded treewidth. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, pages 586–593, 2016.

[Schrijver, 2003] A. Schrijver. *Combinatorial Optimization Polyhedra and Efficiency*, volume 24. Springer-Verlag Berlin Heidelberg, 2003.

[Varma and Yoshida, 2019] N. Varma and Y. Yoshida. Average sensitivity of graph algorithms. *CoRR*, abs/1904.03248, 2019.