Individually Stable Dynamics in Coalition Formation over Graphs

Angelo Fanelli[1], **Laurent Gourvès**[1], **Ayumi Igarashi**[2], **Luca Moscardelli**[3]

[1]Université Paris-Dauphine, Université PSL, CNRS, LAMSADE, 75016, Paris, France,
[2]The University of Tokyo, Japan,
[3]University of Chieti-Pescara, Italy
angelo.fanelli@cnrs.fr, laurent.gourves@dauphine.fr, igarashi@mist.i.u-tokyo.ac.jp, luca.moscardelli@unich.it

Abstract

Coalition formation over graphs is a well studied class of games whose players are vertices and feasible coalitions must be connected subgraphs. In this setting, the existence and computation of equilibria, under various notions of stability, has attracted a lot of attention. However, the natural process by which players, starting from any feasible state, strive to reach an equilibrium after a series of unilateral improving deviations, has been less studied. We investigate the convergence of dynamics towards individually stable outcomes under the following perspective: what are the most general classes of preferences and graph topologies guaranteeing convergence? To this aim, on the one hand, we cover a hierarchy of preferences, ranging from the most general to a subcase of additively separable preferences, including individually rational and monotone cases. On the other hand, given that convergence may fail in graphs admitting a cycle even in our most restrictive preference class, we analyze acyclic graph topologies such as trees, paths, and stars.

1 Introduction

Coalition formation is an important and widely investigated issue in artificial intelligence. In many economic, social, and political settings, individuals carry out activities in groups rather than by themselves. *Hedonic games*, introduced by Drèze and Greenberg (1980) and later developed in (Banerjee, Konishi, and Sönmez 2001; Bogomolnaia and Jackson 2002; Cechlárová and Romero-Medina 2001), are among the most important game-theoretic approaches to the study of coalition formation problems. An outcome for these games is a partition of the players into coalitions, over which the players have utilities. A player's utility depends on the coalition she belongs to, and is not affected by how the participants of other coalitions are partitioned.

The standard model of hedonic games does not impose any restrictions on which coalitions may form. However, in reality, we often encounter *network* constraints on coalition formation, i.e., entities can communicate and cooperate only if they are connected. Such restrictions can be naturally described by means of undirected graphs, giving life to *graph hedonic games*, in which players are identified with vertices, communication links with edges, and feasible coalitions with connected subgraphs (Myerson 1977; Demange 2004; Igarashi and Elkind 2016).

In this paper, we study graph hedonic games under the perspective of *individual stability* (IS for short), a natural notion of stability introduced by Drèze and Greenberg (1980). In particular, a player i performs an *IS deviation* to coalition T whenever i prefers coalition $T \cup \{i\}$ to her current coalition, and also all players in T do not prefer T to $T \cup \{i\}$. Roughly speaking, an IS deviation is a *Nash deviation* (i.e., the deviating player strictly benefits) with the additional constraint that all players in coalition T have to "accept" player i. A partition is individually stable if no player has an IS deviation available. Real life examples of IS deviations exist: a nation can enter the NATO, or the EU, only if its members unanimously agree.

Note that stability is mostly concerned with the final state of the coalition formation process and one frequently ignores how these desirable partitions can actually be reached. Essentially, speaking about stability often implicitly assumes that there is a central authority knowing the preferences of all players and computing a stable partition. However, the existence of a stable state does not imply that players, starting from an initial configuration, eventually reach stability after a finite sequence of improving deviations. In the well known stable marriage problem, a stable matching always exists under mild assumptions, and the centralized algorithm of Gale and Shapley (1962) outputs such a state. However, the natural process of sequentially eliminating the existing blocking pairs may loop (Knuth 1997). Examples of this kind motivate the study of decentralized coalition formation processes operated by autonomous entities.

This article focuses on IS dynamics. We analyze the process in which the players interact and, possibly, reach an individually stable coalition structure. More precisely, we aim at determining the most general possible classes of preferences and graph topologies guaranteeing convergence of IS dynamics in graph hedonic games (for any instance and initial state). To this aim, we consider the following natural classes of preferences, that we list from the most general to the most particular: *general* preferences (no restrictions except the ones that are usual in hedonic games), *individually rational* (IR for short) preferences (each player likes any coalition she belongs to at least as much as if she were alone), *monotone* preferences (each player likes any super-

	general	individually rational (IR)	monotone	LAS
cycles	✗	✗	✗	✗ (Ex. 2.2)
trees	✗	✗	✗ (Ex. 5.1)	✓ (Th. 5.2)
paths	✗ (small modification of Ex. 3.3)	✗ (Ex. 3.3 for ≥ 4 coalitions) ✓ (Th. 3.4 for ≤ 3 coalitions)	✓ (Th. 3.2)	✓ (Th. 3.2)
stars	✗ (Ex. 4.3)	✓ (Th. 4.1)	✓ (Th. 4.1)	✓ (Th. 4.1)

Table 1: Overview of the convergence results. The "✗" indicates that the dynamics are not guaranteed to converge, while "✓" indicates that the dynamics under individual stability notion converge. The existence of an individually stable partition is always guaranteed except when the graph contains a cycle and the preferences are general (Igarashi and Elkind 2016).

set of any coalition S she belongs to at least as much as S), and *local additively separable* (LAS for short) preferences. LAS preferences specialize the well studied case of *additively separable* (AS for short) preferences where each player i assigns a value $v_i(j)$ to any other player j, and a player prefers the coalition maximizing the sum of values towards its members. In LAS preferences, which apply to graph hedonic games, $v_i(j)$ is non negative, and it can be strictly positive only if i and j are neighbours in the graph.

Contributions. Our results are summarized in Table 1. On the negative side, general preferences do not guarantee convergence even for the very basic topologies of paths and stars. Analogously, when considering graphs containing cycles, even the most particular class of preferences under study, that is LAS preferences, does not ensure convergence. On the positive side, while IR preferences ensure convergence for stars, monotone preferences are needed in order to secure convergence on paths[1] and LAS preferences are needed for guaranteeing convergence on trees. For LAS preferences or when the graph is a star, we also provide worst case bounds on the number of steps that the dynamics need before reaching a stable outcome.

Related work. The book chapter by Aziz and Savani (2016) gives a comprehensive overview of hedonic games (HGs for short); see also (Hajduková 2006) for an earlier survey on coalition formation games. Various notions of stability have been considered for HGs, such as core,[2] Nash, or individual stability. These concepts refer to states that exclude a certain kind of deviations, so they are naturally related to the dynamics where, starting from some initial partition, the corresponding deviation is repeatedly performed if it is possible. When it converges, such a decentralized process constitutes a plausible explanation of the formation of coalitions.

One of the best known settings is that of HGs with *symmetric* AS preferences ($v_i(j) = v_j(i)$ for every pair i, j of players) for which the IS dynamics always converge (Bogomolnaia and Jackson 2002). In fact, in such games, the sum of all players' utilities always increases after the deviation of a player to her preferred coalition. Gairing and Savani (2010), however, showed that finding an IS state is PLS-complete, meaning that it is unlikely that the IS dynamics converge after a polynomial number of steps.

In a significant part of the literature on cooperative games, some graph constraints are imposed on the coalitions that can form. The classical result of Demange (2004) in non-transferable utility games translates to the fact that a core stable outcome always exists in a hedonic game on a tree. Igarashi and Elkind (2016) strengthened the result and showed that there is an outcome that satisfies both core and individual stability; further, they provided a polynomial time algorithm to compute an individually stable outcome whenever the graph is a tree.

Concerning the existence of IS states, note that when the graph is connected, having IR preferences implies that the state where all players are in the same coalition is IS. For general preferences, an IS partition is guaranteed to exist when the graph is a forest while it may not exist when the graph contains a cycle (Igarashi and Elkind 2016).

There is a growing literature on the dynamics in HGs. Brandt, Bullinger, and Tappe (2022) studied the convergence of dynamics associated with relaxed notions of IS deviations where, instead of requiring unanimous consents, the deviation of a player is doable if it is accepted by a majority of the members of the welcoming coalition. In another closely related work, Brandt, Bullinger, and Wilczynski (2023) studied the convergence of IS dynamics in anonymous HGs, hedonic diversity games, fractional HGs, and dichotomous HGs. Their results do not readily compare to Table 1 since their games don't have graph connectivity constraints and their players' preferences are very different. For instance, individual rationality is not necessarily satisfied in hedonic diversity games, fractional HGs, and dichotomous HGs. Boehmer, Bullinger, and Kerkmann (2023) studied the dynamics in HGs where the utilities of players change over time, depending on the history of the coalition formation process. Hoefer, Vaz, and Wagner (2018) considered dynamics towards core stable states in a general hedonic coalition formation game with various constraints of visibility and externality, where the players have *correlated preferences*, i.e., all members of a coalition have the same utility.

Omitted proofs and examples can be found in the extended version of this work (Fanelli et al. 2024).

2 Model

A *graph hedonic game* is defined on a finite graph (N, L) where N is a set of $n \geq 2$ players, and L is a set of undi-

[1] When starting from partitions with at most 3 coalitions, individually rational preferences also guarantee convergence for paths.

[2] When considering *core stability*, an outcome is stable if and only if there exists no coalition that could make all its members better off.

rected edges between players. Players are able to cooperate if and only if they are connected in the graph (N, L). Let $\mathcal{F}$ be the set of all nonempty subsets S of N such that the subgraph induced by S is connected. Each player $i \in N$ has a preference ordering $\succeq_i$ over the subsets in $\mathcal{F}(i) := \{ S \in \mathcal{F} \mid i \in S \}$. The subsets of N are referred to as *coalitions*. A coalition is said to be *feasible* if it belongs to $\mathcal{F}$. A partition π of N is such that $\bigcup_{S \in \pi} S = N$ and for every pair of distinct $S, T \in \pi$, we have that $S \cap T = \emptyset$. π is said to be *feasible* if $\pi \subseteq \mathcal{F}$. An *outcome* or *state* (we interchangeably use these terms) of a graph hedonic game is a feasible partition. For player $i \in N$ and a partition π of N, we denote by $\pi(i)$ the coalition to which i belongs. We assume without loss of generality that (N, L) is connected (otherwise each connected component can be treated separately).

Preferences

Fix any player $i \in N$, and $S, T \in \mathcal{F}(i)$. $S \succeq_i T$ means that S is at least as good as T from player i's viewpoint. We write $S \succ_i T$ to express that i strictly prefers S over T, whereas $S \sim_i T$ means that i is indifferent between S and T.

As is standard in hedonic games, we always assume that $\succeq_i$ is complete, transitive, and reflexive (Aziz and Savani 2016). The preference relation $\succeq_i$ is said to be *general* if no further assumption is made on it. The preference relation $\succeq_i$ is said to be *individually rational* (IR) if $S \succeq_i \{i\}$ holds for every $S \in \mathcal{F}(i)$. The preference relation $\succeq_i$ is said to be *monotone* if $S \succeq_i T$ always holds when $T \subseteq S$.

In the well studied case of *additively separable preferences*, every player i has a value $v_i(j)$ for being in the same coalition as player j. Player i has *utility* $\sum_{j \in S \setminus \{i\}} v_i(j)$ when she is in coalition S. The utility is 0 when a player is alone. Then, $S \succeq_i T$ holds when $\sum_{j \in S \setminus \{i\}} v_i(j) \geq \sum_{j \in T \setminus \{i\}} v_i(j)$. In this article we also consider *local additively separable* (LAS in short) preferences, a special case of additively separable preferences where the players' values are non symmetric ($v_i(j)$ can differ from $v_j(i)$), non negative, and $v_i(j)$ can be positive only if i and j are neighbors in (N, L), i.e., $v_i(j) > 0 \Rightarrow (i, j) \in L$.

The hierarchy of the above preference relations is

$$\text{General} \supseteq \text{Individually Rational} \supseteq \text{Monotone} \supseteq \text{LAS}$$

since LAS preferences are monotone, and monotone preferences are IR. An instance of the graph hedonic game is said to be $\mathcal{P}$ with $\mathcal{P} \in \{$general, individually rational, monotone, LAS$\}$ when the preferences of *all* the players are $\mathcal{P}$.

The dynamics under individual stability

Consider a feasible partition π, a player i along with her coalition $S = \pi(i)$, and $T \in \pi \cup \{\emptyset\} \setminus \{\pi(i)\}$. Player i *wants to deviate* from S to T if $T \cup \{i\} \in \mathcal{F}$ and $T \cup \{i\} \succ_i S$.[3] A player $j \in T$ *accepts* a deviation of i to T if $T \cup \{i\} \succeq_j T$. Thus, an *IS deviation* by i from S to T is possible if i wants it, and all players in T accept it. As a result of player i's deviation, S and T become $S \setminus \{i\}$ and $T \cup \{i\}$, respectively.

Definition 2.1. *A feasible partition π of N is said to be individually stable (IS) if no player $i \in N$ has an IS deviation to another coalition in $\pi \cup \{\emptyset\}$.*

This article focuses on the dynamics associated with individual stability (a.k.a. IS dynamics) as described in Algorithm 1. The deviations are sequential and players keep on deviating if the current partition is not individually stable. If, in the dynamics, several players are eligible for an IS deviation, then we suppose that one of them is chosen arbitrarily.

Algorithm 1: IS dynamics

Require: a graph hedonic game $(N, L, (\succeq_i)_{i \in N})$ and an initial feasible partition π_0
Ensure: A feasible partition π of N
1: $\pi \leftarrow \pi_0$
2: **while** there exists an IS deviation of $i \in N$ from $\pi(i)$ to $T \in \pi \cup \{\emptyset\}$ **do**
3: $\pi \leftarrow (\pi \setminus \{\pi(i), T\}) \cup \{T \cup \{i\}\} \cup \{ S \mid S$ is a maximal connected subset of $\pi(i) \setminus \{i\} \}$.
4: **end while**
5: **return** π

When a player i leaves her coalition $\pi(i) \in \mathcal{F}(i)$, the graph induced by $\pi(i) \setminus \{i\}$ is not necessarily connected. In the IS dynamics, it is assumed that the members of $\pi(i) \setminus \{i\}$ reconfigure themselves in a minimum number of feasible coalitions by forming inclusionwise maximal connected subsets of $\pi(i) \setminus \{i\}$ (cf. Line 3). The motivation behind this assumption is to consider the minimal changes in $\pi(i)$ caused by the departure of i.

The IS dynamics consist of successive *better moves*, i.e., a player does not necessarily join her most preferred coalition, within the set of coalitions that would accept her.

A *sequence* in the IS dynamics is an ordered list of states $\langle \pi_0, \ldots, \pi_k \rangle$ where each π_t is obtained from π_{t-1} by a single IS deviation. The sequence $\langle \pi_0, \ldots, \pi_k \rangle$ is *cyclic* (equivalently, $\langle \pi_0, \ldots, \pi_k \rangle$ is a *cycle*) if $\pi_0 = \pi_k$. We say that the IS dynamics *converge* if its input does not admit any cyclic sequence. Otherwise, we say that the IS dynamics *cycle*.

The IS dynamics can cycle in a graph which is a cycle, even if the preferences are LAS (cf. Example 2.2).

Example 2.2. *In this instance $N = \{a, b, c\}$, $L = \{(a, b), (b, c), (a, c)\}$, $v_a(b) = v_b(c) = v_c(a) = 1$, and any other value is 0, so the preferences are:*

- *a:* $\{a, b, c\} \sim \{a, b\} \succ \{a, c\} \sim \{a\}$
- *b:* $\{a, b, c\} \sim \{b, c\} \succ \{a, b\} \sim \{b\}$
- *c:* $\{a, b, c\} \sim \{a, c\} \succ \{b, c\} \sim \{c\}$

The initial partition π_0 is $\{\{a, c\}, \{b\}\}$. Player a moves from $\{a, c\}$ to $\{b\}$ giving $\{\{a, b\}, \{c\}\}$. Player b moves from $\{a, b\}$ to $\{c\}$ giving $\{\{a\}, \{b, c\}\}$. Player c moves from $\{b, c\}$ to $\{a\}$ giving $\{\{a, c\}, \{b\}\}$, which is π_0.

Example 2.2 relies on a cycle of 3 vertices but a similar non-convergence result can be shown when (N, L) is a cycle with $n > 3$ vertices (cf. (Fanelli et al. 2024)).

[3]This condition alone defines a *Nash deviation* for player i, and a feasible partition is *Nash stable* if no player has a Nash deviation available.

Example 2.2 tells us that even if we consider our most restricted type of preferences (i.e., LAS preferences), convergence may fail if the underlying graph contains a cycle. For this reason, in the remainder of this article, we consider trees with LAS preferences and restricted subfamilies of trees (i.e., paths and starts) with monotone and IR preferences.

3 Paths

This section is devoted to the case in which (N, L) is a path. To this respect, we provide a complete picture of the convergence of IS dynamics. In particular, we first prove one of our main technical results: for monotone preferences, the IS dynamics always converge. Then, we exhibit an example of non convergence in which players have individually rational preferences and we show that, under individually rational preferences, the IS dynamics always converge when starting from a partition with at most three coalitions.

Some results of this section will rely on the following lemma.

Lemma 3.1. *The number of coalitions does not increase during the IS dynamics when (N, L) is a path and players have individually rational preferences.*

Monotone preferences

Let us note that individual stability and Nash stability are equivalent when the instance is monotone, since no player refuses that another enters her coalition.

We shall prove that, under monotone preferences, the IS dynamics always converge on paths. Roughly speaking, the proof assumes that a cycle exists in the dynamics and exhibits a contradiction as follows. Consider any player i, being the rightmost of coalition C, joining in the dynamics the coalition C' on her right. We prove that, in order for player i to go back to her original coalition C, the rightmost player in C', say player i', has to join coalition C'', being on the right of i'. By iteratively repeating this argument, we obtain that the rightmost player of the path should also join the coalition on her right: a contradiction, given that no coalition exists on her right. It is therefore possible to prove the following theorem.

Theorem 3.2. *Under monotone preferences, the IS dynamics always converge on paths.*

Proof sketch. We first need some additional notation. The input graph is a path P where the players are named $1, \ldots, n$ from left to right. A feasible coalition in P is a sequence of consecutive players denoted by $[\ell, r]_P$, where ℓ and r are the leftmost and and rightmost players, respectively.

Given a cycle $D = \langle \pi_0, \ldots, \pi_\alpha \rangle$ in the IS dynamics, let $|D| = \alpha + 1$ denote its length, i.e., the number of deviations. Therefore, D is composed of IS deviations $m_0, \ldots, m_{|D|-1}$, where, for every $i = 0, \ldots, |D| - 1$, deviation m_i at time t_i is from state π_i to state $\pi_{(i+1) \mod |D|}$. Given a cycle D in the IS dynamics, and any integers $a, b \in \{0, \ldots, |D| - 1\}$, we denote by $[a, b]_D$ the set of integers defined as follows: if $b \geq a$, then $[a, b]_D = \{a, \ldots, b\}$, otherwise, i.e., if $b < a$, $[a, b]_D = \{a, \ldots, |D| - 1\} \cup \{0, \ldots, b\}$. Finally, we say that integer a is *closer (resp., farther) with respect to time t* than integer b if $(a - t) \mod |D| < (b - t) \mod |D|$ (resp., if $(a - t) \mod |D| > (b - t) \mod |D|$).[4]

Assume by contradiction that there exists a cycle $D = \langle \pi_0, \ldots, \pi_{|D|-1} \rangle$ in the IS dynamics. First of all, notice that, since the number of coalitions can never increase after a deviation of D (by Lemma 3.1), the number of coalitions of all states in D has to be the same: let k be this number. We therefore obtain that, for every $i = 0, \ldots, |D| - 1$, state π_i is composed of k coalitions named $C_1^i, \ldots, C_k^i$ from left to right. In the following, we always assume that superscripts of C are modulo $|D|$. Moreover, for any $j = 1, \ldots, k$ and $i = 0, \ldots, |D| - 1$, let ℓ_j^i and r_j^i denote the leftmost and rightmost player in C_j^i, respectively, i.e., $C_j^i = [\ell_j^i, r_j^i]_P$.

In this sketch, for ease of exposition, we assume that the leftmost player in P making a deviation in D belongs to the first coalition. Let a_1 be any agent moving from coalition 1 to coalition 2 in D: there exists a time $s_1 \in \{0, \ldots, |D| - 1\}$ such that deviation m_{s_1} is performed by player a_1 moving from coalition $C_1^{s_1}$ to coalition $C_2^{s_1+1}$.

For any $j = 2, \ldots, k - 1$, let a_j be the rightmost player in coalition $C_j^{s_{j-1}+1}$ (i.e., a_j is the rightmost player of the coalition in which player a_{j-1} arrives at time $s_{j-1} + 1$) and let s_j be the closest time with respect to time s_{j-1} in which a_j moves from coalition $C_j^{s_j}$ to coalition $C_{j+1}^{s_j+1}$. Moreover, for any $j = 1, \ldots, k - 2$, let t_j be the farthest time with respect to s_j in which player a_j comes back to a coalition of index j, i.e., she moves from coalition $C_{j+1}^{t_j}$ to coalition $C_j^{t_j+1}$.

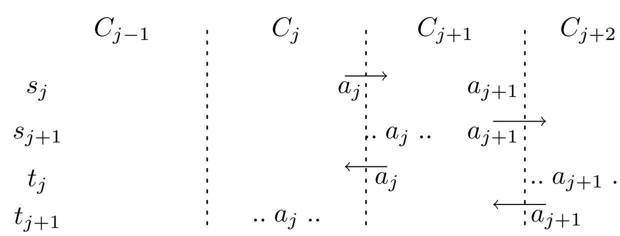


Figure 1: Claim (i) of the proof by induction.

It is possible to prove the following claims by induction on $j = 1, \ldots, k - 1$:

- (i) if $j \leq k - 2$, then it holds that (see Figure 1):
  - (i.a) player a_{j+1} is in coalition $C_x^{t_j}$ with $x \geq j + 2$;
  - (i.b) $s_{j+1} \in [s_j, t_j]_D$ and $t_{j+1} \in [t_j, s_j]_D$;
  - (i.c) player a_j is both in coalition $C_{j+1}^{s_{j+1}}$ and in coalition $C_x^{t_{j+1}}$ with $x \leq j$;
- (ii) if $j = k - 1$, then player a_{k-1} moves to a coalition of index k (by claim (i.a) holding for $j = k - 2$), but

[4] We are assuming that $r = x \mod |D|$ is defined according to the floored division, i.e., when the quotient is defined as $q = \left\lfloor \frac{x}{|D|} \right\rfloor$, and thus the remainder $r = x - q|D|$ is always non-negative even if x is negative.

she cannot go back to a coalition of index $k-1$, essentially because player n cannot abandon the last coalition: a contradiction to the fact that D is a cycle in the IS dynamics. □

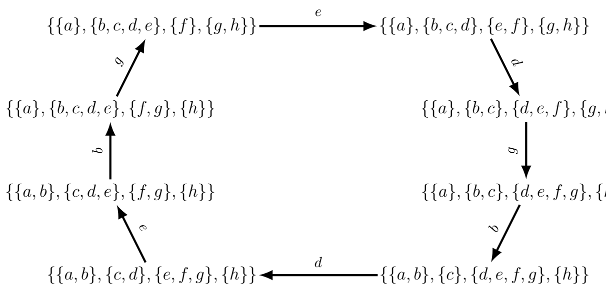
Individually rational preferences

In this section, we consider the case when the graph is a path and agents have individually rational preferences. We first present an example for which the IS dynamics may not converge when the initial partition contains four coalitions.

Example 3.3. *There are* 8 *players,* a, b, c, d, e, f, g, h *aligned on a path in this order. The preferences are given as follows.*

- *a:* $\{a, b\} \succ \{a\}$
- *b:* $\{b, c, d, e\} \succ \{a, b\} \succ \{b, c\} \succ \{b\}$
- *c:* $\{b, c, d, e\} \succ \{c, d, e\} \succ \{c, d\} \succ \{c\}$
- *d:* $\{b, c, d, e\} \succ \{c, d, e\} \succ \{c, d\} \succ \{d, e, f, g\} \succ \{d, e, f\} \succ \{b, c, d\} \succ \{d\}$
- *e:* $\{d, e, f, g\} \succ \{d, e, f\} \succ \{e, f\} \succ \{b, c, d, e\} \succ \{c, d, e\} \succ \{e, f, g\} \succ \{e\}$
- *f:* $\{d, e, f, g\} \succ \{d, e, f\} \succ \{e, f\} \succ \{f\}$
- *g:* $\{d, e, f, g\} \succ \{g, h\} \succ \{f, g\} \succ \{g\}$
- *h:* $\{g, h\} \succ \{h\}$

Note that all coalitions which do not appear are ranked at the same positions as the singleton coalitions. Then, we obtain a cycle as depicted below (deviating players are mentioned above the arrows).



The example can be easily extended to a larger number of coalitions by adding extra coalitions on the right of the last coalition, whose agents do not participate in the dynamics.

Nevertheless, with an initial partition consisting of at most three coalitions and players having individually rational preferences, we are able to prove the following theorem, establishing the convergence of the IS dynamics. This bound on the number of coalitions is tight, as demonstrated by Example 3.3.

Theorem 3.4. *Suppose that the graph* (N, L) *is a path* P *and* $|\pi_0| \leq 3$*. Under individually rational preferences, the IS dynamics always converge.*

Note that for preferences that are not necessarily monotone, we cannot rely on coalition size to reason about player preferences as we did in the proof of Theorem 3.2. However, under the assumptions of the above theorem, deviations occur only between the central coalition and the left (or right) coalition, where the left-most (or right-most) player of the coalition is fixed. We demonstrate that, if we assume the existence of a cycle in the IS dynamics, then players return to the central coalition only if the size of the central coalition decreases, thus yielding a contradiction.

The assumption that the preferences are IR is crucial for Theorem 3.4. Without having IR preferences, the IS dynamics on paths may not converge even when the initial state consists of one coalition: Take Example 3.3 and add $\{a\} \succ_a N$ and $\{f\} \succ_f N \setminus \{a\}$. Start with the grand coalition N. After two deviations (by a and f), one would reach the state in the top left of the cycle in Example 3.3.

4 Stars

This section deals with the special case where (N, L) is a star.

We first show that the IS dynamics always converge under IR preferences.

Theorem 4.1. *Suppose that* (N, L) *is a star. Under individually rational preferences, the IS dynamics converge in a number of steps which is* $\mathcal{O}(n^2)$*.*

In the following, we show that it is possible to obtain a similar result of convergence also for general preferences, by imposing that the dynamics start from a state verifying a suitable property. To this respect, we will say that a coalition $S \subseteq N$ is *individually rational* (IR) if every player $i \in S$ weakly prefers S to $\{i\}$. Moreover, a partition π of N is said to be IR if every $S \in \pi$ is individually rational.

Theorem 4.2. *Suppose that* (N, L) *is a star and the initial state is individually rational. Under general preferences, the IS dynamics converge in a number of steps which is* $\mathcal{O}(n^2)$*.*

It is worth noting that starting from an IR state is a weaker requirement than imposing IR preferences; in fact, when players have IR preferences, any state (including the initial one) is IR. Note that without the assumption that the initial state is IR (and therefore also without the assumption of IR preferences), the IS dynamics may not converge on a star as illustrated by the following example.

Example 4.3. *Consider a star with center* d *and leaves* a, b, c*. Player* d *is indifferent between all coalitions and the preferences of the other players are given as follows:*

- $a : \{a, b, d\} \succ \{a\} \succ \{a, c, d\} \succ \{a, d\} \succ \{a, b, c, d\}$
- $b : \{b, c, d\} \succ \{b\} \succ \{a, b, d\} \succ \{b, d\} \succ \{a, b, c, d\}$
- $c : \{a, c, d\} \succ \{c\} \succ \{b, c, d\} \succ \{c, d\} \succ \{a, b, c, d\}$

A cyclic sequence of states can be as follows: $\{\{a\}, \{b, c, d\}\} \xrightarrow{c} \{\{a\}, \{b, d\}, \{c\}\} \xrightarrow{a} \{\{a, b, d\}, \{c\}\} \xrightarrow{b} \{\{a, d\}, \{c\}, \{b\}\} \xrightarrow{c} \{\{a, c, d\}, \{b\}\} \xrightarrow{a} \{\{c, d\}, \{b\}, \{a\}\} \xrightarrow{b} \{\{a\}, \{b, c, d\}\}$.

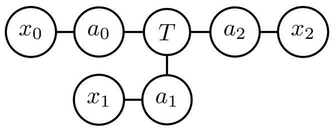
Example A.5 in (Fanelli et al. 2024) shows that the bounds on the time of convergence provided in Theorems 4.1 and 4.2 are asymptotically tight.

Finally, even when the players' preferences are IR, the IS dynamics may cycle in a star with an extra node connected to a leaf (cf. Example A.6 in (Fanelli et al. 2024)).

5 Trees

In this section, we assume that (N, L) is a tree. We first exhibit an example of non convergence where the valuations are non negative and additively separable, which falls in the case of monotone preferences. Then, we move on to *local* additively separable preferences and show that the IS dynamics always converge. At the end of the section, some additional results holding for the special case of (N, L) being a path, under LAS preferences, are provided.

Example 5.1. *Consider the following tree.*



Suppose the players have additively separable preferences with the following values: for any $i = 0, 1, 2$, a_i has value 1 for x_i, value 2 for a_{i+1}, and 0 otherwise (subscripts are modulo 3). Players T and x_i have value 0 for any other player, for any $i = 0, 1, 2$. Since all the values are non negative, the described preferences are monotone. The IS dynamics may cycle as follows.

- $\pi_0 = \{\{x_0, a_0\}, \{T, a_1\}, \{x_1\}, \{a_2, x_2\}\}$.
- $\pi_1 = \{\{x_0\}, \{T, a_0, a_1\}, \{x_1\}, \{a_2, x_2\}\}$.
- $\pi_2 = \{\{x_0\}, \{T, a_0\}, \{a_1, x_1\}, \{a_2, x_2\}\}$.
- $\pi_3 = \{\{x_0\}, \{T, a_0, a_2\}, \{a_1, x_1\}, \{x_2\}\}$.
- $\pi_4 = \{\{x_0, a_0\}, \{T, a_2\}, \{a_1, x_1\}, \{x_2\}\}$.
- $\pi_5 = \{\{x_0, a_0\}, \{T, a_1, a_2\}, \{x_1\}\{x_2\}\}$.

Here, a sequence of IS deviations is made by a_0, a_1, a_2, a_0, a_1, and finally by a_2, resulting in the initial state π_0.

Note that in Example 5.1, one can duplicate the a_i players and reproduce a cycle in the dynamics where all the values are either 1 or 0 (cf. Appendix).

Local additively separable preferences

LAS preferences are considered in this section. Under LAS preferences the IS dynamics can cycle if (N, L) includes a cycle (cf. Example 2.2), but the IS dynamics always converge when (N, L) is acyclic.

Theorem 5.2. *In a graph hedonic game with LAS preferences over a tree the IS dynamics always converge.*

Proof sketch. We denote by $N(i)$ the set of neighbours of i (we assume $i \notin N(i)$) and by $L(i)$ the set of edges incident to i. For every feasible partition π we define $L_\pi = \{(i, j) \in L : \pi(i) = \pi(j)\}$ and $\overline{L}_\pi = L \setminus L_\pi$ and we refer to them as the set of *built* and *broken* edges in π, respectively. For every player i and feasible partition π we define $L_\pi(i) = \{(i, j) : (i, j) \in L(i) \cap L_\pi\}$ and $\overline{L}_\pi(i) = L(i) \setminus L_\pi(i)$, that are the set of built and broken edges incident to i, and we denote by $u_i(\pi)$ the utility of i in π.

The fact that (N, L) is a tree directly implies that every feasible partition π satisfies the following property.

Claim 5.3. *Let π be any feasible partition. Given any $i \in N$ then, for every $j, j' \in N(i)$ such that $j' \neq j$ and $(i, j) \in \overline{L}_\pi(i)$, it holds that $\pi(j) \neq \pi(j')$.*

The deviation of a player i in a feasible partition π can be interpreted as selecting a neighboring vertex j that is incident to an edge in $\overline{L}_\pi(i)$. This observation enables us to describe the IS dynamics in a tree as outlined in Algorithm 2. In order to prove the theorem, we incorporate a labeling scheme ℓ, where $\ell : L \to N \cup \{\perp\}$ is a function labeling the edges of the graph as specified in Algorithm 2.

Algorithm 2: IS dynamics for trees with labeling

Require: a graph hedonic game $(N, L, (\succeq_i)_{i \in N})$, such that (N, L) is a tree, and an initial feasible partition π^0
Ensure: A feasible partition π of N
1: $t \leftarrow 0$
2: **for** each $e \in L$ **do**
3: $\quad \ell^t(e) \leftarrow \perp$
4: **end for**
5: **while** there exists $(\alpha^t, \beta^t) \in \overline{L}_{\pi^t}(\alpha^t)$ such that α^t has an IS deviation from $\pi^t(\alpha^t)$ to $\pi^t(\beta^t)$ **do**
6: $\quad \pi^{t+1} \leftarrow (\pi^t \setminus \{\pi^t(\alpha^t), \pi^t(\beta^t)\}) \cup \{\pi^t(\beta^t) \cup \{\alpha^t\}\} \cup \{ S \mid S$ is a maximal connected subset of $\pi^t(\alpha^t) \setminus \{\alpha^t\} \}$.
7: $\quad$ **for** each $e \in L$ **do**
8: $\quad\quad$ **if** $e \in L_{\pi^t}(\alpha^t) \cup \{(\alpha^t, \beta^t)\}$ **then**
9: $\quad\quad\quad \ell^{t+1}(e) \leftarrow \alpha^t$
10: $\quad\quad$ **else**
11: $\quad\quad\quad \ell^{t+1}(e) \leftarrow \ell^t(e)$
12: $\quad\quad$ **end if**
13: $\quad$ **end for**
14: $\quad t \leftarrow t + 1$
15: **end while**
16: **return** π

Algorithm 2 works as follows. At time step $t \geq 0$ player α^t performs a deviation from $\pi^t(\alpha^t)$ to $\pi^t(\beta^t)$ (see Figure 8 in (Fanelli et al. 2024)); π^{t+1} denotes the partition obtained after this deviation (line 6). As a consequence of Claim 5.3, (α^t, β^t) is the only edge in $L_{\pi^{t+1}}(\alpha^t)$; therefore we say that α^t *builds* (α^t, β^t) and *breaks* all the edges in $L_{\pi^t}(\alpha^t)$ at time t. Moreover, as a consequence of Claim 5.3 and LAS preferences, $u_{\alpha^t}(\pi^{t+1}) = v_{\alpha^t}(\beta^t)$.

Algorithm 2 incorporates also a labeling scheme. At every time step $t \geq 0$, it assigns a label $\ell^t(e)$ to each edge $e = (i, j) \in L$; this label gets value in $\{\perp, i, j\}$. The label tracks which of the two endpoints of the edge made the most recent deviation in previous time steps, thereby determining the current status (built or broken) of the edge. Namely, at time $t = 0$ the label of each edge $e \in L$ is set to $\perp$ (line 3), which means that none of the endpoints of e has performed any deviation. At time $t+1 \geq 1$ the labels of all the edges in $L_{\pi^t}(\alpha^t) \cup \{(\alpha^t, \beta^t)\}$ are set to α^t (line 9), while the labels of all the remaining edges remain unchanged (line 11). The labeling scheme satisfies the following property.

Claim 5.4. *For every $i \in N$ and $t \geq 0$, there exists at most one edge e in $L_{\pi^t}(i)$ such that $\ell^t(e) = i$.*

Now, let us examine how the utility of any player i changes throughout the dynamics given by an execution of Algorithm 2. Note that the utility of i is influenced only by the deviations of i and the deviations of the players in $N(i)$. Moreover, if $v_i(j) = 0$, i would never perform a deviation in which she builds an edge with j. Hence, the utility of i strictly increases after each deviation by i, it may increase after a neighbour j builds an edge with i (it does not change if $v_i(j) = 0$), it may decrease after a neighbor j breaks an edge with i (it does not change if $v_j(i) = 0$), while it strictly decreases after a neighbor j breaks the edge with i when its label is equal to i (in fact, since the edge has been previously built by i, it must hold that $v_i(j) > 0$).

Next, we will bound the number of times a player i can deviate by building an edge with a given neighbor j. In order to do so, we need to introduce some notation. Let $T_Q(i)$ denote the set of time steps in which i performs a deviation by building an edge connecting herself with any node in $Q \subseteq N(i)$. For $j \in N(i)$, we simplify the notation by writing $T_j(i)$ instead of $T_{\{j\}}(i)$. Trivially, we have $|T_Q(i)| = \sum_{j \in Q} |T_j(i)|$. We denote by $T(i)$ the set of time steps in which i performs a deviation, i.e., $T(i) = T_{N(i)}(i)$. Since i may also break edges with some neighbors during each deviation, for $j' \in N(i)$ and $Q \subseteq N(i) \setminus \{j'\}$, we denote by $T_Q^{j'}(i)$ the set of time steps in which i performs a deviation by building an edge with any node in Q while breaking the edge with j' when its label is equal to j'. For $j \in N(i)$, we simplify the notation by writing $T_j^{j'}(i)$ instead of $T_{\{j\}}^{j'}(i)$. Trivially, we have $|T_Q^{j'}(i)| = \sum_{j \in Q} |T_j^{j'}(i)|$. We define $T^{j'}(i) = T_{N(i)\setminus\{j'\}}^{j'}(i)$, that is the set of time steps in which i breaks the edge with $j' \in N(i)$ when its label is equal to j'. Also in this case we have $|T^{j'}(i)| = \sum_{j \in N(i)\setminus\{j'\}} |T_j^{j'}(i)|$. We are ready to show a bound on the number of times i builds an edge.

Lemma 5.5. *For every $i \in N$ and $Q \subseteq N(i)$, $|T_Q(i)| \leq |Q| \left(1 + \sum_{q \in N(i)} |T^i(q)|\right)$.*

Now, let us consider a player r and let $\hat{G}$ be the tree (N, L) rooted in r. Our goal is to bound the number of deviations of r. We denote by C_i the set of children of i, by D_i the set of players in the subtree of $\hat{G}$ rooted in i (including i) and by $p(i)$ the parent of $i \neq r$. Notice that $C_r = N(r)$, $D_r = N$, $C_i = N(i) \setminus \{p(i)\}$ for every $i \neq r$, and $C_i = \emptyset$ and $D_i = \{i\}$ for every leaf i. For every i, let d_i be the maximum distance between i and any leaf in D_i ($d_i = 0$ for every leaf i). For every path $\sigma_j^i = \langle j = q_0, q_1, \ldots, q_k = i \rangle$ such that $k \geq 0$ and $q_{h+1} = p(q_h)$ for every $h = 0 \ldots, k-1$, we define $m_j^i = \prod_{h=0}^{k} |C_{q_h}|$ (notice that $m_j^i = 0$ for every leaf j).

Lemma 5.6. *For every $i \in N \backslash \{r\}$, $|T_{C_i}^{p(i)}(i)| \leq \sum_{j \in D_i} m_j^i$.*

Lemma 5.7. *For the root r, $|T(r)| \leq \sum_{j \in N} m_j^r$.*

Proof. Since, $N(r) = C_r$, from Lemma 5.5 we have

$$|T(r)| = |T_{C_r}(r)| \leq |C_r|\big(1 + \sum_{q \in C_r} |T^r(q)|\big).$$

Moreover, since $T^r(q) = T_{C_q}^{p(q)}(q)$, we can apply Lemma 5.6 to the previous inequality and obtain

$$|T(r)| \leq |C_r|\big(1 + \sum_{q \in C_r} \sum_{j \in D_q} m_j^q\big) = \sum_{j \in D_r} m_j^r.$$

□

Given the arbitrariness of the choice of r, this concludes the proof of Theorem 5.2. □

Corollary 5.8. *In graph hedonic games with LAS preferences, the IS dynamics over a path converge within $2n^2$ deviations.*

The following proposition shows that the IS dynamics can require an exponential number of deviations before converging on trees under LAS preferences, by providing an instance with a player performing a number of deviations matching the bound of Lemma 5.7, and thus also showing that this bound is tight. Finally, Ex. A.11 in (Fanelli et al. 2024) shows that the IS dynamics can require $\Omega(n^2)$ steps for converging on a path under LAS preferences, thus proving that the bound of Corollary 5.8 is asymptotically tight.

Proposition 5.9. *In a graph hedonic game with LAS preferences over a tree, the IS dynamics may require an exponential number of deviations before converging.*

6 Conclusion and Future Work

This article draws a picture of convergence issues of the IS dynamics in graph hedonic games. Our results, summarized in Table 1, depend on the graph topology and a hierarchy of preferences. Many open questions are left for future works.

In most of the cases, we could upper bound the worst-case number of steps that the IS dynamics may need before converging. However, this question remains open for the cases covered by Theorems 3.2 and 3.4.

The IS dynamics rely on *better* responses but note that our counterexamples with cyclic sequences hold also with *best* responses. However, the results showing a lower bound to the number of deviations needed for convergence (Example A.5 in (Fanelli et al. 2024) and Proposition 5.9) do not extend to the case of best responses. We in fact conjecture that under best responses the time of convergence can significantly improve.

When a player i leaves her coalition $\pi(i)$ during the IS dynamics, we suppose that the members of $\pi(i) \setminus \{i\}$ reconfigure themselves into a minimum number of coalitions, but a different reconfiguration scheme can be assumed if, for example, $\pi(i) \setminus \{i\}$ is replaced by singletons. An interesting question is whether the reconfiguration has a significant impact on the convergence of the IS dynamics.

This work deals with convergence for any instance and any initial state but if convergence is not always guaranteed, then it is worth determining the complexity of deciding it for a given instance, from a given initial state.

Finally, a convergence study combining graph topologies and properties of the preferences can be conducted on deviations other than the IS deviations.

Acknowledgments

We thank anonymous reviewers of AAAI 2025 for helpful comments. Ayumi Igarashi acknowledges support from JST FOREST under grant number JPMJFR226O. Laurent Gourvès is supported by Agence Nationale de la Recherche (ANR), project THEMIS ANR-20-CE23-0018. Luca Moscardelli acknowledges support by the PNRR MIUR project FAIR - Future AI Research (PE00000013), Spoke 9 - Green-aware AI, WP 9.2 "Interactions of green-aware agents", cascade sub-project "Existence, Complexity and Efficiency of Stable Solutions in Green-Oriented Games" – ECOGAMES, CUP D23C24000210006, and by the GNCS group of INdAM.

References

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*, 356–376. Cambridge University Press.

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

Boehmer, N.; Bullinger, M.; and Kerkmann, A. M. 2023. Causes of Stability in Dynamic Coalition Formation. In *Proc. 37th AAAI*, 5499–5506. AAAI Press.

Bogomolnaia, A.; and Jackson, M. 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 *Proc. 36th AAAI*, 4867–4874. AAAI Press.

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

Cechlárová, K.; and Romero-Medina, A. 2001. Stability in coalition formation games. *International Journal of Game Theory*, 29(4): 487–494.

Demange, G. 2004. On Group Stability in Hierarchies and Networks. *Journal of Political Economy*, 112(4): 754–778.

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

Fanelli, A.; Gourvès, L.; Igarashi, A.; and Moscardelli, L. 2024. Individually Stable Dynamics in Coalition Formation over Graphs. *CoRR*, abs/2408.11488.

Gairing, M.; and Savani, R. 2010. Computing Stable Outcomes in Hedonic Games. In *Proc. 3rd SAGT*, 174–185. Springer.

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

Hajduková, J. 2006. Coalition formation Games: a survey. *International Game Theory Review*, 8(4): 613–641.

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

Igarashi, A.; and Elkind, E. 2016. Hedonic Games with Graph-restricted Communication. In *Proc. 15th AAMAS*, 242–250. ACM.

Knuth, D. E. 1997. *Stable Marriage and Its Relation to Other Combinatorial Problems: An Introduction to the Mathematical Analysis of Algorithms*. CRM proceedings & lecture notes. American Mathematical Society.

Myerson, R. B. 1977. Graphs and Cooperation in Games. *Mathematics of Operations Research*, 2(3): 225–229.