Technische Universität München

TUM School of Computation, Information and Technology

Computing Desirable Outcomes in Coalition Formation

Martin Bullinger

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:		Prof. Dr. Andreas Wiese
Prüfer*innen der Dissertation:	1.	Prof. Dr. Felix Brandt
	2.	Prof. Dr. Edith Elkind
	3.	Prof. Dr. Michele Flammini

Die Dissertation wurde am 28.11.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 20.03.2023 angenommen.

COMPUTING DESIRABLE OUTCOMES IN COALITION FORMATION

MARTIN BULLINGER

E-MAIL:
martin.bullinger@tum.de

This thesis was created based on a template designed by Hans Georg Seedig. It was typeset using LaTeX and the ClassicThesis style by André Miede, combined with the ArsClassica package by Lorenzo Pantieri and some minor own modifications (mostly by Hans Georg Seedig). The text is set in Palatino with math in Euler, both due to Hermann Zapf. Headlines are set in Iwona by Janusz M. Nowacki, the monospace font is Bera Mono designed by Bitstream, Inc. All graphics were created using TikZ by Till Tantau.

ABSTRACT

The interaction of agents that have the ability to gather in coalitions is a central concern in multi-agent systems since the beginnings of game theory. Coalition formation is typically formalized by so-called hedonic games. In these games, a finite set of agents has to be partitioned into disjoint coalitions, i.e., groups or teams, while they can express preferences about this subdivision. The desirability of coalitions is measured by means of solution concepts. These concepts capture, for instance, the stability of a coalition structure, i.e., its resistance against deviations of agents, or optimality, i.e., global guarantees.

The first set of results concerns classical solution concepts in established classes of hedonic games. We bridge several gaps in the literature by resolving the complexity of various solution concepts, such as contractual Nash stability in additively separable hedonic games, individual stability in symmetric fractional hedonic games, or strong popularity in roommate games. In cardinal classes of hedonic games, we perform comprehensive studies of Pareto optimality and popularity.

We then go over to challenging various paradigms in the literature by making several conceptual contributions. First, we complement static solution concepts by a distributed view. Given a game, instead of asking whether a certain partition exists, we ask whether we can reach specific partitions by means of simple dynamics induced by instabilities. We gain new insights in various classes of hedonic games, such as anonymous hedonic games, hedonic diversity games, or fractional hedonic games. We also analyze the speed of convergence of the obtained dynamics.

Second, we propose stability concepts based on consent through majority votes. These offer a compromise between classical concepts which either neglect any form of consent or require unanimous consent for performing deviations. We study the new solution concepts in additively separable hedonic games and some natural subclasses. The latter are based on the sole distinction of two types of utilities corresponding to friends and enemies.

Third, we propose a way to integrate loyalty, a form of empathy-based incentives, into a given cardinal hedonic game. By iteratively applying the loyalty operation, we converge to a limit game which satisfies egalitarianism at a local level. The limit game contains Pareto-optimal partitions in the core, i.e., they satisfy group stability. However, the loyal game variants lead to intractabilities with respect to preference elicitation and stability.

Besides the main focus of the thesis on hedonic games, we also study a game-theoretic model of residential segregation, which encapsulates Schelling's homophily incentives. In this model, a set of agents is assigned to locations in a metropolitan area. By interpreting the neighborhood of an agent as their coalition, we obtain a model of overlapping coalitions.

For our segregation model, we propose novel concepts of optimality connected to utilitarian social welfare and Pareto optimality. We compute mostly tight bounds for the welfare obtained by all optimality notions, and show how to compute corresponding assignments of agents to locations efficiently.

We conclude the thesis with the consideration of a series of further directions including some specific open problems.

ZUSAMMENFASSUNG

Ein zentrales Anliegen in Multiagentensystemen seit den Anfängen der Spieltheorie ist die Interaktion von Agenten, die sich zu Koalitionen zusammenfinden können. Koalitionsbildung wird typischerweise mittels so genannter hedonischer Spiele formalisiert. In diesen Spielen muss eine endliche Menge von Akteuren in disjunkte Koalitionen, d.h. Gruppen oder Teams, aufgeteilt werden, wobei die Akteure Präferenzen über diese Aufteilung äußern können. Dabei wird die Begehrtheit von Koalitionen mit Hilfe von Lösungskonzepten bestimmt. Diese Konzepte erfassen beispielsweise die Stabilität einer Koalitionsstruktur, d.h. ihre Resistenz gegenüber Koalitionswechseln der Akteure, oder deren Optimalität, also das Erfüllen globaler Anforderungen.

Die erste Reihe an Ergebnissen betrifft klassische Lösungskonzepte in etablierten Klassen hedonischer Spiele. Wir schließen mehrere Lücken in der Literatur, indem wir die Komplexität verschiedener Lösungskonzepte bestimmen wie beispielsweise vertragsgeschützter Nash-Stabilität in additiv separablen hedonischen Spielen, individueller Stabilität in symmetrischen fraktionalen hedonischen Spielen sowie starker Popularität in Zimmerpartner-Spielen. In kardinalen Klassen hedonischer Spiele stellen wir ausführliche Betrachtungen zu Pareto-Optimalität und Popularität an.

Wir gehen dann dazu über, diverse Paradigmen in der Literatur durch verschiedene konzeptionelle Beiträge in Frage zu stellen. Zunächst ergänzen wir die gängigen statischen Lösungskonzepte durch eine distributive Sichtweise. Anstatt die Existenz bestimmter Partitionen zu untersuchen, betrachten wir, ob gute Partitionen mittels einfacher, durch Instabilitäten induzierter Dynamiken erreicht werden können. Wir gewinnen neue Einsichten in verschiedene Klassen von hedonischen Spielen wie anonyme hedonische Spiele, hedonische Diversitätsspiele oder fraktionale hedonische Spiele. Des Weiteren analysieren wir die Konvergenzgeschwindigkeit der so erhaltenen Dynamiken.

Zweitens schlagen wir Stabilitätskonzepte vor, die auf Zustimmung durch Mehrheitsentscheidung beruhen. Diese bieten einen Kompromiss zwischen klassischen Konzepten, die entweder jede Form der Zustimmung vernachlässigen oder einstimmige Zustimmung für die Durchführung von Abweichungen verlangen. Wir untersuchen die neuen Lösungskonzepte in additiv separablen hedonischen Spielen und natürlichen Teilklassen derer. Letztere beruhen auf der alleinigen Unterscheidung von zwei Arten von Nützlichkeit, die Freunden und Feinden entsprechen.

Drittens schlagen wir einen Weg vor, um Loyalität, eine Form von Empathie-basierten Anreizen, in ein gegebenes kardinales hedonisches Spiel zu integrieren. Durch die wiederholte Anwendung der Loyalitätsoperation konvergieren wir zu einem Grenzspiel, das Egalitarismus auf lokaler Ebene begünstigt. Dieses Grenzspiel enthält Pareto-optimale Partitionen im Core; diese erfüllen also Gruppenstabilität. Die auf Loyalität basierenden Varianten des Grundspiels bringen jedoch Berechenbarkeitsprobleme in Bezug auf Stabilität und die allgemeine Präferenzbeschreibung mit sich.

Abgesehen vom Schwerpunkt der Arbeit auf hedonischen Spielen untersuchen wir auch ein spieltheoretisches Modell von Segregation, das Schellings Homophilie-Anreize aufgreift. In diesem Modell werden einer Gruppe von Akteuren Wohnorte zugewiesen. Indem wir die Nachbarschaft eines Akteurs als seine Koalition interpretieren, erhalten wir ein Koalitionsmodell mit überlappenden Koalitionen.

Bezüglich dieses Modells schlagen wir neue Optimalitätskonzepte vor, die eine Brücke zwischen utilitaristischer Wohlfahrt und Pareto-Optimalität schlagen. Wir berechnen vorwiegend exakte Garantien für die Wohlfahrt aller Optimalitätsbegriffe und zeigen, wie man entsprechende Wohnortszuweisungen von Akteuren effizient berechnen kann.

Wir schließen mit der Betrachtung einer Reihe zukünftiger Forschungsrichtungen inklusive konkreter offener Probleme.

CORE PUBLICATIONS

This thesis is based on the following core publications in their original, published form.

REACHING DESIRABLE OUTCOMES

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

[2] Finding and recognizing popular coalition structures. *Journal of Artificial Intelligence Research*, 74:569–626, 2022 (with F. Brandt).†

[3] Boundaries to single-agent stability in additively separable hedonic games. In *Proceedings of the 47th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 26:1–26:15, 2022.

DEVELOPING NOVEL CONCEPTS

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

[5] Single-agent dynamics in additively separable hedonic games. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, pages 4867–4874, 2022 (with F. Brandt and L. Tappe).

---

* This paper was also presented at the *6th World Congress of the Game Theory Society (GAMES)*, 2021.

† This paper was also presented at the *8th International Workshop on Computational Social Choice (COMSOC)*, 2021. An earlier version of the paper appeared in the *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 195–203, 2020.

‡ This paper was also presented at the *3rd AAMAS Workshop on Games, Agents, and Incentives*, 2021, and the *15th Journées d'Intelligence Artificielle Fondamentale (JIAF)*, 2021. The full version of the paper can be found on arXiv at https://arxiv.org/abs/2211.09571.

[6] Loyalty in cardinal hedonic games. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 66–72, 2021 (with S. Kober).

BEYOND DISJOINT COALITIONS

[7] Welfare guarantees in Schelling segregation. *Journal of Artificial Intelligence Research*, 71:143–174, 2021 (with W. Suksompong and A. A. Voudouris).§

§ An earlier version of this paper appeared in the *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, pages 5236–5243, 2021.

FURTHER PUBLICATIONS

This is a list of further publications by the author, which are not part of this thesis.

[8] On the indecisiveness of Kelly-strategyproof social choice functions. *Journal of Artificial Intelligence Research*, 73:1093–1130, 2022 (with F. Brandt and P. Lederer).[¶]

[9] Network creation with homophilic agents. In *Proceedings of the 31th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 151–157, 2022 (with P. Lenzner and A. Melnichenko).

[10] Causes of stability in dynamic coalition formation. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, 2023. Forthcoming (with N. Boehmer and A. M. Kerkmann).

[11] Topological distance games. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, 2023. Forthcoming (with W. Suksompong).[||]

¶ This paper was also presented at the *8th International Workshop on Computational Social Choice (COMSOC)*, 2021. An earlier version of the paper appeared in the *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 251–259, 2021.

|| This paper was presented at the *12th Day(s) on Computational Game Theory*, 2022.

ACKNOWLEDGMENTS

First of all, I would like to thank Felix Brandt for supervising my thesis. You were always there as my scientific advisor and have left me the space I needed to pursue my own projects. Your attitude towards research will be great guidance for my scientific ambitions.

I also want to express my sincere gratitude to a number of people who have supported me on my way to a finished thesis. I want to thank

- my colleagues in the PAMAS group, and especially my two office mates Christian who convinced me to take the bike from Munich to Garching every day and Matthias who made every birthday worth coming. Thank you Chris, René, Patrick, Christian, Anaëlle, Florian, and Johannes for the great time we had.
- my co-authors Anaëlle, Alex, Warut, Patrick, Stefan, Leo, Anna, Pascal, Anna, Niclas, Johannes, Donghao, and of course Felix, with whom I had the honor to work on amazing problems.
- Edith Elkind and Michele Flammini for agreeing to examine my thesis.
- Pascal Lenzner and Tobias Friedrich who I had the pleasure to visit at HPI Potsdam, and Vijay Vazirani who will host me at UCI in spring.
- my mentors Stefan Weltge at TUM and Piotr Faliszewski and Edith Elkind at AAMAS.
- my amazing proof readers Alex, Almut, Louise, and Stefan. I also want to thank Katharina for our many valuable discussions. You are great scientists and friends.
- many fantastic friends who have shared my long path at TUM since the beginning of my bachelor's degree. You are too many to name you all, I just want to thank Stefan for beating me at table soccer once in a while.
- some of the most important coalitions of my life: my amazing roommates, my choir, my family, and mia cara Klara.

You are awesome!

CONTENTS

Part I

SYNTHESIS OF CONTRIBUTIONS

1 INTRODUCTION

Today's world would be unthinkable without the extensive use of algorithms. They help us every day to make decisions or to provide us computational power of which humans are incapable. Even older than the broad field of artificial intelligence, which concerns us today, is game theory, which recently celebrated its hundredth birthday (Borel, 1921; von Neumann, 1928). In short, game theory deals with strategic decision making during the interaction of several individuals. Today, an important aspect of game theory is the algorithmic guidance of these processes as researched, for instance, in the areas of algorithmic game theory and multi-agent systems.

An important aspect of game theory is the emergence of collaboration. In our daily lives, group and team structures are ubiquitous. We live together with our family or friends, we form teams in our working life, and in our free time, we join sport clubs or music groups. The key question is how we establish the right groups or teams. This question is closely related to asking for what we actually want to achieve as individuals.

All of the examples discussed before have in common that we care about who we spend time with. We know our most reliable friends, our most productive co-workers, and who can entertain us the most. The central goal is that the established cooperations should reflect our opinions or preferences about who we would like to form a coalition with.

However, we experience every day that it is not easy to find the groups we want to belong to. Often, there simply is an abundance of options. Formal models of coalition formation try to give a framework to abstract the task of finding suitable cooperations. This helps provide a precise description of the problem and can guide decision processes.

We start by introducing some basic terminology to create a language enabling us to discuss the formation of groups or teams more precisely. Let us refer to the individuals participating in the coalition formation process as a set of *agents*. The goal is to subdivide the agents into different groups which we call *coalitions*. Together, all coalitions yield a *coalition structure* (or a *partition* to be mathematically precise). In a *game*, agents can specify preferences about the coalitions they are part of. The desirability of coalition structures according to the preferences in a specific game is then determined by so-called *solution concepts*, which can be viewed as (desirable) properties of outcomes.

Given the large number of coalitions which an agent can be part of, specifying preferences is not a straightforward task. Therefore, formal models can give guidance for how to express an opinion instead of providing a ranking of all possible options. Usually, this will lead to a loss in generality because some complicated preferences might not be expressible in a certain model. However, this also opens up the possibility to focus on certain aspects of a specific coalition formation process.

For instance, imagine that you are entrusted with the task of forming groups at a workshop where the participants do not know each other in advance. Then, it would be unreasonable if the agents have to take the identity of other agents into account when expressing their preferences. Still, instead of forming arbitrary teams, coalitions could be selected based on preferences about coalition sizes.

In this thesis, we will study several classes of coalition formation games with the goal of computing desirable coalition structures by means of algorithms. In this course, we will also go beyond standard solution concepts. We will investigate new ways of defining desirability, and we will even leave paradigms of the classical theory by dissolving from static solution concepts. Moreover, we will gain new insights in the effects of empathy and homophily on the interpersonal interaction.

1.1 HEDONIC GAMES: THE PREDOMINANT MODEL OF COALITION FORMATION

Since its early days, the treatment of cooperation is an integral part of game theory. In the beginning—and in particular in the seminal book *Theory of Games and Economic Behavior* by von Neumann and Morgenstern (1944)—cooperation was usually analyzed with the aid of a characteristic function that associates a value to every possible coalition. Instead of finding a good coalition structure, this approach aims at rewarding individual contributions appropriately if all agents collaborate as a whole. The value associated with the set of all agents can be distributed among the agents as their payoffs, and the payoff obtained by a certain agent is called their *utility*, which an agent seeks to maximize. Since the distribution of payoffs can be performed arbitrarily, we obtain a model with *transferable* utility.

It was only much later that this view was extended by not taking the coalition structure of the agents as exogenously given. This idea was used in the investigation of the bargaining set by Aumann and Maschler (1964) and then advocated by Aumann and Drèze (1975) in more generality. In their work, the new object of study is the subdivision of the agents and the focus of an agent is their membership in a coalition within a coalition structure. In the sequel, Drèze and Greenberg (1980) go one step further, and start to break away from the assumption of transferable utility between agents. Instead of maximizing payoff, an agent is now interested in the identity of their coalition, that is, their members. In other words, this shifts attention from a cardinal consideration of earning utility to an ordinal consideration of (groups of) individuals. Drèze and Greenberg (1980) coin the term *hedonic games*, where an agent is only concerned about their own coalition while externalities, that is, the composition of the coalition structure outside their own coalition, do not play a role.

It took about twenty years until the framework of hedonic games achieved increased attention starting with important contributions by Banerjee et al. (2001), Cechlárová and Romero-Medina (2001), and Bogomolnaia and Jackson (2002). Hajduková (2006) surveys these early results. Since then, hedonic games have been constantly and extensively studied. In this thesis, we consider coalition formation in the framework of hedonic games.

The two key challenges in hedonic games are finding suitable preference representations, and defining appropriate solution concepts.

The first goal is to consider games with a meaningful interpretation describing a real-life scenario. Over the years, a rich variety of classes of hedonic games has been introduced, capable of modeling settings, such as research team formation (Alcalde and Revilla, 2004), task allocation (Saad et al., 2011), or community detection (Aziz et al., 2019). Many of these classes will be defined formally in Section 2.1.

Second, we want to discuss central ideas of solution concepts. Most of these can be classified as either notions of stability or notions of optimality. The landscape of stability notions in hedonic games is wide but they have in common that agents act based on decisions leading to immediate improvements. This follows the idea of the *core*, a prominent solution concept in transferable utility settings, where an outcome is in the core if there is no set of agents that improves by acting on its own. In more generality, stability means that there is no single agent or set of agents that can perform a beneficial deviation to improve their coalitions.

In contrast, notions of optimality usually take a global perspective. A weak concept of optimality is the classical notion of Pareto optimality which demands that every outcome that is preferred by some agent is worse for another agent. Moreover, if utilities are expressed cardinally, we have a common scale of utilities and can aggregate them to global measures. For instance, we can consider the sum or the minimum of the utilities of the agents in a given coalition structure. These aggregated values are referred to as the utilitarian or egalitarian social welfare. We obtain further optimality notions by considering coalition structures maximizing social welfare.

1.2 SCHELLING SEGREGATION: A MODEL WITH OVERLAPPING COALITIONS

A long-term goal in social sciences has been to model and understand the emergence of segregation such as racial or sexual segregation. This adds the influence of homophily, that is, the desire of humans to be surrounded by like-minded others, to a cooperative scenario.

More than 50 years ago, Thomas Schelling was the first to propose a simple model of segregation (Schelling, 1969; Schelling, 1971). He considers a setting with a set of agents of different *types*, which represent similarity, for example, according to the agents' ethnological background or socio-economic status. The agents live in a street—modeled as a line graph—or a metropolitan area—modeled as a grid graph. Agents are assumed to possess a certain level of homophily in the sense that they seek to live in a neighborhood in which at least a τ fraction of agents has the same type. Whenever an agent is unhappy because of living in a neighborhood of less than a τ fraction of similar agents, they move to another location where they are happy, provided that such a location exists. The surprising result by Schelling is that even a small bias of $\tau \approx \frac{1}{3}$ suffices for the agents to segregate, that is, to eventually live in large clusters of similar agents.

Over the years, this phenomenon has been confirmed in numerous simulations (see, e.g., Clark and Fossett, 2008; Easley and Kleinberg, 2010, Chapter 4). The individual micro-motives by the agents cause an (undesired) macro-behavior on a global scale (Schelling, 2006).

After Schelling's original publications, it has taken a while before the first theoretical breakthroughs. Young (1998) paves the road by considering a probabilistic variant of Schelling's segregation process with the aid of a Markov chain. He considers a set of agents located on a cycle and finds that the stable states correspond to total segregation into intervals of same-type agents. Subsequently, Zhang (2004) extends this result to two dimensions using more involved techniques from stochastic evolutionary game theory. In the sequel, significant progress was also made in the original, unperturbed version of Schelling's model with the goal of measuring the size of monochromatic areas evolving through segregation (Brandt et al., 2012; Barmpalias et al., 2014; Barmpalias et al., 2015; Immorlica et al., 2017).

More recently, a stream of research has started to study Schelling segregation in a game-theoretic setting from a strategic point of view (see, e.g., Chauhan et al., 2018; Echzell et al., 2019; Agarwal et al., 2021). In these so-called *Schelling games*, a set of agents is to be placed on an arbitrary topology graph, that is, any undirected and unweighted graph, which can model simple metropolitan structures as in Schelling's original work, but also more complex structures. In comparison to coalition formation games, the output of a Schelling game is therefore an *assignment* of agents to nodes of a graph, and agents seek to maximize a utility function reflecting the homophily incentive proposed by Schelling.

This model is quite related to coalition formation. However, the essential difference is that the coalition formation process is constrained by the topology graph. This is reflected both in the coalition structures and in the set of available deviations. First, instead of partitioning a set of agents into disjoint coalitions, the immediate *neighborhood* of an agent acts as their coalition, yielding a setting with overlapping coalitions. In particular, if the topology graph of the Schelling game consists only of cliques, we can represent the game by common classes of hedonic games. Second, the available deviations of an agent are also constrained by the topology graph because an agent can only deviate by relocating to another empty node.

1.3 CONTRIBUTION

The majority of the results in this thesis are of algorithmic nature and aim at the possibility of implementing desirable solutions. We examine a specific solution concept and game, and ask fundamental algorithmic questions, such as

> *How can we compute a coalition structure (or assignment) that satisfies the solution concept?*

or, if the existence of outcomes satisfying the solution concept is an issue,

> *How can we decide whether a coalition structure (or assignment) that satisfies this solution concept exists?*

Usually, we ask these questions for a certain class of hedonic games or for the domain of Schelling games. The answer is then of one of two natures. On the one hand, we find positive results in the sense of efficient, that is, polynomial-time, algorithms. On the other hand, we establish computational boundaries in the form of intractabilities, that is, NP-hardness or similar results.[1]

In the case of hardness, we are interested to investigate further how close we are to computational feasibility. Hence, we ask questions like

> *Does a hardness result still hold under stronger restrictions on the class of instances?*

or, more positively

> *Under which further conditions can we provide an efficient algorithm?*

1 We assume that the reader is familiar with classical concepts from computational complexity. For more background on that topic, we refer to, e.g., the text book by Arora and Barak (2009).

For researching the latter question, there are several canonical approaches for the development of algorithms, such as

- *Can we approximate good solutions?*
- *Can we weaken our demands on the solution?* or
- *Can we obtain better results for a subdomain of inputs?*

The first two of these approaches lead to conceptually different yet related solutions, while the third approach yields exact but partial solutions. We will see examples for all of these three approaches. Notably, in the course of determining the computational complexity of various solution concepts, we will close several gaps in the literature.

Additionally, the thesis contains various conceptual contributions. First, we propose novel solution concepts. In the domain of coalition formation, we consider majority-based stability concepts. These offer a compromise between unanimous and completely neglected consent by coalitions involved in deviations. In the domain of Schelling games, we bridge the gap between Pareto optimality and welfare optimality by suggesting appropriate notions of optimality.

Second, the results in this thesis are among the pioneering work on the consideration of *dynamics* based on single-agent deviations in the domain of hedonic games. Instead of determining a fixed coalition structure in an instance which satisfies a notion of stability, we consider an arbitrary initial coalition structure and its instabilities, that is, possible deviations by single agents. If an agent has an incentive to perform a deviation, this leads to a transition to a new coalition structure where we can look for instabilities once again. Continuing this process leads to sequences of coalition structures emerging from sequences of deviations. The key questions that we will ask about the arising dynamics concern their possible or necessary convergence. In addition, we also discuss their running time.

Third, we define *loyalty*, a notion of empathy, in coalition formation. In contrast to much of the literature where new game classes are proposed, loyalty is a concept which is decoupled from utility restrictions and which can be applied to every game (where utilities are expressed cardinally). Given a benchmark game, we obtain its loyal variants, each featuring a different degree of empathy. As we will see, a very high degree of loyalty corresponds to a local version of egalitarianism. For the loyal variants, we study established solution concepts answering similar algorithmic questions as before.

2 MATHEMATICAL MODELS

In this chapter, we introduce the formal models that this thesis builds upon. The majority of this chapter deals with hedonic games, our framework to reason about coalition formation. We will start with a general introduction and have a look at important specific classes of hedonic games. We then discuss desirable outcomes, mostly measured by notions of stability and optimality. Next, we introduce the framework for two novel conceptual ideas, namely dynamics and loyalty. In the final section, we broaden the picture and introduce the formal model of Schelling segregation.

2.1 HEDONIC GAMES

We model coalition formation in the framework of hedonic games going back to Drèze and Greenberg (1980).

Given a positive integer $i \in \mathbb{N}$, we use the common notation $[i] := \{1, \dots, i\}$. Let $N := [n]$ be a finite set of *agents*. Any subset of N is called a *coalition*. We denote the set of all coalitions containing agent $i \in N$ by $\mathcal{N}_i := \{C \subseteq N : i \in C\}$. A *coalition structure* (or simply *partition*) is a subset $\pi \subseteq 2^N$ which partitions N, that is, $\bigcup_{C \in \pi} C = N$, and for every pair $C, D \in \pi$, it holds that $C = D$ or $C \cap D = \emptyset$. Given an agent $i \in N$ and a partition π, let $\pi(i)$ denote the coalition of i, i.e., the unique coalition $C \in \pi$ with $i \in C$. Two prominent partitions are the *singleton partition* where every agent forms a coalition on their own, and the *grand coalition* where the agents form one single joint coalition.

Definition 2.1 *(Drèze and Greenberg, 1980)*

A *(ordinal) hedonic game* is defined by a tuple $(N, \succsim)$ where $\succsim = (\succsim_i)_{i \in N}$. Thereby, $\succsim_i$ is a weak order, that is, a complete, reflexive, and transitive binary relation, over $\mathcal{N}_i$ which represents the preferences of agent i.

We also call $\succsim_i$ agent i's *preference order*. Given two partitions π and π', we say that agent i *(weakly) prefers* π over π' if and only if $\pi(i) \succsim_i \pi'(i)$.

Given a preference order $\succsim_i$, we denote its strict part and indifference part by $\succ_i$ and $\sim_i$, respectively. A preference order is said to be *strict* if it is antisymmetric. Moreover, a hedonic game $(N, \succsim)$ is said to be *globally ranked* if there exists a weak order $\succsim_G$ over 2^N such that, for each $i \in N$, the order $\succsim_i$ is the restriction of $\succsim_G$ to $\mathcal{N}_i$ (Farrell and Scotchmer, 1988; Abraham et al., 2007b).

In some cases, we have cardinal representations of the preference orders. Then, we also represent a hedonic game in the form (N, u) where $u = (u_i)_{i \in N}$ specifies a *utility function* $u_i \colon \mathcal{N}_i \to \mathbb{Q}$ for every agent i. If a hedonic game is given in this representation, we also speak of a *cardinal hedonic game*. The preference order of an agent is then determined by comparing the utilities of coalitions. In other words, a cardinal hedonic game (N, u) is associated with its ordinal variant $(N, \succsim)$, where, for every agent $i \in N$ and every pair of coalitions $C, D \in \mathcal{N}_i$, it holds that $C \succsim_i D$ if and only if $u_i(C) \geqslant u_i(D)$.

We say that a utility function has *binary* utility values if it only attains the values 0 and 1.

Since $|\mathcal{N}_i| = 2^{n-1}$, the preferences of the agents are rarely given explicitly, but rather in some concise representation. In the following, we discuss several representations leading to various classes of hedonic games.

2.1.1 Preference Representation in Hedonic Games

Representing preferences in hedonic games is a tradeoff between expressiveness and efficient encoding. Since providing a complete list of all preferences is infeasible, one idea is to restrict attention to coalitions that seem particularly interesting. A coalition $C \in \mathcal{N}_i$ is said to be *individually rational* if $C \succsim_i \{i\}$, that is, C is weakly preferred to being in a singleton coalition. A hedonic game is encoded with *individually rational lists of coalitions* (IRLC) if, for every agent i and pairs of coalitions $C, D \in \mathcal{N}_i$ with $\{i\} \succ_i C$ and $\{i\} \succ_i D$, it holds that $C \sim_i D$ (Ballester, 2004). In other words, it suffices to specify for every agent an incomplete preference list which contains exactly their individually rational coalitions.

Example 2.2 *(Brandt and Bullinger, 2022)*

We define a simple hedonic game $(N, \succsim)$ in IRLC representation. Let the agent set be $N = \{a, b, c, d\}$ and preferences be given as

- $N \succ_a \{a, b\} \succ_a \{a, c\} \succ_a \{a, d\} \succ_a \{a\}$,
- $N \succ_b \{b, c\} \succ_b \{b, a\} \succ_b \{b, d\} \succ_b \{b\}$,
- $\{c, a\} \succ_c \{c, b\} \succ_c \{c, d\} \succ_c N \succ_c \{c\}$, and
- $\{d, a\} \sim_d \{d, b\} \sim_d \{d, c\} \succ_d N \succ_d \{d\}$.

The only individually rational coalitions are of size 1, 2, and 4. Coalitions of size 3 are equally bad and less preferred than being in a singleton coalition. Consider the coalition structures $\pi_1 = \{\{a, b\}, \{c, d\}\}$ and $\pi_2 = \{\{a, c\}, \{b, d\}\}$. Then π_1 is strictly preferred by agents a and b while π_2 is strictly preferred by agent c. Agent d is indifferent between the two partitions. $\triangleleft$

Let us discuss the IRLC representation. On the one hand, it is not fully expressive, that is, cannot express every preference order over $\mathcal{N}_i$, because a strict comparison of two coalitions that are both not individually rational is not possible. On the other hand, if agents have simple preferences, then a succinct representation is possible. For instance, it can be the case that only coalitions of bounded size are feasible and therefore ranked as individually rational.

Definition 2.3 *(Irving, 1985; Brandt and Bullinger, 2022)*
A hedonic game is called a

- *roommate game* if only coalitions of size at most 2 are individually rational, and
- *flatmate game* if only coalitions of size at most 3 are individually rational.

In principle, coalitions of size larger than 2 or 3 are not disallowed in an output partition in roommate games or flatmate games. This seems counterintuitive if the output is, for instance, supposed to be a matching, but this is merely a cosmetic issue. A coalition which is less preferred than the singleton coalition by all of its members is not part of a partition selected by any reasonable solution concept. Hence, we can just disregard it in all relevant situations.

The IRLC representation has another, technical advantage, namely that it facilitates polynomial-time reductions and therefore paves the way for hardness results. For reductions to run in polynomial time, it only matters whether the reduced instances are sufficiently small with respect to the encoding size of their source instances. It is unimportant whether there are exponentially large instances if they never occur as reduced instances. This idea, which was first applied by Ballester (2004), is, for instance, frequently used in Publication 2.

Another common idea to obtain feasible preference relations is to only allow agents to express succinct information about their preferences. This approach leads to an efficient encoding at the cost of some expressiveness. It is therefore all the more important that the obtained game classes have a meaningful interpretation. A very simple idea of this type is that agents only rank single agents instead of all possible coalitions (Cechlárová and Romero-Medina, 2001). Then, one can lift the preferences about individuals to preferences about coalitions by comparing, for example, the best or the worst agents of different coalitions.

In this thesis, we consider anonymous hedonic games and hedonic diversity games, two classes of hedonic games that follow a similar spirit. In each of them, agents have a complete ranking over all coalitions which is elicited from succinct secondary information. First, we define anonymous hedonic games where only coalition sizes matter.

Definition 2.4 *(Bogomolnaia and Jackson, 2002)*

> A hedonic game $(N, \succsim)$ is called an *anonymous hedonic game* if, for every agent $i \in N$, there exists a weak order $\succsim_i^S$ over integers in $[n]$ (S for sizes) such that $\pi(i) \succsim_i \pi'(i)$ if and only if $|\pi(i)| \succsim_i^S |\pi'(i)|$.

In hedonic diversity games, agents' preferences are also based on a ranking over numbers. The idea is that the set of agents is subdivided into two sets of different types. The valuation of a coalition then depends on the proportion of the two types in the coalition. Hedonic diversity games are motivated by situations like the meeting of senior and junior researchers at a conference dinner or the joint engagement of politicians and scientists in a climate task force.

Definition 2.5 *(Bredereck et al., 2019)*

> A hedonic game $(N, \succsim)$ is called a *hedonic diversity game* if preferences are elicited as follows. Assume that the agents are divided into two different *types*, called red and blue agents. They are represented by the subsets $R \subseteq N$ and $B \subseteq N$, respectively, such that $N = R \cup B$ and $R \cap B = \emptyset$. For each agent $i \in N$, there exists a weak order $\succsim_i^F$ over $\{\frac{p}{q} : p \in [|R|] \cup \{0\}, q \in [n]\}$ (F for fractions) such that $\pi(i) \succsim_i \pi'(i)$ if and only if $\frac{|R \cap \pi(i)|}{|\pi(i)|} \succsim_i^F \frac{|R \cap \pi'(i)|}{|\pi'(i)|}$.

An anonymous game (or hedonic diversity game) is said to be *single-peaked* if, for each agent $i \in N$ and each triple of integers $x, y, z \in [n]$ (or $x, y, z \in \{\frac{p}{q} : p \in |R| \cup \{0\}, q \in [n]\}$) with $x > y > z$ or $z > y > x$, it holds that $x \succ_i^S y$ implies $y \succsim_i^S z$ (or $x \succ_i^F y$ implies $y \succsim_i^F z$).[2] In other words, single-peakedness means that the preferences of every agent peak at a most preferred coalition size (or ratio), and then the liking of coalition sizes (or ratios) monotonically decays for both increasing and decreasing sizes (or ratios).

Note that, in contrast to the preference representations considered in this thesis, there also exist fully expressive representations of hedonic games apart from providing complete preference rankings as lists. For instance, Elkind and Wooldridge (2009) propose an encoding by Boolean formulas. Similar to the IRLC encoding, this encoding has exponential size in the worst case, but simple preferences can be represented succinctly.

2.1.2 Cardinal Classes of Hedonic Games

A prominent modeling paradigm for hedonic games is that we can elicit preferences from cardinal utilities for single agents, that is from utility functions of the form $u_i^S : N \to \mathbb{Q}$ (S for single). We will now introduce three classes of cardinal hedonic games (and some subclasses)

2 In Publication 4, single-peakedness is called *natural* single-peakedness.

based on this assumption. The idea is similar to obtaining a full ranking over coalitions from information about single agents as in the ordinal classes by Cechlárová and Romero-Medina (2001). However, since we have cardinal values, we have more opportunities to aggregate preferences. A particularly natural aggregation methods is to take the sum of utilities.

Definition 2.6 *(Bogomolnaia and Jackson, 2002)*

A cardinal hedonic game (N, u) is called an *additively separable hedonic game* if, for every agent $i \in N$, there exists a function $u_i^S \colon N \to \mathbb{Q}$ such that $u_i(C) = \sum_{j \in C} u_i^S(j)$ for all $C \in \mathcal{N}_i$.

We therefore also represent an additively separable hedonic game as the game (N, u^S) where $u^S = (u_i^S)_{i \in N}$ is the vector of single-agent utility functions. An additively separable hedonic game (N, u^S) is said to be *symmetric* if, for all pairs of agents $i, j \in N$, it holds that $u_i^S(j) = u_j^S(i)$.

There are a few interesting subclasses of additively separable hedonic games in which only a single positive and negative utility value is attained. The interpretation is that an agent simply distinguishes agents of positive and negative utility as friends and enemies. Still, the ratio between the absolute values of the negative and positive utility matters, and can express a priority.

Definition 2.7 *(Dimitrov et al., 2006)*

An *appreciation-of-friends game* is an additively separable hedonic game (N, u^S) such that, for every pair of agents $i, j \in N$, it holds that $u_i^S(j) \in \{n, -1\}$. A *friends-and-enemies game* is an additively separable hedonic game (N, u^S) such that, for every pair of agents $i, j \in N$, it holds that $u_i^S(j) \in \{1, -1\}$.

While appreciation-of-friends games give priority to maximizing friends, subject to which the minimization of enemies is relevant, friends-and-enemies games consider these two objectives with equal importance. Similar to appreciation-of-friends games, there also exist *aversion-to-enemies games*, where the range of the utility functions is $\{1, -n\}$, and therefore priority is given to avoiding enemies.

In contrast to the additive utility aggregation, it is also possible to take the average utility. Dependent on whether an agent also considers themselves, we obtain two classes of hedonic games. If we consider the whole coalition size, we obtain fractional hedonic games.

Definition 2.8 *(Aziz et al., 2019)*

A cardinal hedonic game (N, u) is called a *fractional hedonic game* if, for every agent $i \in N$, there exists a function $u_i^S \colon N \to \mathbb{Q}$ such that the following two conditions hold:

- It holds that $u_i^S(i) = 0$.
- It holds that $u_i(C) = \frac{\sum_{j \in C} u_i^S(j)}{|C|}$ for all $C \in \mathcal{N}_i$.

If we do not consider the agent themselves in the denominator, then the fraction corresponds to the expected utility when selecting another agent in one's coalition uniformly at random. This yields modified fractional hedonic games.

DEFINITION 2.9 *(Olsen, 2012)*
A cardinal hedonic game (N, u) is called *modified fractional hedonic game* if, for every agent $i \in N$, there exists a function $u_i^S \colon N \to \mathbb{Q}$ such that the following two conditions hold:

- It holds that $u_i^S(i) = 0$ and $u_i(\{i\}) = 0$.
- It holds that $u_i(C) = \frac{\sum_{j \in C} u_i^S(j)}{|C|-1}$ for all $C \in \mathcal{N}_i \setminus \{\{i\}\}$.

2.2 DESIRABLE PARTITIONS

It remains to define goals for our coalition formation games. These are formulated in the form of solution concepts, that is, properties that may or may not be satisfied by a certain partition. In general, there exist notions of stability and optimality. The former concerns the prospect of agents maintaining their coalitions from an individual perspective, while the latter deals with desirability from a global perspective. In addition, there are some solution concepts, such as popularity and individual rationality, that have a flavor from both worlds. An overview of most solution concepts for hedonic games considered in this thesis is given in Figure 2.1.

Throughout all of this and the next section, we implicitly assume that we operate on a fixed hedonic game $(N, \succsim)$.

A partition π is said to be *individually rational* if, for every agent $i \in N$, it holds that $\pi(i) \succsim_i \{i\}$, that is, every agent's coalition is individually rational. Individual rationality can be seen as a basic requirement regarding both stability and optimality. On the one hand, no agent in an individually rational partition has an incentive to perform a deviation towards forming a singleton coalition. On the other hand, individual rationality gives a mild guarantee of quality to every agent.

2.2.1 Classical Notions of Stability

Stability concepts are based on incentives of single agents or groups of agents to perform deviations. We start with the former. A *single-agent deviation* performed by agent i transforms a partition π into a partition π' where $\pi(i) \neq \pi'(i)$ and, for all agents $j \neq i$, it holds that $\pi(j) \setminus \{i\} = \pi'(j) \setminus \{i\}$. We write $\pi \xrightarrow{i} \pi'$ to denote a single-agent deviation performed by agent i transforming a partition π into a partition π'.

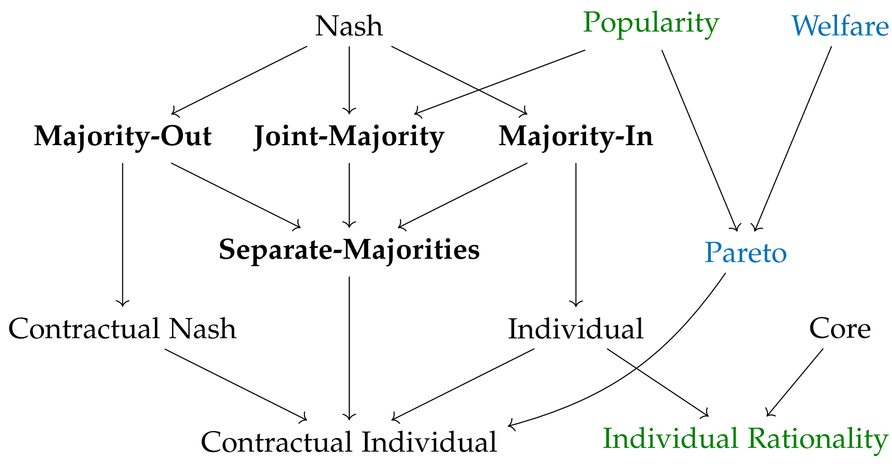


Figure 2.1: Overview of solution concepts for hedonic games (Aziz and Savani, 2016; Brandt et al., 2022a). Notions of stability are in black while notions of optimality are colored blue. Solution concepts with the flavor of both stability and optimality are colored green. The bold concepts are novel and introduced in Publication 5. The arrows display implications.

Our paradigm for stability are myopic rational agents who only engage in a deviation if it immediately makes them better off. This is reflected in the definition of our first stability concept, Nash stability, the analogue of Nash equilibrium in non-cooperative game theory (Nash, 1950).

DEFINITION 2.10 *(Bogomolnaia and Jackson, 2002)*
A *Nash deviation* is a single-agent deviation $\pi \xrightarrow{i} \pi'$ making agent i better off, i.e., $\pi'(i) \succ_i \pi(i)$. A partition which allows for no Nash deviation is said to be *Nash-stable*.

Nash stability is a strong concept and comes with the drawback that only the preferences of the deviating agent are of relevance. Therefore, various refinements have been proposed which additionally require the consent of the abandoned and the welcoming coalition. We follow the approach of Publication 5 and introduce consent-based stability concepts using the notion of favor sets.

Consider a coalition $C \subseteq N$ and an agent $i \in N$. The *favor-in set* of C with respect to i is the set of agents in C (excluding i) that strictly favor having i inside of C rather than outside, i.e., $F_{\text{in}}(C, i) = \{j \in C \setminus \{i\} \colon C \cup \{i\} \succ_j C \setminus \{i\}\}$. Similarly, the *favor-out set* of C with respect to i is the set of agents in C (excluding i) that strictly favor having i outside of C rather than inside, i.e., $F_{\text{out}}(C, i) = \{j \in C \setminus \{i\} \colon C \setminus \{i\} \succ_j C \cup \{i\}\}$. Note that in both of the preceding definitions, the agent i may or may not be a member of coalition C.

Unanimous consent was already an important ingredient of stability in the early work by Drèze and Greenberg (1980). They motivated consent based on an example by Meade (1972) in the context of labor markets. There, it is reasonable that a partner can only join or abandon a firm provided the consent of all other present partners. The concepts in the form used in this thesis were formulated by Bogomolnaia and Jackson (2002) and Dimitrov and Sung (2007).

DEFINITION 2.11 *(Bogomolnaia and Jackson, Dimitrov and Sung)*
An *individual deviation* is a Nash deviation $\pi \xrightarrow{i} \pi'$ such that $F_{out}(\pi'(i), i) = \emptyset$. Similarly, a *contractual deviation* is a Nash deviation $\pi \xrightarrow{i} \pi'$ such that $F_{in}(\pi(i), i) = \emptyset$. A single-agent deviation that is both an individual and a contractual deviation is called a *contractual individual deviation*.

A partition is said to be *individually stable*, *contractually Nash stable*, or *contractually individually stable* if it allows for no individual deviation, contractual deviation, or contractual individual deviation, respectively.

We move on to stability based on group deviations. The solution concept that is ubiquitous since the birth of cooperative game theory is the core (von Neumann and Morgenstern, 1944). Essentially, it is the equivalent of Nash stability when considering group deviations. For hedonic games, it was first considered by Banerjee et al. (2001) and Bogomolnaia and Jackson (2002).

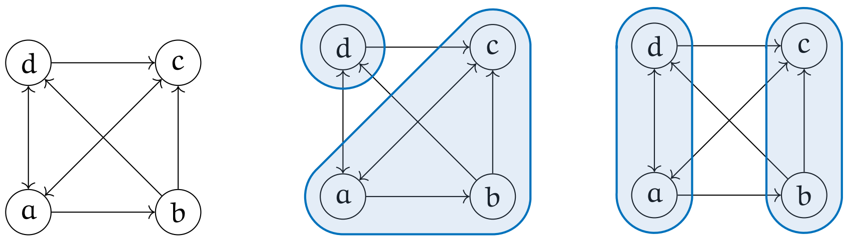
DEFINITION 2.12 *(Banerjee et al., 2001)*
A coalition $C \subseteq N$ is called a *blocking coalition* with respect to a partition π if, for all $i \in C$, $C \succ_i \pi(i)$. A partition π is said to be in the *core* if there exists no blocking coalition with respect to π.

EXAMPLE 2.13
We demonstrate the introduced stability concepts with the aid of an example. Consider the friends-and-enemies game (N, u^S) where $N = \{a, b, c, d\}$ and $u^S_a(b) = u^S_a(c) = u^S_a(d) = u^S_b(c) = u^S_b(d) = u^S_c(a) = u^S_d(a) = u^S_d(c) = 1$. All other pairwise utilities are assumed to be -1. The game is visualized in Figure 2.2a. There, a directed edge from agent v to agent w means that $u^S_v(w) = 1$.

First, consider the partition $\pi_1 = \{\{a, b, c\}, \{d\}\}$ as depicted in Figure 2.2b. It holds that π_1 is individually rational, individually stable, and in the core. Individual rationality follows because the utility of all agents is non-negative.

For individual stability, note that a and c are not better off by joining d or forming a singleton coalition. Moreover, b is not allowed to join d because d blocks this. Finally, d cannot join $\{a, b, c\}$ because c blocks this.

For membership in the core, we first realize that the only better coalition for agent a would be the grand coalition N. However, forming N is not beneficial for agent c. Hence, a cannot be part of a blocking coalition. But this immediately implies



(a) Friendship relation. (b) Partition π_1. (c) Partition π_2.

Figure 2.2: Friends-and-enemies game of Examples 2.13 and 2.15. On the left, the friendship relation is visualized. Partition π_1 in the middle is an individually stable partition in the core. Partition π_2 at the right is contractually Nash stable and even majority-out stable.

that c cannot be part of a blocking coalition, either, because the only coalition left yielding non-negative utility is the singleton coalition $\{c\}$ which is no improvement. For similar reasons, the remaining agents cannot be part of a blocking coalition. Hence, π_1 is in the core.

On the other hand, d can perform a contractual Nash deviation to join $\{a, b, c\}$. Hence, π_1 is not contractually Nash stable, and therefore also not Nash stable.

Second, consider the partition $\pi_2 = \{\{a, d\}, \{b, c\}\}$. Since c receives a utility of -1 for their coalition, this partition is not individually rational, and therefore neither in the core nor individually stable.

However, the partition is contractually Nash stable. Indeed, agents a, c, and d are not allowed to leave their respective coalitions. Moreover, b has no incentive to form a singleton coalition or to join $\{a, d\}$ because this would decrease their utility. $\triangleleft$

2.2.2 Majority-Based Stability

While Nash deviations do not take into account any agent except the deviator, unanimous consent is a strong restriction, especially in settings with medium-size or large coalitions such as joining a club, choir, or orchestra. Therefore, it is natural to consider hybrid concepts where abandoned or welcoming agents are not neglected but still do not have the power to veto decisions. A good compromise seems to take decisions by a simple majority vote. This was introduced in Publication 5 and further studied in Publication 3. Our majority-based stability concepts are an important special case in the class of threshold stability notions as introduced by Gairing and Savani (2019).

Definition 2.14 *(Brandt et al., 2022a)*

A Nash deviation $\pi \xrightarrow{i} \pi'$ is called a *majority-in deviation* if it holds that $|F_{in}(\pi'(i), i)| \geqslant |F_{out}(\pi'(i), i)|$ and a *majority-out deviation* if it holds that $|F_{out}(\pi(i), i)| \geqslant |F_{in}(\pi(i), i)|$. A single-agent deviation that is both a majority-in deviation and a majority-out deviation is called a *separate-majorities deviation*.

A partition is said to be *majority-in stable*, *majority-out stable*, or *separate-majorities stable* if it allows for no majority-in deviation, majority-out deviation, or separate-majorities deviation, respectively.

Finally, it is possible to relax separate-majorities stability by performing one joint vote instead of two separate votes. A Nash deviation $\pi \xrightarrow{i} \pi'$ is called a *joint-majority deviation* if $|F_{in}(\pi'(i), i)| + |F_{out}(\pi(i), i)| \geqslant |F_{out}(\pi'(i), i)| + |F_{in}(\pi(i), i)|$. A partition is said to be *joint-majority stable* if it allows for no joint-majority deviation.

Example 2.15

We continue Example 2.13. Despite being individually stable, partition π_1 is not majority-in stable. Indeed, d is allowed to join the coalition $\{a, b, c\}$ because a and b favor this deviation. Hence, this yields a majority-in deviation.

On the other hand, partition π_2 is even majority-out stable. The partners of a, c, and d have the power to prevent any of their potential deviations. In addition, as before, b can not even perform a Nash deviation.

2.2.3 Notions of Optimality

A very common concept of optimality is Pareto optimality. It was already studied in the early work by Drèze and Greenberg (1980), and later by Bogomolnaia and Jackson (2002).

Definition 2.16

A partition π' *Pareto-dominates* a partition π if, for all agents $i \in N$, it holds that $\pi'(i) \succsim_i \pi(i)$ and, for some agent $j \in N$, it holds that $\pi'(j) \succ_j \pi(j)$. A partition is said to be *Pareto-optimal* if it is not Pareto-dominated by any other partition.

Clearly, it is undesirable for the whole set of agents to be in a Pareto-dominated partition because some agents may be better off while no other agent is harmed. Still, Pareto-optimal partitions are not necessarily stable. While contractual individual stability is a refinement of Pareto optimality, there is no logical relationship between the other stability concepts introduced in the previous section and Pareto optimality.

Also, there usually exist many Pareto-optimal partitions, which can differ a lot in quality. For instance, it is easy to see that, for every agent, there exists a Pareto-optimal partition where they are in a most preferred coalition. In addition, if we are given a cardinal hedonic game, we can also define the (utilitarian) social welfare associated with a partition. Maximizing this welfare also yields Pareto-optimal partitions.

DEFINITION 2.17

Let a cardinal hedonic game (N, u) be given and consider a partition π. The *(utilitarian) social welfare* of partition π is defined as $SW(\pi) := \sum_{i \in N} u_i(\pi(i))$. A partition maximizing the social welfare is said to be *welfare-optimal*.

2.2.4 Popularity

In this section, we introduce the concept of popularity, which was first considered by Gärdenfors (1975) in the domain of matching. While popularity of matchings is a thoroughly considered concept (see, e.g., the book chapter by Cseh, 2017), it has only been researched little in the domain of coalition formation (Aziz et al., 2013b; Kerkmann et al., 2020). Publication 2 presents both new results for roommate and flatmate games, as well as the first in-depth consideration of popularity in additively separable hedonic games and fractional hedonic games.

Roughly speaking, a partition π is popular if it is impossible to present another partition π' that would beat π in a majority vote among the agents. Therefore, popularity corresponds to (weak) Condorcet winners from social choice theory. Given two partitions π and π', let $N(\pi, \pi')$ be the set of agents who prefer π over π', i.e., $N(\pi, \pi') := \{i \in N : \pi(i) \succ_i \pi'(i)\}$. Then, the *popularity margin* of π and π' is defined as $\phi(\pi, \pi') := |N(\pi, \pi')| - |N(\pi', \pi)|$. We say that π is *more popular* than π' if $\phi(\pi, \pi') > 0$.

DEFINITION 2.18 *(Gärdenfors, 1975)*

A partition π is said to be *popular* if, for all partitions π', it holds that $\phi(\pi, \pi') \geqslant 0$. A partition π is called *strongly popular* if, for all partitions π' with $\pi' \neq \pi$, it holds that $\phi(\pi, \pi') > 0$.

In other words, a partition π is popular if and only if no partition is more popular than π, and strongly popular if and only if π is more popular than every other partition. Note that there can be at most one strongly popular partition in a hedonic game, if any.

Popularity combines ideas from both stability and optimality. On the one hand, a majority of agents cannot organize a deviation based on a vote to change the status quo. On the other hand, possibly even more inherent, popularity takes a global perspective by comparing any two partitions. It therefore gives a guarantee against every other partition, which can be interpreted as a form of optimality.

Popularity is a refinement of certain majority-based stability concepts. Among these, joint-majorities stability can be seen as a natural

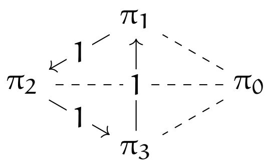


Figure 2.3: Popularity margins in Example 2.20. We use arrows for positive popularity margins, and dashed lines for a popularity margin of 0.

local version of popularity because it is based on a vote among the two coalitions involved in a single-agent deviation.

The existence of popular partitions is not guaranteed in many hedonic games. However, this can be circumvented by introducing randomization and considering probability distributions over partitions (instead of deterministic outcomes).

A *mixed partition* is a set $\{(\pi_1, p_1), \dots, (\pi_k, p_k)\}$, where $\pi_1, \dots, \pi_k$ are pairwise different partitions and $(p_1, \dots, p_k)$ is a probability distribution. Given two mixed partitions $p = \{(\pi_1, p_1), \dots, (\pi_k, p_k)\}$ and $q = \{(\sigma_1, q_1), \dots, (\sigma_l, q_l)\}$, we define the *popularity margin* of p and q as their expected popularity margin, i.e.,

$$\phi(p, q) = \sum_{i=1}^{k} \sum_{j=1}^{l} p_i q_j \phi(\pi_i, \sigma_j).$$

The definition of popularity carries over to mixed partitions.[3]

Definition 2.19 *(Kavitha et al., 2011)*
A mixed partition p is said to be *mixed popular* if, for all mixed partitions q, it holds that $\phi(p, q) \geqslant 0$.

Note that mixed popularity corresponds to the randomized voting rule *maximal lotteries* in the domain of social choice (see, e.g., Fishburn, 1984; Brandl and Brandt, 2020). We demonstrate popularity and mixed popularity in an example.

Example 2.20 *(Brandt and Bullinger, 2022)*
We consider the hedonic game in Example 2.2. There, the Pareto-optimal partitions, which are the only relevant partitions for any type of popularity, are $\pi_0 = \{N\}$, $\pi_1 = \{\{a, b\}, \{c, d\}\}$, $\pi_2 = \{\{a, c\}, \{b, d\}\}$, and $\pi_3 = \{\{a, d\}, \{b, c\}\}$. Their popularity margins are depicted in Figure 2.3. In particular, π_0 is the only (deterministic) popular partition, and there is no strongly popular partition. Also, $p = \{(\pi_1, 1/3), (\pi_2, 1/3), (\pi_3, 1/3)\}$ is a mixed popular mixed partition. It holds that $\phi(p, \pi_0) = \phi(p, \pi_1) = \phi(p, \pi_2) = \phi(p, \pi_3) = 0$. ◁

3 Note that it is unreasonable to consider a concept like mixed strong popularity. Mixed strongly popular partitions are always degenerate in that they encompass a single partition with probability one. Hence, the randomized and deterministic notion coincide.

2.3 DISTRIBUTED COALITION FORMATION

Classical stability concepts are in some sense static. For instance, the existence problem is only concerned with the availability of a desirable solution, but it is unclear how such a solution should be obtained. Therefore, we seek to complement the static view by a dynamical view where agents take advantageous myopic decisions, and we investigate whether this leads to a dynamics reaching a stable partition. The basic idea is that notions of stability naturally give an incentive to deviate.

While dynamics were studied before in the domain of matching (see, e.g., Roth and Vande Vate, 1990; Abeledo and Rothblum, 1995; Brandt and Wilczynski, 2019), their consideration in coalition formation has only received increasing attention during the last years (Bilò et al., 2018; Hoefer et al., 2018; Carosi et al., 2019). Publication 4 is among the pioneering work on dynamics in hedonic games.

Given any partition that is not stable with respect to a single-agent stability notion, there exists an agent that can perform a beneficial deviation. After this deviation, we can once again look for possible deviations. Iterating this process, we obtain an execution of a dynamics.

Definition 2.21 *(Brandt et al., 2021)*

A sequence $(\pi_k)_{k\geqslant 0}^{K}$ is called an *execution of the IS dynamics* (IS for individual stability) if, for every k with $1 \leqslant k \leqslant K$, there exists an agent $d_k \in N$ such that $\pi_{k-1} \xrightarrow{d_k} \pi_k$ is an individual deviation. The partition π_0 is then called the *starting partition*.

In the previous definition, we define dynamics based on individual stability, which yields the type of dynamics investigated in Publication 4. In addition, it is also possible to consider dynamics for any other stability concept. In the course of this thesis, we also briefly investigate JMS, SMS, MIS, and MOS dynamics which are based on joint-majority, separate-majorities, majority-in, and majority-out deviations, respectively.

In Definition 2.21, we allow the case $K = \infty$, corresponding to an infinite execution of the IS dynamics. The key algorithmic questions about dynamics concern their convergence. Given a hedonic game, we say that the IS dynamics *possibly converges* from starting partition π_0 if there exists a finite execution of an IS dynamics with starting partition π_0 that terminates in an individually stable partition. Similarly, given a hedonic game, we say that the IS dynamics *necessarily converges* from starting partition π_0 if every execution of an IS dynamics with starting partition π_0 is finite.[4] We say that the IS dynamics *necessarily* (or *possibly*) *converges* in a given hedonic game,

4 In Publication 4, possible convergence is referred to by the *existence of a path to stability* and necessary convergence is called *guaranteed convergence*.

if it necessarily (or possibly) converges from every starting partition. By contrast, if there exists an infinite execution of the IS dynamics from some starting partition, we say that the IS dynamics may *cycle*.

Necessary convergence of dynamics is independent of the selection of deviations in the cases where there are multiple choices. More precisely, neither the agent that performs the deviation nor the exact deviation performed by this agent matters. In some cases, however, it can be useful to restrict possible deviations. We introduce one arguably weak selection rule that is useful in the context of hedonic diversity games.

Assume that a hedonic diversity game $(N, \succsim)$ is given. Recall that the agents are divided into two sets R and B of red and blue agents. We call a coalition $C \subseteq N$ *homogeneous* if it consists only of agents of one type, i.e., $C \subseteq R$ or $C \subseteq B$. Now, we say that a deviation satisfies *solitary homogeneity* if, whenever the target coalition of the deviator is homogeneous, then it is a singleton coalition (Brandt et al., 2022b). Note that whenever an agent can perform an IS deviation, then they can also perform a deviation satisfying solitary homogeneity. Indeed, they can form the homogeneous singleton coalition instead of joining existing homogeneous coalitions.

2.4 LOYALTY IN CARDINAL HEDONIC GAMES

In most of the literature discussed so far, the analysis of hedonic games has taken a clear path: Consider a certain class of games and investigate a specific solution concept on this class. Different classes of hedonic games have mostly been considered in isolation and similar computational questions have been answered again and again. In Publication 6, we take a different route. Before we apply a solution concept to a game, we *modify* the game in a meaningful way. The original game serves as a benchmark game, while the utility functions in the modified game integrate empathy incentives. In this way, instead of gaining insight into a single class of games, we learn about the change of the coalition formation process as a whole if agents change their behavior. Such an approach has been studied in the non-cooperative game theory literature with the goal to measure the degree of empathy or altruism needed to challenge the selfishness of players leading to phenomena like the prisoner's dilemma (see, e.g., Apt and Schäfer, 2014). Similar ideas have also been discussed in economics (Mueller, 1986) and network design (Elias et al., 2010).

In the cooperative game theory literature, our model are altruistic hedonic games as defined by Kerkmann et al. (2022). These games assume a structure with friends and enemies among the agents as in appreciation-of-friends games introduced in Definition 2.7. Then, an agent's utility is aggregated by taking the sum or minimum utility of an agent and their friends in their own coalition.

Our notions of loyalty as developed in Publication 6 differ in two key aspects from the model by Kerkmann et al. (2022). First, we generalize by allowing more general underlying games. The only restriction we impose is the ability to express single-agent utilities by cardinal values. Moreover, loyalty can have several degrees. The loyal variant of a hedonic games is once again a hedonic game. Hence, we can consider the loyal variant's loyal variant, and can iterate this idea.

In this section, we assume that we operate on a fixed cardinal hedonic game (N, u). Given an agent $i \in N$, we define their *loyalty set* as $L_i = \{j \in N \setminus \{i\} \colon u_i(\{i, j\}) > 0\}$. This set specifies the agents having an effect on the cardinal utilities of an agent under loyalty. We use this set to define the loyal variant of a game.

Definition 2.22 *(Bullinger and Kober, 2021)*
Let a cardinal hedonic game (N, u) be given. Then, its *loyal variant* (N, u^L) is given by utility functions $u_i^L \colon \mathcal{N}_i \to \mathbb{Q}$ defined by $C \mapsto \min_{j \in C \cap (L_i \cup \{i\})} u_j(C)$ for every agent $i \in N$.

In other words, the utility function of an agent in the loyal variant is defined by the minimum utility in the benchmark game of the agent themselves and agents within their coalition towards which they express loyalty. In the special case where the benchmark game is an appreciation-of-friends game, this coincides with the minimum-based equal-treatment preferences defined by Kerkmann et al. (2022) in the context of altruistic hedonic games. As we mentioned before, the loyal variant is a cardinal hedonic game itself, and we can consider its own loyal variant. Hence, we can recursively define several loyal variants. Given a cardinal hedonic game (N, u), we define its k*-fold loyal variant* (N, u^k) by setting $(N, u^1) = (N, u^L)$, and, for $k \geqslant 1$, defining the $(k+1)$-fold loyal variant (N, u^{k+1}) as the loyal variant of (N, u^k).

A natural question is whether this process terminates. Indeed, our first result in Section 3.2.3 will establish convergence after at most n steps.

2.5 SCHELLING SEGREGATION

In this section, we introduce a game-theoretic model of Schelling segregation. First, we define the formal model, and subsequently the solution concepts considered in this thesis.

2.5.1 Schelling Games on Graphs

The first game-theoretic approach to Schelling segregation is due to Chauhan et al. (2018). In their model, strategic agents seek to optimize an objective that is composed of reaching a threshold of similar agents in their neighborhood (as in Schelling's original model) and of occupying a favorable location. Closely thereafter, Agarwal et al. (2021) introduced a variant of this game called *Schelling games on graphs*. This model combines homophily and location incentives by assigning a characteristic to every agent. The characteristic of an agent is that they are either strategic or stubborn. A strategic agent seeks to maximize a homophily incentive while a stubborn agent blocks locations. The latter can be viewed as a very strong location incentive. Stubborn agents seem to be very restrictive and blur the behavior caused by pure homophily.[5] Therefore, we consider Schelling games on graphs without stubborn agents to focus on the analysis of segregation caused by homophily.

DEFINITION 2.23 *(Agarwal et al., 2021)*
A *Schelling instance* is a tuple (N, G) consisting of the following two components.

- First, $N := [n]$ is a set of $n \geqslant 2$ *agents*. Each agent has one of two different *types*, red or blue.
- Second, $G = (V, E)$ is a simple undirected graph called *topology graph* (or simply *topology*), which satisfies $|V| \geqslant n$.

We assume for the remainder of this section that we are given a Schelling instance (N, G). Denote by r and b the number of red and blue agents, respectively. The distribution of agents into types is called *balanced* if $|r - b| \leqslant 1$. We say that two agents $i, j \in N$ with $i \neq j$ are *friends* if i and j are of the same type. For each $i \in N$, we denote the set of all friends of agent i by $F(i)$.

The output of a Schelling game is an assignment of the agents in N to nodes of G such that there are no collisions. Formally, an *assignment* is an n-tuple $\boldsymbol{v} = (v(1), \dots, v(n)) \in V^n$ such that $v(i) \neq v(j)$ for all $i, j \in N$ with $i \neq j$. A node $v \in V$ is *occupied* by agent i if $v(i) = v$. For a given assignment $\boldsymbol{v}$ and an agent $i \in N$, let $N_i(\boldsymbol{v}) = \{j \in N : \{v(i), v(j)\} \in E\}$ be the set of *neighbors* of agent i. Let $f_i(\boldsymbol{v}) = |N_i(\boldsymbol{v}) \cap F(i)|$ be the number of neighbors of i in $\boldsymbol{v}$ who are their friends.

5 For instance, stubborn agents are essential in several proofs establishing hardness results (see, e.g., Agarwal et al., 2021, Theorems 3.2 and 4.2). Publication 7 and Kreisel et al. (2022) show that stubborn agents are not necessary for these results.

DEFINITION 2.24 *(Agarwal et al., 2021)*
A *Schelling game* is a Schelling instance where agent i's utility for assignment $\mathbf{v}$ is specified as

$$u_i(\mathbf{v}) = \begin{cases} 0 & \text{if } N_i(\mathbf{v}) = \emptyset \\ \frac{f_i(\mathbf{v})}{|N_i(\mathbf{v})|} & \text{otherwise.} \end{cases}$$

Hence, whenever an agent has a neighbor, then their utility is defined as the fraction of their friends in their neighborhood.

2.5.2 Solution Concepts for Schelling Games

In contrast to most of the game-theoretic literature on Schelling segregation focusing on notions of stability, we study notions of optimality. First, *welfare optimality* and *Pareto optimality* are defined analogous to the corresponding concepts for hedonic games.

DEFINITION 2.25
The *(utilitarian) social welfare* of an assignment $\mathbf{v}$ is defined as

$$SW(\mathbf{v}) := \sum_{i \in N} u_i(\mathbf{v}).$$

In addition, an assignment $\mathbf{v}$ is said to be *welfare-optimal* if it maximizes social welfare.

An assignment $\mathbf{v}$ is said to be *Pareto-optimal* if there exists no assignment $\mathbf{v}'$ such that, for all agents $i \in N$, it holds that $u_i(\mathbf{v}') \geqslant u_i(\mathbf{v})$ and, for some agent $j \in N$, it holds that $u_j(\mathbf{v}') > u_j(\mathbf{v})$.

Denote by $SW_R(\mathbf{v})$ and $SW_B(\mathbf{v})$ the sum of the utilities of the red and blue agents, respectively; we have $SW_R(\mathbf{v}) + SW_B(\mathbf{v}) = SW(\mathbf{v})$.

Now, we introduce two novel concepts of optimality that are tailored for the study of Schelling games. Given two vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ of the same length k, we say that $\mathbf{w}_1$ *weakly dominates* $\mathbf{w}_2$ if for each $i \in \{1, \dots, k\}$, the ith element of $\mathbf{w}_1$ is at least as large as that of $\mathbf{w}_2$. We say that $\mathbf{w}_1$ *strictly dominates* $\mathbf{w}_2$ if at least one of these inequalities is strict.

For an assignment $\mathbf{v}$, denote by $\mathbf{u}(\mathbf{v})$ the vector of length n consisting of the agents' utilities $u_i(\mathbf{v})$, sorted in non-increasing order. Similarly, denote by $\mathbf{u}_R(\mathbf{v})$ and $\mathbf{u}_B(\mathbf{v})$ the corresponding sorted vectors of length r and b for the red and blue agents, respectively.

DEFINITION 2.26 *(Bullinger et al., 2021)*
An assignment $\mathbf{v}$ is said to be

- *group-welfare dominated* by an assignment $\mathbf{v}'$ if $SW_X(\mathbf{v}') \geqslant SW_X(\mathbf{v})$ for $X \in \{R, B\}$ and at least one of the inequalities is strict, and

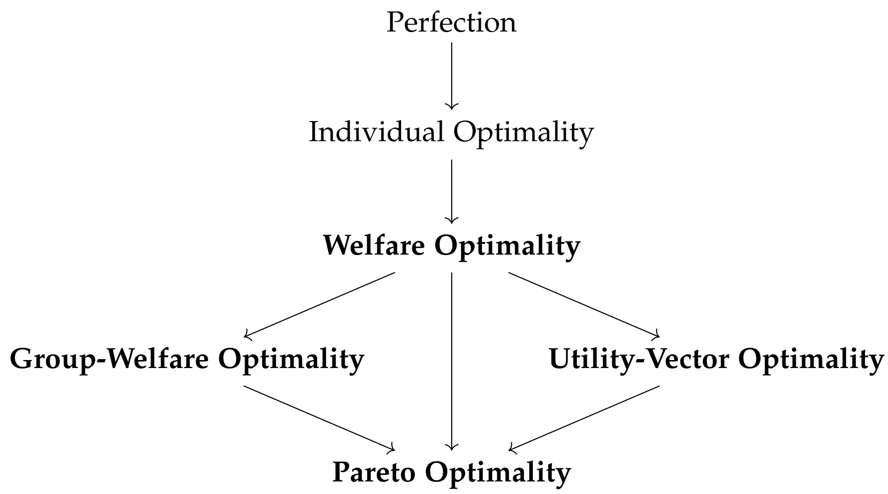


Figure 2.4: Implication relations among optimality notions considered for Schelling segregation in Publication 7. The bold concepts are guaranteed to exist.

- *utility-vector dominated* by an assignment $\mathbf{v}'$ if $\mathbf{u}(\mathbf{v}')$ strictly dominates $\mathbf{u}(\mathbf{v})$.

An assignment $\mathbf{v}$ is *group-welfare optimal* if it is not group-welfare dominated by any other assignment. Similarly, an assignment $\mathbf{v}$ is *utility-vector optimal* if it is not utility-vector dominated by any other assignment.

In Publication 7, we show that group-welfare optimality and utility-vector optimality are independent in the sense that neither notion implies the other one. Moreover, there is a close relationship to Pareto optimality because an assignment $\mathbf{v}$ is Pareto-optimal if and only if there is no other assignment $\mathbf{v}'$ such that $\mathbf{u}_X(\mathbf{v}')$ weakly dominates $\mathbf{u}_X(\mathbf{v})$ for $X \in \{R, B\}$ and at least one of the dominations is strict. Utility-vector optimality relaxes this domination by considering one joint vector. Hence, we obtain the relationships between the optimality concepts as depicted in Figure 2.4.

We quantify the welfare guarantee provided by each optimality notion (e.g., Pareto optimality) by a worst-case ratio of an optimal assignment compared to a welfare-optimal assignment. Such price concepts are a standard measure in game theory, and are based on the idea of the price of anarchy (Koutsoupias and Papadimitriou, 1999). In the domain of hedonic games, the price of Pareto optimality was introduced by Elkind et al. (2020) and subsequently studied in Publication 1. Balliu et al. (2022) study the price of Pareto optimality in social distance games. To define this concept, let $\mathbf{v}^*(N, G)$ denote a welfare-optimal assignment for the instance (N, G).

Definition 2.27 *(Bullinger et al., 2021)*

Given a property P of assignments and a Schelling instance (N, G), the *price of* P for that instance is defined as the ratio between the maximum social welfare (of any assignment) and the minimum social welfare of an assignment satisfying P:

$$\text{Price of } P \text{ for instance } (N, G) = \frac{\mathrm{SW}(\boldsymbol{v}^*(N, G))}{\min_{\boldsymbol{v} \in P(N,G)} \mathrm{SW}(\boldsymbol{v})},$$

where $P(N, G)$ is the set of all assignments satisfying P in the instance (N, G).[6] The *price of* P for a class of instances is then defined as the supremum price of P over all instances in that class.

Clearly, assignments which are welfare-optimal—and therefore also assignments satisfying weaker optimality notions—are guaranteed to exist. Therefore, their price notion is well-defined. In addition, we also consider two strengthenings of welfare optimality. These do not guarantee existence—and therefore we do not consider their price—but they turn out to be useful in the computational analysis of optimality notions. Note that the maximum utility that an agent can attain in a Schelling game is 1.

Definition 2.28 *(Bullinger et al., 2021)*
An assignment $\boldsymbol{v}$ is called *perfect* if $u_i(\boldsymbol{v}) = 1$ for all $i \in N$. Moreover, an assignment $\boldsymbol{v}$ is called *individually optimal for agent* i if $u_i(\boldsymbol{v}) \geqslant u_i(\boldsymbol{v}')$ for all assignments $\boldsymbol{v}'$. An assignment is called *individually optimal* if it is individually optimal for all agents.

These properties are at the top of the hierarchy in Figure 2.4.

6 We interpret the ratio $\frac{0}{0}$ in this context to be equal to 1.

3 SUMMARY OF PUBLICATIONS

In this chapter, the results of the publications, which this thesis is based on, are summarized.

3.1 COMPUTING DESIRABLE OUTCOMES

We start by considering the solution concepts introduced in Section 2.2 in hedonic games. The focus is on the computational problem of finding a partition that satisfies a given solution concept. In many cases, such partitions are not guaranteed to exist, and we then consider the existence and verification of the desired partitions. The methodology for obtaining hardness results is discussed in detail in Section 4.3.

3.1.1 Single-Agent Stability

First, we consider classical stability notions in cardinal classes of hedonic games. Single-agent stability concepts in these classes have already been studied extensively (see, e.g., Bogomolnaia and Jackson, 2002; Sung and Dimitrov, 2010; Aziz et al., 2013b; Brandl et al., 2015). Nevertheless, in Publications 3 and 4, we were able to bridge a few gaps in the literature.

In additively separable hedonic games, the intractability of Nash and individual stability are resolved in an early computational study (Sung and Dimitrov, 2010), while there exists an efficient algorithm to compute contractually individually stable partitions whose existence is guaranteed (Aziz et al., 2013b). We complete the picture by proving the intractability of contractual Nash stability. Notably, in contrast to other complexity results, such as Theorem 3.2 or Theorem 3.22 later on, the challenge to prove this theorem is not in finding a first No-instance. Indeed, a simple No-instance encompassing only 4 agents was already provided by Sung and Dimitrov (2007). The difficulty is rather in using No-instances and further auxiliary agents in the right way to make a reduction work.

Theorem 3.1 *(Bullinger, 2022)*

> *It is NP-complete to decide whether an additively separable hedonic game contains a contractually Nash stable partition.*

Under symmetric utilities, the non-existence of stable partitions in additively separable hedonic games with respect to single-agent stability concepts ceases. Even Nash-stable partitions are guaranteed to exist (Bogomolnaia and Jackson, 2002). By contrast, the behavior of fractional hedonic games is very different. There exist simple symmetric games where Nash-stable outcomes fail to exist, which leads to computational intractability (Brandl et al., 2015, Theorems 2 and 5). We strengthen this non-existence by providing a symmetric fractional hedonic game with 15 agents and involved utilities where no individually stable partition exists. We discuss the computer-aided techniques that were applied for finding this game in Section 4.2. The non-existence of individually stable partitions leads to a new hardness result.

Theorem 3.2 *(Brandt et al., 2021)*

> There exists a symmetric fractional hedonic game without an individually stable partition. Moreover, it is NP-complete to decide whether a symmetric fractional hedonic game contains an individually stable partition.

3.1.2 Notions of Optimality

Next, we consider Pareto optimality and welfare optimality. Recall that optimal partitions always exist, and we therefore seek to efficiently compute one. In addition, even partitions that are both Pareto-optimal and individually rational are guaranteed to exist. Indeed, the singleton partition is individually rational, and one can subsequently transition to Pareto-dominating partitions to reach a Pareto-optimal and individually rational partition from the singleton partition.

We start with the consideration of symmetric modified fractional hedonic games. Interestingly, computing Pareto-optimal partitions is related to solving a combinatorial problem that generalizes maximum cardinality matchings, namely maximum clique matchings. A *clique matching* is a partition of the vertices of a graph such that every set in the partition induces a clique. A clique is called *non-trivial* if it encompasses at least two vertices. Then, a *maximum clique matching* is a clique matching such that the number of vertices covered by non-trivial cliques is maximized. Computing a maximum clique matching can be done in polynomial time (Hell and Kirkpatrick, 1984). The algorithm that yields the next theorem computes several maximum clique matchings as a subroutine.

Theorem 3.3 *(Bullinger, 2020)*

> For symmetric modified fractional hedonic games, there exists a polynomial-time algorithm that computes a Pareto-optimal and individually rational partition. Moreover, the partition produced by this algorithm is a 2-approximation of the maximum social welfare.

Interestingly, the complexity of computing a welfare-optimal partition is open in modified fractional hedonic games, even if the game is assumed to be symmetric. The algorithm in the previous theorem is the best known approximation. However, it is shown in Publication 1 that it suffices to consider partitions consisting of coalitions of size at most 3 to find a welfare-optimal partition. Moreover, it is possible to efficiently compute welfare-optimal outcomes if utility values are assumed to be binary. This improves upon a result by Elkind et al. (2020, Theorem 5.8) who showed that, under the same conditions, Pareto-optimal partitions yield a 2-approximation of social welfare. The key insight is to exploit the combinatorial structure of optimal partitions, and especially that maximum clique matchings on the graph induced by utility values of 1 induce welfare-optimal partitions.

THEOREM 3.4 *(Bullinger, 2020)*
Let a symmetric modified fractional hedonic game (N, u) with binary utility values be given and consider a partition π. Then, a partition is welfare-optimal if and only if it is Pareto-optimal. In particular, for symmetric modified fractional hedonic games with binary utility values, a welfare-optimal partition can be computed in polynomial time.

If we consider additively separable or fractional hedonic games instead, a theorem analogous to Theorem 3.3 is unlikely to be true due to computational boundaries. Note that in the next theorem and in all subsequent results where partitions satisfying the considered solution concept are guaranteed to exist, the hardness holds for the respective search problem and is obtained via a Turing reduction.

THEOREM 3.5 *(Bullinger, 2020)*
For symmetric additively separable hedonic games and symmetric fractional hedonic games, computing a Pareto-optimal and individually rational partition is NP-hard.

In fact, for both of these games, it is intractable to compute Pareto-dominating partitions (cf. Publication 1). However, despite the intractability in the previous theorem, both of the involved solution concepts can be satisfied separately. Clearly, the singleton partition is individually rational. Moreover, Pareto-optimal partitions can be computed for large subclasses of additively separable hedonic games and fractional hedonic games. The corresponding algorithms in Publication 1 are similar to serial dictatorships: One by one, agents restrict the set of available partitions to their most preferred options.

THEOREM 3.6 *(Bullinger, 2020)*
For symmetric additively separable hedonic games and fractional hedonic games with binary utility values, a Pareto-optimal partition can be computed in polynomial time.

3.1.3 Popularity

The complexity of popularity has been extensively studied in the domain of matchings and in particular in marriage games (see, e.g., Gärdenfors, 1975; Abraham et al., 2007a; Biró et al., 2010). As with many matching problems, two approaches have turned out to be particularly fruitful, namely exploiting the combinatorial and the geometric structure of matchings. The former is related to augmenting paths and their structure in the non-bipartite case including so-called blossoms discovered by Edmonds (1965b). The second is concerned with the polyhedral structure of matchings, that is, the study of the matching polytope (Edmonds, 1965a).

Both approaches also have proved to be useful in the course of this thesis. While the first approach was applied in the study of Pareto optimality in Theorem 3.4, the geometric view via the matching polytope was very powerful in the study of popularity.

The general idea is to enhance Edmonds' matching polytope with constraints that capture the essence of popularity. Since there are exponentially many matchings that are potentially more popular than a given candidate matching, this leads to a super-polynomial number of constraints. Hence, tractability of the obtained linear program depends on whether we can efficiently solve the separation problem to apply an algorithm such as the Ellipsoid method (Khachiyan, 1979). Fortunately, this is possible by applying an algorithm by McCutchen (2008), which was originally developed to compute the so-called unpopularity margin of a given matching.

This approach was first applied by Kavitha et al. (2011) to compute mixed popular matchings in marriage games. While these authors enhance the bipartite matching polytope, we leverage the non-bipartite matching polytope instead to generalize their result. An important subtlety when dealing with mixed popular matchings in the matching polytope is that we operate with an aggregated *fractional* form of matchings. Hence, an important task is to decompose the aggregated matching as an explicit probability distribution over deterministic matchings. While a succinct such decomposition in principle exists due to Carathéodory's theorem, it can be efficiently constructed by means of the Ellipsoid method (Grötschel et al., 1981, Theorem 3.9) or even by a combinatorial algorithm (Padberg and Wolsey, 1984).

THEOREM 3.7 *(Brandt and Bullinger, 2022)*

Mixed popular matchings in roommate games with weak preferences can be found in polynomial time.

This approach is very universal and gives linear programming-based alternatives to combinatorial algorithms such as the preference refinement algorithm by Aziz et al. (2013a) for computing Pareto-optimal matchings[7] or the algorithm by Biró et al. (2010) to verify popular matchings. Moreover, we can provide an efficient algorithm for finding strongly popular matchings in roommate games—whenever they exist—by computing multiple mixed popular matchings. This resolves a repeatedly mentioned open problem (see, e.g., Biró et al., 2010; Manlove, 2013).

Theorem 3.8 *(Brandt and Bullinger, 2022)*

Computing a strongly popular matching or deciding that no such matching exists in roommate games (with weak preferences) can be done in polynomial time.

Notably, the positive results in the preceding two theorems stand in contrast with the intractability of popularity in roommate games (Faenza et al., 2019; Gupta et al., 2019). In other words, mixed popularity and strong popularity admit better computational tractability for weak preferences than ordinary popularity for strict preferences only.

The consideration of popularity in classes of hedonic games that allow coalitions of size at least 3 is scarce. Some work considers the verification problem for popularity and strong popularity and the existence problem for strong popularity (Kerkmann et al., 2020; Kerkmann and Rothe, 2020). Apart from an erroneous proof by Aziz et al. (2013b),[8] the existence problem for popularity had seemingly not been considered, yet. We provide the first results regarding this existence problem as well as the first computational boundaries for mixed popularity.

We start with the natural extension from roommate games to flatmate games. Once we transition to these games, the contrast in complexity between popularity and mixed or strong popularity in roommate games vanishes, and all related problems become intractable. In fact, even if we assume all preferences to be strict and globally ranked, we obtain several hardness results.

Theorem 3.9 *(Brandt and Bullinger, 2022)*

Consider the class of flatmate games with strict and globally ranked preferences. Then, it is

- coNP-hard to decide whether there exists a strongly popular partition,
- NP-hard to compute a mixed popular partition,
- coNP-hard to decide whether there exists a popular partition,

7 This works because every matching in the support of a mixed popular matching is Pareto-optimal (cf. Publication 2). Note that this already follows from a general observation about weak Condorcet winners (Fishburn, 1984, Proposition 3).

8 In Theorem 3.10, we correct this by proving a stronger statement.

- coNP-complete to verify a strongly popular partition,
- coNP-complete to verify a mixed popular partition, and
- coNP-complete to verify a popular partition.

Interestingly, all of these hardness results follow from one common, yet involved, approach. In essence, we provide a flatmate game that resembles the structure of a complete binary tree (of depth logarithmic in the input size of the instance). Each of the reductions uses this game (or multiple copies of it) as a ground structure, and we attach gadgets appropriate for the respective reductions. We discuss this proof technique in more detail in Section 4.3.2.

Moreover, there is a strong connection regarding the verification problem for mixed popularity and popularity. Based on the natural embedding, that is, the interpretation of popular partitions as mixed popular partitions where one specific partition is selected with probability 1, it follows that hardness of verification of popular partitions implies hardness of verification of mixed popular partitions in any domain (cf. Publication 2). Hence, while mixed popularity can be an escape route from computational boundaries with respect to the existence problem (which is the case in roommate games, as we have discussed after Theorem 3.8), it cannot be an escape route with respect to the verification problem.

If we consider cardinal classes of hedonic games such as additively separable hedonic games and fractional hedonic games, then we obtain similar results.

THEOREM 3.10 *(Brandt and Bullinger, 2022)*

Consider the class of symmetric additively separable hedonic games or the class of symmetric fractional hedonic games with non-negative utilities. Then, it is

- coNP-hard to decide whether there exists a strongly popular partition,
- NP-hard to compute a mixed popular partition,
- NP-hard to decide whether there exists a popular partition,
- coNP-hard to decide whether there exists a popular partition,
- coNP-complete to verify a strongly popular partition,
- coNP-complete to verify a mixed popular partition, and
- coNP-complete to verify a popular partition.

Finally, it is worth mentioning that the hardness results for the existence problems considered in this section are unlikely to be tight. Naturally, these existence problems belong to the complexity class Σ_2^p, and it is very well possible that they are even complete for this class. An indication towards this conjecture is that deciding about the existence of popular partitions in symmetric additively separable hedonic games and symmetric fractional hedonic games with non-negative utilities is complete for both NP and coNP. Hence, membership in any of these classes would cause the collapse of the polynomial hierarchy to the first level.

3.2 DEVELOPING NOVEL CONCEPTS

In the previous section, we have investigated classic desiderata in prominent classes of hedonic games. Now, we broaden the picture by challenging three paradigms of the classical literature on coalition formation. First, we consider coalition formation in a distributed manner, where the goal is to reach a desirable—in our case stable—partition by executing dynamics. Second, we oppose the unanimous consent of individual stability and contractual Nash stability by considering stability concepts based on majority consent. Interestingly, the positive results about majority-based stability concepts also concern dynamics. Third, we discuss how the agents' behavior changes under loyalty.

3.2.1 Distributed Coalition Formation

In this section, we consider algorithmic questions regarding the behavior of IS dynamics.

Convergence of IS Dynamics

We start by presenting convergence results for IS dynamics in a large variety of classes of hedonic games.

We start with the consideration of anonymous hedonic games. For these games, if preferences are single-peaked, Bogomolnaia and Jackson (2002) show that partitions exist that simultaneously satisfy good properties with respect to stability and optimality. They provide an efficient algorithm which is identical to a specific run of IS dynamics started from the singleton coalition. Thereby, they find an individually stable and weakly Pareto-optimal partition. We show that the convergence of the IS dynamics is no coincidence. In this class of games, IS dynamics converge regardless of the starting partition and regardless of the performed deviations.

Theorem 3.11 *(Brandt et al., 2021)*

The IS dynamics necessarily converges in anonymous hedonic games if the agents' preferences are strict and single-peaked.

In the full version of Publication 4, we are able to show that the result still holds if preferences may be weak (Brandt et al., 2022b).

In hedonic diversity games, dynamics were also known as a means of convergence. In fact, Boehmer and Elkind (2020) prove that individually stable partitions are guaranteed to exist in these games with the aid of dynamics: they define a partition with promising properties regarding individual stability. Then, they prove possible convergence of the IS dynamics from this partition. Due to this strong existence result, it would be interesting to see if we can obtain a similarly strong convergence result for dynamics. Unfortunately, this is not the case. In contrast to evidence collected from simulations by Boehmer and Elkind (2020), dynamics only necessarily converge under strong restrictions.[9]

Theorem 3.12 *(Brandt et al., 2022b)*

The IS dynamics may cycle in hedonic diversity games even if any three of the following restrictions apply:

1. preferences are single-peaked,
2. preferences are strict,
3. the starting partition is the singleton partition, and
4. all deviations satisfy solitary homogeneity.

The first two conditions of Theorem 3.12 restrict the considered class of hedonic games, while the last two conditions restrict the dynamics. From these, solitary homogeneity of deviations seems to be the weakest condition because it only defines a tie-breaking rule for selecting the next deviation in very specific cases. Interestingly, this small condition is an important ingredient for the following convergence result when all four of these conditions are satisfied.

Theorem 3.13 *(Brandt et al., 2022b)*

The IS dynamics necessarily converges in hedonic diversity games if agents have strict and single-peaked preferences and if the dynamics starts from the singleton partition and all deviations satisfy solitary homogeneity.

Finally, we consider fractional hedonic games. As a first result, we obtain convergence if we assume binary utility values and run the dynamics from the singleton partition.

9 The next two statements are taken from the full version of the article, correcting an error in Publication 4.

PROPOSITION 3.14 *(Brandt et al., 2021)*

The IS dynamics starting from the singleton partition necessarily converges in $\mathcal{O}(n^2)$ steps in symmetric fractional hedonic games with binary utility values. In these games, it may take $\Theta(n\sqrt{n})$ steps.

If we assume non-symmetric utilities, guaranteed convergence depends on acyclicity of the preference structure. To state this theorem, we define the notion of a utility graph. Given a cardinal hedonic game (N, u), we call the directed graph $G = (N, A)$ where $(i, j) \in A$ if $u_i(\{i, j\}) > 0$ its *utility graph*.

PROPOSITION 3.15 *(Brandt et al., 2021)*

Consider an asymmetric[10] fractional hedonic game. Then, the IS dynamics starting from the singleton partition necessarily converges if and only if the game's utility graph is acyclic. Moreover, under acyclicity, it converges in $\mathcal{O}(n^4)$ steps.

The previous results specify conditions under which convergence in fractional hedonic games is possible, but these conditions seem quite strong. However, recall that we have already seen in Theorem 3.2 that individually stable partitions are not guaranteed to exist, even if we impose symmetry of utilities. Hence, there also exist symmetric fractional hedonic games where the dynamics is doomed to cycle. We even obtain computational boundaries for deciding on the necessary and possible convergence of dynamics.

THEOREM 3.16 *(Brandt et al., 2021)*

It is NP-hard to decide whether the IS dynamics necessarily (or possibly) converges in symmetric fractional hedonic games.

Naturally, since individually stable partitions are not guaranteed to exist in general anonymous hedonic games and hedonic diversity games, similar problems are reasonable to pose in these classes. Indeed, we also obtained computational boundaries for anonymous hedonic games and hedonic diversity games (cf. Publication 4).

Speed of Convergence

So far, we were only concerned about whether dynamics converge or not. A related question asks for the speed of convergence. A side product of Propositions 3.14 and 3.15 is the polynomial running time of dynamics in certain fractional hedonic games. These results were obtained by investigating so-called potential functions, i.e., functions that assign a score or a multi-dimensional vector to every partition. The potential functions in these proofs are rather simple, by defining scores or vectors associated with partitions or the coalitions constituting a partition, respectively. Convergence then follows because the

10 Recall that asymmetry of utilities is *not* the contrary of symmetry.

associated scores and vectors change monotonically after every individual deviation. Proof strategies based on potential functions are discussed in Section 4.1.

We also obtain polynomial bounds on the running time of IS dynamics in anonymous hedonic games and hedonic diversity games. Obtaining these bounds requires a far more involved use of potential functions. These results are only covered in the full version of Publication 4 (Brandt et al., 2022b).

The potential function for anonymous hedonic games is composed of values associated with both agents and coalitions in a temporary partition. Notably, these values may depend on the whole IS dynamics until reaching this partition, i.e., on an entire sequence of partitions. The key insight is that the potential function is non-decreasing and strictly increasing whenever an agent deviates towards a larger coalition. The values composing the potential function mimic the drift of agents towards larger coalitions. We therefore strengthen Theorem 3.11 by even obtaining a polynomial running time.

Theorem 3.17 *(Brandt et al., 2022b)*

> The IS dynamics necessarily converges in $\mathcal{O}(n^3)$ steps in anonymous hedonic games if the agents' preferences are strict and single-peaked.

In hedonic diversity games, the running time of the IS dynamics can be polynomially bounded if we assume the same restrictions as in Theorem 3.13. The proof of this statement is highly interesting. In contrast to most results in the hedonic games literature, it establishes a reduction to a seemingly different class of hedonic games. After tedious and non-trivial manipulations, we can show that only coalitions with very specific proportions of the agent types play a role in an execution of the dynamics. These can be related to coalition sizes, and therefore to anonymous hedonic games. This allows us to apply Theorem 3.17.

Theorem 3.18 *(Brandt et al., 2022b)*

> Under the restrictions of Theorem 3.13, the IS dynamics necessarily converges in $\mathcal{O}(n^5)$ steps in hedonic diversity games.

So far, we have seen several examples of polynomial running time in classes of hedonic games where dynamics are guaranteed to converge. This impression must be handled with care. Even if we know that dynamics are guaranteed to converge, this does not necessarily imply fast convergence. For instance, dynamics based on Nash stability are guaranteed to converge in symmetric additively separable hedonic games according to the potential function argument by Bogomolnaia and Jackson (2002). However, computing a Nash stable or an individually stable partition is PLS-complete (Gairing and Savani, 2019). By exploiting the properties of the PLS-reduction, one can show that this immediately implies the existence of exponentially long dynamics based on Nash stability (cf. Publication 5). Moreover, by modifying an example provided by Monien and Tscheuschner (2010), we can even give an explicit example of an IS dynamics of exponential length.

Proposition 3.19 *(Brandt et al., 2022a)*

> The IS dynamics in symmetric additively separable hedonic games may be of exponential length.

A related question to asking for fast convergence is asking for fastest convergence, i.e., for the smallest number of steps until dynamics possibly converge. In very recent work, we provide NP-hardness results for several dynamics based on both single-agent and group deviations in additively separable hedonic games (Boehmer et al., 2023). Because of hardness results like Theorem 3.16, similar results seem to be likely in other classes of hedonic games.

3.2.2 Stability Based on Majority Consent

We now discuss results on majority-based stability concepts in subclasses of additively separable hedonic games. All existence results follow from the convergence of dynamics. If stability is based on majority consent by both the abandoned and welcoming coalition, we obtain an existence result.

Theorem 3.20 *(Brandt et al., 2022a)*

> The JMS dynamics (and therefore SMS dynamics) necessarily converges in appreciation-of-friends games and in friends-and-enemies games.

Moreover, we obtain another positive result regarding the existence of majority-in stable partitions in appreciation-of-friends games.

Theorem 3.21 *(Brandt et al., 2022a)*

> The MIS dynamics starting from the grand coalition necessarily converges in appreciation-of-friends games.

While majority-based stability notions lead to positive results in natural classes of restricted additively separable hedonic games, we also obtain strong computational boundaries. The key challenge for the proof of Theorem 3.22 is to construct sophisticated No-instances for the respective problems. One indication for their complexity is the number of agents required for them. Indeed, the smallest No-instances found for majority-out stability in appreciation-of-friends games and for majority-out and majority-in stability in friends-and-enemies games consist of 16, 23, and 183 agents, respectively.

Theorem 3.22 *(Bullinger, 2022)*

> It is NP-complete to decide whether

- an appreciation-of-friends game contains a majority-out stable partition,
- a friends-and-enemies game contains a majority-out stable partition, and
- a friends-and-enemies game contains a majority-in stable partition.

Interestingly, Theorem 3.21 compared with the first result of Theorem 3.22 shows that—despite their conceptual duality—the behavior of majority-out and majority-in stability can be very different. Moreover, the results in Theorem 3.22 lie at the boundary of computational feasibility. When we move to stability concepts based on unanimous consent, we have existence and efficient computability based on converging dynamics in appreciation-of-friends games and friends-and-enemies games (cf. Publication 5). In addition, if we further restrict the considered classes of games, we also obtain positive results for majority-based stability notions. In fact, appreciation-of-friends games in which every agent has at most one friend admit a majority-out stable partition, and if the friend or enemy relations are complete, then majority-out stable or majority-in stable partitions are guaranteed to exist in friends-and-enemies games (cf. Publication 3). We will use the first statement from Theorem 3.22 to illustrate the general proof technique for obtaining hardness reductions for stability notions in Section 4.3.1.

3.2.3 Cardinal Hedonic Games under Loyalty

In this section, we consider results from Publication 6 concerning loyal variants of cardinal hedonic games. Our first result is that the process of transitioning to the loyal variant converges after a finite number of steps.

Proposition 3.23 *(Bullinger and Kober, 2021)*
Let a cardinal hedonic game (N, u) be given. Then, there exists a cardinal hedonic game (N, u^E) such that the vector u^k of utility functions of its k-fold loyal variant satisfies $u^k = u^E$ for all $k \geqslant n$ where $n = |N|$.

The utility functions of the limit game have a compact representation based on computing minimum utilities locally. Therefore, the game (N, u^E) is called the *locally egalitarian variant*.

We are ready to discuss computational results regarding the loyal variants of hedonic games. Interestingly, considering the games modified under loyalty adds complexity to eliciting preferences. It is even hard to determine best coalitions. Formally, the problem of, given a cardinal hedonic game (N, u), an agent $i \in N$, and a rational number

$q \in \mathbb{Q}$, deciding if there exists a coalition $C \in \mathcal{N}_i$ with $u_i(C) \geqslant q$, is called `BestCoalition`.

THEOREM 3.24 *(Bullinger and Kober, 2021)*
Let $k \geqslant 1$. Then,

1. `BestCoalition` can be solved in polynomial time for the k-fold loyal variant (or locally egalitarian variant) of symmetric modified fractional hedonic games.
2. `BestCoalition` is NP-complete for the k-fold loyal variant (or locally egalitarian variant) of symmetric appreciation-of-friends games, and

A key result of Publication 6 is the guaranteed existence of partitions satisfying both notions of stability and optimality if the degree of loyalty is sufficiently high. We present this theorem for the classes of cardinal hedonic games known so far. A more general version of the theorem is part of Publication 6.

THEOREM 3.25 *(Bullinger and Kober, 2021)*
There exists a Pareto-optimal coalition structure in the core of the locally egalitarian variant of symmetric additively separable, fractional, and modified fractional hedonic games.

Interestingly, the guaranteed existence of nice partitions yields no guarantee for their efficient computability.[11]

THEOREM 3.26 *(Bullinger and Kober, 2021)*
The following statements hold.

1. Let $k \geqslant 1$. A partition in the core can be computed in polynomial time for the k-fold loyal variant (or locally egalitarian variant) of symmetric modified fractional hedonic games.
2. Let $k \geqslant 2$. Computing a partition in the core is NP-hard for the k-fold loyal variant (or locally egalitarian variant) of symmetric appreciation-of-friends games with non-empty core.

Moreover, in contrast to the existence of partitions in the core of the locally egalitarian variant, non-existence becomes an issue in the loyal variants.

THEOREM 3.27 *(Bullinger and Kober, 2021)*
Let $k \geqslant 1$. Deciding whether the core is non-empty is NP-hard for the k-fold loyal variant of symmetric additively separable hedonic games.

11 This result is a rare exception in the hedonic games literature where the existence and efficient computability of desirable partitions usually go hand in hand.

An interesting open problem from Publication 6 is whether the core is always non-empty in the loyal variant of an appreciation-of-friends game. These games are not covered by the hardness results in Theorems 3.26 and 3.27. In particular, the second part of Theorem 3.26 leaves out the case $k = 1$.

3.3 OPTIMALITY IN SCHELLING SEGREGATION

In this section, we present results on Schelling segregation in the framework of Schelling games on graphs. We consider welfare optimality, group-welfare optimality, utility-vector optimality, and Pareto optimality. We start with the consideration of social welfare. Recall that the utility of every agent is bounded by 1, and therefore the maximum welfare in any instance of a Schelling game with n agents is bounded by n. Consider the function $g\colon \mathbb{N} \to \mathbb{Q}$ defined by

$$g(n) = \begin{cases} \frac{n(n-2)}{2(n-1)} & \text{if } n \text{ is even;} \\ \frac{n-1}{2} & \text{if } n \text{ is odd.} \end{cases}$$

The following theorem determines an optimal welfare guarantee that can be attained in polynomial time.

THEOREM 3.28 *(Bullinger et al., 2021)*
For any Schelling game with n agents, there exists an assignment with social welfare at least $g(n)$. Moreover, the bound $g(n)$ cannot be improved. Even more, we can compute an assignment with social welfare $g(n)$ in polynomial time.

The tightness of the bound $g(n)$ means that, for every $n \in \mathbb{N}$, there exists a Schelling game with n agents such that every assignment in this game has social welfare of at most $g(n)$. The proof idea for this theorem is easy to describe. A random assignment of the agents to nodes already satisfies the bound $g(n)$ in expectation. To obtain a deterministic, polynomial-time algorithm, one can derandomize the procedure of randomly assigning agents to nodes by the *method of conditional probabilities*. To this end, we subsequently select optimal remaining positions for each agent with respect to the expected welfare obtained when assigning the remaining agents at random—a probabilistic quantity that can be computed deterministically and efficiently. This algorithmic idea is well known from the approximation of the MAXCUT problem (see, e.g., Mitzenmacher and Upfal, 2005, Chapter 6; Williamson and Shmoys, 2011, Chapter 5).

Also, the tightness of the bound is rather simple and follows from considering a clique graph as a topology where the type distribution of the agents is balanced. Clearly, in such an instance, every assignment has equal social welfare.

Still, the context of this theorem is very interesting. First, the Freeman segregation index, a standard measure for the strength of segregation in social sciences, defines the absence of segregation based on the social welfare of an expected assignment (Freeman, 1978). In this respect, Theorem 3.28 can be interpreted as saying that non-segregated states are a good approximation of the maximum social welfare. In fact, since $g(n) \geqslant \frac{n}{2} - 1$ for every $n \in \mathbb{N}$, the algorithm in Theorem 3.28 almost yields a 2-approximation of the maximum social welfare. While it is unknown if we can achieve a better approximation factor, we know that we meet computational boundaries when we seek to maximize social welfare. The next result generalizes and strengthens earlier hardness results by Agarwal et al. (2021).

Theorem 3.29 *(Bullinger et al., 2021)*

The following problem is NP-complete: Given a Schelling game and a rational number q, decide whether there exists an assignment with social welfare at least q.

The hardness holds even for the class of instances where the number of agents is equal to the number of nodes and the topology is a regular graph.

Next, we consider other notions of optimality. By computing the upper and lower bounds for their welfare guarantees, we gain information about their worst-case behavior. The results are summarized in the next theorem.

Theorem 3.30 *(Bullinger et al., 2021)*

The following statements hold for the class of Schelling games.

- The price of group-welfare optimality is $\Theta(n)$.
- The price of utility-vector optimality is $\Theta(n)$.
- The price of Pareto optimality satisfies $\Omega(n)$ and $\mathcal{O}(n\sqrt{n})$.

In other words, for all other welfare notions, optimal assignments can suffer a significant welfare loss in the worst case. Moreover, group-welfare optimality and utility-vector optimality guarantee constant social welfare, i.e., a social welfare of $\Theta(1)$, in every instance. Whether Pareto optimality also yields constant social welfare is unknown, and the upper bound on the price of Pareto optimality only comes from a guaranteed welfare of $\frac{1}{\sqrt{n}}$ of Pareto-optimal assignments. Still, there is evidence that Pareto-optimal assignments could also assure constant welfare because this statement is true in restricted instances of Schelling games. Note that these restrictions either leave the topology graph or the agent distribution unconstrained.

Theorem 3.31 *(Bullinger et al., 2021)*

Consider a Schelling game.

- If the topology graph is a tree, then every Pareto-optimal assignment has a social welfare of at least $\frac{n}{n-1}$.
- If the distribution of the agents is balanced, then every Pareto-optimal assignment has a social welfare of at least 1.

Note that the bound for tree graphs is tight in the sense that, for every $n \in \mathbb{N}$, there exists a Pareto-optimal assignment on a star graph of social welfare $\frac{n}{n-1}$. However, the corresponding agent distribution is highly unbalanced, having only 2 agents of one of the two types.

In the remaining section, we consider the efficient computability of optimal assignments. The next theorem states that perfection, the strongest optimality concept in Figure 2.4, is intractable.

Theorem 3.32 *(Bullinger et al., 2021)*
It is NP-complete to decide whether there exists a perfect assignment in a Schelling game.

This directly implies intractability of all notions for which existence is guaranteed: there exists a perfect assignment if and only if every optimal assignment is perfect. Hence, if we were able to compute an optimal assignment, then we could decide if a perfect assignment exists by checking whether every player receives a utility of 1 in an optimal assignment.

It is also possible to infer intractability for other weakenings of perfection such as maximizing Nash welfare or egalitarian welfare. Furthermore, an inspection of the reduction of this proof reveals that deciding about the existence of individually optimal assignments is also hard (cf. Publication 7).

Corollary 3.33 *(Bullinger et al., 2021)*
Computing a utility-vector optimal (or Pareto-optimal) assignment in a Schelling game is NP-hard. Also, deciding whether there exists an individually optimal assignment is NP-complete.

In a similar way, it is also possible to apply Theorem 3.32 to prove hardness of welfare optimality and group-welfare optimality. However, the proof of Theorem 3.32 crucially depends on the existence of empty nodes (otherwise, perfect partitions are not possible). For welfare optimality, Theorem 3.29 already proves a stronger statement, and we can infer an intractability result for group-welfare optimality under the same restrictions.

Theorem 3.34 *(Bullinger et al., 2021)*
Computing a group-welfare optimal assignment in a Schelling game is NP-hard, even for the class of Schelling games where the number of agents is equal to the number of nodes and the topology is a regular graph.

For an equal number of agents and nodes, Theorem 3.32 breaks down because then every instance that contains two types of agents and a connected topology is a No-instance. In addition, in this case, we obtain a complete characterization of the instances containing an individually optimal assignment, which can be checked in polynomial time (cf. Publication 7).

Perfection and individual optimality are strong concepts of optimality which are hard to satisfy at all. It is therefore natural to ask for weaker notions that demand less. To this end, we consider *positive agents* under some assignment, that is, agents for which there exists at least one agent of the same type in their neighborhood (cf. Publication 7).

There are simple topologies—for example, star graphs—for which there exists no assignment such that every agent is positive. On the other hand, it is rather straightforward to provide an assignment which makes a majority of the agents positive by assigning the majority type to a connected subgraph of the topology. In addition, if the topology is a tree, we can decide if there exists an assignment such that all agents are positive by means of dynamic programming. Finally, if the topology graph has a minimum degree of 2, it is always possible to assign the agents in a way such that every agent is positive. Hence, if the topology satisfies a mild property that is very natural in dense metropolitan areas, then it might not be possible to compute optimal assignments but we can at least ensure every agent to live next to a friend (cf. Publication 7).

4 METHODOLOGY

In Chapter 3, we gave an overview of the results in this thesis while the details of the proofs were kept rather brief. Here, we focus on the methodology and explore important methods that were applied and proposed to obtain many results in this thesis. Instead of stating complete proofs, we present different themes of proofs that can be used as role models in similar settings. Often it is easier to understand *that* a solution technique works rather than *why* it works. In this chapter, we aim at providing a better understanding of the latter.

4.1 ANALYSIS OF DYNAMICS VIA POTENTIAL FUNCTIONS

In this section, we discuss techniques for proving the necessary convergence of dynamics. While several proof methods can be applied in general—for instance, Theorem 3.11 is shown by contraction—a typical proving scheme for the convergence dynamics is the use of so-called *potential functions* (cf. Propositions 3.14 and 3.15 and Theorems 3.17 and 3.18).

The idea of potential functions is similar to local search as captured in the complexity class PLS (Johnson et al., 1988). Within the space of possible solutions, we can transition to neighboring solutions. In our case, a solution is a partition, and a transition can happen based on a deviation. Every solution is associated with a number, its *potential*, which can be interpreted as its value. While the goal of local search is to find a solution achieving the maximum value in its neighborhood, our goal is to find a valuation scheme such that transitions in the neighborhood, that is, deviations, increase the value. We seek to interpret dynamics as specific local search algorithms.

To set the stage, we develop the formal framework for these ideas. Consider a fixed hedonic game $(N, \succeq)$ and let Π be the set of all partitions of $n = |N|$ agents. We can define the *deviation graph* as the directed graph $DG = (\Pi, A)$ with vertex set Π where there exists a directed edge $(\pi, \pi') \in A$ if and only if there exists an agent $d \in N$ such that $\pi \xrightarrow{d} \pi'$ is an individual deviation. Of course, one can define a deviation graph for any type of deviation, but since the focus of our analysis of dynamics lies on individual deviations, we state the definition for this case. Now, the execution of an IS dynamics is simply a walk in the deviation graph. Moreover, the necessary convergence of

IS dynamics in a certain game is equivalent to acyclicity of the game's deviation graph.

A *potential function* is a function $\Phi\colon \Pi \to \mathbb{R}$. Potential functions are a tool for proving necessary convergence of dynamics. The goal is to prove that there exists a potential function Φ that is increasing for any transition in the deviation graph. Formally, a potential function is said to be *increasing* if, for every edge $(\pi, \pi') \in A$ of the deviation graph, it holds that $\Phi(\pi') > \Phi(\pi)$. To summarize, we obtain the following proposition.

PROPOSITION 4.1

Let a hedonic game $(N, \succsim)$ be given. The following are equivalent:

1. The IS dynamics on $(N, \succsim)$ necessarily converges.
2. The deviation graph of $(N, \succsim)$ is acyclic.
3. The game $(N, \succsim)$ possesses an increasing potential function.

The proposition follows from interpreting the potential function as a topological ordering of the deviation graph (Kahn, 1962).

An influential early result concerning potential functions in cardinal hedonic games was already shown by Bogomolnaia and Jackson (2002).

PROPOSITION 4.2 *(Bogomolnaia and Jackson, 2002)*

Consider a symmetric additively separable hedonic game. If $\pi \xrightarrow{d} \pi'$ is a Nash deviation, then $SW(\pi') > SW(\pi)$.

Essentially, this proposition means that the social welfare is an increasing potential function with respect to Nash deviations. This idea is quite fundamental and is still applied in similar settings (Bilò et al., 2022; Bullinger and Suksompong, 2023).

As discussed before, many of the proofs in this thesis make use of potential functions. Instead of discussing the details of the mostly complex proofs, we outline general themes of the potential functions in increasing complexity.

Simple Potential Functions

In Proposition 3.14, we show necessary convergence of the IS dynamics starting from the singleton partition in symmetric fractional hedonic games with binary utility values. The key structural insight of the proof is that the partition in every step of the dynamics is a collection of cliques (in the undirected graph with edges induced by the binary and symmetric utilities). Unfortunately, using the social welfare of partitions as a potential function is not possible. In fact, if a partition π in a fractional hedonic game consists of cliques only, then

$\mathrm{SW}(\pi) = \sum_{C\in\pi} |C| - 1$. Hence, the only possibility for increasing the social welfare is if the abandoned coalition was a singleton coalition.

However, a similar idea works. Indeed, the function $\Phi\colon \Pi \to \mathbb{R}$, $\pi \mapsto \sum_{i\in N} |\pi(i)|$ is an increasing potential function. For an individual deviation $\pi \xrightarrow{d} \pi'$, it holds that

$$\begin{aligned}\Phi(\pi') - \Phi(\pi) &= \underbrace{|\pi'(d)| - |\pi(d)|}_{\text{change by } d} + \underbrace{(|\pi'(d)| - 1)}_{\text{new coalition}} - \underbrace{(|\pi(d)| - 1)}_{\text{old coalition}} \\ &= 2(|\pi'(d)| - |\pi(d)|) > 0.\end{aligned}$$

The computation uses the fact that d abandons a clique coalition and joins a larger clique coalition. Interestingly, this potential is equivalent to considering the social welfare of π if the single-agent utilities were aggregated as if the game was an additively separable hedonic game. Also, note that the method cannot generalize to arbitrary symmetric weights due to the non-existence of individually stable partitions in such fractional hedonic games (Theorem 3.2).

Lexicographic Comparison of Ordered Vectors

Another common technique for finding potential functions is to consider ordered vectors with respect to lexicographic comparison. These vectors can either be seen as multi-valued potential functions, or can be aggregated to single-valued potential functions at the cost of having a large range of values for the potential function. Note that the latter does not cause computational problems because the potential functions are an implicit underlying structure of a game.

For instance, Aziz et al. (2019, Theorem 5.6) use this technique to prove the existence of core partitions in a restricted class of fractional hedonic games. There, they associate to a partition the vector of utilities of the agents in non-decreasing order. Then, they show that a partition which lexicographically maximizes this vector is in the core.

The proof of Proposition 3.15 features the combination of two potential functions, each of which is based on an ordered vector. The first vector is the vector of coalition sizes in non-increasing order. We call this vector $\boldsymbol{v}^S$. The second vector, called $\boldsymbol{v}^T$, is based on a topological order of the agents, which exists for acyclic utility graphs (Kahn, 1962).

The convergence of the dynamics follows from an interesting interplay of these two vectors. While $\boldsymbol{v}^S$ need not be monotonic with respect to the lexicographic ordering, $\boldsymbol{v}^T$ is decreasing weakly in every step of the dynamics. The problem is that $\boldsymbol{v}^T$ may have equal lexicographic score after a deviation. However, $\boldsymbol{v}^S$ is strictly increasing (lexicographically) whenever this happens. Hence, we can use the combination of both vectors to bound the running time of the dynamics.

Global Bounds of Non-Increasing Potential Functions

While the necessary convergence of dynamics is equivalent to the existence of an increasing potential function (cf. Proposition 4.1), it can be favorable to analyze *non-increasing* potential functions. These can be simpler to understand, and we might be able to control the number steps of the dynamics in which a non-increasing potential function decreases.

For instance, in Publication 5, we develop a method for using the social welfare as a potential function in games where the social welfare can decrease after a deviation. The key insight is that there is a global trend towards an increase of the social welfare. In other words, rather than proving local improvements of the potential function, we show an overall increase of the potential function dependent on the length of the dynamics. This is based on a purely combinatoric interpretation of dynamics based on single-agent deviations.

Beyond Static Potential Functions

Potential functions need not be defined to be static in that they do not have to be functions of the form $\Phi\colon \Pi \to \mathbb{R}$. One can even define a potential function based on the *history* of the specific execution of the dynamics, that is, the potential function can have the set of finite sequences of partitions as its domain. This is particularly useful if we do not have control over the starting partition of the dynamics.

An example is the potential function developed in the proof of Theorem 3.17 dealing with anonymous hedonic games under single-peaked preferences. It consists of a history-dependent sum of values associated with each coalition and each agent in a partition. The potential function reflects the structure of the single-peaked preferences that we can elicit from the history of the dynamics. Since we only learn about this structure as soon as an agent is affected by an individual deviation for the first time, the potential function depends on the exact sequence of deviations until the currently considered partition.

While the use of history-based potential functions can be very powerful, they have the drawback that they are a lot harder to comprehend. For instance, the above discussed potential function from Theorem 3.17 emulates the drift of agents towards the peak of their single-peaked preferences. However, the update rules are complex and fine-tuned to cover all relevant incidences during an execution of the dynamics.

4.2 FINDING COMPLEX EXAMPLES BY LINEAR PROGRAMMING

Linear programming is a standard tool in mathematical optimization (see, e.g., Grötschel et al., 1993). In this thesis, linear programs have been applied in two fundamental ways.

First, linear programs give rise to efficient algorithms, and can therefore be used as a black box for obtaining theoretical guarantees. This has been useful for determining the complexity of computing mixed popular matchings in roommate games (cf. Theorem 3.7).

In this section, we discuss the second application of linear programming, namely, how to make use of an explicit linear program to construct a specific hedonic game. This method was used to prove the first part of Theorem 3.2, that is, the existence of a fractional hedonic game without an individually stable partition. For convenience, we repeat the statement here.

THEOREM 4.3 *(Brandt et al., 2021)*
There exists a symmetric fractional hedonic game without an individually stable partition.

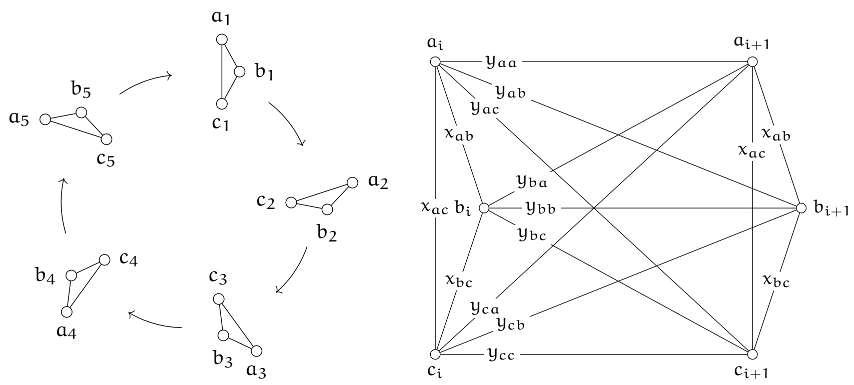
Setting Up a Linear Program

Intuitively, it is clear that stability based on unilateral deviations depends on a set of inequalities because a partition is stable if and only if no agent can perform a beneficial deviation. Each of these deviations can be modeled as an inequality. Additionally, binary variables of an integer program can represent membership in joint coalitions. This idea leads to a way to tackle a problem like "Given a partition, decide if there exists a beneficial deviation." However, this approach is not helpful if we do not even know the game, for which we want to analyze the existence of stable partitions. In other words, we aim at setting up a method for solving the much more complicated quantified statement

> "There exists a game, such that for every partition, there exists some deviation."

The natural way to use linear programs is to resolve existential questions. Hence, we focus on finding a specific game, and therefore use the variables of a linear program to encode single-agent utilities. However, it is not clear how to encode the universal quantifier, which determines that the game should not contain a stable partition.

The approach taken towards the proof of Theorem 4.3 is to combine an educated human guess about the structure of the desired game with exact specifications of such a game computed by a linear program. Instead of directly finding our game, we apply linear programming to prove the following weaker statement.



(a) Agents are intended to deviate along a cycle such that agents in N_i join the triple coalition N_{i+1}.

(b) The utilities between and within the triples are the variables of a linear program.

Figure 4.1: Towards a symmetric fractional hedonic game without an individually stable partition (Brandt et al., 2021).

Proposition 4.4
: There exists a symmetric fractional hedonic game where IS dynamics can cycle.

We use linear inequalities for the deviations of some specific candidate dynamics in a guessed candidate game. If the linear program has a feasible solution, then we have found a game with a cyclic IS dynamics. Otherwise, we do not have a definite answer about the existence of a counterexample. We just know that there does not exist a game with a certain structure such that a specific execution of IS dynamics can cycle.

Consider a fractional hedonic game (N, u) with agent set $N = \bigcup_{i \in [5]} N_i$, where $N_i = \{a_i, b_i, c_i : i \in [5]\}$. The agent set is partitioned into five triples of agents. We would like to determine the single-agent values u^S by means of a linear program. Since we are looking for symmetric utilities, we write $u(i,j)$ for $u_i^S(j) = u_j^S(i)$.

The goal is to find utilities such that a specific sequence of partitions evolves through individual deviations. Therefore, define the partitions

- $\pi_0 = \{N_1 \cup N_2, N_3, N_4, N_5\}$,
- $\pi_1 = \{N_1 \cup \{b_2, c_2\}, N_3 \cup \{a_2\}, N_4, N_5\}$,
- $\pi_2 = \{N_1 \cup \{c_2\}, N_3 \cup \{a_2, b_2\}, N_4, N_5\}$, and
- $\pi_3 = \{N_1, N_2 \cup N_3, N_4, N_5\}$.

Note that $\pi_0 \xrightarrow{a_2} \pi_1 \xrightarrow{b_2} \pi_2 \xrightarrow{c_2} \pi_3$ is a sequence of unilateral deviations. The deviations only rely on the single-agent utilities between the agents in $N_1 \cup N_2$ and the agents in $N_2 \cup N_3$. The partitions π_0 and π_3 only differ by an index shift of the agent. Hence, this dynamics can be extended to an infinite cycle if the game always looks the same locally. The structure of the game and the dynamics are depicted in Figure 4.1a.

To enable infinite dynamics, we assume the following constraints on the utilities:

- For $v, w \in \{a, b, c\}$ with $v \neq w$ and $i, j \in [5]$, we assume that $u(v_i, w_i) = u(v_j, w_j)$.
- For $v, w \in \{a, b, c\}$ and $i, j \in [5]$, we assume that $u(v_i, w_{i+1}) = u(v_j, w_{j+1})$.

There, we read indices modulo 5 mapping to the representative in [5]. The first constraint assures that the utilities within each agent set N_i are identical, and the second constraint assures that the utilities between each pair of subsequent sets N_i and N_{i+1} are identical. As a consequence, if we can find utilities satisfying these constraints and such that $(\pi_k)_{k=0}^3$ is an execution of the IS dynamics, then the IS dynamics can cycle in the game (N, u).

We are ready to define the variables of our linear program.

- Let the variables x_{ab}, x_{ac}, and x_{bc} represent the symmetric utilities $u(a_1, b_1)$, $u(a_1, c_1)$, and $u(b_1, c_1)$, respectively.
- For $v, w \in \{a, b, c\}$, let the variable y_{vw} represent $u(v_1, w_2)$.

Hence, variables of the type x_{vw} capture utilities among a triple of agents, and variables of the type y_{vw} capture utilities between two subsequent triples. The variables of the linear program are shown in Figure 4.1b. Note that the variables do not encompass all pairwise utilities of the fractional hedonic game (N, u) but just all utilities relevant to the coalitions in the dynamics.

Next, we can specify the constraints corresponding to the dynamics $(\pi_k)_{k=0}^3$. To this end, let $i \in [3]$ and suppose that d_i is the deviating agent leading to partition π_i. Then,

- we want that $u_{d_i}(\pi_i(d_i)) > u_{d_i}(\pi_{i-1}(d_i))$, and
- for $v \in \pi_i(d_i) \setminus \{d_i\}$, we want that $u_v(\pi_i(v)) \geqslant u_v(\pi_{i-1}(v))$.

The first, strict inequality assures that each deviation is a Nash deviation. The second set of inequalities guarantees the consent of the agents in the welcoming coalition, and therefore that each deviation even is an individual deviation. Note that the strict inequality can be transformed into a weak inequality by making use of a slack variable.

Each of these constraints can be written as a linear constraint using only variables x_{vw} and y_{vw}. This might seem counterintuitive at first glance, because the utilities of fractional hedonic games are fractions that include coalition sizes. However, this is not a problem. We already know the coalition sizes of all coalitions involved in deviations, so these are only constants in the inequalities. For instance,

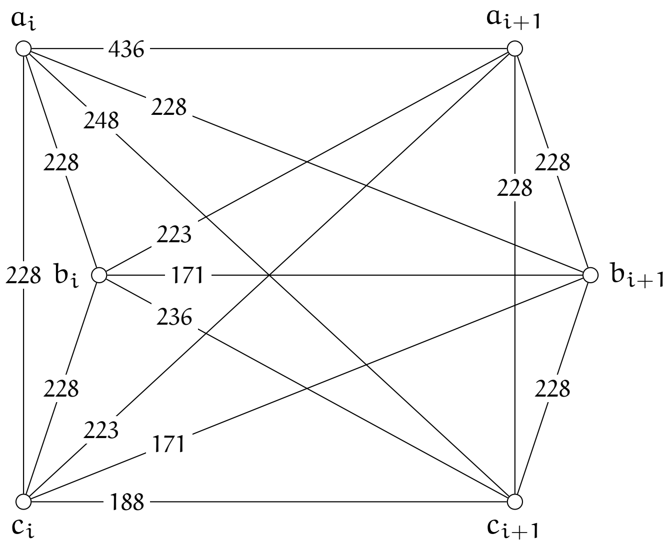


Figure 4.2: Utilities found by linear programming (Brandt et al., 2021).

the inequality $u_{a_2}(N_3 \cup \{a_2\}) > u_{a_2}(N_1 \cup N_2)$ is equivalent to the inequality

$$\frac{1}{4}\left(y_{aa} + y_{ab} + y_{ac}\right) > \frac{1}{6}\left(x_{ab} + x_{ac} + y_{aa} + y_{ba} + y_{ca}\right).$$

This fully specifies our linear program.

Finding the Counterexample

As mentioned before, the variables of the linear program do not encompass all single-agent utilities of a fractional hedonic game with 15 agents. In fact, we just find values for the variables that correspond to our specific cyclic dynamics.

Hence, to extract an actual game from solutions of the linear program, we still have the freedom to select values for the utilities $u(v, w)$ for $v \in N_i$ and $w \in N_j$ with $|i-j| \in \{2, 3\}$. We set these utilities to sufficiently large negative numbers such that agents with a mutual negative utility cannot be in the same coalition in any individually rational partition (one can think of a utility of $-\infty$). This is convenient because it significantly reduces the candidates for stable partitions in a case analysis. Additionally, it is even necessary to have negative utilities because in a fractional hedonic game with non-negative utilities, the grand coalition is Nash-stable and therefore individually stable.

So far, we do not know whether the linear program of the previous section is actually useful. If it is infeasible, we gain no insight towards proving Theorem 4.3. However, if it has a feasible solution, we at least find a proof of Proposition 4.4.

In fact, a feasible solution of our linear program exists and is depicted in Figure 4.2. Hence, we find the single-agent utilities of a fractional hedonic game (N, u), where the utilities are defined by the variables of the linear program and large negative utilities between agents of non-adjacent triples.

It is not hard to confirm that the found fractional hedonic game admits a cycling dynamics. For instance, it holds that

$$u_{a_2}(N_3 \cup \{a_2\}) = \frac{1}{4}(436 + 228 + 248) = 228$$
$$> 223 = \frac{1}{6}(228 + 228 + 436 + 223 + 223) = u_{a_2}(N_1 \cup N_2).$$

This means that $\pi_0 \xrightarrow{a_2} \pi_1$ is a Nash deviation. Moreover, it holds that

$$u_{a_3}(N_3 \cup \{a_2\}) = 223 \geqslant 152 = u_{a_3}(N_3),$$
$$u_{b_3}(N_3 \cup \{a_2\}) = 171 \geqslant 152 = u_{b_3}(N_3), \text{ and}$$
$$u_{c_3}(N_3 \cup \{a_2\}) = \frac{704}{3} \geqslant 152 = u_{c_3}(N_3).$$

Thus, this deviation is even an individual deviation. Similarly, $\pi_1 \xrightarrow{b_2} \pi_2$ and $\pi_2 \xrightarrow{c_2} \pi_3$ are individual deviations.

Interestingly, the constructed game is even sufficient to prove Theorem 4.3. The proof is a tedious case distinction for excluding every partition from being individually stable. It can be found in the full version of Publication 4 (Brandt et al., 2022b).

4.3 HARDNESS REDUCTIONS IN HEDONIC GAMES

Many results in the hedonic games literature and in particular in this thesis concern the computational intractability of solution concepts. As discussed in Section 2.1.1, Ballester (2004) set the stage for deriving hardness results by representing preferences by individually rational lists of coalitions. This approach is appropriate for hedonic games with an ordinal representation of preferences. Subsequently, computational questions regarding hedonic games have been a constant object of study (see, e.g., Sung and Dimitrov, 2010; Aziz et al., 2013b; Woeginger, 2013; Peters and Elkind, 2015). In this section, we discuss important approaches for obtaining hardness results used in this thesis.

4.3.1 Hardness of Single-Agent Stability

An important proving scheme is due to Sung and Dimitrov (2010) who considered hardness in cardinal classes of hedonic games. This approach is for instance applied by Aziz et al. (2013b) or in Publications 3 and 4 to obtain Theorems 3.1, 3.2 and 3.22.

Assume that we want to prove hardness of the existence problem corresponding to some solution concept in a certain class of hedonic games. The general strategy of Sung and Dimitrov (2010) consists of the following five steps.

1. Find a No-instance, that is, a hedonic game in the considered class of games where the studied solution concept cannot be satisfied.
2. Encode the combinatorial structure of some NP-hard problem as a hedonic game of the given class.
3. Leverage the No-instance from Step 1 as gadget.
4. Add various auxiliary agents.
5. Prove the correspondence of Yes-instances of the source problem and instances satisfying the solution concept.

Of course, these five steps can be very different in difficulty and sophistication. Sometimes, even finding a first counterexample is extremely complex. For instance, as we have seen in the previous section, the fractional hedonic game constructed in the proof of Theorem 3.2 has a sophisticated structure of utilities, or the friends-and-enemies game for the third statement of Theorem 3.22 encompasses 183 agents. Moreover, it might even be the case that it is favorable not to use simple No-instances as gadgets as more complex No-instances can have additional properties that facilitate reductions. Examples for this incident are the reductions concerning additively separable hedonic games in Theorem 3.10.

In addition, usually there are various ways for the encoding in Step 2, and another challenge can be to find the right trade-off between the complexity of the encoding and the complexity of the auxiliary structures added in Step 4. As an example, there exist very simple additively separable hedonic games without a contractually Nash stable partition (Sung and Dimitrov, 2007), while the difficulty for deriving Theorem 3.1 is the appropriate usage of auxiliary agents.

In the following, we exemplify the reduction technique by outlining the proof of the first statement of Theorem 3.22. All steps in this reduction have moderate complexity, while none of the steps are trivial. Here, we want to derive the following theorem.

THEOREM 4.5 *(Bullinger, 2022)*

It is NP-complete to decide whether an appreciation-of-friends game contains a majority-out stable partition.

Step 1: Constructing a No-Instance

It is not immediately clear why appreciation-of-friends games do not always contain a majority-out stable partition. For instance, these games always contain a majority-in stable partition (cf. Theorem 3.21).

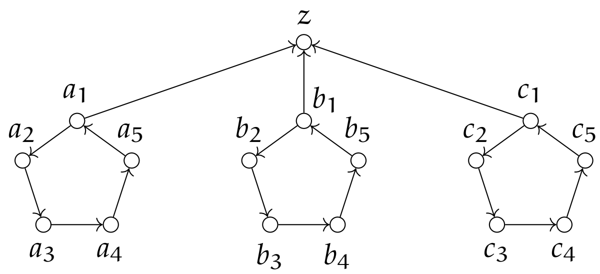


Figure 4.3: Appreciation-of-friends game without a majority-out stable partition (Bullinger, 2022). The depicted (directed) edges represent friends, i.e., a utility of n, whereas missing edges represent a utility of -1.

We now present the No-instance from Publication 3. Consider an appreciation-of-friends game (N, u) where $N = \{z\} \cup \{a_i, b_i, c_i : i \in [5]\}$. The utilities u^S for single agents are depicted in Figure 4.3. Whenever there is a directed edge from agent i to agent j, then $u_i^S(j) = n$; otherwise, $u_i^S(j) = -1$.

Let us discuss the main insights into why this game has no majority-out stable partition. The key is that the friendship relation of a (direct) 5-cycle is already very restrictive. To see this, consider the subgame (M, u^M) of (N, u) where $M = \{a_i : i \in [5]\}$ and u^M is the restriction of u to M. The following statement holds.

Proposition 4.6

The unique majority-out stable partition in the game (M, u^M) is the grand coalition.

Proof. Let π be a majority-out stable partition of (M, u^M).

Define $C := \pi(a_1)$ and assume for contradiction that $|C| < 5$. Then, there exists $i \in [5]$ such that $u^M_{a_i}(C) \leqslant 0$. For the remainder of the proof, we interpret indices modulo 5 mapping to their representative in $[5]$.

It holds that a_i has a Nash deviation to join $\pi(a_{i+1})$. This deviation is even a majority-out deviation unless $\pi(a_i) = \{a_i, a_{i-1}\}$. Therefore, since π is majority-out stable, it holds that $\pi(a_i) = \{a_i, a_{i-1}\}$.

Then, a_{i-2} has a Nash deviation to join C and, by majority-out stability, $\pi(a_{i-2}) = \{a_{i-2}, a_{i-3}\}$. But then a_{i-4} is in a singleton coalition and can perform a majority-out deviation to join $\pi(a_{i-3})$.

Hence, we derive a contradiction and it must be the case that $|C| = 5$. Note that the grand coalition is majority-out stable. □

We obtain the No-instance (N, u) by taking three copies of the 5-cyclic game and adding an additional agent z who is a friend of one agent in each of the cycles. Now, the agent z does not have any friends, and can therefore enforce to be in a small coalition in any majority-out stable partition. On the other hand, agents like a_1 who are in a 5-cycle and have agent z as a friend are in a constant conflict:

If they form a coalition with agents in their cycle, then this coalition has to be large (the intuition is that, as in Proposition 4.6, this coalition has to contain a whole cyclic structure). Hence, the coalition of z is smaller and a_1 wants to join. Otherwise, if a_1 is in a coalition with z, then a_2 cannot be forced to stay in their coalition unless this coalition is very small. Hence, a_1 would again perform a deviation. This tension cannot be resolved in any partition, and therefore there exists no majority-out stable partition.

Step 2: Encoding a Covering Problem

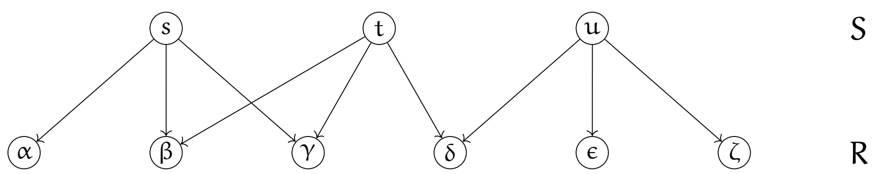
Many hardness reductions in hedonic games use covering problems as source instances. The reason is that there exist good ways to represent coverings by hedonic games. Here, we present a reduction from the NP-complete problem Exact Cover by 3-Sets (Karp, 1972). An instance of Exact Cover by 3-Sets consists of a tuple (R, S), where R is a finite ground set together with a set S of 3-element subsets of R. A Yes-instance is an instance so that there exists a subset $S' \subseteq S$ which partitions R.

A natural way for encoding a covering problem is to use the friendship relation of an appreciation-of-friends game to mimic the inclusion structure of an instance (R, S) of Exact Cover by 3-Sets. We can define an appreciation-of-friends game (N^1, u^1) with agent set $N^1 = R \cup S$ and single-agent utilities $u^1_x(y) = |N^1|$ if and only if $x \in S$ and $y \in R$ with $y \in x$. We refer to agents corresponding to elements in R and S as R-agents and S-agents, respectively.

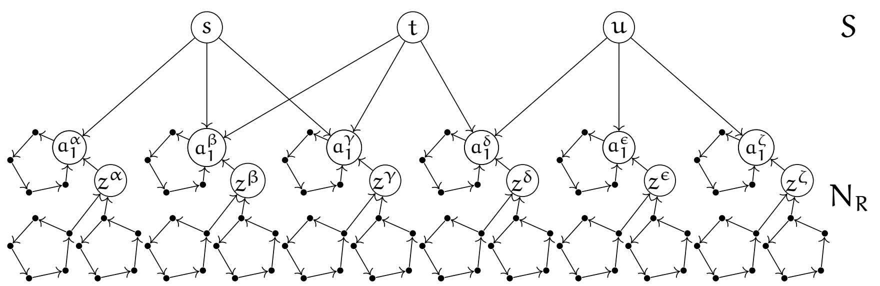
We illustrate Steps 2 to 5 in Figure 4.4. There, we consider the source instance (R, S) where $R = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta\}$ and $S = \{s, t, u\}$ where $s = \{\alpha, \beta, \gamma\}$, $t = \{\beta, \gamma, \delta\}$, and $\{\delta, \epsilon, \zeta\}$. The game (N^1, u^1) is depicted in Figure 4.4a.

Step 3: Leveraging the No-Instance

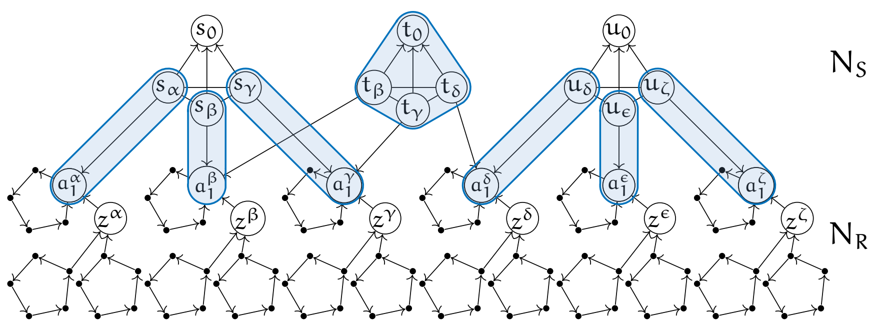
The game (N^1, u^1) is merely a representation of the combinatorial structure of a covering problem. It is not relevant if the source instance is a Yes-instance or a No-instance. In fact, the game (N^1, u^1) always contains a majority-out stable partition. This can be seen by running an MOS dynamics starting from the singleton coalition. First, let every S-agent s join an R-agent r with $r \in s$. This yields a partition π_1. Now, let the MOS dynamics continue arbitrarily. Note that, from this point onwards, only S-agents can perform a deviation. R-agents are in singleton coalitions or in coalitions with S-agents only, who would block their deviations. Hence, whenever they are not in a singleton coalition, they are not allowed to leave. Moreover, consider the vector of length $|S|$ containing the coalition sizes of each S-agent in decreasing size. Whenever an S-agent performs a deviation, she will decrease her coalition size, and therefore the vector of coalition sizes



(a) Step 2: Encoding a covering instance.



(b) Step 3: Using No-instance as gadget.



(c) Steps 4 and 5: Using auxiliary agents and finding a correspondence with Yes-instances (Bullinger, 2022).

Figure 4.4: Steps 2 to 5 of the reduction. We illustrate the reduction for the source instance (R, S) of Exact Cover by 3-Sets where $R = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta\}$ and $S = \{s, t, u\}$ where $s = \{\alpha, \beta, \gamma\}$, $t = \{\beta, \gamma, \delta\}$, and $\{\delta, \epsilon, \zeta\}$. The ground structure in Figure (a) represents the membership of elements in R in sets in S. Then, in Figure (b), every element in R is replaced by a gadget which is a copy of the found No-instance. Finally, in Figure (c), the sets in S are replaced by auxiliary agents. We then outline the partition corresponding to an exact cover of R with the sets s and u.

decreases lexicographically. This process has to terminate eventually, and the dynamics reaches a majority-out stable partition.

Thus, we need to create the chance that reduced instances are No-instances. To this end, we replace every R-agent by a copy of the No-instance from Step 1. Hence, we obtain a new game (N^2, u^2) where $N^2 = S \cup N_R$ with $N_R = \bigcup_{r \in R} N^r$ such that N^r corresponds to a copy of the agent set from the No-instance. The game is illustrated in Figure 4.4b. We retain the combinatorial structure of the covering instance by associating the copy of a_1 with the element in R and wiring the friendship of S-agents to these agents.

Step 4: Adding Auxiliary Agents

By leveraging No-instances for every element in R, we have added many local causes of instability to our base game. Hence, the agents of a set N^r for $r \in R$ need to form coalitions with agents outside of N^r in every majority-out stable partition. However, there are only few agents who have an incentive to form such coalitions, namely the agents in S. Moreover, we want that the formation of coalitions with agents in N^r corresponds to a covering.

We reach this goal by replacing the S-agents by a set of auxiliary agents. More precisely, we define a game (N^3, u^3) where $N^3 = N_S \cup N_R$. The set N_R is as in the previous step while $N_S = \bigcup_{s \in S} N^s$ where $N^s = \{s_0\} \cup \{s_r \colon r \in s\}$ for $s \in S$. Now, we split the origin of the three friendship relations of a previous S-agent s by having them originate from each of the three agents representing the elements in s. The game (N^3, u^3) is our final reduced instance and is depicted in Figure 4.4c.

Step 5: Correspondence of Yes-Instances

All of the previous steps had the purpose of setting up the construction of the reduced instances. However, the essence of hardness results is of course proving correctness of the provided reduction. In particular, this would require to map Yes-instances of the source problem to Yes-instances of the reduced problem.

We want to give some intuition of why this works for the presented reduction. Our running example, for which we depict the reduced instance in Figure 4.4 is a Yes-instance because the set R can be partitioned by $\{s, u\}$. There exists a majority-out stable partition in the reduced instance, which in particular contains the coalitions $\{a_1^\alpha, s_\alpha\}$, $\{a_1^\beta, s_\beta\}$, $\{a_1^\gamma, s_\gamma\}$, $\{a_1^\delta, u_\delta\}$, $\{a_1^\epsilon, u_\epsilon\}$, $\{a_1^\zeta, u_\zeta\}$, and $\{t_0, t_\beta, t_\gamma, t_\delta\}$. These coalitions are also highlighted in Figure 4.4c. The first six coalitions force an agent from each No-instance to a coalition that they cannot leave. Together, they prevent any danger of instability. The final coalition, $\{t_0, t_\beta, t_\gamma, t_\delta\}$, contains the agents corresponding to the set of $t \in S$, which is not contained in the partition of R.

As in this example, agents in a set N^s for $s \in S$ have two choices in a majority-out stable partition: Either they form a coalition of their own, or each of the agents s_r for $r \in s$ forms a coalition of size 2 with the agent a_1^r. In this way, we can simultaneously prevent the instabilities caused by each of the gadgets corresponding to elements in R if and only if we can partition R by elements in S.

4.3.2 A Generic Reduction Technique for Complex Concepts

As mentioned before, the reduction method presented in the previous section is quite universal and was applied in many different places within this thesis. However, axioms that naturally belong to complexity classes such as Σ_2^P which are above NP in the polynomial hierarchy sometimes require more involved techniques. One of these axioms is popularity. In this section, we show a reduction techniques that turns out to be extremely useful in the analysis of popularity. It is applied in all proofs of Theorems 3.9 and 3.10, but for simplicity, we limit attention to the following statement.

Theorem 4.7

> In flatmate games with strict and globally ranked preferences, it is coNP-hard to decide whether there exists a strongly popular partition.

A key challenge in the analysis of popularity is that the verification of popular partitions is often coNP-complete. As a consequence, proving that a reduced instance is a Yes-instance, that is, that it contains a popular partition, can be tedious. The central idea for dealing with this problem is that we construct a game that encodes some NP-complete problem such that decisions about popularity, independent of whether the source instance was a Yes- or No-instance, only depend on a small number of agents.

We visualize the scheme of the reduction in Figure 4.5. There, the circular vertices display agents, whereas edges indicate individually rational coalitions. The straight edges without arrow tips display a specific partition called π^*, which turns out to be the only reasonable candidate partition for popularity. The arrows indicate changes necessary in this partition to improve an agent's coalition. In the formal definition of the game given in Publication 2, all individually rational coalitions are of size at most 3, and we therefore obtain a flatmate game.

The ground structure of this technique is a binary tree as depicted in Figure 4.5 in blue. All coalitions of π^* of size 3 are a "vertex" in this tree. The root of the tree is at the bottom of the figure, and every vertex that is not a leaf has two children, which are to the left and right above it. More specifically, there are even two interleaved binary trees, where the second tree is the ground structure for the empty vertices. For simplicity, we only depict one of the trees.

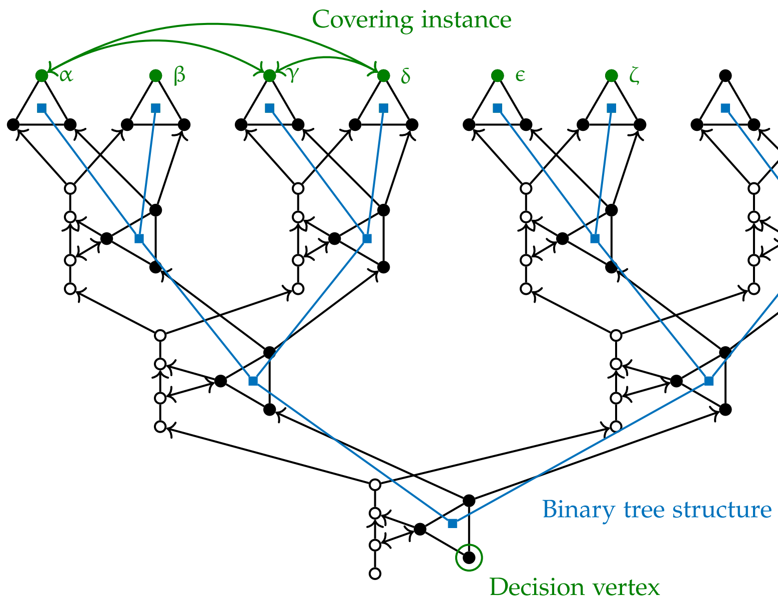


Figure 4.5: Reduction for simplifying complex decisions regarding the verification of popularity (Brandt and Bullinger, 2022). The source instance of a covering problem is attached to the top layer. Then the decision about the popularity of a good candidate partition is propagated downwards along a structure resembling a binary tree. In the end, decisions regarding the popularity of the candidate partition only depend on the agent at the decision vertex.

Now consider an instance (R, S) of Exact Cover by 3-Sets. In Figure 4.5, we have $R = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta\}$ and $\{\alpha, \gamma, \delta\} \in S$. This instance is encoded in the top layer of the reduced instance, and every element of the ground set R corresponds to a leaf of the binary tree. Note that the binary tree has more leaves than there are elements in R. This is always the case because the number of leaves of the binary tree is a power of 2 whereas $|R|$ is divisible by 3. In the actual reduction, we take the smallest size of a binary tree where more than half of the leaves correspond to elements of R.

The sets in S represent preferred coalitions for the agents representing the elements in R, which is for instance indicated for the set $\{\alpha, \gamma, \delta\}$. The key property of the reduction is that all the relevant information about popularity *propagates* to the root node of the tree. Hence, we reduce the analysis of the complex verification problem to the inspection of the root node. More specifically, the partition π^* is popular, both when the source instance is a Yes- and a No-instance.

The important properties of this construction that facilitate reductions are summarized as follows. If the source instance is a No-instance, then the partition π^* is strongly popular with popularity margin at least 2 against any other partition. On the other hand, if the source instance is a Yes-instance, then we can move the agents in the top layer to coalitions corresponding to a partition of R with sets in S.[12] However, to challenge the partition π^* as a whole, we must propagate coalition changes downwards in a very specific way. We can find a partition π' such that $\phi(\pi^*, \pi') = 1$ where the decision vertex at the bottom corresponds to an agent in a singleton coalition in π'.

The final step to establish Theorem 4.7 is to use an auxiliary agent z that would like to form a coalition with the agent corresponding to the decision vertex. Hence, we can consider the partition π^* augmented with a singleton coalition for z. Then, if the source instance is a No-instance, this partition is still strongly popular. On the other hand, due to z, if the source instance is a Yes-instance, there exists now another partition against which this partition only achieves a popularity margin of 0.

To obtain the other statements of Theorem 3.9, we can use multiple copies of the reduced game in Figure 4.5 or attach other gadgets to the decision vertex. It is possible to show that the obtained reductions yield flatmate games with strict and globally ranked preferences, but we omit the details here.

12 The actual construction has further auxiliary agents that deal with agents in the top layer that do not correspond to an element in R.

5 FURTHER DIRECTIONS

We have seen that hedonic games can cover a diverse set of coalition formation scenarios. These encompass applications like selecting roommates, allocating teams, or forming groups under diversity constraints. Moreover, formal models of coalition formation can contribute to a better understanding of the dynamical evolvement of coalition structures or the emergence of segregation.

The key goal of this thesis was to provide algorithms for obtaining coalition structures that satisfy a notion of desirability, such as stability or optimality. While we have bridged a few gaps in the literature, we have also opened up new research directions by considering stability constrained by majority decisions, dynamics based on individual decisions, or empathy in the form of loyalty. Moreover, most of the contributions leading to this thesis leave some specific problems open. We will conclude the thesis by discussing some general research directions and concrete open problems in order to give guidance for future work.

A first general path is to consider majority-based stability concepts in other contexts. While we obtain a good understanding of these concepts in additively separable hedonic games, some initial results by Tappe (2021) show their hardness in unrestricted fractional hedonic games. It would be interesting to see if we can guarantee their existence and obtain efficient algorithms in restricted classes of fractional hedonic games. In addition, studying majority-based stability in further classes of hedonic games may lead to intriguing discoveries.

A second venue is to follow the approach of Publication 6 and to move away from static utility models. One could consider games related to a given benchmark game that integrate novel incentives, for example, based on other notions of empathy. Towards this direction, we very recently propose a possibility to combine dynamics and utility modification by having agents change their utilities based on previously performed deviations (Boehmer et al., 2023). There is an abundance of possibilities to changes agents' incentives, and identifying the most meaningful ones could be a fruitful challenge.

In the remainder of this section, we will discuss some specific problems that were left open in the course of this thesis.

The complexity of some solution concepts remains open. First, while the algorithm in Theorem 3.3 is a 2-approximation to maximizing social welfare in modified fractional hedonic games, it is unknown whether welfare-maximizing partitions can be computed efficiently.

Open Problem 5.1

> Determine the complexity of computing a partition maximizing social welfare in modified fractional hedonic games.

On the one hand, maximizing social welfare in other classes of cardinal hedonic games, such as additively separable or fractional hedonic games, is NP-hard (Aziz et al., 2013b, Theorem 4; Aziz et al., 2015, Theorem 3), even under further utility restrictions. However, welfare-maximizing partitions can be computed in polynomial time in symmetric modified fractional hedonic games with binary utility values (Monaco et al., 2018, Theorem 3.5) and, consist only of partitions of size 2 and 3 for arbitrary utilities (cf. Publication 1). The answer could go towards either direction.

Second, it remains to determine the exact complexity of the existence problem for popularity in various classes of hedonic games. Since Theorem 3.10 contains both NP- and coNP-hardness, it feels that this problem could be Σ_2^P-complete.

Conjecture 5.2

> Consider the class of symmetric additively separable or symmetric fractional hedonic games. It is Σ_2^P-complete to decide whether there exists a popular partition in these games.

Similar open problems concern strong popularity, and the consideration of flatmate games.

Next, we consider the core of the loyal variant of appreciation-of-friends games. Recall that, according to Theorem 3.27, the core can be empty in the loyal variant of an additively separable hedonic game, and the corresponding decision problem is NP-hard. Moreover, even though we do not know whether partitions in the core of the k-fold loyal variant of appreciation-of-friends games for $k \geqslant 2$ are guaranteed to exist, we face computational boundaries once again (cf. Theorem 3.26). However, the core is non-empty for the loyal variant of an additively separable hedonic game if the friendship structure is a tree (cf. Publication 6). In addition, the core in ordinary appreciation-of-friends games is always non-empty (Dimitrov et al., 2006). Hence, there is evidence towards either direction for the next open problem.

Open Problem 5.3

> Invetigate whether the core can be empty in the loyal variant of an appreciation-of-friends game. Moreover, determine the complexity of computing a partition in the core of such a game.

Note that Open Problem 5.3 is equivalent to asking for elements in the core of altruistic hedonic games under minimum-based equal-treatment preferences, an open problem mentioned by Kerkmann et al. (2022).

Our next open problem concerns dynamics in hedonic games. In Publication 4, we have mainly considered necessary convergence. In particular, despite the existence and efficient computability of individually stable partitions in hedonic diversity games, IS dynamics may cycle under strong restrictions (cf. Theorem 3.12). In Publication 4, we also discuss restrictions for necessary cycling, which preclude possible convergence. However, in particular in the following cases, it might be possible to complement the negative results of Theorem 3.12 by more positive results.

OPEN PROBLEM 5.4

Investigate whether IS dynamics in hedonic diversity games possibly converge if

1. the dynamics start from the singleton partition, or
2. preferences are strict and single-peaked.

We finally state a conjecture concerning the welfare guarantee of Pareto-optimal assignments in Schelling games. In Theorem 3.29, there is a gap for the price of Pareto optimality. We have a lower bound of $\Omega(n)$, whereas the upper bound of $\mathcal{O}(n\sqrt{n})$ is not matching. However, we have an upper bound of $\mathcal{O}(n)$ in restricted subclasses of Schelling games (cf. Theorem 3.31). Moreover, the price of group-welfare optimality and utility-vector optimality, natural refinements of Pareto optimality, is $\Theta(n)$. Hence, there is multiple evidence for the next conjecture.

CONJECTURE 5.5

The price of Pareto optimality in Schelling games satisfies $\Theta(n)$.

Together, we have seen that there is still a lot of potential for exciting future research in coalition formation.

Part II

ORIGINAL PUBLICATIONS

6 CORE PUBLICATION 1: PARETO-OPTIMALITY IN CARDINAL HEDONIC GAMES

SUMMARY

Coalition formation is an important research topic in multi-agent systems and typically studied in the framework of hedonic games. We study cardinal classes of hedonic games in which a set of agents expresses their preferences about potential coalitions with the aid of cardinal utilities. These classes are additively separable, fractional, and modified fractional hedonic games. Each of these classes can model different aspects of dividing a society into groups.

The desirability of coalition structures in hedonic games is usually measured by considering so-called solution concepts. Among these, Pareto optimality and individual rationality are among the most natural requirements. A coalition structure is Pareto-optimal if any coalition structure that is better for some agent is worse for another agent. Pareto optimality can therefore be seen as a global guarantee of a coalition structure. A stronger guarantee is given by welfare optimality which demands maximality with respect to the utilitarian welfare of a coalition structure. On the other hand, individual rationality of a coalition structure means that no agent prefers to be on their own over being in their designated coalition. Therefore, individual rationality gives an incentive to engage in a coalition formation process at all.

For all of the above classes of games, we provide algorithms that find Pareto-optimal coalition structures under some natural restrictions. While the output is also individually rational for modified fractional hedonic games, combining both Pareto optimality and individual rationality leads to an NP-hardness for symmetric additively separable and symmetric fractional hedonic games. In addition, we prove that welfare-optimal and Pareto-optimal partitions coincide for symmetric modified fractional hedonic games with binary utility values, resolving an open problem. For general modified fractional hedonic games, our algorithm returns a 2-approximation for maximizing social welfare. While we leave the complexity of computing welfare-optimal coalition structures as a an open problem, we can show that finding welfare-optimal coalition structures in modified fractional hedonic games only requires the consideration of coalitions containing at most three agents.

REFERENCE

M. Bullinger. **Pareto-optimality in cardinal hedonic games**. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 213–221, 2020.
DOI: https://dl.acm.org/doi/10.5555/3398761.3398791

INDIVIDUAL CONTRIBUTION

This is a single-authored publication and I, Martin Bullinger, am responsible for all parts of this publication.

COPYRIGHT AGREEMENT

The right to present this paper in a doctoral thesis has been granted by the publisher, the International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), in the copyright form presented below. There, IFAAMAS grants permission to "*Personal reuse of all or portions of the above article/paper in other works of their own authorship.*" This form can also be found at https://www.ifaamas.org/AAMAS/aamas07/IFAAMASCopyrightForm.pdf (accessed November 10, 2022).

INTERNATIONAL FOUNDATION FOR AUTONOMOUS AGENTS AND MULTIAGENT SYSTEMS (IFAAMAS)
COPYRIGHT FORM

Title of Article/Paper: ____

Publication in Which Article Is to Appear: ________________________________

Author’s Name(s): ____

Please type or print your name as you wish it to appear in print

(Please read and sign Part A only, unless you are a government employee and created your article/paper as part of your employment. If your work was performed under Government contract, but you are not a Government employee, sign Part A and see item 5 under returned rights.)

PART A – Copyright Transfer Form

The undersigned, desiring to publish the above article/paper in a publication of the International Foundation for Autonomous Agents and MultiAgent Systems (IFAAMAS), hereby transfer their copyrights in the above article/paper to the International Foundation for Autonomous Agents and MultiAgent Systems (IFAAMAS), in order to deal with future requests for reprints, translations, anthologies, reproductions, excerpts, and other publications.

This grant will include, without limitation, the entire copyright in the article/paper in all countries of the world, including all renewals, extensions, and reversions thereof, whether such rights current exist or hereafter come into effect, and also the exclusive right to create electronic versions of the article/paper, to the extent that such right is not subsumed under copyright.

The undersigned warrants that they are the sole author and owner of the copyright in the above article/paper, except for those portions shown to be in quotations; that the article/paper is original throughout; and that the undersigned right to make the grants set forth above is complete and unencumbered.

If anyone brings any claim or action alleging facts that, if true, constitute a breach of any of the foregoing warranties, the undersigned will hold harmless and indemnify IFAAMAS, their grantees, their licensees, and their distributors against any liability, whether under judgment, decree, or compromise, and any legal fees and expenses arising out of that claim or actions, and the undersigned will cooperate fully in any defense IFAAMAS may make to such claim or action. Moreover, the undersigned agrees to cooperate in any claim or other action seeking to protect or enforce any right the undersigned has granted to IFAAMAS in the article/paper. If any such claim or action fails because of facts that constitute a breach of any of the foregoing warranties, the undersigned agrees to reimburse whomever brings such claim or action for expenses and attorneys’ fees incurred therein.

Returned Rights

In return for these rights, IFAAMAS hereby grants to the above authors, and the employers for whom the work was performed, royalty-free permission to:
1. Retain all proprietary rights other than copyright (such as patent rights).
2. Personal reuse of all or portions of the above article/paper in other works of their own authorship.
3. Reproduce, or have reproduced, the above article/paper for the author’s personal use, or for company use provided that IFAAMAS copyright and the source are indicated, and that the

copies are not used in a way that implies IFAAMAS endorsement of a product or service of an employer, and that the copies per se are not offered for sale. The foregoing right shall not permit the posting of the article/paper in electronic or digital form on any computer network, except by the author or the author's employer, and then only on the author's or the employer's own web page or ftp site. Such web page or ftp site, in addition to the aforementioned requirements of this Paragraph, must provide an electronic reference or link back to the IFAAMAS electronic server, and shall not post other IFAAMAS copyrighted materials not of the author's or the employer's creation (including tables of contents with links to other papers) without IFAAMAS's written permission.
4. Make limited distribution of all or portions of the above article/paper prior to publication.
5. In the case of work performed under U.S. Government contract, IFAAMAS grants the U.S. Government royalty-free permission to reproduce all or portions of the above article/paper, and to authorize others to do so, for U.S. Government purposes.

In the event the above article/paper is not accepted and published by IFAAMAS, or is withdrawn by the author(s) before acceptance by IFAAMAS, this agreement becomes null and void.

Author's Signature	Date
Employer for whom work was performed	Title (if not author)

PLEASE FAX THIS SIGNED COPYRIGHT FORM TO:
Jay Modi
FAX: 215-895-0545
Department of Computer Science
Drexel University
Philadelphia, PA, 19104

INTERNATIONAL FOUNDATION FOR AUTONOMOUS AGENTS AND MULTIAGENT SYSTEMS
c/o Professor Ed Durfee
University of Michigan
EECS Department
2260 Hayward Street
Ann Arbor, MI 48109 USA

PART B – U.S. Government Employee Certification

This will certify that all authors of the above article/paper are employees of the U.S. Government and performed this work as part of their employment, and that the article/paper is therefore not subject to U.S. copyright protection. The undersigned warrants that they are the sole author/translator of the above article/paper, and that the article/paper is original throughout, except for those portions shown to be in quotations.

U.S. Government Employee Authorized Signature	Date
Name of Government Organization	Title (if not author)

Pareto-Optimality in Cardinal Hedonic Games

Martin Bullinger
Technische Universität München
München, Germany
bullinge@in.tum.de

ABSTRACT

Pareto-optimality and individual rationality are among the most natural requirements in coalition formation. We study classes of hedonic games with cardinal utilities that can be succinctly represented by means of complete weighted graphs, namely additively separable (ASHG), fractional (FHG), and modified fractional (MFHG) hedonic games. Each of these can model different aspects of dividing a society into groups. For all classes of games, we give algorithms that find Pareto-optimal partitions under some natural restrictions. While the output is also individually rational for modified fractional hedonic games, combining both notions is NP-hard for symmetric ASHGs and FHGs. In addition, we prove that welfare-optimal and Pareto-optimal partitions coincide for simple, symmetric MFHGs, solving an open problem from Elkind et al. [9]. For general MFHGs, our algorithm returns a 2-approximation to welfare. Interestingly, welfare-optimal partitions in MFHGs only require coalitions of at most three agents.

KEYWORDS

Coalition Formation; Hedonic Games; Pareto-optimality; Welfare

ACM Reference Format:
Martin Bullinger. 2020. Pareto-Optimality in Cardinal Hedonic Games. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 9 pages.

1 INTRODUCTION

Coalition formation is a central problem in multi-agent systems and has been extensively studied, ever since the publication of von Neumann and Morgenstern's *Theory of Games and Economic Behavior* in 1944. In coalition formation, every agent of a group seeks to be in a desirable coalition. As an important special case therein, in clustering problems, a society is observed, for which a structuring into like-minded groups, or communities, is to be identified [16]. Coalition scenarios can be modeled by letting the agents submit preferences, subject to which the happiness of an individual agent with her coalition, or the like-mindedness of a coalition, can be measured. The goal of every individual agent is to maximize the value of her coalition.

In many settings, it is natural to assume that an agent is only concerned about her own coalition, i.e., externalities are ignored. As a consequence, much of the research on coalition formation now concentrates on these so-called *hedonic games* [7]. Still, the number of coalitions an agent can be part of is exponential in the number of agents, and therefore it is desirable to consider expressive, but succinctly representable classes of hedonic games. This can be established by encoding preferences by means of a complete directed weighted graph where an edge weight $v_x(y)$ is a cardinal value or utility that agent x assigns to agent y. Still, this underlying structure leaves significant freedom, how to obtain (cardinal) preferences over coalitions. An especially appealing type of preferences is to have the sum of values of the other individuals as the value of a coalition. This constitutes the important class of additively separable hedonic games (ASHGs) [6].

In ASHGs, an agent is willing to accept an additional agent into her coalition as long as her valuation of this agent is non-negative and in this sense, ASHGs are not sensitive to the intensities of single-agent preferences. In particular, if all valuations are non-negative, forming the grand coalition consisting of all agents is a best choice for every agent. In contrast, fractional hedonic games (FHGs) define preferences over coalitions by dividing the sum of values by the size of the coalition [1]. This incentivizes agents to form dense cliques, and therefore appropriately models like-mindedness in the sense of clustering problems.

On the other hand, agents in FHGs may improve with a new agent whose valuation is below the average of their current coalition partners. To avoid this, it is also natural to define the value of a coalition as the average value of other agents, i.e., the denominator in the definition of FHGs is replaced by the size of the coalition minus 1. This defines the class of modified fractional hedonic games (MFHGs) [17]. For all measures of stability and optimality that have been investigated, these games do not guarantee the formation of large cliques and are therefore less suitable for clustering problems. In fact, partitions of agents into coalitions satisfying many desirable properties can be computed in polynomial time, often simply by computing maximum (weight) matchings. However, these games ensure a certain degree of homogeneity of the agents, as agents tend to contribute uniformly to each other's utility in stable coalitions.

Having selected a representation for modeling the coalition formation process, one needs a measure for evaluating the quality of a partition. Various such measures, also called *solution concepts*, have been proposed in the literature. Most of them aim to guarantee a certain degree of stability—preventing single agents or groups from agents to break apart from their coalitions—and optimality—guaranteeing a globally measured outcome that is good for the society as entity. A good overview of solution concepts is by Aziz and Savani [5]. The most undisputed measure of optimality is *Pareto-optimality*, i.e., there should be no other partition, such that every agent is weakly, and some agent strictly, better off. Apart from its optimality guarantee, Pareto-optimality can also be seen as a measure of stability, because a Pareto-optimal partition disallows an agent to propose a partition she prefers without having another agent vetoing this proposal. A stronger notion of optimality is that

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

of *(utilitarian) welfare-optimality*, which aims to maximize the sum of utilities of all agents.

Pareto-optimal outcomes still might be extremely disadvantageous for single agents that receive large negative utility in order to give small positive utility to another agent. Therefore, it is also desirable that agents receive at least the utility they would receive in a coalition of their own (in all our models this means non-negative utility). This condition is called *individual rationality*. Clearly, a Pareto-optimal and individually rational outcome can always be found by the simple local search algorithm that starts with the partition of the agents into singleton coalitions and moves to Pareto-improvements as long as these exist. In general, this basic algorithm need not run in polynomial time and the output need not be welfare-optimal. It can even occur that no welfare-optimal partition is individually rational. As we will see, it is even often NP-hard to compute a Pareto-optimal and individually rational partition (Theorem 5.2, Theorem 6.4).

We study Pareto-optimality in all three classes of games, and give polynomial-time algorithms for computing Pareto-optimal partitions in important subclasses, including symmetric ASHGs and MFHGs. In addition, we prove that welfare-optimal and Pareto-optimal partitions coincide for simple symmetric MFHGs, closing the bounds on the price of Pareto-optimality for this class of games left by Elkind et al. [9]. In the weighted case, our algorithm for Pareto-optimality in MFHGs gives a 2-approximation of welfare and its output is always individually rational. While we can prove that even in the weighted case, welfare-optimality is attained by a partition consisting only of coalitions of size two and three, the complexity of computing a welfare-optimal partition remains open. On the other hand, we prove that computing a Pareto-optimal and individually rational partition is NP-hard for symmetric ASHGs and FHGs, thus extending a result by Aziz et al. [3] for general ASHGs. Note that symmetry is a significant restriction for hardness reductions, because non-symmetric games allow for the phenomenon of non-mutual interest.

2 RELATED WORK

Hedonic games were first introduced by Drèze and Greenberg [7]. Since then, a great amount of research has been devoted to the study of algorithmic and mathematical properties of axiomatic concepts regarding stability and optimality, representability of preferences, and the discovery of well-behaved, yet expressive classes of hedonic games. The survey by Hajduková [12] gives a critical overview of preference representations and conditions that allow for the existence and efficient computability of central stability notions, such as Nash and core stability.

Pareto-optimality can be studied in many classes of hedonic games by exploiting a strong relationship between Pareto-optimality and perfection, i.e., partitions that put every agent in one of her most preferred partitions [2]. This gives rise to the preference refinement algorithm (PRA) which finds Pareto-optimal partitions under certain conditions by means of a perfection-oracle. The resulting Pareto-optimal partitions are even individually rational. The assumptions required for the algorithm include the efficient computation of preference refinements, which is not possible for ASHGs and FHGs. Indeed, during the PRA, one has to search a set of hedonic games that interpolates between two preference profiles and can contain games not implementable as an ASHG or FHG, respectively. In fact, for ASHGs and FHGs, perfect partitions can be computed in polynomial time (Theorem 5.4, Theorem 6.6), while computing Pareto-optimal and individually rational partitions is NP-hard (Theorem 5.2, Theorem 6.4).

For cardinal hedonic games, stability and optimality have been studied to some extent. Gairing and Savani [10, 11] settled the complexity of the individual-player stability notions of Drèze and Greenberg for symmetric ASHGs by treating them as local search problems. For FHGs, hardness and approximation results for welfare are given by Aziz et al. [4], while Aziz et al. [1] study stability. Monaco et al. [15] show the tractability of some stability notions and welfare-optimality for simple symmetric MFHGs, and the computability of partitions in the core for weighted MFHGs. The approximation results for FHGs and most of the positive results for MFHGs rely on computing specific matchings.

Pareto-optimality for cardinal hedonic games was mainly studied by Elkind et al. [9] in terms of the price of Pareto-optimality (PPO), a worst-case ratio of Pareto-optimal and welfare-optimal partitions. The focus lies on simple symmetric graphs and their main result is to bound the PPO between 1 and 2 for simple, symmetric MFHGs. Since we show that, for these games, every Pareto-optimal partition is welfare-optimal, we will close this gap. Pareto-optimality for ASHGs was considered by Aziz et al. [3]. However, they only dealt with a restricted class of preferences that guarantees unique top-ranked coalitions and therefore one can apply a simple serial dictatorship algorithm.

3 PRELIMINARIES

The primary ingredient of our model is a set of agents N that assign hedonic preferences over *partitions* of N (also called *coalition structures*), where the only information of a partition an agent is interested in, is her own coalition. Preferences of agent i are therefore given over $\mathcal{N}_i = \{C \subseteq N \colon i \in C\}$, i.e., the subsets of agents including herself, by valuation functions $v_i : \mathcal{N}_i \to \mathbb{R}$. We investigate a partition π of the agents for notions of optimality and stability, most importantly *Pareto-optimality*. Given a partition π, we denote by $\pi(i)$ the partition of agent i and the utility she received from this partition by $v_i(\pi) = v_i(\pi(i))$. A partition π' is a *Pareto-improvement* over π if, for all agents $i \in N$, $v_i(\pi') \geq v_i(\pi)$ and there exists an agent $j \in N$ with $v_j(\pi') > v_j(\pi)$. In this case, we also say that π' *Pareto-dominates* π. A partition π is *Pareto-optimal* if it is not Pareto-dominated.

A stronger requirement is that of *(utilitarian) welfare-optimality*. For a subset $M \subseteq N$ of agents, we denote $v_M(\pi) = \sum_{i \in M} v_i(\pi)$. The *social welfare* of a coalition C is $\mathcal{SW}(C) = v_C(\pi)$. The social welfare of a partition π is $\mathcal{SW}(\pi) = \sum_{C \in \pi} \mathcal{SW}(C) = \sum_{i \in N} v_i(\pi)$. A partition π is called *welfare-optimal* if it maximizes the function $\mathcal{SW}$ amongst all partitions of agents. Welfare-optimal partitions are Pareto-optimal.

A partition π is *individually rational* for agent i if $v_i(\pi) \geq v_i(\{i\})$, i.e., agent i does not prefer to stay alone. In addition, π is *individually rational* if it is individually rational for every agent. Partitions that are welfare-optimal or individually rational and Pareto-optimal always exist.

Preferences are succinctly represented by a family of cardinal utility functions $(v_i)_{i \in N}$ where $v_i : N \to \mathbb{R}$ with $v_i(i) = 0$ that can be aggregated to preferences over coalitions. A natural representation is by means of a complete, directed, and weighted graph $G = (N, E, v)$ where the weights are defined by the utility functions. A game is called *symmetric* if, for all pairs of agents $i, j \in N$, $v_i(j) = v_j(i)$. In this case, the underlying graph is symmetric and we denote $v(e) = v(i, j) = v_i(j) = v_j(i)$ for a 2-elementary set of agents $e = \{i, j\}$. A game is called *simple* if $v_i(j) \in \{0, 1\}$ for all agents $i, j \in N$. A hedonic game with simple and symmetric preferences can therefore be represented by an unweighted and undirected (but incomplete) graph.

We define the aggregated utilities for ASHGs, FHGs, and MFHGs for partition π and agent i by

$$v_i^{ASHG}(\pi) = \sum_{j \in \pi(i)} v_i(j)$$

$$v_i^{FHG}(\pi) = \frac{v_i^{ASHG}(\pi)}{|\pi(i)|}, \text{ and}$$

$$v_i^{MFHG}(\pi) = \begin{cases} \frac{v_i^{ASHG}(\pi)}{|\pi(i)|-1} & \text{for } \pi(i) \neq \{i\} \\ 0 & \text{for } \pi(i) = \{i\}. \end{cases}$$

If it is clear from the context which game is considered, we omit the superscripts of the utility functions.

We use the following notation from graph theory. For an arbitrary graph $G = (V, E)$, a vertex set $W \subseteq V$ and an edge set $F \subseteq E$, denote by $G[W]$ and $G[F]$ the subgraph of G induced by W and F, respectively, and denote by $E(G)$ its edge set.

4 MODIFIED FRACTIONAL HEDONIC GAMES

In this section we focus on symmetric MFHGs. The analysis of Pareto-optimality on this class of games relies on an extension of maximum matchings to cliques. Given a graph, a set C of vertex-disjoint cliques, each of size at least 2, is called a *clique matching*. A vertex that is part of any clique in C is called *covered* or *matched.* We are interested in clique matchings that cover a maximum number of vertices. We call the corresponding search problem `MaxCliqueMatching`. Interestingly, a clique matching is maximum if and only if it is *inclusion-maximal w.r.t. vertices,* i.e., there exists no other clique matching covering a strict superset of vertices (Theorem 4.5), and can be computed in polynomial time. We can further simplify the analysis to the special case that only triangles and edges are allowed.

Given a graph, a set of vertex disjoint cliques, each of size 2 or 3, is called a *3-clique matching*. By splitting larger cliques, `MaxCliqueMatching` is equivalent to `Max3CliqueMatching`, i.e., finding a maximum 3-clique matching. Given a 3-clique matching C, we denote by $M(C)$ and $T(C)$ its cliques of size 2 (edges) and size 3 (triangles), respectively.

Theorem 4.1 (Hell and Kirkpatrick [13]). *The problem* `Max3CliqueMatching` *can be solved in polynomial time.*

We prove that `MaxCliqueMatching` is equivalent to finding a Pareto-optimal partition on an MFHG. Note that a relationship between clique matchings and *simple* symmetric MFHGs was already exploited by Monaco et al. [15] for computing welfare-optimal partitions.

Theorem 4.2. `MaxCliqueMatching` *is equivalent to finding a Pareto-optimal partition on a symmetric MFHG (under Turing reductions). Moreover, if we can solve* `MaxCliqueMatching` *in polynomial time, we can even compute a Pareto-optimal and individually rational partition for a symmetric MFHG in polynomial time.*

Proof. Assume first that we are given an algorithm to find a Pareto-optimal partition on a symmetric MFHG and let $G = (V, E)$ be an instance of `MaxCliqueMatching`. We transform G into an MFHG with the underlying weighted symmetric graph $G' = (V, E', v)$ where $E' = \{e \subseteq V : |e| = 2\}$ and

$$v(e) = \begin{cases} 1 & \text{if } e \in E \\ -\Delta - 1 & \text{else,} \end{cases}$$

where Δ is the maximum degree of a vertex in G.

Let π be a Pareto-optimal partition of vertices into coalitions for the symmetric MFHG with underlying graph G'. Define $C = \{P \in \pi : |P| \geq 2\}$. Then, C consists of cliques in G. Otherwise, by construction of the utilities, there is one agent who receives negative utility. Assume for contradiction that there exists a coalition $P \in C$ such that some agents in P receive negative utility. Let $S = \{p \in P : v(p, p') = 1 \text{ for all } p' \in P \backslash \{p\}\}$ be the set of agents that receive non-negative utility from all other agents in P. Since some agents receive negative utility, there exists an agent $q \in P \backslash S$. But then, the coalition $S_q = S \cup \{q\}$ forms a clique in G and the partition $\pi' = (\pi \backslash \{P\}) \cup \{S_q\} \cup \{\{p'\} : p' \in P \backslash S_q\}$ is a Pareto improvement over π. Hence, π consists only of cliques and, by design of the MFHG utilities, assigns utility 1 to agents in a clique of size at least 2, and 0 to agents in singleton coalitions. Consequently, C is inclusion-maximal w.r.t. vertices, because every clique matching that covers strictly more agents gives rise to a Pareto-improvement that assigns utility 1 to a strict superset of agents that already receive utility 1. Hence, we can solve `MaxCliqueMatching` by computing a Pareto-optimal partition of the MFHG induced by G'.

Conversely, assume that we can solve `MaxCliqueMatching`. Consider a symmetric MFHG induced by a complete weighted graph $G = (N, E, v)$. Algorithm 1 computes a Pareto-optimal partition in polynomial time given an algorithm `MaxCliqueMatching` that computes a vertex-maximal clique matching in polynomial time. The idea is to restrict attention to the unweighted subgraph induced by edges with the largest positive weight still available.

The running time of the algorithm is clearly polynomial, including polynomially many calls of `MaxCliqueMatching`. We prove its correctness. Let π be the output of the algorithm. First, note that all non-singleton coalitions are cliques in G with identical positive utility within each clique. Hence, the output is individually rational.

For the proof of Pareto optimality, we assume that the while loop took m iterations and we subdivide $\pi = \mathcal{S} \cup \bigcup_{k=1}^{m} C_k$, where C_k is the clique matching in iteration k, and $\mathcal{S}$ consists of the singleton coalitions that are added to π after the while loop. We will show by induction over m that if the algorithm uses m iterations of the while loop, then the output is Pareto-optimal. Let π' be any coalition such that, for all agents $i \in N$, $v_i(\pi) \leq v_i(\pi')$. We will prove that this implies, for all agents $i \in N$, $v_i(\pi) = v_i(\pi')$.

Input: Symmetric MFHG induced by graph $G = (N, E, v)$
Output: Pareto-optimal and individually rational partition π

$\pi \leftarrow \emptyset, A \leftarrow N, G_r \leftarrow G[\{e \in E: v(e) > 0\}]$
while $E(G_r) \neq \emptyset$ **do**
  $v_{\max} \leftarrow \max\{v(e): e \in E(G_r)\}$
  $E_H \leftarrow \{e \in E(G_r): v(e) = v_{\max}\}$
  $H \leftarrow G_r[E_H]$
  $C \leftarrow$ `MaxCliqueMatching`(H)
  $\pi \leftarrow \pi \cup C$
  $A \leftarrow \{a \in A: a$ not covered by $C\}$
  $G_r \leftarrow G_r[A]$
return $\pi \cup \{\{a\}: a \in A\}$

Algorithm 1: Pareto-optimal partition of a sym. MFHG

If $m = 0$, G contains no edges of positive weight and therefore, for all agents $i \in N$, $v_i(\pi) = 0 \geq v_i(\pi')$. For the induction step, let $m \geq 1$. Let H be the auxiliary graph of the first while loop. Note that within π, agents in C_1 can only be matched with agents in H since they receive the highest possible utility of any agent in any coalition in G and every other agent diminishes their MFHG utility. In particular, they cannot be better off. Since C_1 is a vertex-maximal clique matching on H, no agent in H not covered by C_1 can be in a coalition with an agent in C_1 in π'. Define $W = \{i \in N: i$ not covered by $C_1\}$ and consider $\hat{G} = G[W]$. Then $\hat{\pi} = \mathcal{S} \cup \bigcup_{k=1}^{m-1} C_{k+1}$ is a possible outcome of the algorithm for $\hat{G}$. Furthermore, π' restricted to agents in W is weakly better for any agent in W than $\hat{\pi}$. Hence, by induction, also no agent outside C_1 can be better off. □

Even though Pareto-optimal outcomes may be worse than welfare-optimal outcomes by an arbitrarily large factor, the output of the above algorithm guarantees significant social welfare.

Theorem 4.3. *Let a symmetric MFHG be given. Let π be the partition computed by Algorithm 1 for this game. Then, for any welfare-optimal partition π^*, it holds that $2\mathcal{SW}(\pi) \geq \mathcal{SW}(\pi^*)$.*

Proof. Consider a symmetric MFHG induced by graph $G = (N, E, v)$. Let π be a partition computed by Algorithm 1 for this game and let π^* be welfare-optimal. We will show that, for any coalition $C \in \pi^*$, it holds that $2v_C(\pi) \geq v_C(\pi^*)$, which implies the assertion.

Before we prove this, we make the observation that, for each edge $\{x, y\} = e \in E$, it holds that $v_x(\pi) \geq v(x, y)$ or $v_y(\pi) \geq v(x, y)$. Indeed, if $v(x, y) \leq 0$, then this follows from the individual rationality of π. If we reach a maximum weight $v_{\max} \leq v(x, y)$ in the while loop and $x \notin G_r$ or $y \notin G_r$ then $v_x(\pi) > v(x, y)$ or $v_y(\pi) > v(x, y)$. Otherwise, we reach an iteration where $v_{\max} = v(x, y)$ and any maximum clique matching matches at least one of them.

Now let any coalition $C \in \pi^*$ be given where $|C| = k$. The next step is to sort the agents in C by means of Algorithm 2. The resulting order places the agents essentially in decreasing value w_i of an incident edge of high utility. Let $(c_1, \dots, c_k)$, $(w_1, \dots, w_k)$ be an outcome of this algorithm. We claim that for every $i \in \{1, \dots, k\}$, $v_{c_i}(\pi^*) \leq v_{c_i}(\pi) + \sum_{1 \leq j < i} \frac{w_j}{k-j}$. Note that for all $1 \leq j < i$,

Input: Coalition $C \in \pi^*$
Output: Order $(c_1, \dots, c_k)$ of C and weights $(w_1, \dots, w_k)$

$H \leftarrow C, F \leftarrow \{\{x, y\} \in E: x, y \in C, v(x, y) > 0\}, j \leftarrow 1$
while $F \neq \emptyset$ **do**
  $v_{\max} \leftarrow \max\{v(e): e \in F\}$
  Choose $\{x, y\} \in \operatorname{argmax}\{v(x, y): \{x, y\} \in F\}$ with $v_x(\pi) \geq v(x, y)$
  $c_j \leftarrow x$, $w_j \leftarrow v_{\max}$
  $H \leftarrow H \backslash \{x\}$, $F \leftarrow \{e \in F: x \notin e\}$, $j \leftarrow j + 1$
Order $H = (c_j, \dots, c_k)$ arbitrarily, $w_i \leftarrow 0$ for $i = j, \dots, k$
return $(c_1, \dots, c_k), (w_1, \dots, w_k)$

Algorithm 2: Special ordering for weight distribution

$v(c_i, c_j) \leq w_j$ and for all $j > i$, $v(c_i, c_j) \leq w_i$. Hence,

$$v_{c_i}(\pi^*) = \sum_{j \neq i} \frac{v(c_i, c_j)}{k-1} \leq \sum_{j>i} \frac{w_i}{k-1} + \sum_{j<i} \frac{w_j}{k-1}$$
$$\leq \sum_{j>i} \frac{w_i}{k-i} + \sum_{j<i} \frac{w_j}{k-j} = w_i + \sum_{j<i} \frac{w_j}{k-j} \leq v_{c_i}(\pi) + \sum_{j<i} \frac{w_j}{k-j}.$$

We infer that

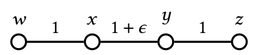
$$v_C(\pi^*) = \sum_{i=1}^{k} v_{c_i}(\pi^*) \leq \sum_{i=1}^{k} \left(v_{c_i}(\pi) + \sum_{j<i} \frac{w_j}{k-j} \right)$$
$$= v_C(\pi) + \sum_{i=1}^{k} \sum_{j<i} \frac{w_j}{k-j} = v_C(\pi) + \sum_{j=1}^{k} (k-j) \frac{w_j}{k-j} \leq 2v_C(\pi).$$

□

Note that the approximation guarantee of the theorem extends to the case of non-symmetric weights, because the symmetrization $v'(x, y) = \frac{1}{2}(v_x(y) + v_y(x))$ preserves the welfare.

Moreover, the factor of 2 is the best possible approximation guarantee of Algorithm 1. Let $\epsilon > 0$ and the complete graph on vertex set $V = \{w, x, y, z\}$ be given with weights as

$$v(e) = \begin{cases} 1 + \epsilon & \text{if } e = \{x, y\} \\ 1 & \text{if } e \in \{\{w, x\}, \{y, z\}\} \\ 0 & \text{else.} \end{cases}$$



Then, the output of Algorithm 1 is $\pi = \{\{x, y\}, \{w\}, \{z\}\}$ with $\mathcal{SW}(\pi) = 2 + 2\epsilon$ while $\pi^* = \{\{w, x\}, \{y, z\}\}$ is welfare-optimal with $\mathcal{SW}(\pi^*) = 4$.

In the special case of simple symmetric games, the output is welfare-optimal. The proof uses the characterization of Pareto-optimal partitions by Elkind et al. [9, Lemma 15] that have observed that all coalitions in such partitions are stars or cliques. The proof shows that some welfare-optimal partition is a clique matching and therefore, all maximum clique matchings are welfare-optimal. We omit the proof due to space restrictions.

Theorem 4.4. *Let a simple symmetric MFHG be given. Let π be a clique matching of the underlying unweighted graph. Then, π is* *welfare-optimal if and only if it is a maximum clique matching. In particular, the output of Algorithm 1 is welfare-optimal.*

As we have seen in the above example, Pareto-optimal partitions need not be welfare-optimal. However, this was shown by Elkind et al. [9] for simple, symmetric MFHGs induced by a bipartite graph, and we will extend their result to simple symmetric MFHGs. While their results rely on estimating the welfare of partitions in terms of minimum vertex covers, we will exploit a combinatorial description of `Max3CliqueMatching`. We will develop it using terminology closely related to the famous blossom algorithm by Edmonds [8] to show the close relationship of computing maximum cardinality matchings and maximum size 3-clique matchings.

The blossom algorithm deals with odd cycles by finding subgraphs called *flowers*. Let a graph $G = (V, E)$ together with a matching M be given. A path is called *M-alternating* if it alternately uses edges of M and outside M. An *M-augmenting* path is an M-alternating path starting and ending at vertices not covered by M. An *M-stem* is an M-alternating path of even length. The uncovered endpoint is its *root*, the covered endpoint is its *tip*. An *M-blossom* is an odd cycle $C = (B, E_B)$ of G such that all vertices except one are covered by $M \cap E_B$. Let $C = (B, E_B)$ be an M-blossom such that $b \in B$ is uncovered and let $b_1, b_2 \in B$ be the neighbors of b on C. B is called *M-chordal* if $\{b_1, b_2\} \in E$. An *M-(chordal) flower* is the union of a stem and a (chordal) blossom that intersect exactly in the tip of the stem. An example of a chordal flower together with an augmentation is given in Figure 1. If the matching is clear from the context, we will not specify it for the previous notation.

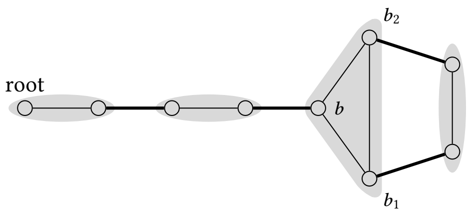


Figure 1: The bold matching indicates a chordal flower that can be augmented via the gray clique cover.

Theorem 4.5 (Hell and Kirkpatrick [13]). *Let $G = (V, E)$ be a graph. Then C is a maximum size 3-clique matching if and only if*

1. *There exists no $M(C)$-augmenting path.*
2. *There exists no $M(C)$-alternating path starting at an uncovered vertex and ending at a vertex covered by a triangle.*
3. *There exists no $M(C)$-chordal flower.*

In particular, a clique matching is maximum if and only if is vertex-maximal.

Note that compared to the classical characterization of maximum cardinality matchings, the second condition allows for improvement by deleting a triangle, and the third one by creating one in order to improve a 3-clique matching.

Theorem 4.6. *Let a simple, symmetric MFHG be given. Let π be a partition of the agents. Then, π is welfare-optimal if and only if it is Pareto-optimal.*

Proof sketch. Let an MFHG induced by a unweighted graph $G = (V, E)$ be given. Clearly, welfare-optimal partitions are Pareto-optimal.

For the reverse implication, let π be a Pareto-optimal partition which consists of cliques and stars only [9, Lemma 15]. By splitting larger cliques, we may assume that all cliques are of size 2 and 3 (splitting the cliques yields a partition with identical utilities). Let $\mathcal{S} \subseteq \pi$ be its star-coalitions. For any star S with $k_S \geq 2$ leaves, let c^S be its center and $l_1^S, \dots, l_{k_S}^S$ be its leaves. We define a new partition $\pi' = \{C \in \pi : C \notin \mathcal{S}\} \cup \{\{c^S, l_1^S\}, \{l_2^S\}, \dots, \{l_{k_S}^S\} : S \in \mathcal{S}\}$. Then, $\mathcal{SW}(\pi') = \mathcal{SW}(\pi)$. The core of the proof is that π' is a maximum clique matching and therefore by Theorem 4.4 welfare-optimal.

Define the set $M_{\mathcal{S}} = \{\{c^S, l_1^S\} : S \in \mathcal{S}\}$, i.e., the set of edges that are added to the partition π'. Then, $M(\pi') = M(\pi) \cup M_{\mathcal{S}}$ is the set of 2-cliques of the 3-clique matching π'.

We will prove that the conditions of the combinatorial characterization of Theorem 4.5 are satisfied. First, assume that there exists an $M(\pi')$-augmenting path P. An illustration of this step is given after the proof with the aid of Figure 2.

For $e \in M_{\mathcal{S}}$, we denote by S_e the star which the edge e originates from. An edge $e \in M_{\mathcal{S}} \cap P$ is called *exterior* if c^{S_e}, the center vertex of S_e, is the second or second-last vertex on the path. Otherwise, we call the edge *interior*.

For an exterior edge e, we denote by $t(e)$ the endpoint of P that is a neighbor of c^{S_e} on P. The first step is to modify P. In a second step, the modified path will yield a Pareto-improvement over π. An exterior edge e is called *saturated* if $t(e)$ is a leaf of S_e. An interior edge e is called *saturated* if all leaves of S_e are covered by P.

First, we may assume that every exterior edge is saturated or there exists only one exterior edge e^* which corresponds to a star S with two leaves and l_2^S is the endpoint of P that is not $t(e^*)$.

To this end, assume first that there exist two exterior edges e and f originating from stars S and T, respectively. Replacing $t(e)$ and $t(f)$ by l_2^S and l_2^T leaves both edges saturated. Otherwise, if e^* is the only exterior edge originating from star S, and is not saturated, then S has only two leaves, or we can replace $t(e^*)$ by a leaf of S uncovered by P. In addition, if l_2^S is not the other endpoint of P, we can replace $t(e^*)$ by l_2^S. This establishes the claim.

Second, we show that we can additionally assume that all interior edges are saturated. Indeed, if e is an interior edge and l is a leaf of S_e not covered by P, then we can replace P by the augmenting path that starts with l and proceeds on P with e. Assume that the path ends in an exterior edge f. If f is not saturated, we replace $t(f)$ by $l_2^{S_f}$. Otherwise, we follow the path to the end. In any case, this procedure yields a path P' such that all exterior edges are as after the first step and all interior edges are saturated.

We will show how to obtain a Pareto-improvement over π from P. Label the vertices of the path $p_0, \dots, p_m$ for some (odd) integer m. If the first matching-edge of the path is exterior and saturated, we delete p_0 and p_1 from the path. If the last matching-edge is exterior and saturated, we delete p_{m-1} and p_m from the path. This leaves a path P' on vertices $p'_0, \dots, p'_{m'}$ such that all leaves corresponding to stars of edges in $M_{\mathcal{S}} \cap P'$ are covered by P'. Let $\mathcal{T}$ be the set of star coalitions T such that $c^T \notin \{p'_0, \dots, p'_{m'}\}$, but some leaf of T is a endpoint of P'.

Consider $\pi' = \{C \in \pi \colon C \cap \{p_0, \ldots, p_m\} = \emptyset\} \cup \{T \backslash \{p'_0, p'_{m'}\} \colon T \in \mathcal{T}\} \cup (P' \backslash M(\pi'))$, which is a Pareto-improvement over π. Hence, π is not Pareto-optimal, which is a contradiction. Therefore, the first condition of the combinatorial characterization is satisfied.

An example for this case is given in Figure 2. The path P is formed by the straight lines and the partition π' by the bold edges. Dashed edges indicate leaves of the stars S and T. The edge $\{c^S, l_1^S\}$ is exterior and saturated, and the edge $\{c^T, l_1^T\}$ is interior and saturated. The gray partition yields a Pareto-improvement over π.

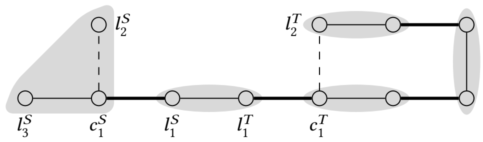


Figure 2: Example of a Pareto-improvement that can be constructed from an $M(\pi')$-augmenting path.

The proofs that the second and third condition hold are similar, and use that the first condition is already proved.

Hence, π' is a maximum clique matching and is therefore welfare-optimal. Since $\mathcal{SW}(\pi') = \mathcal{SW}(\pi)$, it follows that π is welfare-optimal. □

As a corollary, we obtain efficient verification of Pareto-optimality for simple symmetric MFHGs.

Theorem 4.7. *The problem of verifying Pareto-optimality can be done in polynomial time for simple symmetric MFHGs.*

Proof. Let an MFHG be given and a partition π that is to be checked for Pareto-optimality. Simply compute a partition π^* via Algorithm 1 and compare their social welfare. By Theorem 4.2 and Theorem 4.1, this runs in polynomial time. By Theorem 4.4, π^* is welfare-optimal. Finally, by Theorem 4.6, comparing social welfare checks for Pareto-optimality. □

For general, weighted MFHGs, it is still of interest as to whether one can also find a welfare-optimal partition in polynomial time. While this question remains open, we can at least narrow down the search to coalitions of small size. Then, a weighted version of `Max3CliqueMatching` might give rise to an efficient algorithm.

The proof of the proposition relies on the fact that we can split a coalition C of size at least 4 into an edge e and the remainder $C \backslash e$ such that $\mathcal{SW}(e) + \mathcal{SW}(C \backslash e) \geq \mathcal{SW}(C)$. The edge e maximizes a cleverly chosen objective function that relies on the weight of e, the weight of the cut between e and $C \backslash e$, and the welfare of $C \backslash e$.

Proposition 4.8. *Let a partition π of the agents of a general MFHG be given. Then, there exists a partition π' with $|C| \leq 3$ for all $C \in \pi'$ with $\mathcal{SW}(\pi') \geq \mathcal{SW}(\pi)$. In particular, there exists a welfare-optimal partition consisting of coalitions of size at most* 3.

5 ADDITIVELY SEPARABLE HEDONIC GAMES

In this section, we will survey Pareto-optimality on ASHGs. Positive results exist so far only for a very restrictive class that does not allow for 0-weights, and unique top-ranked coalitions are therefore guaranteed [3, Theorem 11]. We extend this result to a very general class of ASHGs that includes symmetric ASHGs.

An ASHG is called *mutually indifferent* if $v_i(j) = 0$ implies $v_j(i) = 0$ for every pair of agents i, j. Note that every symmetric ASHG is mutually indifferent.

Theorem 5.1. *A Pareto-optimal outcome for mutually indifferent ASHGs can be computed in polynomial time. In particular, a Pareto-optimal outcome for symmetric ASHGs can be computed in polynomial time.*

Proof. Consider Algorithm 3. The algorithm can be seen as a variant of serial dictatorship where in every coalition formed through a dictator d_i, the dictator asks the agents in her coalition to improve using a strict rank order such that none of higher rank becomes worse off.

Input: Mutually indifferent ASHG induced by $G = (N, E, v)$
Output: Pareto-optimal partition π

$\pi \leftarrow \emptyset, D \leftarrow N, i \leftarrow 1$
while $D \neq \emptyset$ **do**
  Pick $d_i \in D$
  $C_i \leftarrow \{d_i\} \cup \{j \in D \colon v_{d_i}(j) > 0\}, I_i \leftarrow \{j \in D \colon v_{d_i}(j) = 0\}$
  $H_i \leftarrow C_i \backslash \{d_i\}$
  while $H_i \neq \emptyset \wedge I_i \neq \emptyset$ **do**
    Pick $h \in H_i$
    $C_i \leftarrow C_i \cup \{j \in I_i \colon v_h(j) > 0\}$
    $H_i \leftarrow (H_i \cup \{j \in I_i \colon v_h(j) > 0\}) \backslash \{h\}$
    $I_i \leftarrow \{j \in I_i \colon v_h(j) = 0\}$
  $\pi \leftarrow \pi \cup \{C_i\}, D \leftarrow D \backslash C_i, i \leftarrow i + 1$
return π

Algorithm 3: Pareto-optimality for mutual indifference

The running time is polynomial since every edge in the graph underlying the ASHG is checked at most once.

For correctness, denote by $C_1, \ldots, C_k$ the coalitions that form in the order of the algorithm, and I_i the respective indifference sets (some may be empty) at the end of the inner while-loop. Note that for all $i \in 1, \ldots, k, a \in C_i, b \in I_i$ holds that $v_a(b) = 0$. For correctness, assume that π' is a partition such that for every agent $i \in N$, $v_i(\pi') \geq v_i(\pi)$. If $|C_i| > 1$, then $C_i \subseteq \pi'(d_i) \subseteq C_i \cup I_i$. Hence, no agent in such a coalition will be better off. In addition, no agent in a singleton coalition $C_i = \{d_i\} \in \pi$ will be better off, since they can only form coalitions with other singleton coalitions or with coalitions such that they are in the set I_j, both of which give them 0 utility. Therefore, no agent's utility has improved. □

Slight modifications of the above algorithm give computational tractability of Pareto-optimality even for more general classes of ASHGs. The same algorithm works for the class of ASHGs such that $v_i(j) = 0$ implies $v_j(i) \leq 0$ for every pair of agents i, j. Hence, the only edges remaining for the full domain of ASHGs are *critical* edges of the form $\{i, j\}$ such that $v_i(j) > 0$ while $v_j(i) = 0$. One idea towards obtaining an algorithm for a more general class of ASHGs is to use a pivoting rule that selects dictators. This allows, for example, for a positive result for the class of ASHGs such that the critical edges form a directed acyclic graph (using a topological order on the agents for a pivoting rule).

The outcome of the algorithm can, however, have an arbitrarily large gap to the maximum social welfare that is obtained in a welfare-optimal outcome. In addition, all but one agent may obtain a worst coalition. On the other hand, a Pareto-optimal and individually rational outcome does always exist, but computing such a partition is intractable. The following is a strengthening of a result by Aziz et al. [3] who dealt with the whole class of ASHGs and established weak NP-hardness.

The reduction is from the NP-complete problem `Exact3Cover` [14]. An instance (R, S) of `Exact3Cover` (X3C) consists of a ground set R together with a set S of 3-element subsets of R. A 'yes' instance is an instance so that there exists a subset $S' \subseteq S$ that partitions R.

Theorem 5.2. *Finding a Pareto-optimal and individually rational partition for symmetric ASHGs is (strongly) NP-hard, even if all weights are integers bounded from above by* 3.

Proof sketch. We provide a Turing reduction, illustrated in Figure 3, from X3C. Given an instance (R, S) of X3C, we construct the symmetric ASHG with agent set $N = R \cup V$ where $V = \{s_i : i = 1, \dots, 5, s \in S\}$ consists of 5 copies of agents for the sets in S. Preferences are given by weights v as

- $v(i, j) = 0$, $i, j \in R, i \neq j$
- $v(i, s_1) = 2$, $s \in S, i \in s$
- $v(s_1, s_2) = v(s_1, s_3) = v(s_2, s_4) = v(s_3, s_5) = v(s_4, s_5) = 3$, $v(s_2, s_3) = 0$, $s \in S$, and
- all other weights are set to -13.

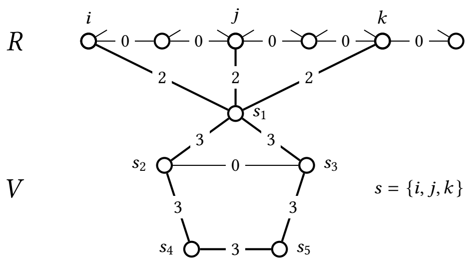


Figure 3: ASHG for the reduction. Indicated edges between R-agents have weight 0, other omitted edges have weight -13.

We will argue that if we can compute a Pareto-optimal and individually rational partition, we can decide X3C. Amongst all individually rational partitions, the highest utility that the agents can obtain is 2, 6, and 3 for agents in R, $\{s_1 : s \in S\}$, and $\{s_i : i = 2, \dots, 4, s \in S\}$, respectively. It can be shown that there exists an individually rational partition that attains these bounds for every agent if and only if there exists a 3-partition of R through sets in S.

Hence, we can solve X3C in polynomial time by computing a Pareto-optimal and individually rational partition for the corresponding ASHG, and check whether every agent receives the utility of a best partition amongst individually rational partitions. □

By applying a local search algorithm that starts with the singleton partition, we obtain the following corollary.

Corollary 5.3. *Finding a Pareto-improvement is NP-hard for symmetric ASHGs.*

Finally, it is interesting to see why the strong relation between Pareto-optimality and perfection exploited by Aziz et al. [2] does not hold for ASHGs. The preference refinement algorithm computes an individually rational, Pareto-optimal partition given an oracle that decides whether there exists a perfect partition and, in the case there exists one, can compute one. While the former problem is NP-hard, the latter can be solved in polynomial time by forming a top-ranked coalition, adding requisite agents, and adding outside agents that need an inside agent. This theorem even holds for the more general class of separable hedonic games [3, Theorem 9].

Theorem 5.4. *The problem of, given a general ASHG, computing a perfect partition or deciding that no such partition exists, can be done in polynomial time.*

6 FRACTIONAL HEDONIC GAMES

The serial dictatorship version used for ASHGs in Algorithm 3 implicitly exploits the fact that ASHGs have the property that top-ranked coalitions in subgames are the restrictions of top-ranked coalitions in the original game. This is not the case anymore for FHGs. However, the set of top-ranked coalitions can be described using the following observation.

Proposition 6.1. *Let an FHG be given based on a graph $G = (N, E, v)$ and let $d \in N$. Let C be a top-ranked coalition of d and set $\mu = v_d(C)$. Then, $C = \{a \in N \setminus \{d\} : v_d(a) > \mu\} \cup W$ for some $W \subseteq \{a \in N \setminus \{d\} : v_d(a) = \mu\}$.*

Hence, to obtain the top-ranked coalitions of an agent, one can order the other agents in decreasing value and add them until another agent is not beneficial. If there are agents that give exactly the utility of a top-ranked coalition, they may or may not be added.

We will now show efficient computability of Pareto-optimal partitions for certain classes of FHGs. An FHG satisfies the *equal affection condition* if $v_x(y), v_x(z) > 0$ implies $v_x(y) = v_x(z)$ for every triple of agents x, y, and z. An FHG is called *generic* if $v_x(y) \neq v_x(z)$ for every triple of agents x, y, and z, i.e., the utilities over the remaining set of agents are pairwise distinct for every agent.

Since the equal affection condition guarantees unique top-ranked coalitions for every agent, we obtain the following theorem, which applies in particular to simple FHGs.

Theorem 6.2. *Finding a Pareto-optimal partition for FHGs satisfying the equal affection condition can be done in polynomial time.*

Another variant of serial dictatorship finds Pareto-optimal partitions on generic FHGs.

Theorem 6.3. *Finding a Pareto-optimal partition for generic FHGs can be done in polynomial time.*

Proof sketch. Let an FHG be based on the graph $G = (N, E, v)$. We give an algorithm based on serial dictatorship that exploits a dynamically created hierarchy for the dictatorship. The next dictator is chosen based on the top choices of the previous dictator.

By Proposition 6.1, we know the structure of the top-ranked coalitions of an agent. Let an agent set $M \subseteq N$ be given, that induces the FHG on the subgraph $G[M]$, and let $d \in M$. There exists a unique smallest top-ranked coalition, which we denote by $T_d(M)$. Furthermore, for a generic FHG, there exist at most two top-ranked coalitions. Denote in this case by $t_d(M)$ the number of such coalitions for agent d in the subgame and if $t_d(M) = 2$, let $\alpha_d(M) \in M$ be the unique agent such that $T_d(M) \cup \{\alpha_d(M)\}$ is the other top-ranked coalition.

We are ready to formulate the recursive algorithm $\mathcal{A}$ that computes a Pareto-optimal partition by adding iteratively coalitions to a partial partition. The actual Pareto-optimal partition is obtained by choosing an arbitrary first dictator $d \in N$ and calling $\mathcal{A}(N, \emptyset, d)$.

Input: Non-empty agent set M, partial partition π, pivot agent d
Output: Partition π

if $t_d(M) = 1$ **then**
  if $T_d(M) = M$ **then**
    return $\pi \cup \{T_d(M)\}$
  else
    Pick $d_{\text{new}} \in M \backslash T_d(M)$
    Execute $\mathcal{A}(M \backslash T_d(M), \pi \cup \{T_d(M)\}, d_{\text{new}})$
else
  if $v_{\alpha_d(M)}(T_d(M) \cup \{\alpha_d(M)\}) > v_{\alpha_d(M)}(M \backslash T_d(M))$ **then**
    if $T_d(M) \cup \{\alpha_d(M)\} = M$ **then**
      return $\pi \cup \{T_d(M) \cup \{\alpha_d(M)\}\}$
    else
      Pick $d_{\text{new}} \in M \backslash (T_d(M) \cup \{\alpha_d(M)\})$
      Execute $\mathcal{A}(M \backslash (T_d(M) \cup \{\alpha_d(M)\}), \pi \cup \{T_d(M) \cup \{\alpha_d(M)\}\}, d_{\text{new}})$
  else if $v_{\alpha_d(M)}(T_d(M) \cup \{\alpha_d(M)\}) < v_{\alpha_d(M)}(M \backslash T_d(M))$ or $v_{\alpha_d(M)}(T_d(M) \cup \{\alpha_d(M)\}) = v_{\alpha_d(M)}(M \backslash T_d(M)) = 0$ **then**
    Execute $\mathcal{A}(M \backslash T_d(M), \pi \cup \{T_d(M)\}, \alpha_d(M))$
  else
    Pick $d_{\text{new}} \in \text{argmax}\{v_{\alpha_d(M)}(x) : x \in M \backslash T_d(M)\}$
    if $\alpha_d(M) \in T_{d_{\text{new}}}(M \backslash T_d(M))$ **then**
      Execute $\mathcal{A}(M \backslash T_d(M), \pi \cup \{T_d(M)\}, d_{\text{new}})$
    else
      Execute $\mathcal{A}(M \backslash (T_d(M) \cup \{\alpha_d(M)\}), \pi \cup \{T_d(M) \cup \{\alpha_d(M)\}\}, d_{\text{new}})$

Algorithm 4: Pareto-optimal partition for generic FHG by the recursive algorithm $\mathcal{A}$

By the top-ranked coalition structure of agents in generic FHGs, every step of Algorithm 4 can be executed and as argued after Proposition 6.1, top-ranked coalitions can be efficiently computed. Hence, the algorithm runs in polynomial time.

It can be checked that $\pi = \mathcal{A}(N, \emptyset, d)$ returns a Pareto-optimal partition, provided that the input evolves from a generic FHG. □

Finally, similar statements as for ASHGs hold for FHGs. The proofs are similar.

Theorem 6.4. *Finding a Pareto-optimal and individually rational partition for symmetric FHGs is NP-hard.*

Theorem 6.5. *Finding a Pareto-improvement is NP-hard for symmetric FHGs.*

Theorem 6.6. *The problem of, given an FHG, computing a perfect partition or deciding that no such partition exists, can be done in polynomial time.*

7 CONCLUSION

We have investigated Pareto-optimality in three types of cardinal hedonic games. The main findings and important related results are collected in Table 1. We can efficiently find Pareto-optimal partitions in symmetric MFHGs and AHSGs, and reasonable classes of FHGs including simple FHGs. The key insight for MFHGs is the equivalence with an extension of matchings to cliques. The combinatorial view of the problem allowed us to completely understand Pareto-optimal outcomes on simple, symmetric MFHGs, where they coincide with welfare-optimal outcomes. This motivates the study of the weighted case, where Pareto-optimal outcomes have no guarantee on the welfare, and yet our algorithm returns a 2-approximation to social welfare. The complexity of welfare-optimization in the weighted case is an interesting open problem. We are at least able to prove that coalitions of size 2 and 3 suffice, which is in line with the research on any other solution concept for MFHGs.

	Pareto optimality		*Welfare optimality*	
	PO	PO∧IR	Deterministic	Approximation
MFHG	P (sym, Thm. 4.2)	P (sym, Thm. 4.2)	P (0/1 sym, [15])	2 (sym, Thm. 4.3)
ASHG	P (sym, Thm. 5.1)	NP-h (sym, Thm. 5.2)	NP-h (sym, [3])	open
FHG	P (0/1 sym, Thm. 6.2)	NP-h (sym, Thm. 6.4)	NP-h (0/1 sym, [4])	4 (sym, [4])

Table 1: Complexity of Pareto- and welfare-optimality for cardinal hedonic games. Preference restrictions are given in parenthesis, where (0/1) sym denotes (simple) symmetric preferences. For welfare-optimality, the best known efficiently attainable approximation ratio is given.

The key technique for positive results on ASHGs and FHGs are refinements of serial dictatorship algorithms. Further enhancements, e.g., with respect to the order of selection of the dictators, might yield even better results. On the other hand, computing Pareto-optimal outcomes that satisfy further properties will often be intractable. Computational hardness is obtained if we require individual rationality in addition. Since it is even hard to compute Pareto-improvements, local search heuristics based on Pareto-optimality cannot be exploited.

Partitions on simple MFHGs can simultaneously satisfy high demands in terms of stability and optimality. Interesting further directions for research therefore concern weighted MFHGs as well as Pareto-optimality on the general domains of cardinal hedonic games.

ACKNOWLEDGMENTS

This material is based on work supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/12-1. The author would like to thank Felix Brandt, Katharina Moench, and Anaëlle Wilczynski for feedback on early versions of the paper and is indebted to three anonymous reviewers who have provided very valuable feedback.

REFERENCES

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

[2] H. Aziz, F. Brandt, and P. Harrenstein. 2013. Pareto Optimality in Coalition Formation. *Games and Economic Behavior* 82 (2013), 562–581.

[3] H. Aziz, F. Brandt, and H. G. Seedig. 2013. Computing Desirable Partitions in Additively Separable Hedonic Games. *Artificial Intelligence* 195 (2013), 316–334.

[4] H. Aziz, S. Gaspers, J. Gudmundsson, J. Mestre, and H. Täubig. 2015. Welfare Maximization in Fractional Hedonic Games. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*. AAAI Press, 461–467.

[5] H. Aziz and R. Savani. 2016. Hedonic Games. In *Handbook of Computational Social Choice*, F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia (Eds.). Cambridge University Press, Chapter 15.

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

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

[8] J. Edmonds. 1965. Paths, Trees and Flowers. *Canadian Journal of Mathematics* 17 (1965), 449–467.

[9] E. Elkind, A. Fanelli, and M. Flammini. 2016. Price of Pareto optimality in hedonic games. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Conference on Artificial Intelligence (AAAI), 475–481.

[10] M. Gairing and R. Savani. 2010. Computing stable outcomes in hedonic games. In *Proceedings of the 3rd International Symposium on Algorithmic Game Theory (SAGT) (Lecture Notes in Computer Science)*, Vol. 6386. Springer-Verlag, 174–185.

[11] M. Gairing and R. Savani. 2011. Computing Stable Outcomes in Hedonic Games with Voting-Based Deviations. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. IFAAMAS, 559–566.

[12] J. Hajduková. 2006. Coalition formation games: A survey. *International Game Theory Review* 8, 4 (2006), 613–641.

[13] P. Hell and D. G. Kirkpatrick. 1984. Packings by cliques and by finite families of graphs. *Discrete Mathematics* 49, 1 (1984), 45–59.

[14] R. M. Karp. 1972. Reducibility among Combinatorial Problems. In *Complexity of Computer Computations*, R. E. Miller and J. W. Thatcher (Eds.). Plenum Press, 85–103.

[15] G. Monaco, L. Moscardelli, and Y. Velaj. 2018. Stable Outcomes in Modified Fractional Hedonic Games. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. IFAAMAS, 937–945.

[16] M. E. J. Newman. 2004. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter and Complex Systems* 38, 2 (2004), 321–330.

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

[18] J. von Neumann and O. Morgenstern. 1947. *Theory of Games and Economic Behavior* (2nd ed.). Princeton University Press.

7

CORE PUBLICATION 2: FINDING AND RECOGNIZING POPULAR COALITION STRUCTURES

SUMMARY

An important aspect of multi-agent systems concerns the formation of coalitions that are stable or optimal in some well-defined way. The notion of popularity has recently received a lot of attention in this context. A partition is popular if there is no other partition in which more agents are better off than worse off. A stronger notion of popularity is strong popularity which demands that, in every other partition, there are more agents worse off than better off. In addition, there exists a randomized version of popularity called mixed popularity. Mixed popularity is particularly attractive because existence is guaranteed by the Minimax Theorem, while there exist simple coalition formation games in which popular or strongly popular partitions need not exist.

In this paper, we study all three notions of popularity in a variety of coalition formation settings. Extending previous work on marriage games, we show that mixed popular partitions in roommate games can be found efficiently via linear programming and a separation oracle. This approach is quite universal, leading to efficient algorithms for verifying whether a given partition is popular and for finding strongly popular partitions (resolving an open problem).

By contrast, we prove that both problems become computationally intractable when moving from coalitions of size 2 to coalitions of size 3, even when preferences are strict and globally ranked. Moreover, we show that finding and recognizing popular, strongly popular, and mixed popular partitions in symmetric additively separable hedonic games and symmetric fractional hedonic games are hard.

Together, our results indicate strong boundaries to the tractability of popularity in both ordinal and cardinal models of hedonic games.

It is worth mentioning that, based on the observation that the verification problem is hard, it is not straightforward to analyze the problem of computing popular partitions. To facilitate our hardness results, we develop a reduction technique that reduces the complex decision of verifying popular partitions to the consideration of very few agents. The key idea of this technique is to let decisions about popularity "propagate" through a structure similar to a binary tree such that all important information can be accessed via the root of the tree.

REFERENCE

F. Brandt and M. Bullinger. **Finding and recognizing popular coalition structures**. *Journal of Artificial Intelligence Research*, 74:569–626, 2022.
DOI: https://doi.org/10.1613/jair.1.13470

INDIVIDUAL CONTRIBUTION

I, Martin Bullinger, am the main author of this publication. In particular, I am responsible for the joint development and conceptual design of the research project, proofs and write-up of all results, and the joint write-up of the remaining manuscript.

COPYRIGHT AGREEMENT

The right to present this paper in a doctoral thesis has been granted by the publisher, AI Access Foundation, in the copyright form presented below. There, it is stated that "*The author retains the right to use this material in future works of his or her own authorship.*" This form can also be found at http://jair.org/public/resources/copyright.pdf (accessed November 10, 2022).

AI Access Foundation

Transfer of Copyright

Title of Contribution:

Author:

The undersigned (Author), desiring to publish the article named above, in the Journal of Artificial Intelligence Research, hereby grants and assigns exclusively to the AI Access Foundation (Publisher) all rights of copyright in this article, and the exclusive right to copy and distribute the article throughout the world, and the authority to exercise or to dispose of all subsidiary rights in all countries and in all languages.

The author retains the right to use this material in future works of his or her own authorship. The author retains the right to post an electronic copy of the article on his or her web site, on web sites operated by the author's institution, and on scholarly repositories, provided there is no commercial purpose involved, and with appropriate bibliographic citation to the JAIR article.

The Author agrees that the contribution does not contain any material created by others, or from other copyrighted works, unless the written consent of the owner of such material is attached.

The undersigned warrants that s/he has full power to make this agreement, and has not previously granted, assigned, or encumbered any of the rights granted and assigned herein; that the material submitted for publication is the work of the author, and is original, unpublished, and not previously the subject of any application for copyright registration (except material for which the Publisher has written grants of permission to include, as described above).

Author	Date

Journal of Artificial Intelligence Research 74 (2022) 569-626 Submitted 11/2021; published 06/2022

Finding and Recognizing Popular Coalition Structures

Felix Brandt BRANDTF@IN.TUM.DE
Martin Bullinger BULLINGE@IN.TUM.DE
Technical University of Munich,
Boltzmannstr 3, 85748 Munich, Germany

Abstract

An important aspect of multi-agent systems concerns the formation of coalitions that are stable or optimal in some well-defined way. The notion of *popularity* has recently received a lot of attention in this context. A partition is popular if there is no other partition in which more agents are better off than worse off. In this paper, we study popularity, strong popularity, and mixed popularity (which is particularly attractive because existence is guaranteed by the Minimax Theorem) in a variety of coalition formation settings. Extending previous work on marriage games, we show that mixed popular partitions in roommate games can be found efficiently via linear programming and a separation oracle. This approach is quite universal, leading to efficient algorithms for verifying whether a given partition is popular and for finding strongly popular partitions (resolving an open problem). By contrast, we prove that both problems become computationally intractable when moving from coalitions of size 2 to coalitions of size 3, even when preferences are strict and globally ranked. Moreover, we show that finding popular, strongly popular, and mixed popular partitions in symmetric additively separable hedonic games and symmetric fractional hedonic games is NP-hard. Together, these results indicate strong boundaries to the tractability of popularity in both ordinal and cardinal models of hedonic games.

1. Introduction

Coalitions and coalition formation have been a central concern of game theory, ever since the publication of von Neumann and Morgenstern's *Theory of Games and Economic Behavior* in 1944. The traditional models of coalitional game theory, in particular TU (transferable utility) and NTU (non-transferable utility) coalitional games, involve a formal specification of what each group of agents can achieve on their own. Drèze and Greenberg (1980) noted that in many situations this is not feasible, possible, or even relevant to the coalition formation process, as, e.g., in the formation of social clubs, teams, or societies. Instead, in coalition formation games, the agents' preferences are defined directly over the coalition structures, i.e., partitions of the agents in disjoint coalitions. Formally, coalition formation can thus be considered as a special case of the general social choice setting, where the agents entertain preferences over a special type of alternatives, namely coalition partitions of themselves, from which one or more need to be selected. In most situations it is natural to assume that an agent's appreciation of a partition only depends on the coalition he is a member of and not on how the remaining agents are grouped. Popularized by Bogomolnaia and Jackson (2002), much of the work on coalition formation now concentrates on these so-called *hedonic games*.

The main focus in hedonic games has been on finding and recognizing partitions that satisfy various notions of *stability*—such as Nash stability, individual stability, or core stability—or optimality—such as Pareto optimality, utilitarian welfare maximality, or egalitarian welfare maximality (see Aziz & Savani, 2016, for an overview). In this paper, we focus on the notion of *popularity* (Gärdenfors, 1975), which has the flavor of both stability and optimality. A partition is popular if there is no other partition that is preferred by a majority of the agents. Moreover, a partition is strongly popular if it is preferred to every other partition by some majority of agents. Popularity thus corresponds to the notion of weak and strong Condorcet winners in social choice theory, i.e., candidates that are at least as good as any other candidate in pairwise majority comparisons. Just like stability notions, popularity is based on the idea that a subset of agents breaks off in order to increase their well-being. However, since the new partition has to make at least as many agents better off than worse off, popularity also has the flavor of optimality. According to the standard reference *Algorithmics of Matching Under Preferences*, "popular matchings [...] have been an exciting area of research in the last few years" (Manlove, 2013, p. 333). A recent survey on popular matchings is provided by Cseh (2017).

In contrast to Pareto optimal partitions, popular partitions are not guaranteed to exist. We therefore also consider *mixed* popular partitions, as proposed by Kavitha, Mestre, and Nasre (2011) and whose existence follows from the Minimax Theorem. A mixed popular partition is a probability distribution over partitions p such that there is no other mixed partition q such that the expected number of agents who prefer the partition returned by p to q is at least as large as the other way round. Mixed popular partitions are a special case of *maximal lotteries*, a randomized voting rule that has recently gathered increased attention in social choice theory (Fishburn, 1984; Brandl, Brandt, & Seedig, 2016; Brandl & Brandt, 2020; Brandl, Brandt, & Stricker, 2022).

We study the computational complexity of popular, strongly popular, and mixed popular partitions in a variety of hedonic coalition formation settings including additively separable hedonic games, fractional hedonic games as well as hedonic games where the coalition size is bounded. The latter includes flatmate games (which only allow coalitions of up to three agents) and roommate games (which only allow coalitions of up to two agents). Our main findings can also be found in Tables 1 and 2 in the conclusion and are summarized as follows.

- Generalizing earlier results by Kavitha et al. (2011), we show how mixed popular partitions in roommate games can be computed in polynomial time via linear programming and a separation oracle on a subpolytope of the matching polytope for non-bipartite graphs.[1] This stands in contrast to a recent result showing that computing popular partitions in roommate games is NP-hard (Faenza, Kavitha, Power, & Zhang, 2019; Gupta, Misra, Saurabh, & Zehavi, 2019).

- As corollaries we obtain that verifying popular partitions (Biró, Irving, & Manlove, 2010), finding Pareto optimal partitions (Aziz, Brandt, & Harrenstein, 2013a), and finding strongly popular partitions can all be done in polynomial time in roommate games, even when preferences admit ties. The latter statement resolves an acknowledged open problem.[2]

1. The results by Kavitha et al. (2011) only hold for house allocation and marriage markets and require extra work to be extended to roommate markets. See Section 2 for more details.
2. See, for example, Biró et al. (2010) and Manlove (2013): "A third open problem is the complexity of finding a strongly popular matching (or reporting that none exists), for an instance of RPT [Roommate

- We provide the first negative computational results for mixed popular partitions and strongly popular partitions by showing that finding these partitions in flatmate games is NP-hard. Moreover, it turns out, that verifying whether a given partition is popular, strongly popular, or mixed popular in flatmate games is coNP-complete. All of these results hold for strict and globally ranked preferences, where coalitions appear in the same order in each individual preference ranking. This is interesting insofar as finding popular partitions in roommate games becomes tractable under the same restrictions.
- We prove that computing popular, strongly popular, and mixed popular partitions is NP-hard in symmetric additively separable hedonic games and symmetric fractional hedonic games. Furthermore, we show coNP-completeness of all corresponding verification problems.
- Many of our hardness reductions follow a general scheme that might be of broader interest beyond the scope of popularity. Specifically, we merely embed the combinatorial structure of an NP-hard problem (in our case the incidence structure of a covering instance) into the leaves of an object similar to a binary tree. Within this tree, we can propagate all relevant information for the property under consideration (e.g., popularity) to the root agent of the tree, which then acts as a decision taker in our reduction. Hence, checking exponentially many partitions relevant to popularity reduces to checking one specific agent.

2. Related Work

Gärdenfors (1975) first proposed the notions of popularity and strong popularity in the context of marriage games. He showed that popular matchings (or “majority assignments” in his terminology) need not exist when preferences are weak, but that existence is guaranteed for strict preferences because every stable matching is popular. As a consequence, the well-known Gale-Shapley algorithm efficiently identifies popular matchings in marriage games with strict preferences. Kavitha and Nasre (2009), Huang and Kavitha (2011), and Kavitha (2014) provide efficient algorithms for computing popular matchings that satisfy additional properties such as rank maximality or maximum cardinality. For weak preferences, computing popular matchings is NP-hard, even when all agents belonging to one side have strict preferences (Biró et al., 2010; Cseh, Huang, & Kavitha, 2015).

In the more restricted setting of house allocation (henceforth housing games), Abraham, Irving, Kavitha, and Mehlhorn (2007) proposed efficient algorithms for finding popular allocations of maximum cardinality for both weak and strict preferences. Mahdian (2006) proved an interesting threshold for the existence of popular allocations: if there are n agents and the number of houses exceeds αn with $\alpha \approx 1.42$, then the probability that there is a popular allocation converges to 1 as n goes to infinity.

For roommate games with weak preferences, NP-hardness of computing popular matchings follows from the above-mentioned hardness results for marriage games. It was recently

Problem with Ties]” (Biró et al., 2010, p. 107); “Our last open problem concerns the complexity of the problem of finding a strongly popular matching, or reporting that none exists, given an instance of SRTI [Stable Roommates with Ties and Incomplete lists], which is unknown at the time of writing” (Manlove, 2013, p. 380).

shown that this problem is still NP-hard when preferences are strict (Gupta et al., 2019; Faenza et al., 2019; Cseh & Kavitha, 2018). Also, finding a maximum-cardinality popular matching in instances where popular matchings are guaranteed to exist is NP-hard (Brandl & Kavitha, 2018).

There are less results on *strongly* popular matchings. It is known from Gärdenfors (1975) that a strongly popular matching has to be a unique popular matching and that every strongly popular matching is stable in roommate and marriage games. Based on these insights, Biró et al. (2010) showed that strongly popular matchings in roommate games and marriage games with strict preferences can be found efficiently by first computing an arbitrary stable matching and then checking whether it is strongly popular. The case of weak preferences was left open and little progress has been made since then. Király and Mészáros-Karkus (2017) recently gave an algorithm for finding strongly popular matchings in marriage games where preferences are strict, except that agents belonging to one side may be completely indifferent. In housing games, a matching is strongly popular if and only if it is a unique perfect matching. Hence, strongly popular matchings in housing games can be found in polynomial time. All of the above mentioned results on strong popularity, including the open problem, follow from our Corollary 3.

Mixed popular matchings were introduced by Kavitha et al. (2011) who also showed how to compute a *fractional* popular matching in housing games and marriage games, which can then be translated into a mixed popular matching via a Birkhoff-von Neumann decomposition. This is possible in bipartite settings because every fractional matching is implementable as a probability distribution over deterministic matchings. When moving from marriage markets to roommate markets, this does not hold anymore. For example, a matching involving three agents where every pair of agents is matched with probability 1/2 is not implementable. Huang and Kavitha (2017) have shown that in marriage games with strict preferences, the popular matching polytope is half-integral and that half-integral mixed popular matchings can be computed in polynomial time. No such matchings are guaranteed to exist when preferences are weak. They also apply the same techniques to roommate games in order to compute an optimal half-integral solution over the bipartite matching polytope in the case of strict preferences. However, the resulting solutions may again fail to be implementable. Apart from that, their methods heavily rely on computing stable matchings, which may be intractable when preferences are weak. By contrast, our results in Section 4.2.1 are based on the matching polytope for non-bipartite graphs via odd-set constraints and allow both to deal with ties and to efficiently compute a solution that is implementable using LP methods (Proposition 5).[3] The axiomatic properties of mixed popular matchings such as efficiency and strategyproofness were investigated by Aziz, Brandt, and Stursberg (2013c), Brandt, Hofbauer, and Suderland (2017), and Brandl, Brandt, and Hofbauer (2017).

To the best of our knowledge, popularity, strong popularity, and mixed popularity have not been studied for coalition formation settings that go beyond coalitions of size 2 except for

3. The journal version of the paper by Huang and Kavitha (2017), which appeared after the conference version of our paper, also independently considers the non-bipartite matching polytope and briefly outlines how to compute mixed popular matchings (Huang & Kavitha, 2021). However, some important subtleties such as how to retain deterministic matchings from the fractional solution (our Proposition 5) are not considered.

a theorem by Aziz, Brandt, and Seedig (2013b, Th. 15) who claimed that checking whether a partition is popular in ASHGs is NP-hard and that verifying whether a partition is popular is coNP-complete. However, the proof of the first statement is incorrect.[4] We substantially modified the reduction to prove a stronger statement and independently proved a stronger statement for the verification problem.

3. Preliminaries

Let N be a finite set of agents. A *coalition* is a non-empty subset of N. By $\mathcal{N}_i$ we denote the set of coalitions agent i belongs to, i.e., $\mathcal{N}_i = \{S \subseteq N : i \in S\}$. A *coalition structure*, or simply a *partition*, is a partition π of the agents N into coalitions, where $\pi(i)$ is the coalition agent i belongs to. A *mixed partition* is a set $p = \{(\pi_1, p_1), \dots, (\pi_k, p_k)\}$, where π_i s a partition for every $i \in \{1, \dots, k\}$, and $(p_1, \dots, p_k)$ represents a probability distribution. A mixed partition is interpreted as a randomization over partitions.

A *hedonic game* is a pair $(N, \succsim)$, where $\succsim = (\succsim_i)_{i \in N}$ is a *preference profile* specifying the preferences of each agent i as a complete and transitive *preference relation* $\succsim_i$ over $\mathcal{N}_i$. If $\succsim_i$ is also anti-symmetric we say that i's preferences are *strict*. Otherwise, we say that preferences are *weak*. We denote by $S \succ_i T$ if $S \succsim_i T$ but not $T \succsim_i S$—i.e., *i strictly prefers* S to T—and by $S \sim_i T$ if both $S \succsim_i T$ and $T \succsim_i S$—i.e., i is *indifferent* between S and T. In hedonic games, agents are only concerned about their own coalition. Accordingly, preferences over coalitions naturally extend to preferences over partitions as follows: $\pi \succsim_i \pi'$ if and only if $\pi(i) \succsim_i \pi'(i)$.

Sometimes, we consider strict preferences, which are obtained from weak preferences by breaking ties arbitrarily. To express such preferences succinctly, given a set X of alternatives, we denote by $X^{\succ}$ an arbitrary, but fixed strict preference order of the alternatives in X. For example, $a \succ \{b, c\}^{\succ} \succ d$ could be replaced by $a \succ b \succ c \succ d$. For simplicity, one can assume that ties are broken lexicographically. When referring to index sets, such as sets of players, we use the shorthand $[k]$ for $\{1, \dots, k\}$ and $[k, l]$ for $\{k, \dots, l\}$.

Two basic properties of partitions are Pareto optimality and individual rationality. Given a hedonic game $(N, \succsim)$, a partition π is *Pareto optimal* if there is no partition π' such that $\pi' \succsim_j \pi$ for all agents j and $\pi' \succ_i \pi$ for at least one agent i. A coalition $S \in \mathcal{N}_i$ is *individually rational* for agent i if she prefers the coalition to staying alone, i.e., $C \succsim_i \{i\}$. A Partition π is *individually rational* if $\pi(i) \succsim_i \{i\}$ for all $i \in N$. The rationale behind individual rationality is that agents cannot be forced into a coalition.

Individual rationality is also the crucial ingredient of a succinct representation of hedonic games where only the preferences over individually rational coalitions are considered (Ballester, 2004). A hedonic game $(N, \succsim)$ is represented by *Individually Rational Lists of Coalitions* (IRLC) via the game $(N, \succsim')$ where $\succsim'$ is a preference profile such that $\succsim'_i$ is the restriction of $\succsim_i$ to individually rational coalitions in $\mathcal{N}_i$. In this case, $(N, \succsim)$ is called a *completion* of $(N, \succsim')$. This representation of games is useful to obtain meaningful hardness results because the size of the naive representation of a hedonic game is exponential in the

4. The reduction fails because for a Yes-instance of Exact 3-Cover, the partition π claimed to be popular for the ASHG it maps to is not popular: the partition $\pi' = \{\{y^s, z_1^s, z_2^s\} : s \in S\} \cup \{\{b_1^r, a_2^r\} : r \in R\} \cup \{\{b_2^r, a_1^r, a_3^r\} : r \in R\}$ is more popular.

number of agents while the IRLC representation may only require polynomial space if the number of individually rational coalitions is small enough.

In order to define popularity and strong popularity, let $N(\pi, \pi')$ be the set of agents who prefer π over π', i.e., $N(\pi, \pi') = \{i \in N \colon \pi(i) \succ_i \pi'(i)\}$, where π, π' are two partitions of N. For any subset $M \subseteq N$ of agents and partitions π, π' of N, $\phi_M(\pi, \pi') = |N(\pi, \pi') \cap M| - |N(\pi', \pi) \cap M|$ is called the *popularity margin on M with respect to π and π'*. If $M = \{i\}$ is a singleton set, we use the shorthand notation ϕ_i instead of $\phi_{\{i\}}$. On top of that, we define the *popularity margin* of π and π' as $\phi(\pi, \pi') = \phi_N(\pi, \pi')$. Then, π is called *more popular* than π' if $\phi(\pi, \pi') > 0$. Furthermore, π is called *popular* if, for all partitions π', $\phi(\pi, \pi') \geq 0$, i.e., no partition is more popular than π. Also, π is called *strongly popular* if, for all partitions $\pi' \neq \pi$, $\phi(\pi, \pi') > 0$, i.e., π is more popular than every other partition. Note that there can be at most one strongly popular partition in any hedonic game.

For a hedonic game $(N, \succsim)$ in IRLC representation, a partition π is called *popular* if it is popular in the completion of $(N, \succsim)$ where, for each agent, all coalitions that are not individually rational are gathered in a single indifference class that is less preferred than the singleton coalition. This definition of popularity generalizes the definition of popularity that is used for marriage games by Kavitha et al. (2011), and adds the appropriate perspective on individual rationality.[5] Note that a popular partition need not be individually rational.

Many hedonic games do not admit a popular partition. However, existence can be guaranteed by introducing randomization via mixed partitions, i.e., probability distributions over partitions. Let therefore two mixed partitions $p = \{(\pi_1, p_1), \dots, (\pi_k, p_k)\}$ and $q = \{(\sigma_1, q_1), \dots, (\sigma_l, q_l)\}$ be given, where $(p_1, \dots, p_k), (q_1, \dots . q_l)$ are probability distributions. We define the popularity margin of p and q as their expected popularity margin, i.e.,

$$\phi(p, q) = \sum_{i=1}^{k} \sum_{j=1}^{l} p_i q_j \phi(\pi_i, \sigma_j).$$

Clearly, the definition of popularity carries over to the extension of ϕ. As first observed by Kavitha et al. (2011), mixed popular partitions always exist, because they can be interpreted as maximin strategies of a symmetric zero-sum game (see also Fishburn, 1984; Aziz et al., 2013c).

Proposition 1. *Every hedonic game admits a mixed popular partition.*

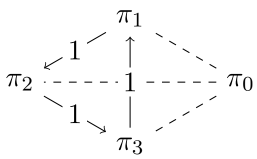
Proof. Every hedonic game can be viewed as a finite two-player symmetric zero-sum game where the rows and columns of the two players are indexed by all possible partitions $\pi_1, \dots, \pi_{B_{|N|}}$ and the entry at position (i, j) of the game matrix is $\phi(\pi_i, \pi_j)$. There, $B_{|N|}$ denotes the Bell number. By the Minimax Theorem (von Neumann, 1928), the value of this game is 0 and therefore, any maximin strategy, whose existence is guaranteed, is popular. $\square$

5. The IRLC representation ignores preferences over coalitions that are not individually rational. However, in contrast to core stability or Nash stability, these preferences can affect whether a partition is popular or not. In order to circumvent this problem, one could strengthen the definition of popularity by requiring that a coalition needs to be popular for *all* extensions of the IRLC represented preferences. All our results also hold for this notion, because we construct individually rational partitions for which the two notions of popularity coincide.

Before stating and proving our results, we illustrate the most important concepts by means of an example.

Example 1. Consider a hedonic game $(N, \succsim)$ with $N = \{a, b, c, d\}$ where preferences are given in IRLC representation:

- $N \succ_a \{a,b\} \succ_a \{a,c\} \succ_a \{a,d\} \succ_a \{a\}$
- $N \succ_b \{b,c\} \succ_b \{b,a\} \succ_b \{b,d\} \succ_b \{b\}$
- $\{c,a\} \succ_c \{c,b\} \succ_c \{c,d\} \succ_c N \succ_c \{c\}$
- $\{d,a\} \sim_d \{d,b\} \sim_d \{d,c\} \succ_d N \succ_d \{d\}$



Then, the Pareto optimal partitions (which are the only ones relevant for popularity) are $\pi_0 = \{N\}$, $\pi_1 = \{\{a,b\},\{c,d\}\}$, $\pi_2 = \{\{a,c\},\{b,d\}\}$, and $\pi_3 = \{\{a,d\},\{b,c\}\}$. Their popularity margins are depicted right of the preferences, where a dashed line denotes indifference with respect to popularity. In particular, π_0 is the only (deterministic) popular partition and there is no strongly popular partition. Further, the mixed partition $p = \{(\pi_1, 1/3), (\pi_2, 1/3), (\pi_3, 1/3)\}$ is mixed popular. It holds that $\phi(p, \pi_0) = \phi(p, \pi_1) = \phi(p, \pi_2) = \phi(p, \pi_3) = 0$.

4. Results

Our results are divided into three subsections. We first show some basic properties and relationships between the different notions of popularity. Then, we analyze popularity in ordinal hedonic games (such as flatmate and roommate games) and cardinal hedonic games (such as additively separable and fractional hedonic games), respectively.

4.1 Basic Relationships

Clearly, a strongly popular partition is also popular and a popular partition, interpreted as a probability distribution with singleton support, is mixed popular. Furthermore, every coalition structure in the support of a mixed popular partition is Pareto optimal. This already follows from a more general statement by Fishburn (1984, Prop. 3). We give a simple proof for completeness.

Proposition 2. *Let $p = \{(\pi_1, p_1), \dots, (\pi_k, p_k)\}$ be a mixed popular partition. Then, for every $i \in [k]$ with $p_i > 0$, π_i is Pareto optimal.*

Proof. Let $p = \{(\pi_1, p_1), \dots, (\pi_k, p_k)\}$ be a mixed popular partition and fix $i \in [k]$ such that $p_i > 0$. Assume for contradiction that π_i' is a Pareto improvement over π_i. Define $p' = \{(\pi_1, p_1), \dots, (\pi_{i-1}, p_{i-1}), (\pi_i', p_i), (\pi_{i+1}, p_{i+1}), \dots, (\pi_k, p_k)\}$. Note that $\phi(\pi_i', p) = \sum_{j=1, j\neq i}^k p_j \phi(\pi_i', \pi_j) + p_i \phi(\pi_i', \pi_i) \geq \sum_{j=1, j\neq i}^k p_j \phi(\pi_i, \pi_j) + p_i \phi(\pi_i', \pi_i) > \sum_{j=1, j\neq i}^k p_j \phi(\pi_i, \pi_j) + p_i \phi(\pi_i, \pi_i) = \phi(\pi_i, p)$.

Then, $\phi(p', p) = \sum_{j=1, j\neq i}^k p_j \phi(\pi_j, p) + p_i \phi(\pi_i', p) > \sum_{j=1, j\neq i}^k p_j \phi(\pi_j, p) + p_i \phi(\pi_i, p) = \phi(p, p) = 0$.

Hence, p is not mixed popular, a contradiction. $\square$

We thus have the following relationships between strong popularity (sPop), popularity (Pop), partitions in the support of any mixed popular partition (supp(mPop)), and Pareto optimality (PO):

$$\text{sPop} \implies \text{Pop} \implies \textbf{supp(mPop)} \implies \textbf{PO}.$$

The concepts printed in boldface are guaranteed to exist. As a consequence, hardness results for computing Pareto optimal partitions imply hardness of computing mixed popular partitions (though not for popular partitions since they need not exist). Mixed popular partitions also satisfy probabilistic strengthenings of Pareto optimality based on stochastic dominance and pairwise comparisons (Aziz, Brandl, Brandt, & Brill, 2018).

The existence problems for popular and strongly popular partitions are naturally contained in the complexity class Σ_2^p. The verification problems are contained in coNP. The following relationship turns out to be helpful for deducing the complexity of verifying mixed popular partitions from the respective result for popular partitions.

Proposition 3. *Let a class of hedonic games be given such that the verification problem of popular partitions is coNP-hard. Then, the verification problem of mixed popular partitions is coNP-hard.*

Proof. Let $\mathcal{C}$ be a class of hedonic games and let (G, π) be an instance of the deterministic verification problem, i.e. $G \in \mathcal{C}$ is a hedonic game and π a partition of the agents of G. By linearity of $\pi' \mapsto \phi(\pi, \pi')$, π is popular if, and only if, it is mixed popular. Hence, the embedding of the deterministic into the mixed case gives the desired reduction for coNP-hardness. □

Hence, whenever hardness results are obtained for the verification of popularity, they transfer automatically to mixed popularity. Conversely, polynomial-time algorithms for mixed popularity can be used to efficiently verify whether a partition is popular.

Also, since partitions have polynomial size (with respect to the number of agents), we can use more popular partitions as polynomial-size certificates to No-instances of the verification problem. This shows membership in coNP in the deterministic case and can also be applied for mixed popularity. Indeed, whenever there exists a more popular mixed coalition, then there exists also a more popular deterministic one. If p is a mixed partition for a game G and $p' = \{(\pi'_1, p'_1), \dots, (\pi'_k, p'_k)\}$ is more popular, then $0 < \phi(p', p) = \sum_{i=1}^{k} p'_i \phi(\pi'_i, p)$. Consequently, for some $i \in [k]$, $\phi(\pi'_i, p) > 0$.

Popular partitions are not only Pareto optimal, but it also suffices to compare a partition against Pareto optimal partitions when checking for popularity. This is useful when proving popularity of a given partition, for example in hardness reductions.

Proposition 4. *A partition π is popular if and only if, for all Pareto optimal partitions π', $\phi(\pi, \pi') \geq 0$. In addition, π is strongly popular if and only if, for all Pareto optimal partitions $\pi' \neq \pi$, $\phi(\pi, \pi') > 0$.*

Proof. We show that the respective popularity margin with Pareto optimal partition determine popularity.

This follows from the fact that for every two partitions π, $\hat{\pi}$, and a Pareto optimal Pareto improvement π' of $\hat{\pi}$, it holds that $\phi(\pi, \hat{\pi}) \geq \phi(\pi, \pi')$. If we investigate strong popularity, it can happen that $\pi' = \pi$, but in this case $\phi(\pi, \hat{\pi}) > 0$ by Pareto dominance. □

4.2 Ordinal Hedonic Games

In this section we investigate hedonic games in IRLC representation. Important subclasses of these games are defined by restricting the size of individually rational coalitions using a global constant. We thus obtain *flatmate games* as games in which only coalitions of up to three agents are individually rational and *roommate games* as games in which only coalitions of size 2 are individually rational. More restrictions are obtained by partitioning the set of agents into two groups, say, into males and females, and even further by additionally demanding that one group of agents is completely indifferent, say, by assuming that they are objects such as houses. A *marriage game* is a roommate game where the agents can be partitioned in two sets such that the only individually rational partitions are formed with agents from the other set. A *housing game* is a marriage game where all agents belonging to one set of the partition are completely indifferent. In roommate games (and their subclasses), partitions are referred to as *matchings*. All of these classes permit IRLC representations with size bounded polynomially with respect to the number of the agents. We have the following inclusion relationships.[6]

$$\text{Housing} \subsetneq \text{Marriage} \subsetneq \text{Roommates} \subsetneq \text{Flatmates} \subsetneq \text{IRLC}.$$

Finally we consider a severe preference restriction in coalition formation. A preference profile admits *globally ranked preferences* if there exists a common (global) ranking $\succsim$ of all coalitions in $2^N \setminus \{\emptyset\}$ and each individual preference relation $\succsim_i$ is the restriction of $\succsim$ to $\mathcal{N}_i$.

Under globally ranked preferences, the intractability of computing popular matchings in roommates games with strict preferences (Gupta et al., 2019; Faenza et al., 2019; Cseh & Kavitha, 2018) breaks down. In fact, it is known that under these preferences, every roommate game admits a stable matching, which can furthermore be efficiently computed (Abraham, Leravi, Manlove, & O'Malley, 2008). Since every stable matching also happens to be popular for strict preferences (see Section 2), this implies that computing popular matchings in roommates games becomes tractable. By contrast, all hardness results for flatmate games that will be shown in Section 4.2.2 hold even when preferences are globally ranked. This confirms the robustness of these results and underlines the crucial difference between settings with coalitions of size 2 and coalitions of size 3.

In our reductions, we consider hedonic games in globally ranked IRLC representation that are further restricted. All coalitions C in the reduced instances are either individually rational for all agents in C or for none. Hence, the global ranking of coalitions can be compactly represented by omitting all coalitions C that are ranked below any of the singleton coalitions consisting of one of the members of C. Any such coalition is Pareto dominated and therefore irrelevant for popularity (Proposition 4).

When defining global rankings we will often connect rankings over subsets of coalitions with each other. To simplify the exposition, we introduce the notion of the *join* of two preference relations $\succsim_1$ and $\succsim_2$ over two disjoint sets (of coalitions) C_1 and C_2, respectively, as the preference relation $join(\succsim_1, \succsim_2) = \succsim_1 \cup \succsim_2 \cup C_1 \times C_2$ over the set $C_1 \cup C_2$. In other words, two sets $X, Y \in C_1, C_2$ are in relation $join(\succsim_1, \succsim_2)$ if $X, Y \in C_i$ and $X \succsim_i Y$ for some $i \in [2]$, or if $X \in C_1$ and $Y \in C_2$. We extend this definition recursively to the

6. Note that the inclusion between housing games and marriage games does not hold for strict preferences.

join of relations $\succsim_1, \dots, \succsim_k$ over pairwise disjoint sets $C_1, \dots, C_k$ as $join(C_1, \dots, C_k) = join(join(C_1, \dots, C_{k-1}), C_k)$ for $k \geq 3$. Note that the join operation is not commutative.

4.2.1 Roommate Games

We start by investigating mixed popularity in roommate games by an LP-based approach, which will later have important consequences for popular and strongly popular matchings.

Kavitha et al. (2011) showed that mixed popular matchings in housing games and marriage games can be found in polynomial time. However, as explained in Section 2, their algorithm cannot directly be applied to roommate games. In this section, we show how to obtain an algorithm for the more general class of roommate games.

To introduce our matching notation, we fix a graph $G = (N, E)$ where the vertex set is the set of agents and there is an edge between two vertices if the corresponding coalition of size 2 is individually rational for both agents. For technical reasons, it is useful to restrict attention to the case of perfect matchings, i.e., matchings in which every vertex is matched with some vertex. Similarly to the construction by Kavitha et al. (2011), this can be achieved by introducing worst-case partners w_a for every agent a with $\{a, w_a\} \sim_a \{a\}$. These worst-case partners are not individually rational for all other original agents, and are indifferent among all other agents themselves. They mimic the case when an agent remains unmatched and do not affect the popularity of a partition. In graph-theoretic terms, this is equivalent to adding a loop to every vertex. If some loop is contained in a perfect matching, this means that the agent is matched to herself, or in other words, remains unmatched.

We now establish a relationship between mixed matchings and fractional matchings, where the latter are defined as points in the (perfect) matching polytope $P_{Mat} \subseteq [0,1]^E$, defined as follows (Edmonds, 1965).

$$\begin{aligned} P_{Mat} = \{x \in \mathbb{R}^E \colon & \sum_{e \in E, v \in e} x(e) = 1 \ \forall v \in N, \\ & \sum_{e \in \{\{v,w\} \in E \colon v,w \in C\}} x(e) \leq \frac{|C| - 1}{2} \ \forall C \subseteq N, |C| \text{ odd}, \\ & x(e) \geq 0 \ \forall e \in E\} \end{aligned}$$

The main constraint is often called *odd set constraint* and ensures that, for every odd set of agents C, the weight of the fractional matching restricted to these agents is at most $(|C| - 1)/2$, where this quantity is equal to the maximum cardinality that any matching on the set C may have.

Given a matching M, denote by $\chi_M \in P_{Mat}$ its incidence vector. We obtain a correspondence of mixed matchings and fractional matchings by mapping a mixed matching $p = \{(M_1, p_1), \dots, (M_k, p_k)\}$ to the fractional matching $x_p = \sum_{i=1}^k p_i \chi_{M_i}$. Note that $x_p \in P_{Mat}$ by convexity. Since we only want to operate on the more concise matching polytope, we need to ensure that we can recover a mixed matching efficiently. The following proposition, which is based on general LP theory, can be seen as an extension of the Birkhoff-von Neumann theorem to non-bipartite graphs.

Proposition 5. *Let $G = (N, E)$ be a graph and $x \in P_{Mat}$ a vector in the associated matching polytope. Then, a mixed matching $p = \{(M_1, p_1), \ldots, (M_k, p_k)\}$ such that $x_p = x$ can be found in polynomial time.*

Proof. The separation problem for the matching polytope P_{Mat} can be solved in polynomial time, i.e., the class of matching polytopes is solvable. Therefore, given a graph $G = (N, E)$ and a vector $x \in P_{Mat}$ we can find a convex combination of extreme points of P_{Mat} that yield x in polynomial time (Grötschel, Lovász, & Schrijver, 1981, Th. 3.9). A combinatorial algorithm to address this problem was proposed by Padberg and Wolsey (1984).

Since the extreme points of the matching polytope are the incidence vectors of matchings (Edmonds, 1965), this is a mixed matching whose corresponding fractional matching is x. □

To be able to operate on fractional matchings only, we seek to define popularity of fractional matchings equivalent to popularity of mixed matchings that induce them. Popular fractional matchings can be described as feasible points of a (non-empty) subpolytope of the matching polytope. The separation problem for the subpolytope can be solved efficiently using a modification of McCutchen's algorithm for determining the unpopularity margin of a matching (McCutchen, 2008).

To this end, we need to define the popularity margin for fractional matchings. Given $x, y \in P_{Mat}$, we define their *popularity margin* as

$$\phi(x, y) = \sum_{a \in N} \sum_{i,j \in N_G(a)} x(a, i) y(a, j) \phi_a(i, j)$$

where $N_G(a) = \{v \in N : \{v, a\} \in E\}$ is the neighborhood of a in G and

$$\phi_a(i, j) = \begin{cases} 1 & \text{if } i \succ_a j \\ -1 & \text{if } i \prec_a j \\ 0 & \text{if } i \sim_a j \end{cases}.$$

Imagine that the matchings x and y independently match agent a to agent i and j with probability $x(a, i)$ and $y(a, j)$, respectively. Then, we can interpret the quantity $x(a, i)y(a, j)\phi_a(i, j)$ as the probability of agent a being matched to i through x and to j through y times the characteristic function of agent a's binary preference between these two matching partners. Then, $\sum_{i,j \in N_G(a)} x(a, i)y(a, j)\phi_a(i, j)$ is the expected preference of agent a between matchings x and y, and $\phi(x, y)$ is the expected popularity margin of the preferences of all agents.

Next, we relate the popularity margins of both worlds. The proof of the next proposition is identical to the corresponding statement for marriage games by Kavitha et al. (2011). For the sake of self-containment, we state its proof in the appendix. All other missing proofs can also be found in the appendix.

Proposition 6. *Let p and q be mixed matchings. Then,*

$$\phi(p, q) = \phi(x_p, x_q).$$

In particular, p is popular if and only if for all matchings M, $\phi(x_p, \chi_M) \geq 0$.

As a consequence, mixed popular matchings correspond precisely to the feasible points of the polytope

$$P_{Pop} = \{x \in P_{Mat} : \phi(x, \chi_M) \geq 0 \text{ for all matchings } M\}.$$

It remains to find a feasible point of the popularity polytope P_{Pop}. By adopting the auxiliary graph in McCutchen's algorithm for non-bipartite graphs, we can find a matching M minimizing $\phi(x, \chi_M)$ by solving a maximum weight matching problem (McCutchen, 2008). This solves the separation problem for P_{Pop}.

Proposition 7. *The separation problem for P_{Pop} can be solved in polynomial time.*

We are now ready to prove the following theorem.

Theorem 1. *Mixed popular matchings in roommate games with weak preferences can be found in polynomial time.*

Proof. By Proposition 7 and by means of the Ellipsoid method (Khachiyan, 1979), we can find a fractional popular matching in polynomial time. This can be translated into a mixed popular matching by leveraging Proposition 5. □

Theorem 1 has a number of interesting consequences. Since every mixed popular matching is Pareto optimal, we now have an LP-based algorithm to find Pareto optimal matchings for weak preferences as an alternative to combinatorial algorithms like the Preference Refinement Algorithm by Aziz et al. (2013a).

Corollary 1. *Pareto optimal matchings in roommate games with weak preferences can be found in polynomial time.*

Biró et al. (2010) provided a sophisticated algorithm for verifying whether a given matching is popular. An efficient LP-based algorithm for this problem follows from Theorem 1.

Corollary 2. *It can be verified in polynomial time whether a given matching in a roommate game is popular.*

Finally, the linear programming approach allows us to resolve the open problem of finding strongly popular matchings when preferences are weak.

Corollary 3. *Finding a strongly popular matching or deciding that no such matching exists in roommate games with weak preferences can be done in polynomial time.*

Proof. If a strongly popular matching exists, it is unique. In particular, it is the unique mixed popular matching. Given a (deterministic) matching M, we can check in polynomial time if it is strongly popular. We can apply the reduction of Proposition 7 and check whether the maximum weight matching amongst the matchings *different to M* on the auxiliary graph has negative weight (in which case the matching M is strongly popular) or not. Note that every matching different to M is contained in at least one (incomplete) graph obtained by deleting an edge from M, while M is not contained in any such graph. Hence, we simply compute a maximum weight matching for every graph obtained by deleting exactly one edge from M in the auxiliary graph. The maximum weight matching amongst these matchings has the highest weight amongst matchings different from M.

The algorithm to compute a strongly popular matching if one exists first computes a fractional popular matching. If it does not correspond to a deterministic matching, there exists no strongly popular matching. Otherwise, it is deterministic and, as described above, we can check if it is strongly popular. If this is the case, we return it. If not, there exists no strongly popular matching. □

As shown in the previous proof, the verification problem for strongly popular matchings in roommate games can also be solved efficiently.

4.2.2 Flatmate Games

It turns out that moving from coalitions of size 2 to size 3 renders all search problems related to popular partitions intractable. For mixed popular partitions, we can leverage the relationship to Pareto optimal partitions. Aziz et al. (2013a, Th. 5) have shown that finding Pareto optimal partitions in flatmate games with weak preferences is NP-hard. Since mixed popular partitions are guaranteed to exist (Proposition 1) and satisfy Pareto optimality (Proposition 2), this immediately implies the NP-hardness of computing mixed popular partitions by means of a Turing reduction.[7]

Theorem 2. *Computing a partition in the support of a mixed popular partition in flatmate games with weak preferences is NP-hard.*

For strict preferences, the same method does not work. Pareto optimal partitions can always be found efficiently by serial dictatorship. Therefore, we will give direct reductions that yield hardness for strong popularity and mixed popularity in flatmate games with strict preferences. The reduction for popularity is a bit more involved and will be given afterwards. All of these reductions are based on a common type of flatmate games that evolve from instances of the NP-complete problem Exact 3-Cover (Karp, 1972). An instance (R, S) of *Exact* 3*-Cover* (X3C) consists of a ground set R together with a set S of 3-element subsets of R. A Yes-instance is an instance such that there exists a subset $S' \subseteq S$ that partitions R.

Before presenting the proof, we want to discuss our proof strategy which is very generic and also key to many hardness reductions for cardinal hedonic games in Section 4.3. We want to describe the essential properties satisfied by reduced instances of our reduction. We say that a class of games satisfies *property PP* (for popularity propagation) if there exists a polynomial-time reduction from X3C that constructs for every instance (R, S) a game $(N, \succsim)$ together with a special agent $x \in N$, and a partition π^* such that for every partition $\pi \neq \pi^*$, it holds that

1. $\phi(\pi^*, \pi) \geq 1$,
2. if $\pi^*(x) \cap \pi(x) = \{x\}$, then $\phi(\pi^*, \pi) \geq 3$ or (R, S) is a Yes-instance,
3. for all $y \in N$, $\pi^*(y) \succ_y \{y\}$, and

7. Using the same argument, one can transfer further results on Pareto optimality (Aziz et al., 2013a), e.g., for room-roommate games or three-cyclic matching games.

4. $\pi^*(x) \succ_x C$ for all $C \in \mathcal{N}_x \setminus \{\pi^*(x)\}$.

In addition, if (R, S) is a Yes-instance, then there exists a partition π' with

5. $\phi(\pi^*, \pi') = 1$, and

6. $\pi'(x) = \{x\}$.

The first condition guarantees that π^* is strongly popular and with the second condition, strong popularity is unaffected when adding one or two auxiliary agents that only have an effect on x. The third condition is only needed for the proofs concerning fractional hedonic games with non-negative utility functions, but it also holds for all other classes investigated. It ensures that every agent is part of an individually rational coalition, and in fact prefers her coalition in π^* over staying alone. The forth condition says that x is in her unique top-ranked coalition under the partition π^*. The last two properties ensure that we can obtain a more popular partition by adding auxiliary agents that form a new coalition with x.

In this section, we will exemplify a reduction satisfying property PP for flatmate games. We will first describe the reduced flatmate games, then prove the first two items of property PP in Lemma 1. Then, we provide a lemma for global rankedness of the game, and finally give the actual reductions which implicitly construct the partition π' from property PP.

To this end, consider an instance (R, S) of X3C. Let $k = \min\{k \in \mathbb{N} \colon 2^k \geq |R|\}$ be the smallest power of 2 that is larger than the cardinality of R. We define a flatmate game on vertex set $N = \bigcup_{j=0}^{k} N_j$, where $N_j = \bigcup_{i=1}^{2^j} A_j^i$ consists of 2^j sets of agents A_j^i.

We define the sets of agents as

- $A_k^i = \{a_k^i, b_k^i, c_k^i\}$ for $i \in [|R|]$,
- $A_k^i = \{a_k^i, b_k^i, c_k^i, y_1^i, y_2^i\}$ for $i \in [|R| + 1, 2^k]$, and
- $A_j^i = \{a_j^i, b_j^i, c_j^i, \alpha_j^i, \beta_j^i, \gamma_j^i, \delta_j^i\}$ for $j \in [0, k-1], i \in [2^j]$.

Similar names of agents suggest that these agents are going to play the same role in the reduction. The preferences are designed in a way such that if there exists no 3-partition of R through sets in S, then there exists a unique best partition that assigns more than half of the agents a top-ranked coalition. Otherwise, there exists a partition that puts exactly all the other agents in one of their top coalitions. We order the set R in an arbitrary but fixed way, say $R = \{r^1, \dots, r^{|R|}\}$ and for a better understanding of the proof and the preferences, we label the agents $b_k^i = r^i$ for $i \in [|R|]$. If we view the set of agents N as $k+1$ levels of agents, then the ground set R of the instance of X3C is identified with some specific agents in the top level k. Preferences of the agents are as follows. Recall that $X^{\succ}$ denotes an arbitrary, but fixed strict preference order of the alternatives in X. We define

- $\{y_1^i, y_2^i\} \succ_{y_1^i} \{y_1^i\}$, $i \in [|R| + 1, 2^k]$,
- $\{b_k^i, y_2^i\} \succ_{y_2^i} \{y_1^i, y_2^i\} \succ_{y_2^i} \{y_2^i\}$, $i \in [|R| + 1, 2^k]$,
- $\{a_k^i, b_k^i, c_k^i\} \succ_{a_k^i} \{a_k^i, a_k^{i+1}, \delta_{k-1}^{(i+1)/2}\} \succ_{a_k^i} \{a_k^i\}$, $i \in [2^k]$ odd,

- $\{a_k^i, b_k^i, c_k^i\} \succ_{a_k^i} \{a_k^i, a_k^{i-1}, \delta_{k-1}^{i/2}\} \succ_{a_k^i} \{a_k^i\}$, $i \in [2^k]$ even,
- $\{a_j^i, \beta_j^i, \gamma_j^i\} \succ_{a_j^i} \{a_j^i, b_j^i, c_j^i\} \succ_{a_j^i} \{a_j^i\}$, $j \in [0, k-1], i \in [2^j]$,
- $\{\{b_k^i, b_k^v, b_k^w\} \colon \{r^i, r^v, r^w\} \in S \text{ for } v, w \in [|R|]\}^{\succ} \succ_{b_k^i} \{a_k^i, b_k^i, c_k^i\} \succ_{b_k^i} \{b_k^i\}$, $i \in [|R|]$,
- $\{b_k^i, y_2^i\} \succ_{b_k^i} \{a_k^i, b_k^i, c_k^i\} \succ_{b_k^i} \{b_k^i\}$, $i \in [|R|+1, 2^k]$,
- $\{b_j^i, c_{j+1}^{2i-1}, c_{j+1}^{2i}\} \succ_{b_j^i} \{a_j^i, b_j^i, c_j^i\} \succ_{b_j^i} \{b_j^i\}$, $j \in [0, k-1], i \in [2^j]$,
- $\{a_j^i, b_j^i, c_j^i\} \succ_{c_j^i} \{c_j^i, c_j^{i+1}, b_{j-1}^{(i+1)/2}\} \succ_{c_j^i} \{c_j^i\}$, $j \in [k], i \in [2^j]$ odd,
- $\{a_j^i, b_j^i, c_j^i\} \succ_{c_j^i} \{c_j^i, c_j^{i-1}, b_{j-1}^{i/2}\} \succ_{c_j^i} \{c_j^i\}$, $j \in [k], i \in [2^j]$ even,
- $\{a_0^1, b_0^1, c_0^1\} \succ_{c_0^1} \{c_0^1\}$,
- $\{\alpha_j^i, \beta_j^i\} \succ_{\alpha_j^i} \{\alpha_j^i, \alpha_j^{i+1}, \delta_{j-1}^{(i+1)/2}\} \succ_{\alpha_j^i} \{\alpha_j^i\}$, $j \in [k-1], i \in [2^j]$ odd,
- $\{\alpha_j^i, \beta_j^i\} \succ_{\alpha_j^i} \{\alpha_j^i, \alpha_j^{i-1}, \delta_{j-1}^{i/2}\} \succ_{\alpha_j^i} \{\alpha_j^i\}$, $j \in [k-1], i \in [2^j]$ even,
- $\{\alpha_0^1, \beta_0^1\} \succ_{\alpha_0^1} \{\alpha_0^1\}$,
- $\{\beta_j^i, \gamma_j^i, a_j^i\} \succ_{\beta_j^i} \{\beta_j^i, \alpha_j^i\} \succ_{\beta_j^i} \{\beta_j^i\}$, $j \in [0, k-1], i \in [2^j]$,
- $\{\gamma_j^i, \delta_j^i\} \succ_{\gamma_j^i} \{\beta_j^i, \gamma_j^i, a_j^i\} \succ_{\gamma_j^i} \{\gamma_j^i\}$, $j \in [0, k-1], i \in [2^j]$,
- $\{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\} \succ_{\delta_j^i} \{\delta_j^i, \gamma_j^i\} \succ_{\delta_j^i} \{\delta_j^i\}$, $j \in [0, k-2], i \in [2^j]$, and
- $\{\delta_{k-1}^i, a_k^{2i-1}, a_k^{2i}\} \succ_{\delta_{k-1}^i} \{\delta_{k-1}^i, \gamma_{k-1}^i\} \succ_{\delta_{k-1}^i} \{\delta_{k-1}^i\}$, $i \in [2^{k-1}]$.

The structure of the flatmate game is illustrated in Figure 1 for the case $k = 3$. We will be particularly interested in coalitions of the types $\{a_j^i, b_j^i, c_j^i\}$, $\{\alpha_j^i, \beta_j^i\}$, $\{\gamma_j^i, \delta_j^i\}$, and $\{y_1^i, y_2^i\}$ which are marked by undirected edges. These coalitions form the partition π^* of Lemma 1 that we need later to investigate for strong and mixed popularity in the respective reductions. The directed edges indicate that an agent at the tail of the arrow needs to form a coalition with the agent at the tip of the arrow in order to improve from her coalition of the above type. The ground structure of the set of agents can be viewed as a binary tree of triangles depicted by the circular-shaped vertices. The important property of this tree is that whenever a coalition of the type $\{a_j^i, b_j^i, c_j^i\}$ gets dissolved, there can only be an improvement in popularity for the agents in A_j^i if they propagate changes in the partition upwards within this tree. This is achieved for agents b_j^i directly through the binary tree and for agents a_j^i with help of the auxiliary agents $\{\alpha_j^i, \beta_j^i, \gamma_j^i, \delta_j^i\}$ that are depicted as diamond-shaped vertices.

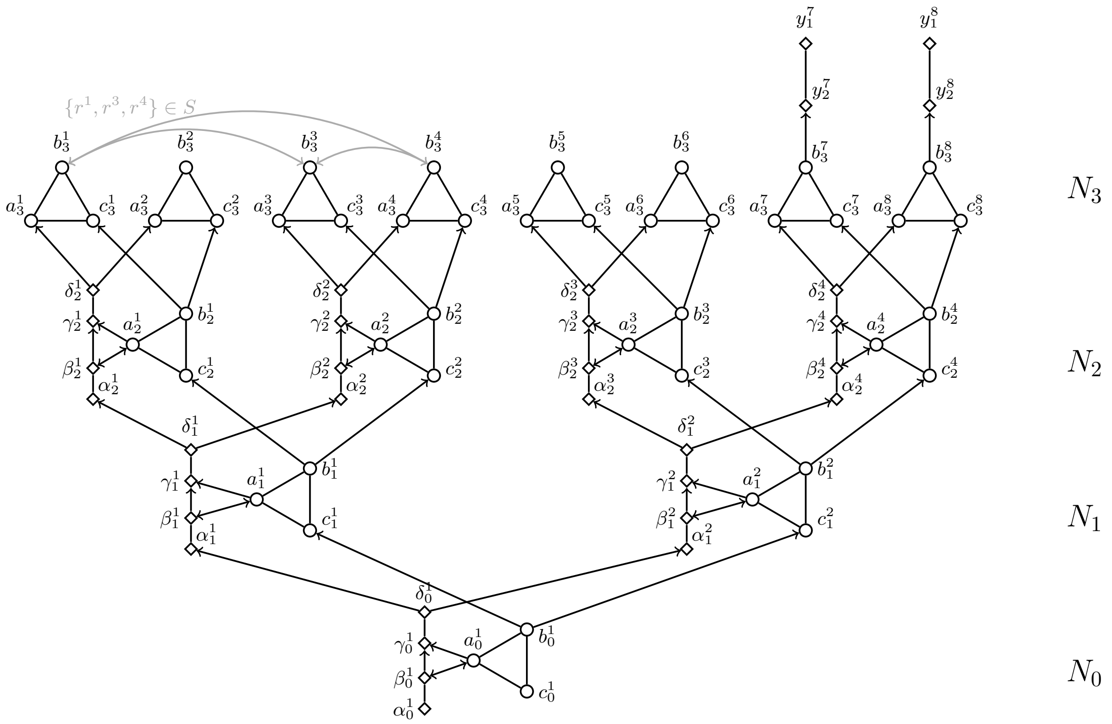


Figure 1: Schematic of the reduction for flatmate games with strict preferences. There is an edge between two agents if they are in the coalition π^* defined in Lemma 1. Directed edges indicate improvements from π^*. The gray edges suggest a 3-elementary set in S.

Lemma 1. *Let an instance (R, S) of X3C be given and define the corresponding flatmate game as above. Consider the partition $\pi^* = \{\{a^i_j, b^i_j, c^i_j\} : j \in [0, k], i \in [2^j]\} \cup \{\{\alpha^i_j, \beta^i_j\}, \{\gamma^i_j, \delta^i_j\} : j \in [0, k-1], i \in [2^j]\} \cup \{\{y^i_1, y^i_2\} : i \in [|R| + 1, 2^k]\}$. Let $\pi \neq \pi^*$ be an arbitrary partition of agents distinct from π^*. Then $\phi(\pi^*, \pi) \geq 1$. In addition, if $c^1_0 \in N(\pi^*, \pi)$, then $\phi(\pi^*, \pi) \geq 3$ or $\{b^i_k : i \in [2^k]\} \subseteq N(\pi, \pi^*)$.*

Proof. Let an instance (R, S) of X3C be given and define the corresponding flatmate game as above. Let π^* be defined as in the lemma and $\pi \neq \pi^*$ another partition. We recursively define the following sets of agents: for $i \in [2^k]$, $T^i_k = A^i_k$ and for $j = k-1, \ldots, 0, i \in [2^j]$, $T^i_j = A^i_j \cup T^{2i-1}_{j+1} \cup T^{2i}_{j+1}$. We will prove the following claim by induction over $j = k, \ldots, 0$.

For every $i \in [2^j]$ holds: Assume there exists an agent $x \in T^i_j$ with $\pi(x) \neq \pi^*(x)$. Then $\phi_{T^i_j}(\pi^*, \pi) \geq 1$. If even $\pi(a^i_j) \neq \pi^*(a^i_j)$, then $\phi_{T^i_j}(\pi^*, \pi) \geq 3 \vee \{b^i_k : i \in [2^k]\} \cap T^i_j \subseteq N(\pi, \pi^*)$.

Note that the claim implies $\phi_{T^i_j}(\pi^*, \pi) \geq 0$ in any case. Clearly, the assertion of the lemma follows from the case $j = 0$.

We frequently use the facts that for all $j \in [0, k-1], i \in [2^j]$,

- $\alpha^i_j \notin N(\pi, \pi^*)$ and if $\beta^i_j \in N(\pi, \pi^*)$, then $\alpha^i_j \in N(\pi^*, \pi)$, and

- $\gamma_j^i \notin N(\pi, \pi^*)$ and if $\delta_j^i \in N(\pi, \pi^*)$, then $\gamma_j^i \in N(\pi^*, \pi)$.

The case $j = k$ and $i \in [2^k]$ is immediate (using a similar fact for agents y_1^i and y_2^i in the case $i \in \{|R| + 1, \dots, 2^k\}$).

For the induction step, let $j \in \{k-1, \dots, 0\}$ and fix $i \in [2^j]$. We will essentially prove that changing the coalitions in A_j^i causes severe loss in popularity, unless we propagate changes to substructures via b_j^i or δ_j^i. Assume first that there exists an agent $x \in T_j^i$ with $\pi(x) \neq \pi^*(x)$ but no such agent in A_j^i. Then, $x \in T_{j+1}^{2i-1} \vee x \in T_{j+1}^{2i}$ and the claim follows by induction. Assume therefore that there exists an agent $x \in A_j^i$ with $\pi(x) \neq \pi^*(x)$. Note that $\phi_{A_j^i}(\pi, \pi^*) \leq 1$.

First consider the case that $\pi(a_j^i) \neq \pi^*(a_j^i)$. If $b_j^i \in N(\pi, \pi^*)$, we can apply induction for T_{j+1}^{2i-1} and T_{j+1}^{2i} and we are done, because by induction $\phi_{T_{j+1}^{2i-1} \cup T_{j+1}^{2i}}(\pi^*, \pi) \geq 4 \vee \{b_k^i \colon i \in [2^k]\} \cap (T_{j+1}^{2i-1} \cup T_{j+1}^{2i}) \subseteq N(\pi, \pi^*)$. We may therefore assume that $b_j^i \in N(\pi^*, \pi)$. Then, $\phi_{A_j^i}(\pi^*, \pi) \geq 3$ or $a_j^i \in N(\pi, \pi^*)$. In the latter case, $\phi_{A_j^i}(\pi^*, \pi) \geq 3$ unless $\delta_j^i \in N(\pi, \pi^*)$. Finally, if $\delta_j^i \in N(\pi, \pi^*)$, then the claim follows by induction for T_{j+1}^{2i-1} and T_{j+1}^{2i}, because $\phi_{T_j^i}(\pi^*, \pi) = \phi_{A_j^i}(\pi^*, \pi) + \phi_{T_{j+1}^{2i-1}}(\pi^*, \pi) + \phi_{T_{j+1}^{2i}}(\pi^*, \pi) \geq 1 + 1 + 1 = 3$.

It remains the case that $\pi(x) \neq \pi^*(x)$ for $x \in \{\alpha_j^i, \gamma_j^i\}$ while $\pi(a_j^i) = \pi^*(a_j^i)$. If $\pi(\alpha_j^i) \neq \pi^*(\alpha_j^i)$, then $\phi_{A_j^i}(\pi^*, \pi) \geq 2$. If $\pi(\gamma_j^i) \neq \pi^*(\gamma_j^i)$, then $\phi_{A_j^i}(\pi^*, \pi) \geq 2$ or $\phi_{A_j^i}(\pi^*, \pi) \geq 0 \wedge \pi(\delta_j^i) = \{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\}$ and the claim follows by induction. □

In the next lemma, we prove that the preferences used in the construction are even globally ranked.

Lemma 2. *Let an instance (R, S) of X3C be given and define the corresponding flatmate game as above. Then, the preferences are globally ranked.*

Proof. The global preferences are composed of preferences $\succ_0, \dots, \succ_k$ over the sets of coalitions $C_0, \dots, C_k$, where C_j is essentially the set of coalitions that is individually rational for some agent in A_j^i for some $i \in [2^j]$. More formally, $C_k = \bigcup_{i=1}^{2^k} \{C \subseteq N \colon \exists v \in A_k^i : C \succsim_v \{v\}\}$ and, for $j = k-1, \dots, 0$, $C_j = \bigcup_{i=1}^{2^j} \{C \subseteq N \colon \exists v \in A_j^i : C \succsim_v \{v\}\} \setminus C_{j+1}$. Note that this separates coalitions by level, and $C_j \cap C_{j'} = \emptyset$ for $j \neq j'$. In particular, coalitions of the types $\{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\}, \{\delta_{k-1}^i, a_k^{2i-1}, a_k^{2i}\}$, and $\{b_j^i, c_{j+1}^{2i-1}, c_{j+1}^{2i}\}$ that involve agents of two levels are added to the coalitions of the higher level. The global ranking is given in succinct form over $\bigcup_{j=0}^k C_j$ as $join(\succ_0, \dots, \succ_k)$. It can be extended to a full global ranking by adding coalitions that are not individually rational for one of its members at the bottom. It remains to specify these subrankings. The preferences over sets of coalitions can always be arbitrary. The ranking $\succ_k$ is given as

$$\begin{aligned}
&\{\{y_1^i, y_2^i\} \colon i \in [|R|+1, 2^k]\}^{\succ} \\
\succ_k &\{\{b_k^i, y_2^i\} \colon i \in [|R|+1, 2^k]\}^{\succ} \\
\succ_k &\{\{y_1^i\}, \{y_2^i\} \colon i \in [|R|+1, 2^k]\}^{\succ} \\
\succ_k &\{\{b_k^i, b_k^v, b_k^w\} \colon \{r^i, r^v, r^w\} \in S \text{ for } v, w \in [|R|]\}^{\succ} \\
\succ_k &\{\{b_k^i\} \colon i \in [2^k]\}^{\succ} \\
\succ_k &\{\{a_k^i, b_k^i, c_k^i\} \colon i \in [2^k]\}^{\succ} \\
\succ_k &\{\{b_{k-1}^i, c_k^{2i-1}, c_k^{2i}\}, \{\delta_{k-1}^i, a_k^{2i-1}, a_k^{2i}\} \colon i \in [2^{k-1}]\}^{\succ} \\
\succ_k &\{\{a_k^i\}, \{c_k^i\} \colon i \in [2^k]\}^{\succ}
\end{aligned}$$

For $j \in [k-1]$, the ranking $\succ_j$ is given as

$$\begin{aligned}
&\{\{\gamma_j^i, \delta_j^i\} \colon i \in [2^j]\}^{\succ} \\
\succ_j &\{\{a_j^i, \beta_j^i, \gamma_j^i\} \colon i \in [2^j]\}^{\succ} \\
\succ_j &\{\{a_j^i, b_j^i, c_j^i\}, \{\alpha_j^i, \beta_j^i\} \colon i \in [2^j]\}^{\succ} \\
\succ_j &\{\{b_{j-1}^i, c_j^{2i-1}, c_j^{2i}\}, \{\delta_{j-1}^i, \alpha_j^{2i-1}, \alpha_j^{2i}\} \colon i \in [2^{j-1}]\}^{\succ} \\
\succ_j &\{\{a_j^i\}, \{b_j^i\}, \{c_j^i\}, \{\alpha_j^i\}, \{\beta_j^i\}, \{\gamma_j^i\}, \{\delta_j^i\} \colon i \in [2^j]\}^{\succ}
\end{aligned}$$

Finally, $\succ_0$ is given as

$$\begin{aligned}
&\{\gamma_0^1, \delta_0^1\} \succ_0 \{a_0^1, \beta_0^1, \gamma_0^1\} \succ_0 \{\{a_0^1, b_0^1, c_0^1\}, \{\alpha_0^1, \beta_0^1\}\}^{\succ} \\
\succ_0 &\{\{a_0^1\}, \{b_0^1\}, \{c_0^1\}, \{\alpha_0^1\}, \{\beta_0^1\}, \{\gamma_0^1\}, \{\delta_0^1\}\}^{\succ}
\end{aligned}$$

The individual preferences are clearly induced by the global ranking. □

We are now ready to apply the two lemmas for the desired reductions.

Theorem 3. *Deciding whether there exists a strongly popular partition in flatmate games is coNP-hard, even if preferences are strict and globally ranked.*

Proof. The reduction is from X3C. Given an instance (R, S) of X3C, we define a hedonic game on agent set $N' = N \cup \{z\}$ where the agents N are as in the above construction with the identical preferences except changing the preferences of c_0^1 to $\{a_0^1, b_0^1, c_0^1\} \succ_{c_0^1} \{c_0^1, z\} \succ_{c_0^1} \{c_0^1\}$, and $\{c_0^1, z\} \succ_z \{z\}$. In particular, for every agent in $N \setminus \{c_0^1\}$, coalitions together with z are not individually rational. Note that $|N'| = 3\sum_{j=0}^{k} 2^j + 4\sum_{j=0}^{k-1} 2^j + 2(2^k - |R| - 1) + 1 = 12 \cdot 2^k - 2 \cdot |R| - 8 = \mathcal{O}(|R|)$ and the reduction is in polynomial time.

Consider the partition $\sigma^* = \{\{a_j^i, b_j^i, c_j^i\} \colon j \in [0,k], i \in [2^j]\} \cup \{\{\alpha_j^i, \beta_j^i\}, \{\gamma_j^i, \delta_j^i\} \colon j \in [0, k-1], i \in [2^j]\} \cup \{\{y_1^i, y_2^i\} \colon i \in [|R|+1, 2^k]\} \cup \{\{z\}\} = \pi^* \cup \{\{z\}\}$ for the partition π^* from Lemma 1. Let $\sigma \neq \sigma^*$ be given and define $\pi = (\sigma \setminus \sigma(z)) \cup \{\sigma(z) \setminus \{z\}\}$, i.e. the partition on the agent set N, where z left her coalition. Note that due to the preferences of agents in N, $\phi(\pi^*, \pi) \leq \phi_N(\sigma^*, \sigma)$. We investigate the popularity margin of σ^* and σ by a case distinction over the possible coalitions for agent z using the knowledge of Lemma 1 about the relationship of the partitions π^* and π. If $\sigma(z) = \{z\}$, then $\phi(\sigma^*, \sigma) = \phi(\pi^*, \pi) \geq 1$. If $\sigma(z) = \{c_0^1, z\}$, then $\phi(\sigma^*, \sigma) \geq -1 + \phi(\pi^*, \pi) \geq -1 + 1 \geq 0$. Otherwise, $\phi(\sigma^*, \sigma) = 1 + \phi(\pi^*, \pi) \geq 1$. It follows directly that σ^* is popular and hence there exists a strongly popular partition if and only if σ^* is strongly popular. We will prove that this is the case if and only if the instance of X3C is a No-instance.

Assume that there exists no 3-partition of R through sets in S. The only case above, where the popularity margin is not strictly positive, is if $\sigma(z) = \{z, c_0^1\}$, but in this case $\pi(c_0^1) = \{c_0^1\}$ and it follows that $\phi(\sigma^*, \sigma) \geq -1 + \phi(\pi^*, \pi) \geq -1 + 3 \geq 2$. Hence, σ^* is strongly popular.

Conversely, assume that there exists a 3-partition $S' \subseteq S$ of R. Define

$$\begin{aligned}\sigma' =&\{\{b_k^v, b_k^w, b_k^x\}\colon \{r^v, r^w, r^x\} \in S'\} \cup \{\{b_k^i, y_2^i\}, \{y_1^i\}\colon i \in [|R|+1, 2^k]\}\\ &\cup \{\{\delta_{k-1}^i, a_k^{2i-1}, a_k^{2i}\}\colon i \in [2^{k-1}]\} \cup \{\{b_j^i, c_{j+1}^{2i-1}, c_{j+1}^{2i}\}, \{a_j^i, \beta_j^i, \gamma_j^i\}\colon j \in [k-1], i \in [2^j]\}\\ &\cup \{\{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\}\colon j \in [k-2], i \in [2^j]\} \cup \{\{\alpha_0^1\}, \{z, c_0^1\}\}.\end{aligned}$$

It is easily checked that $\phi(\sigma', \sigma^*) = 0$.

Indeed, $N(\sigma', \sigma^*) = \{b_k^i\colon i \in [2^k]\} \cup \{\beta_j^i, \delta_j^i, a_j^i\colon j \in [0, k-1], i \in [2^j]\} \cup \{y_2^i\colon i \in [|R|+1, 2^k]\} \cup \{z\}$. Therefore, $|N(\sigma', \sigma^*)| = 2^k + 4\sum_{j=1}^{k-1} 2^j + 2^k - (|R|+1) + 1 = 6 \cdot 2^k - |R| - 4 = \frac{1}{2}|N'|$. Hence, $\phi(\sigma', \sigma^*) \geq 0$ and equality follows from popularity of σ^*. Therefore, there exists no strongly popular partition. □

A similar reduction as in Theorem 3 also works for mixed popularity. Then, however, we need two auxiliary agents to control the switch between a strongly popular and non-popular partition.

Theorem 4. *Computing a mixed popular partition in flatmate games is NP-hard, even if preferences are strict and globally ranked.*

Popular partitions are guaranteed to exist in roommate games with strict and globally ranked preferences (Abraham et al., 2008). We show by means of a counterexample that this is no longer the case when moving from roommate to flatmate games. This example game will serve as a crucial gadget to prove the hardness of computing popular partitions.

Proposition 8. *There exists a flatmate game with strict and globally ranked preferences which does not admit a popular partition.*

Proof. Consider $N = \{x_1, x_2, x_3\} \cup \{z_1^j, z_2^j\colon j \in [4]\}$, and preferences induced by the global ranking $\succ$ given by $\{\{x_1, z_1^j, z_2^j\}\colon j \in [4]\}^{\succ} \succ \{\{x_2, z_1^j, z_2^j\}\colon j \in [4]\}^{\succ} \succ \{\{x_3, z_1^j, z_2^j\}\colon j \in [4]\}^{\succ} \succ (\{\{x_i\}\colon i \in [3]\} \cup \{\{z_k^j\}\colon k \in [2], j \in [4]\})^{\succ}$. We claim that there exists no popular partition. By Proposition 2, we only need to consider Pareto optimal partitions. Let π be any Pareto optimal partition. Then π is individually rational. We will show how to obtain a more popular partition. By the pigeon hole principle, there exists $j \in [4]$ with $\{z_1^j\}, \{z_2^j\} \in \pi$. If there exists $i \in [3]$ with $\{x_i\} \in \pi$, then creating the coalition $\{x_i, z_1^j, z_2^j\}$ is more popular.

Otherwise, we may assume that for some $\{j_1, j_2, j_3\} \subseteq [4]$, $\pi(x_i) = \{x_i, z_1^{j_i}, z_2^{j_i}\}$, for $i \in [3]$. Let $j_4 \in [4] \setminus \{j_1, j_2, j_3\}$ be the remaining index. We obtain a new partition π' by forming $\pi'(x_i) = \{x_i, z_1^{j_{i+1}}, z_2^{j_{i+1}}\}$, leaving $z_1^{j_1}$ and $z_2^{j_1}$ in singleton coalitions.

Then, $N(\pi', \pi) \supseteq \{z_1^{j_i}, z_2^{j_i} \colon i \in [2,4]\}$ while $N(\pi, \pi') \subseteq \{x_1, x_2, x_3, z_1^{j_1}, z_2^{j_1}\}$. Hence, $\phi(\pi', \pi) \geq 1$. □

The idea is to replace the agents x_i of this example by the gadget of Lemma 1 to obtain a hardness result.

Theorem 5. *Deciding whether there exists a popular partition in flatmate games with strict and globally ranked preferences is coNP-hard.*

Proof. Given an instance (R, S) of X3C, we construct the flatmate game $(N, \succsim)$ with strict and globally ranked preferences as follows. We take 3 copies $(N_i, \succsim_i)$ of the game of Lemma 1, where $\succsim_i$ are the strict and globally ranked preferences of Lemma 2. Denote the special partition and agent of the lemma by π_i^* and $x_i = c_{0i}^1$, respectively. Also, denote the set of coalitions ranked by $\succsim_i$ with C_i' and define $C_i = C_i' \setminus \{\{x_i\}\}$. We set $N = N_1 \cup N_2 \cup N_3 \cup \{z_1^j, z_2^j \colon j \in [4]\}$. To define global preferences, we define preferences over $C_4 = \{\{x_i, z_1^j, z_2^j\} \colon i \in [3], j \in [4]\} \cup \{\{x_i\} \colon i \in [3]\} \cup \{\{z_k^j\} \colon k \in [2], j \in [4]\}$.

$$\begin{aligned}
&\{\{x_1, z_1^j, z_2^j\} \colon j \in [4]\}^{\succ} \\
\succ_4 &\{\{x_2, z_1^j, z_2^j\} \colon j \in [4]\}^{\succ} \\
\succ_4 &\{\{x_3, z_1^j, z_2^j\} \colon j \in [4]\}^{\succ} \\
\succ_4 &(\{\{x_i\} \colon i \in [3]\} \cup \{\{z_k^j\} \colon k \in [2], j \in [4]\})^{\succ}
\end{aligned}$$

The global ranking is given over $\bigcup_{j=1}^4 C_j$ as $\succsim = join(\succsim_1, \succsim_2, \succsim_3, \succsim_4)$ in succinct form.

We claim that there exists a popular partition if and only if (R, S) is a No-instance of X3C.

If (R, S) is a No-instance, consider $\pi^* = \bigcup_{i=1}^3 \pi_i^* \cup \{\{z_k^j\} \colon k \in [2], j \in [4]\}$. Let π be any other partition. Let $I = \{i \in [3] \colon \pi^*(x_i) \neq \pi(x_i)$ and define $N' = N_1 \cup N_2 \cup N_3$ and $Z = \{z_1^j, z_2^j \colon j \in [4]\}$. We have $\phi_{N'}(\pi^*, \pi) \geq 3|I|$ (due to Lemma 1) while $\phi_Z(\pi, \pi^*) \leq 2|I|$. Hence, π^* is more popular than π if $|I| \geq 1$. In the case $|I| = 0$, it holds $\phi_{N'}(\pi, \pi^*) \leq 0$ while $\phi_Z(\pi, \pi^*) \leq 0$ and as $\pi \neq \pi^*$, one of the inequalities must be strict.

Now assume that (R, S) is a Yes-instance of X3C and assume for contradiction that π is popular (and hence Pareto optimal). Then, for $i \in [3]$, $i \in I$. Indeed, if $i \notin I$, then π restricted to N_i must be π_i^* (otherwise, π_i^* is more popular). There exists $j \in [4]$ with $\pi(z_1^j) \neq \{x_1, z_1^j, z_2^j\}$ and by Pareto optimality $\{z_1^j\}, \{z_2^j\} \in \pi$. We obtain a more popular partition π' by replacing the coalitions of $N_i \cup \{z_1^j, z_2^j\}$ by the partition of the proof of Theorem 4 for the subgame $(N_i, \succsim_i)$.

It remains the case that $I = [3]$. We may assume that for some $\{j_1, j_2, j_3\} \subseteq [4]$, $\pi(x_i) = \{x_i, z_1^{j_i}, z_2^{j_i}\}$, for $i \in [3]$. Let $j_4 \in [4] \setminus \{j_1, j_2, j_3\}$ be the remaining index. We obtain a new partition π' by removing $z_1^{j_4}, z_2^{j_4}$ from their coalitions and forming $\pi'(x_i) = \{x_i, z_1^{j_{i+1}}, z_2^{j_{i+1}}\}$, leaving $z_1^{j_1}$ and $z_2^{j_1}$ in singleton coalitions.

Then, $N(\pi', \pi) \supseteq \{z_1^{j_i}, z_2^{j_i} \colon i \in [2,4]\}$ while $N(\pi, \pi') \subseteq \{x_1, x_2, x_3, z_1^{j_1}, z_2^{j_1}\}$. Hence, $\phi(\pi', \pi) \geq 1$, a contradiction. □

To conclude the section, we deal with the problem of verifying whether a given partition is popular or strongly popular. The respective results follow directly from the constructions of the hardness of existence.

Theorem 6. *Verifying whether a given partition in a flatmate game with strict and globally ranked preferences is popular is coNP-complete.*

Proof. In the proof of Theorem 5, the partition π^* is popular if and only if (R, S) is a No-instance of X3C. $\square$

Theorem 7. *Verifying whether a given partition in a flatmate game is strongly popular is coNP-complete, even if preferences are strict and globally ranked.*

Proof. In the proof of Theorem 3, the partition π^* is strongly popular if and only if (R, S) is a No-instance of X3C. $\square$

We would like to remark a strong relationship of the existence and verification problems. Our general proof strategy for the coNP-hardness of existence problems is to provide an instance of a game *together with* a partition that is (strongly) popular if and only if the constructed game arises from a No-instance of the NP-hard source problem (this is the partition π^* of property PP). If the game is based on a Yes-instance, there is no (strongly) popular partition. In other words, all relevant questions on (strong) popularity can be answered with this given partition.

Consequently, we actually prove coNP-hardness for a restriction of the verification problem that is only allowed to ask for verification of partitions that have to be (strongly) popular if such a partition exists. Clearly, the hardness of this restricted problem implies both hardness of the verification and the existence problem. The latter follows from the simple reduction that maps tuples (G, π) of a game and a partition to the game G. Instead of giving the reduction for this unifying problem, we prefer not to introduce this restricted verification problem, and to keep the focus on the problems that we are actually interested in. Still, the same phenomenon will occur again for the proofs regarding cardinal hedonic games in the next section.

4.3 Cardinal Hedonic Games

Important subclasses of hedonic games that admit succinct representations are based on cardinal utility functions. For one, there are *additively separable hedonic games* (Bogomolnaia & Jackson, 2002), where the utility that an agent associates with a coalition is the sum of utilities he ascribes to each member of the coalition. On the other hand, there are *fractional hedonic games* (Aziz, Brandl, Brandt, Harrenstein, Olsen, & Peters, 2019), where the sum of utilities is divided by the number of agents contained in the coalition.

In the following, let $v_i(j)$ denote the utility that agent i associates with agent j. Based on these utilities and the underlying class of games, we will deduce the utility $v_i(S)$ that i associates with some coalition $S \in \mathcal{N}_i$. The preferences of i over two coalitions $S, T \in \mathcal{N}_i$ are then given by assuming that $S \succsim_i T$ if and only if $v_i(S) \geq v_i(T)$. A hedonic game $(N, \succsim)$ is an *additively separable hedonic game* (ASHG) if there is $(v_i(j))_{i,j \in N}$ such that, for every agent i, the preferences $\succsim_i$ are induced by the cardinal utilities given by $v_i(S) = \sum_{j \in S} v_i(j)$.

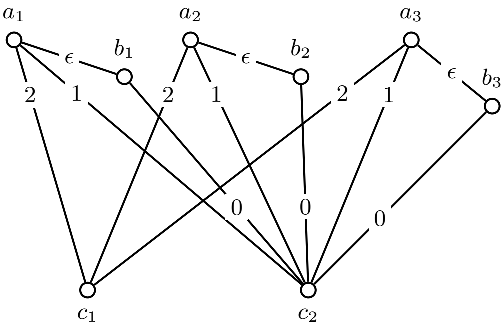


Figure 2: Instance of an additively separable hedonic game with no popular partition. Omitted edges have weight $-K$.

The hedonic game $(N, \succsim)$ is a *fractional hedonic game* (FHG) if there exists $(v_i(j))_{i,j \in N}$ such that, for every agent i, the preferences $\succsim_i$ are induced by the cardinal utilities given by $v_i(S) = (\sum_{j \in S} v_i(j))/|S|$, for $S \subseteq N$. We focus on *symmetric* ASHGs and FHGs, i.e., games for which $v_i(j) = v_j(i)$ for all $i, j \in N$ and denote the symmetric utilities by $v(i, j) = v_i(j) = v_j(i)$.

All hardness results in this section are obtained by rather involved reductions from X3C.

4.3.1 Additively separable hedonic games

We start by having a look at an example of an ASHG that contains no popular partition and that will be used as a gadget in the hardness construction. There are smaller ASHGs without a popular partition, but the instance of the proposition satisfies further properties required for the reduction of Theorem 8 to work. All games considered in this section only contain a single negative weight, whose absolute value is large enough to ensure that certain coalitions will not form.

Proposition 9. *Let $0 < \epsilon < 1$ and $K \geq 4$. Consider the following ASHG, depicted in Figure 2 with agent set $N = \{a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2\}$ and utilities given by $v(a_i, c_1) = 2, v(a_i, c_2) = 1, v(a_i, b_i) = \epsilon, v(b_i, c_2) = 0$ for all $i \in [3]$ and $v(x, y) = -K$ for all other values not defined, yet. Then, there exists no popular partition.*

Proof. Assume for contradiction that π was a popular partition. Then the following facts hold:

- $a_i \notin \pi(a_j), i \neq j$,
- $a_i \notin \pi(b_j), i \neq j$,
- $b_i \notin \pi(b_j), i \neq j$, and
- $c_1 \notin \pi(c_2), c_1 \notin \pi(b_j)$.

In all of these cases, dissolving the coalition in question would be more popular, because all but possibly one agent in the coalition have negative utility and an agent with non-negative utility can only be contained in the coalition if it contains at least 3 agents. Note that K is larger than the sum of positive weights incident to any agent and therefore its utility is negative once it is in a coalition with an agent that gives negative utility.

Now, for every j, exactly one of the following holds: $c_1 \in \pi(a_j)$ or $b_j \in \pi(a_j)$. In fact, both cannot hold as excluded above. If none holds, then $\pi(a_j) \subseteq \{a_j, c_2\}$ and we could delete b_j from her coalition (making no agent worse) and add it to $\pi(a_j)$, resulting in a more popular partition.

Next, for $i \in [2]$, there exists j with $c_i \in \pi(a_j)$. Otherwise, there existed k with $\pi(a_k) \subseteq \{a_k, b_k\}$ and removing b_k and adding c_i is more popular.

Thus, up to symmetry, the only possibility is $\pi = \{\{a_1, c_1\}, \{b_1\}, \{a_2, c_2, b_2\}, \{a_3, b_3\}\}$. But then $\{\{a_2, c_1\}, \{b_2\}, \{a_3, c_2, b_3\}, \{a_1, b_1\}\}$ is more popular. Hence, π was not popular. □

We now discuss the proof strategy for showing that computing popular partitions in symmetric ASHGs is NP-hard.

For a reduction from X3C, given an instance (R, S), we have R-gadgets for every element of the ground set R and S-gadgets for every 3-elementary set in S. The gadgets for elements of R rely on the ASHG of Proposition 9. The gadget for a set $s \in S$ consists of three agents that are very happy in a coalition of their own, but one of them is linked to the R-gadgets corresponding to the agents in s and can simultaneously prevent the agents in these R-gadgets from voting down a partition. This is of course at the expense of the happiness of agents in the S-gadgets and can only happen if all three R-gadgets are simultaneously dealt with. This is where we achieve the correspondence of the covering with 3-partitions, which we can read off from the coalitions of the agents in S-gadgets.

Theorem 8. *Checking whether there exists a popular partition in a symmetric ASHG is NP-hard.*

The verification problem for ASHGs turns out to be coNP-complete. The proof of Theorem 9 is simpler than Aziz et al.'s ((2013b)) proof of a weaker statement for ASHGs that do not have to be symmetric.

Theorem 9. *Checking whether a given partition in a symmetric ASHG is popular is coNP-complete.*

The reductions for coNP-hardness of mixed and strong popularity as well as popularity on ASHGs rely on the idea of property PP which we already employed in Lemma 1. The next lemma establishes this property and is subsequently applied to prove the next four theorems. Note that it is not possible to leverage the relationship of mixed popularity and Pareto optimality, because Pareto optimal partitions can be found in polynomial time for symmetric ASHGs (Bullinger, 2020).

Lemma 3. *The class of symmetric ASHGs satisfies property PP.*

We obtain several hardness results.

Theorem 10. *Checking whether there exists a strongly popular partition in a symmetric ASHG is coNP-hard.*

Theorem 11. *Verifying whether a given partition in a symmetric ASHG is strongly popular is coNP-complete.*

Theorem 12. *Computing a mixed popular partition in a symmetric ASHG is NP-hard.*

We even obtain coNP-hardness of the existence of popular partitions which makes it unlikely that the existence problem for symmetric ASHGs is in NP (otherwise coNP = NP) and, together with Theorem 8, might be seen as evidence that this problem is even Σ_2^p-complete.

Theorem 13. *Checking whether there exists a popular partition in a symmetric ASHG is coNP-hard.*

4.3.2 Fractional hedonic games

We now turn to FHGs. In general, reduction proofs for FHGs tend to be more complicated than for ASHGs, because utility functions are not additive. On top of that, negative utilities have very different consequences in ASHGs and FHGs. In ASHGs with non-negative utility functions, the grand coalition will form under any set of reasonable assumptions because it is the best possible coalition for all agents. The same is not true for FHGs, which incentivize small coalitions by having the size of a coalition in the denominator of utility functions. Hence, in contrast to ASHGs, FHGs are meaningful in the absence of negative utilities and it is therefore desirable to prove hardness results that even hold for non-negative utilities.

Before investigating popularity, we quote a useful proposition about the structure of top-ranked coalitions in FHGs.

Proposition 10 (Bullinger (2020)). *Let a FHG $(N, \succsim)$ be given and let $i \in N$ be an agent. Let μ be the utility of a top-ranked coalition of agent i. Then, the top-ranked coalitions of agent i are precisely the coalitions of the form $\{i\} \cup \{j \in N : v_i(j) > \mu\} \cup W$ for $W \subseteq \{j \in N : v_i(j) = \mu\}$.*

In other words, every top-ranked coalition of agent i consists precisely of all agents j whose utility $v_i(j)$ exceeds a certain threshold.

Now, we consider the existence and verification problem for popular partitions in fractional hedonic games. The strategy is similar to the case of ASHGs. Again, there exist gadgets for every element of R and the sets in S. The R-gadgets rely on rather simple graphs, namely stars.

We define by S_k the star graph with k leaves, i.e., $S_k \cong G$, where $G = (V, E)$ with $V = \{c, l_1, \dots, l_k\}$, $E = \{\{c, l_j\} : j \in [k]\}$. We say that an FHG is induced by S_k if its agent set is $N = V$, and symmetric, binary utilities are given by $v(i, j) = 1$ if $\{i, j\} \in E$ and $v(i, j) = 0$, otherwise, where $i, j \in N$. The next proposition classifies, which star graphs induce FHGs admitting popular partitions. The boundary cases are illustrated in Figure 3.

Proposition 11. *Let $k \in \mathbb{N}$ and consider the FHG induced by S_k.*

For $k \leq 5$, the (sub-)partition (of) $\pi = \{\{c, l_1, l_2, l_3\}, \{l_4\}, \{l_5\}\}$ is popular. For $k \geq 6$, S_k admits no popular partition.

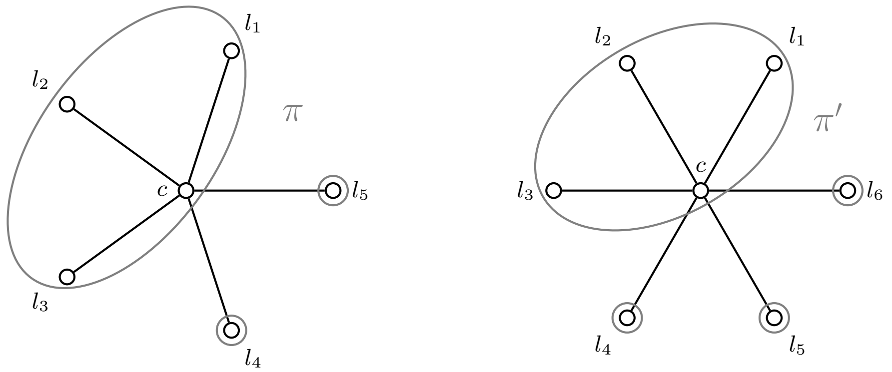


Figure 3: FHGs induced by stars. For stars with 5 leaves, a popular partition π exists (left). This is not the case for stars with more leaves (right). For instance, the grand coalition is more popular than partition π'.

Proof. The first part is easily seen.

For the second assertion, let $k \geq 6$ and assume that π was a popular partition. Then, $|\pi(c)| \leq 4$, since otherwise we obtain a more popular partition if one leaf leaves $\pi(c)$. But in this case, the grand coalition is more popular (having c and at least $k - 3$ leaves better off). □

Using stars as gadgets, we can prove the next theorem.

Theorem 14. *Checking whether there exists a popular partition in a symmetric FHG is NP-hard, even if all utilities are non-negative.*

The hardness proof for the verification problem for FHGs is a more involved version of the proof for ASHGs.

Theorem 15. *Checking whether a given partition in a symmetric FHG is popular is coNP-complete, even if all utilities are non-negative and the underlying graph is bipartite.*

The graphs used in the proof of Theorem 15 have girth 6. This is in contrast to the polynomial-time algorithm by Aziz et al. (2019) for computing the core in FHGs with girth at least 5.

As in the case of ASHGs, we now consider strong and mixed popularity for FHGs. First, we derive property PP for FHGs. The underlying graph is almost identical to the one for ASHGs, which might be surprising, because the utilities for ASHGs and FHGs induced by the same graph will in general cause very different preferences over coalitions. However, all coalitions that actually matter for the particular instance we consider are of size 2 and 3 and therefore the different game models behave very similarly.

Lemma 4. *The class of symmetric FHGs with non-negative utility functions satisfies property PP.*

The proof of the hardness of the existence of strongly popular partitions on FHGs is very similar to the case of ASHGs, but there are some subtle differences regarding the preferences of the additional agent.

Theorem 16. *Checking whether there exists a strongly popular partition in a symmetric FHG is coNP-hard, even if all utilities are non-negative.*

Theorem 17. *Verifying whether a given partition in a symmetric FHG is strongly popular is coNP-complete, even if all utilities are non-negative.*

Theorem 18. *Computing a mixed popular partition in a symmetric FHG is NP-hard, even if all utilities are non-negative.*

As for ASHGs, we can pinpoint the complexity of the existence of popular partitions more exactly. The general proof idea is the same, but the case analyses are simpler, because we can choose positive utilities of the auxiliary agents, which can never help the original agents in the copies of the game in Lemma 4.

Theorem 19. *Checking whether there exists a popular partition in a symmetric FHG is coNP-hard, even if all utilities are non-negative.*

5. Conclusion

We have investigated the computational complexity of finding and recognizing popular, strongly popular, and mixed popular partitions in various types of ordinal hedonic games and cardinal hedonic games. Tables 1 and 2 summarize our results and give an overview of the complexity for computing a respective partition. In the tables, NP-hardness refers to intractability of the corresponding search problem, which follows directly from NP-hardness or coNP-hardness of the existence problem via a Turing reduction. Note that both NP-hardness and coNP-hardness of the existence problem for popularity hold for flatmate games, ASHGs, and FHGs, where the NP-hardness for flatmate games follows from the hardness for roommate games. It is open whether these problems are even Σ_2^p-complete. Whenever we obtain hardness of an existence problem, the corresponding verification problem is coNP-complete. For mixed popularity, this follows from Proposition 3.

Two important factors that govern the complexity of computing these partitions in ordinal hedonic games are whether preferences may contain ties and whether coalitions of size 3 are allowed. When preferences are weak, computing mixed popular and strongly popular partitions is only difficult for representations for which we cannot even compute Pareto optimal partitions efficiently. For strict preferences, however, Pareto optimal partitions can be found efficiently while computing popular, mixed popular, and strongly popular partitions remains intractable. These results are quite robust and all results for flatmate games hold even when preferences are globally ranked, while this restriction allows for tractability of popularity under strict preferences in roommate games. It can be shown that our hardness results remain intact for tripartite matching (with strict and globally ranked preferences), where the agents can be partitioned into three groups and individually rational coalitions

	weak preferences				*strict preferences*			
	PO	mPop	sPop	Pop	PO	mPop	sPop	Pop
IRLC	↑	↑	↑	↑	in P	↑	↑	↑
Flatmates	NP-h.[a]	NP-h. (Th. 2)	NP-h. (Th. 3)		↓	NP-h. (Th. 4)	NP-h. (Th. 3)	
Roommates	in P[b]	in P (Th. 1)	in P (Cor. 3)			in P (Th. 1)	in P[d]	NP-h.[g]
Marriage	↓	↓	↓	NP-h.[e]		↓	↓	in P[f]
Housing				in P[c]	in P	in P[h]	in P	in P[c]

Table 1: Complexity of finding popular and Pareto optimal partitions in various classes of hedonic games. New results are highlighted in gray and implications are marked by gray arrows. NP-hardness of computing a popular or strongly popular partition always follows by a Turing reduction from the existence problem.
[a]: Aziz et al. (2013a, Th. 5), [b]: Aziz et al. (2013a, Th. 7), [c]: Abraham et al. (2007, Th. 3.9), [d]: Biró et al. (2010, Th. 6), [e]: Biró et al. (2010, Th. 11), Cseh et al. (2015, Th. 2), [f]: Gärdenfors (1975, Th. 3), [g]: Gupta et al. (2019, Th. 1.1), Faenza et al. (2019, Th. 4.6), Cseh and Kavitha (2018, Th. 2), [h]: Kavitha et al. (2011, Th. 2); the result by Kavitha et al. also holds for marriage games and weak preferences; these cases are implied by our Th. 1.

	PO	PO/IR	mPop	sPop	Pop
symmetric ASHGs	in P[a]	NP-h.[a]	NP-h. (Th. 12)	NP-h. (Th. 10)	NP-h. (Th. 8, 13)
symmetric FHGs	in P (0/1)[a]	NP-h.[a]	NP-h. (Th. 18)	NP-h. (Th. 16)	NP-h. (Th. 14, 19)

Table 2: Complexity of finding popular and Pareto optimal partitions in cardinal hedonic games. New results are highlighted in gray. NP-hardness of computing a popular or strongly popular partition always follows by a Turing reduction from the existence problem. Pareto optimal partitions in FHGs can be computed in polynomial time for (0/1)-preferences.
[a]: Bullinger (2020, Th. 5.1, 5.1, 6.2, 6.4)

may only contain at most one agent of each group.[8] An interesting avenue for future research is to consider the three notions of popularity in further restrictions of flatmate games such as room-roommate games or three-cyclic matching games.[9] Notably, the related existence problem for stable three-dimensional matchings has also been shown to be NP-hard (Lam & Plaxton, 2019).

Our positive results for roommate games are obtained via a single linear programming approach that unifies a number of existing results and exploits the relationships between the different types of popularity. On the other hand, both in flatmate games and cardinal hedonic games, our hardness results are based on the same central idea, formalized via

8. The reduction for this result changes the source problem for the reduction in Lemma 1 to 3-Dimensional Matching instead of Exact 3-Cover, and consists essentially of finding the tripartition of the agent set in the existing reduction by placing the agents corresponding to the elements of the ground set of a source instance the right way in the top layer in Figure 1.
9. Some advances in this direction were recently made by Cseh and Peters (2022).

property PP. All of these classes of hedonic games contain games with a strongly popular partition together with an agent that can govern the switch between strong popularity and non-popularity by joining different sets of additional auxiliary agents. As a consequence, results for all types of popularity and for both existence and verification problems can be extracted from the same reduction.[10]

Since mixed popular partitions always exist, the natural computational problem is the search problem. We have shown that intractability of this problem can be inferred from corresponding results on Pareto optimality. Moreover, we prove the hardness of computing mixed popular partitions in classes of games in which Pareto optimal partitions can be found efficiently.[11] In all our reductions, it is already hard to compute some (deterministic) partition in the support of a mixed popular partition, i.e., a subset of Pareto optimal partitions.

Acknowledgments

This material is based on work supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/12-1. The authors thank Stefan Weltge and three anonymous reviewers for providing valuable feedback. Preliminary results of this paper were presented at the 19th International Conference on Autonomous Agents and Multi-Agent Systems (May, 2020).

Appendix A. Omitted Proofs

The appendix contains all omitted proofs.

A.1 Ordinal Hedonic Games

We introduce a useful notation for the next two propositions. Given a matching M and an agent a, denote by $M(a)$ the agent, a is matched with.

Proposition 6. *Let p and q be mixed matchings. Then,*

$$\phi(p, q) = \phi(x_p, x_q).$$

In particular, p is popular if and only if for all matchings M, $\phi(x_p, \chi_M) \geq 0$.

Proof. Let p and q be two mixed matchings. By extending them with some matchings of probability 0, we may assume that both are defined on the same set of matchings $M_1, \ldots, M_k$ as $p = \{(M_1, p_1), \ldots, (M_k, p_k)\}$ and $q = \{(M_1, q_1), \ldots, (M_k, q_k)\}$. We derive that

10. The careful reader might have noticed that we do not extract the proofs of the verification problem for ASHGs and FHGs from this general approach. While this is also possible, we have obtained independent proofs which hold for a more restrictive variant of FHGs, and allow the comparison with existing results about the complexity of computing partitions in the core.

11. Note that the complexity of Pareto optimality in FHGs under arbitrary symmetric weights is still open. As indicated in Table 2, Bullinger (2020) only settles the problem for some restricted classes of FHGs including binary utilities.

$$
\begin{aligned}
\phi(p,q) &= \sum_{s,t=1}^{k} p_s q_t \phi(M_s, M_t) \\
&= \sum_{s,t=1}^{k} p_s q_t \sum_{a\in N} \phi_a(M_s(a), M_t(a)) \\
&= \sum_{s,t=1}^{k} p_s q_t \sum_{a\in N} \sum_{i,j\in N_G(a)} \chi_{M_s}(a,i)\chi_{M_t}(a,j)\phi_a(i,j) \\
&= \sum_{a\in N} \sum_{i,j\in N_G(a)} \left(\sum_{s=1}^{k} p_s \chi_{M_s}(a,i)\right)\left(\sum_{t=1}^{k} q_t \chi_{M_t}(a,j)\right)\phi_a(i,j) \\
&= \sum_{a\in N} \sum_{i,j\in N_G(a)} x_p(a,i)x_q(a,i)\phi_a(i,j) \\
&= \phi(x_p, x_q).
\end{aligned}
$$

This proves the desired equality. □

Proposition 7. *The separation problem for P_{Pop} can be solved in polynomial time.*

Proof. Assume that a vector $x \in \mathbb{R}^E$ is given. The separation problem for the matching polytope can be solved in polynomial time. For the popularity constraints, we assign weights w_x to the edges of the underlying graph such that for all matchings M on G, $w_x(M) = \phi(\chi_M, x)$. Therefore, their separation problem turns into finding a maximum weight matching, which can be done in polynomial time.

We define the weights by letting

$$
w_x(i,j) = \sum_{a\in N_G(i)} x(i,a)\phi_i(j,a) + \sum_{a\in N_G(j)} x(j,a)\phi_j(i,a)
$$

and compute

$$
\begin{aligned}
\phi(\chi_M, x) &= \sum_{a\in N} \sum_{i,j\in N_G(a)} \chi_M(a,i)x(a,j)\phi_a(i,j) \\
&= \sum_{a\in N} \sum_{i,j\in N_G(a)} \chi_M(a,i)x(a,j)\phi_a(i,j) \\
&= \sum_{a\in N} \sum_{j\in N_G(a), i=M(a)} x(a,j)\phi_a(i,j).
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
w_x(M) &= \sum_{\{i,j\}\in M}\left[\sum_{b\in N_G(i)} x(i,b)\phi_i(j,b) + \sum_{b\in N_G(j)} x(j,b)\phi_j(i,b)\right]\\
&= \sum_{\{i,j\}\in M}\left[\sum_{b\in N_G(i), j=M(i)} x(i,b)\phi_i(j,b) + \sum_{b\in N_G(j), i=M(j)} x(j,b)\phi_j(i,b)\right]\\
&= \sum_{a\in N, a \text{ matched}} \sum_{j\in N_G(a), i=M(a)} x(a,j)\phi_a(i,j)\\
&= \sum_{a\in N} \sum_{j\in N_G(a), i=M(a)} x(a,j)\phi_a(i,j)
\end{aligned}
$$

The last equation is due to the fact that the inner sum is empty for unmatched agents in M. Putting everything together, we conclude that $\phi(\chi_M, x) = w_x(M)$, which completes the proof.

□

Theorem 4. *Computing a mixed popular partition in flatmate games is NP-hard, even if preferences are strict and globally ranked.*

Proof. We provide a Turing reduction from X3C to the problem of finding a partition in the support of a mixed popular partition together with its probability in this mixed partition.

Given an instance $X3C$, we construct a very similar game as in the proof of Theorem 3. We have $N' = N \cup \{z_1, z_2\}$ where the agents N are as in the above construction with identical preferences, except for changing the preferences of agent c_0^1 to $\{a_0^1, b_0^1, c_0^1\} \succ_{c_0^1} \{c_0^1, z_1, z_2\} \succ_{c_0^1} \{c_0^1\}$, and $\{c_0^1, z_1, z_2\} \succ_{z_i} \{z_i\}$ for $i \in [2]$. By a case distinction similar to the one in the proof of Theorem 3 and using Lemma 1, it follows that the partition $\pi^* = \{\{a_j^i, b_j^i, c_j^i\} \colon j \in [0,k], i \in [2^j]\} \cup \{\{\alpha_j^i, \beta_j^i\}, \{\gamma_j^i, \delta_j^i\} \colon j \in [0,k], i \in [2^j] \text{ odd}\} \cup \{\{y_1^i, y_2^i\} \colon i \in [|R|+1, 2^k]\} \cup \{\{z_1\}, \{z_2\}\}$ is strongly popular if there exists no 3-partition of R through sets in S. Therefore the unique mixed popular partition assigns probability 1 to π^*.

On the other hand, assume that there exist a 3-partition $S' \subseteq S$ of R. Define $\pi = \{\{b_k^v, b_k^w, b_k^x\} \colon \{r^v, r^w, r^x\} \in S'\} \cup \{\{b_k^i, y_2^i\}, \{y_1^i\} \colon i \in [|R|+1, 2^k]\} \cup \{\{\delta_{k-1}^i, a_k^{2i-1}, a_k^{2i}\} \colon i \in [2^{k-1}]\} \cup \{\{b_j^i, c_{j+1}^{2i-1}, c_{j+1}^{2i}\}, \{a_j^i, \beta_j^i, \gamma_j^i\} \colon j \in [k-1], i \in [2^j]\} \cup \{\{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\} \colon j \in [k-2], i \in [2^j]\} \cup \{\{\alpha_0^1\}, \{z_1, z_2, c_0^1\}\}$. It is easily checked that $\phi(\pi, \pi^*) = 1$. Therefore, there exists no mixed popular partition that assigns probability 1 to π^*.

We can solve X3C by computing a partition π in the support of a mixed popular partition and checking its probability in case $\pi = \pi^*$. □

A.2 Additively separable hedonic games

Next, we consider the existence problem for ASHGs.

Theorem 8. *Checking whether there exists a popular partition in a symmetric ASHG is NP-hard.*

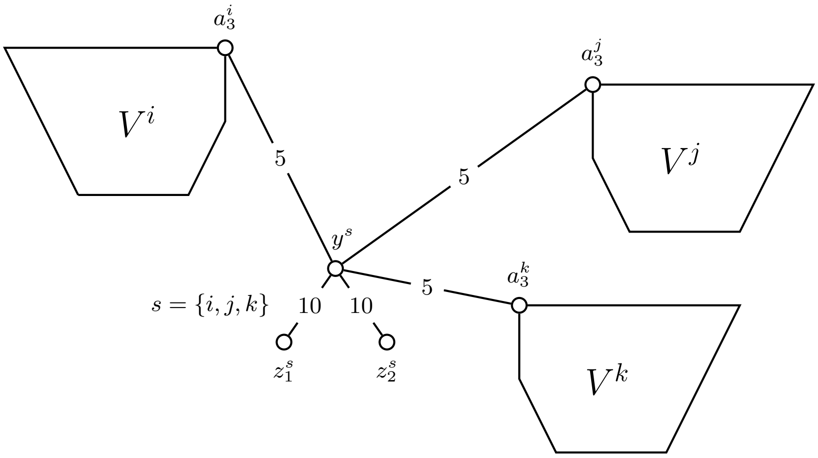


Figure 4: Schematic of the reduction of the existence problem for ASHGs. Edges of weight 0 and of negative weight are omitted.

Proof. The reduction is from X3C to deciding whether there exists a popular partition.

Let (R, S) be an instance of X3C. This can be reduced to an instance $(N, \succsim)$, where $(N, \succsim)$ is an ASHG defined in the following way.

Let $N = \{a_1^r, a_2^r, a_3^r, b_1^r, b_2^r, b_3^r, c_1^r, c_2^r \colon r \in R\} \cup \{y^s, z_1^s, z_2^s \colon s \in S\}$ and edge weights as

- $v(a_i^r, c_1^r) = 2$ and $v(a_i^r, c_2^r) = 1, v(a_i^r, b_i^r) = \epsilon, v(b_i^r, c_2^r) = 0$ for all $i \in [3]$ and $r \in R$,
- $v(a_3^r, a_3^{r'}) = 0, v(b_3^r, a_3^{r'}) = 0, v(b_3^r, b_3^{r'}) = 0$ for all $s \in S$ and $r, r' \in s$,
- $v(a_3^r, y^s) = 5$ and $v(b_3^r, y^s) = 0$ for all $s \in S$ and $r \in R$ such that $r \in s$,
- $v(y^s, z_1^s) = v(y^s, z_2^s) = 10$ and $v(z_1^s, z_2^s) = 0$ for all $s \in S$, and
- $v(x, y) = -40$ for all other valuations not defined.

In order to enable the reduction, we can, for example, choose $\epsilon = \frac{1}{2}$. A schematic of the reduction for a certain set $s = \{i, j, k\} \in S$ is depicted in Figure 4. We abbreviate in the figure and the rest of the proof $V^r = \{a_1^r, a_2^r, a_3^r, b_1^r, b_2^r, b_3^r, c_1^r, c_2^r\}$, where $r \in R$, and $W^s = \{y^s, z_1^s, z_2^s\}$, where $s \in S$. Also denote $V^R = \cup_{r \in R} V^r, W^S = \cup_{s \in S} W^s$ and $A_3 = \{a_3^r \colon r \in R\}$.

We show that there exists a popular partition of $(N, \succsim)$ if and only if (R, S) is a Yes-instance of X3C.

Assume (R, S) is a Yes-instance of X3C. Then, there exists $S' \subseteq S$ such that S' is a partition of R. Consider the partition $\pi = \{\{a_1^r, c_1^r\} \colon r \in R\} \cup \{\{a_2^r, b_2^r, c_2^r\} \colon r \in R\} \cup \{\{b_1^r\} \colon r \in R\} \cup \{\{y^s, a_3^i, a_3^j, a_3^k, b_3^i, b_3^j, b_3^k\} \colon s = \{i, j, k\} \in S'\} \cup \{W^s \colon s \in N \setminus S'\} \cup \{\{z_1^s, z_2^s\} \colon s \in S'\}$. We claim that π is popular.

Assume for contradiction that π' is more popular than π.

We first prove the following two claims:

1. Let $r \in R$ such that for all $s \in S$ with $r \in s$ holds that $y^s \notin \pi'(a_3^r)$. Then, $|N(\pi, \pi') \cap V^r| - |N(\pi', \pi) \cap V^r| \geq 1$.

2. Let $r \in R$. If $|\{y^s : s \in S\} \cap \pi'(a_3^r)| \leq 1$ then, $|N(\pi, \pi') \cap V^r| - |N(\pi', \pi) \cap V^r| \geq 0$. If $|\{y^s : s \in S\} \cap \pi'(a_3^r)| \geq 2$ then, $|N(\pi, \pi') \cap V^r| - |N(\pi', \pi) \cap V^r| \geq -1$.

We start with the proof of the first claim.

Let therefore $r \in R$ such that for all $s \in S$ with $r \in s$ holds that $y^s \notin \pi'(a_3^r)$. Since $r \in R$ is fixed, we omit the superscript r for proving this claim. We know that $a_3 \in N(\pi, \pi')$ and $b_2, b_3 \notin N(\pi', \pi)$ We distinguish several cases:

- First, consider the case that $c_1 \in \pi'(a_1)$. Then, $b_1, a_2 \notin N(\pi', \pi)$. In addition, we may assume $a_1 \notin N(\pi', \pi)$, because otherwise $c_1, c_2 \in N(\pi, \pi')$ and the claim is true.

  If $c_i \in N(\pi', \pi)$, then $c_{3-i} \notin N(\pi', \pi)$ and either $(a_1 \in N(\pi, \pi') \vee a_2 \in N(\pi, \pi')) \wedge b_3 \in N(\pi, \pi')$ or $a_1, a_2 \in N(\pi, \pi')$. In every case, $|N(\pi', \pi)| \leq 2$ and $|N(\pi, \pi')| \geq 3$ and the claim follows.

  Hence, we may assume that $c_i \notin N(\pi', \pi)$ and no agent can be in $N(\pi', \pi)$. In this case, the claim follows.

- Second, assume $c_1 \in \pi'(a_2)$. Then, $a_1, b_2 \in N(\pi, \pi')$. If $a_2 \notin N(\pi', \pi)$, then it has a negative neighbor, i.e., $a_2 \in N(\pi, \pi')$. We have $|N(\pi, \pi')| \geq 4, |N(\pi', \pi)| \leq 3$.

  Hence, $a_2 \in N(\pi', \pi)$. As a consequence, $c_1 \notin N(\pi', \pi)$ and $c_2 \notin N(\pi', \pi) \vee b_1 \notin N(\pi', \pi)$ and we conclude with $|N(\pi, \pi')| \geq 3, |N(\pi', \pi)| \leq 2$.

- Third, assume $c_1 \in \pi'(a_3)$. Then, $a_1, b_3 \in N(\pi, \pi')$. If $c_2 \in \pi'(a_3)$, then $c_1, c_2, a_2 \in N(\pi, \pi')$ and we conclude with $|N(\pi, \pi')| \geq 6$. If $c_2 \notin \pi'(a_3)$, then $\{a_1, a_3, b_3\} \subseteq N(\pi, \pi')$ and $a_2, b_2 \notin N(\pi', \pi)$ and either $b_2 \in N(\pi, \pi')$ or $c_2 \notin N(\pi', \pi)$.

- Finally, assume $c_1 \notin \pi'(a_1) \cup \pi'(a_2) \cup \pi'(a_3)$. Then $a_1, c_1 \in N(\pi, \pi')$ and $a_2 \notin N(\pi', \pi) \vee c_2 \notin N(\pi', \pi)$. Hence, $|N(\pi, \pi')| \geq 3, |N(\pi', \pi)| \leq 2$. This concludes the proof of the first claim.

Before we prove the second claim, we argue that we can assume without loss of generality that for all $r \in R$, $\pi'(a_3^r) \cap V^r \subseteq \{a_3^r, b_3^r\} \vee \{y^s : s \in S\} \cap \pi'(a_3^r) = \emptyset$. Indeed, if both conditions are not met, then leaving with $y^s \in \{y^s : s \in S\} \cap \pi'(a_3^r)$ and forming a coalition with W^s yields a partition π'' with the following properties:

- $|N(\pi'', \pi) \cap (N \setminus W^s)| \geq |N(\pi', \pi) \cap (N \setminus W^s)| - 1$ (Note that the only agent that is not still better off is possibly a_3^r since the other $a_3^{r'}$ are worse off since they would get negative utility in $\pi'(a_3^r)$.),

- $|N(\pi, \pi'') \cap (N \setminus W^s)| \geq |N(\pi, \pi') \cap (N \setminus W^s)| + 1$ (the only candidate is again a_3^r),

- $|N(\pi'', \pi) \cap W^s| \geq |N(\pi', \pi) \cap W^s| + 3$ if $\pi(y^s) \neq W^s$, and

- $|N(\pi, \pi'') \cap W^s| \geq |N(\pi, \pi') \cap W^s| - 3$ if $\pi(y^s) = W^s$.

Other changes in W^s cannot occur at the same time and we conclude $\phi(\pi'', \pi) \geq \phi(\pi', \pi)$ (in fact the inequality is strict).

For the second claim, this means that if some $y^s \in \pi'(a_3^r)$ we can consider π' modified such that y^s leaves her coalition. This can only decrease the size of $N(\pi', \pi) \cap V^r$ if $|\{y^s \colon s \in S\} \cap \pi'(a_3^r)| \geq 2$ and cannot increase the size of $N(\pi, \pi') \cap V^r$ by more than 1. Hence, the claim follows from the first case.

We define the set of critical subsets $s \in S$ as $Y^c = \{s \in S : \exists r \in R \text{ with } y^s \in \pi'(a_3^r)\}$ and the set of happy R gadgets as $R^h = \{r \in R \colon |\{y^s \colon s \in S\} \cap \pi'(a_3^r)| \geq 2\}$.

We know that for every $y^s \in Y^c$ at most 3 of the a_3^r do not satisfy the condition of the first claim. Hence, a total of $\max\{|R| - 3|Y^c| + |R^h|, 0\}$ of the agents a_3^r does so. Putting together the claims yields

$$
\begin{aligned}
&|N(\pi, \pi') \cap V^R| - |N(\pi', \pi) \cap V^R| \\
&\geq \max\{|R| - 3|Y^c| + |R^h|, 0\} - |R^h| \geq |R| - 3|Y^c|.
\end{aligned} \tag{1}
$$

We claim that in addition

$$
|N(\pi', \pi) \cap W^S| - |N(\pi, \pi') \cap W^S| \leq |R| - 3|Y^c|. \tag{2}
$$

The idea to prove this inequality is that every agent y^s has to decide whether the agents in W^s or the a_3^r with $r \in s$ should be happy. Without loss of generality, we can assume that for all $s \in S$, $\pi(y^s) \cap A_3 = \emptyset$ or $\pi(y^s) \cap W^s = \{y^s\}$. Indeed, if both conditions are not met, then leaving with y^s and forming a coalition with W^s yields a partition π'' with $\phi(\pi'', \pi) \geq \phi(\pi', \pi)$.

To prove Equation (2) note that $W^s \subsetneq N(\pi, \pi') \cap W^S$ for every $s \in Y^c$ such that $\pi(y^s) = W^s$. In other words, $|N(\pi, \pi') \cap W^S| \geq 3|\{s \in Y^c \colon \pi(y^s) = W^s\}|$.

In addition, the only agents that get better in W^S can be in a W^s such that $\pi(y^s) \neq W^s$ and $y^s \notin Y^c$. This is, $|N(\pi', \pi) \cap W^S| \leq 3|\{s \notin Y^c \colon \pi(y^s) \neq W^s\}|$.

Combining the inequalities yields

$$
\begin{aligned}
&|N(\pi', \pi) \cap W^S| - |N(\pi, \pi') \cap W^S| \\
&\leq 3(|\{s \notin Y^c \colon \pi(y^s) \neq W^s\}| - |\{s \in Y^c \colon \pi(y^s) = W^s\}|) \\
&= 3(|\{s \notin Y^c \colon \pi(y^s) \neq W^s\}| + |\{s \in Y^c \colon \pi(y^s) \neq W^s\}| \\
&- |\{s \in Y^c \colon \pi(y^s) \neq W^s\}| - |\{s \in Y^c \colon \pi(y^s) = W^s\}|) \\
&= 3|S'| - 3|Y^c| = |R| - 3|Y^c|.
\end{aligned}
$$

Combining Equation (1) and Equation (2) yields $|N(\pi, \pi')| - |N(\pi', \pi)| \geq 0$, contradicting the assumption that π' was more popular than π.

It remains to prove that every popular partition yields a 3-partition of R with sets in S. Therefore, assume that π is a popular partition in $(N, \succsim)$. The partition will be found by checking intersections of $\pi(y^s) \cap A_3$ as captured in the following claims:

1. For all $r \in R$ there exists a unique $s \in S$ with $y^s \in \pi(a_3^r)$. For this s holds that $r \in s$.
2. For all $s \in S$ holds: $(\exists i \in s : a_3^i \in \pi(y^s)) \Rightarrow (\forall j \in s, a_3^j \in \pi(y^s))$.

If the claim is true, $S' = \{s \in S : A_3 \cap \pi(y^s) \neq \emptyset\}$ covers R due to existence and is a partition due to uniqueness and the second claim that ensures that either all three or none of the agents in A_3 corresponding to elements in s are present in a coalition $\pi(y^s)$.

We start to show the existence part of the first claim which will follow directly from the property that $N|_{V^r}$ contains no popular partition (Proposition 9).

Assume for contradiction that there exists a $r \in R$ such that for all $s \in S$ holds $y^s \notin \pi(a_3^r)$. We obtain a more popular partition in two steps. First, we modify π such that for all agents in $v \in V^r$ we split their coalition into $\pi(v) \cap V^r$ and $V^r \setminus \pi(v)$. This cannot decrease the utility of any agent. Application of Proposition 9 yields a more popular partition locally on V^r that can be extended to the whole N via the remaining (modified) coalitions in π.

For the uniqueness part assume for contradiction that there is $r \in R$ and $s \neq s' \in S$ with $\{y^s, y^{s'}\} \subseteq \pi(a_3^r)$. We distinguish two cases.

First, assume that $|\pi(a_3^r) \cap A_3| \leq 3$. Then, there exists (without loss of generality using symmetry amongst s and s') an agent $r' \in R$ with $r' \in s$ and $a_3^{r'} \notin \pi(a_3^r)$. Then, the partition π' obtained from π by removing the agents in W^s from their partitions in π and letting them form a coalition is more popular. Indeed, $|N(\pi, \pi')| \leq 2$ (the two remaining agents a_3^t with $t \neq r'$ and $t \in s$ are the only ones to possibly loose utility) and $W^s \subseteq N(\pi', \pi)$.

Second, assume that $|\pi(a_3^r) \cap A_3| \geq 4$. Then, there exists an agent $u \in A_3 \cap \pi(a_3^r)$ with $u \notin s$. The same partition π' as in the first case yields $|N(\pi, \pi')| \leq 3$ and $|N(\pi', \pi)| \geq |W^s \cup \{u\}| = 4$.

In both cases, we have found a more popular partition, a contradiction.

Finally, for the second claim, in the case that there exists a $s \in S$ with $1 \leq |\{j \in s : a_3^j \in \pi(y^s)\}| \leq 2$, the same rearrangement of coalitions (i.e., forming the coalition W^s) is more popular. $\square$

Theorem 9. *Checking whether a given partition in a symmetric ASHG is popular is coNP-complete.*

Proof. The problem is in coNP, because a more popular partition serves as a polynomial-time certificate for a No-instance.

For hardness, we reduce again from X3C. Given an instance (R, S) of X3C, we assume without loss of generality that $|R| \geq 6$. We define an ASHG $(N, \succsim)$ given by $N = R \cup \{s_1, s_2, s_3 \colon s \in S\} \cup \{b_1, b_2, b_3\}$ and weights as

- $v(i, s_3) = 1$ for $i \in s, s \in S$,
- $v(s_1, s_3) = v(s_2, s_3) = 4$ for $s \in S$,
- $v(s_j, b_j) = 1$ for $s \in S, j \in [2]$,
- $v(b_1, b_3) = v(b_2, b_3) = \alpha$ for $\frac{|R|}{3} - 1 < \alpha < \frac{|R|}{3}$,
- $v(i, j) = 0$ for $i, j \in R$, $v(s_1, s_2) = 0$ for $s \in S$, and $v(b_1, b_2) = 0$, and
- $v(x, y) = -\max\{12, |S| + |R|/3\}$ for all agents $x, y \in N$ such that no utility is defined, yet.

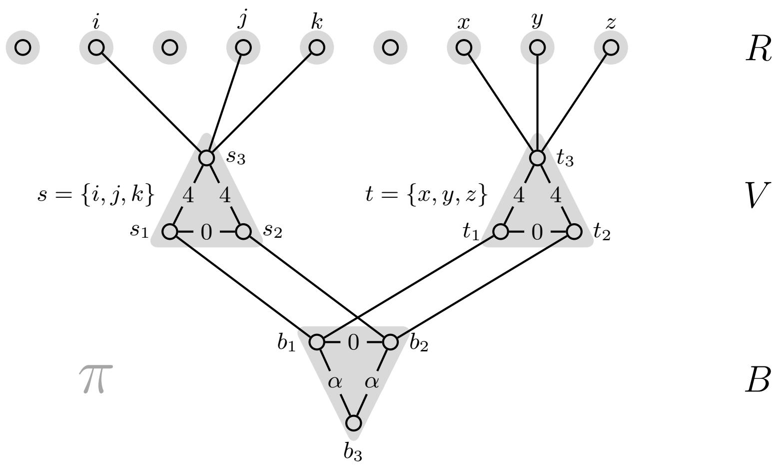


Figure 5: Schematic of the reduction for the verification problem of popular partitions on symmetric ASHGs. Edges without explicit weight have weight 1. Omitted edges for agents in R have weight 0. All other omitted edges have weight -12. The partition π marked in gray is the one under consideration for verification.

One can choose, e.g., $\alpha = (|R| - 1)/3$, but for the reduction, only the above bounds matter. We introduce some useful notation for the proof. Denote $V^s = \{s_1, s_2, s_3\}$ for $s \in S$, $B = \{b_1, b_2, b_3\}$, and $V = \cup_{s \in S} V^s$.

The partition in question is $\pi = \{V^s \colon s \in S\} \cup \{\{r\} \colon r \in R\} \cup \{B\}$. We claim that (R, S) is a Yes-instance of X3C if and only if π is not popular for the ASHG given by G.

If (R, S) is a Yes-instance, there exists a subset $S' \subseteq S$ that partitions R. In particular $|R| = 3|S'|$.

Consider the partition given by $\pi' = \{V^s \colon s \in S \setminus S'\} \cup \{\{s_3, i, j, k\} \colon \{i, j, k\} = s \in S'\} \cup \{\{b_j, s_j \colon s \in S'\} \colon j \in [2]\} \cup \{\{b_3\}\}$.

Then, $N(\pi', \pi) = R \cup \{b_1, b_2\}$ and $N(\pi, \pi') = \cup_{s \in S'} V^s \cup \{b_3\}$. Hence, π' is more popular than π.

Conversely, assume that there exists a more popular partition π' and fix one that maximizes $\phi(\pi', \pi)$. We have to prove that there exists a subset $S' \subseteq S$ that yields a partition of R. Note that the negative weight is chosen so large that agents in a coalition linked by negative utility are always worse off.

First, we claim that for all $s \in S$, $N(\pi', \pi) \cap V^s = \emptyset$. Assume for contradiction that for $j \in [2]$, $s_j \in N(\pi', \pi)$. Then, $\{s_j, s_3, b_j\} \subseteq \pi'(s_j) \subseteq V^s \cup \{b_j\}$. Thus, $s_{3-j}, s_3, b_j, b_3 \in N(\pi, \pi')$.

We form a new coalition π'' from π' by having the coalitions V^s and B (these agents leave their coalitions in π') and all other coalitions remain the same. We consider two cases:

- If $|\pi'(b_{3-j}) \cap V| \leq 1$, then $b_{3-j} \in N(\pi, \pi')$. (We used that $|R| \geq 6$.) We have that $s_3, s_{3-j}, b_1, b_2, b_3 \in N(\pi, \pi') \setminus N(\pi, \pi'')$, $s_2 \in N(\pi', \pi) \setminus N(\pi, \pi'')$ and possibly the agent $t \in \pi'(b_{3-j}) \cap V$ yields $t \in N(\pi', \pi) \cap N(\pi, \pi'')$. Hence, $\phi(\pi'', \pi) > \phi(\pi', \pi)$.

- Otherwise, $\pi'(b_{3-j})\cap V \subseteq N(\pi,\pi')$, but possibly $b_{3-j} \in N(\pi',\pi)\setminus N(\pi,\pi'')$ in addition. However, $\phi(\pi'',\pi) > \phi(\pi',\pi)$ remains valid.

In any case, we derived a contradiction to the maximality condition on π'.

If $s_3 \in N(\pi',\pi)$, then $\{s_1,s_2\} \subseteq \pi'(s_3)$, $s \cap \pi'(s_3) \neq \emptyset$, and $\pi'(s_3) \subseteq V^s \cup s$ (here $s \subseteq R$ is the set of R-agents corresponding to elements of the set s). Hence, forming a coalition π'' by leaving with the agents in s moves these agents and s_1, s_2 out of $N(\pi,\pi')$, while only removing s_3 from $N(\pi',\pi)$. Hence, we again contradict the maximality of $\phi(\pi',\pi)$.

For the rest of the analysis, we narrow down the possible more popular partitions to a very specific situation that corresponds to 3-partitions. The idea is basically that whenever we 'sacrifice' a set V^s of agents, we can improve only 3 agents in R. Due to the boundaries on α, we will cross the threshold, where we can have a popularity margin of precisely 1 exactly at the moment when we gathered $\frac{|R|}{3}$ neighbors for b_1 and b_2 in order to improve these.

We introduce the sets $R_I = R \cap N(\pi',\pi)$ and $S_C = \{s \in S \colon \pi'(s_3) \cap R \neq \emptyset\}$. Our goal is to prove $|R| = |R_I| = 3|S_C|$.

For $s \in S_C$ holds $V^s \subseteq N(\pi,\pi')$ (which follows for s_3 since $s_3 \notin N(\pi',\pi)$). Consequently, $|N(\pi,\pi') \cap V| \geq 3|S_C|$. In addition, $|N(\pi',\pi) \cap R| = |R_I| \leq 3|S_C|$ and $\phi_B(\pi',\pi) \leq 1$.

If $|R_I| < 3|S_C|$, then $\phi(\pi,\pi') = \phi_B(\pi,\pi') + \phi_V(\pi,\pi') + \phi_R(\pi,\pi') \geq -1 + 3|S_C| - (|R_I|) = 3|S_C| - |R_I| - 1 \geq 0$ and π' is not more popular. We conclude that $|R_I| = 3|S_C|$.

Before we conclude the proof, we show two auxiliary claims:

1. If $B \subseteq \pi'(b_3)$, then $b_1 \notin N(\pi',\pi)$ or $b_2 \notin N(\pi',\pi)$.
2. For $j \in [2]$, if $b_j \in N(\pi',\pi)$, then $b_j \in \pi'(b_3)$ or $|\{s \in S \colon s_j \in \pi'(b_j)\}| \geq \frac{|R|}{3}$.

The first claim follows from the fact that if b_j forms a coalition with an agent outside B that gives her positive utility, then b_{3-j} cannot be both in this coalition and improve her utility. The second claim follows from $v_\pi(b_j) = \alpha > \frac{|R|}{3} - 1$.

We are ready to prove $|R| = 3|S_C|$. We consider the agents in B. The only possibility for $\phi(\pi',\pi) > 0$ is that $\phi_B(\pi',\pi) \geq 1$ which can only happen if $\{b_1,b_2\} \subseteq N(\pi',\pi)$. Due to the auxiliary claims, there exists $j \in \{1,2\}$ with $|\{s \in S \colon s_j \in \pi'(b_j)\} \cap \pi'(b_j)| \geq \frac{|R|}{3}$.

If $s^* \in \{s \in S \colon s_j \in \pi'(b_j)\} \setminus S_C$, then $s^*_j \in N(\pi,\pi')$ (using $|R| \geq 6$, i.e., $|\pi'(b_j) \cap \{s \in S \colon s_j \in \pi'(b_j)\}| \geq 2$).[12]

Consequently, $\phi(\pi,\pi') = \phi_B(\pi,\pi') + \phi_V(\pi,\pi') + \phi_R(\pi,\pi') \geq -1 + (3|S_C|+1) - 3|S_C| \geq 0$, a contradiction. Therefore, $\{s \in S \colon s_j \in \pi'(b_j)\} \subseteq S_C$ and $\frac{|R|}{3} \leq |\{s \in S \colon s_j \in \pi'(b_j)\}| \leq |S_C| = \frac{|R_I|}{3} \leq \frac{|R|}{3}$.

Consider the set $S' = S_C$. Then, S_C covers R since $R_I = R$. In addition, since $|R| = 3|S_C|$, every agent $r \in R$ is present in exactly one $s \in S_C$. Hence, S' is a partition of R with sets in S. In total, (R,S) is a Yes-instance of X3C. $\square$

We first prove the existence of the graph that underlies the subsequent reductions for ASHGs. It satisfies similar properties as the flatmate game considered in Lemma 1. However, for the reduction to work, we need two sets of auxiliary agents. The first set corresponds to the 3-elementary sets in S of an instance (R,S) of X3C, while the second set

12. This argument is stronger than what is needed for ASHGs, but it is needed for the case of FHGs.

consists of two agents that allow the agents in the top-level not corresponding to elements of R to improve their coalition.

Lemma 3. *The class of symmetric ASHGs satisfies property PP.*

Proof. Let (R,S) be an instance of X3C. We construct the following game. Let $k = \min\{k \in \mathbb{N}\colon 2^k \geq |R|\}$ define the smallest power of 2 that is larger than the cardinality of R. We define an ASHG on vertex set $N = \{y_1^s, y_2^s\colon s \in S\} \cup \{y_1, y_2\} \cup \bigcup_{j=0}^{k} N_j$, where $N_j = \bigcup_{i=1}^{2^j} A_j^i$ consists of 2^j sets of agents A_j^i.

We define the sets of agents as

- $A_k^i = \{a_k^i, b_k^i, c_k^i\}$ for $i \in [2^k]$, and
- $A_j^i = \{a_j^i, b_j^i, c_j^i, \alpha_j^i, \beta_j^i, \gamma_j^i, \delta_j^i\}$ for $j \in [0, k-1], i \in [2^j]$.

We order the set R in an arbitrary but fixed way, say $R = \{r^1, \dots, r^{|R|}\}$, and for a better understanding of the proof and the preferences, we label the agents $b_k^i = r^i$ for $i \in [|R|]$. If we view the set of agents N as $k+1$ levels of agents, then the ground set R of the instance of X3C is identified with some specific agents in the top level k. We are ready to define the symmetric preferences as

- $v(y_1^s, y_2^s) = 6k + 8$ for all $s \in S$,
- $v(y_2^s, b_k^i) = 2k + 3$ if there exists $s \in S$ with $r^i \in s$,
- $v(y_1, y_2) = 1$,
- $v(y_2, b_k^i) = 2k + 3$, $i \in [|R| + 1, 2^k]$,
- $v(b_k^i, b_k^{i'}) = 0$, $i, i' \in [|R| + 1, 2^k]$,
- $v(b_k^i, b_k^{i'}) = 0$, $i, i' \in [|R|]$,
- $v(a_k^i, b_k^i) = v(a_k^i, c_k^i) = v(b_k^i, c_k^i) = k + 1$, $i \in [2^k]$,
- For $j \in [0, k-1]$, $i \in [2^k]$,
  - $v(a_j^i, b_j^i) = v(a_j^i, c_j^i) = j + 1$, $v(b_j^i, c_j^i) = j + 1.5$,
  - $v(b_j^i, c_{j+1}^{2i-1}) = v(b_j^i, c_{j+1}^{2i}) = j + 1.5$,
  - $v(\alpha_j^i, \beta_j^i) = j + 1$, $v(\beta_j^i, \gamma_j^i) = 0$,
  - $v(\beta_j^i, a_j^i) = j + 1.75$, $v(\gamma_j^i, a_j^i) = j + 1.25$,
  - $v(\gamma_j^i, \delta_j^i) = j + 2$, $v(\delta_j^i, \alpha_{j+1}^{2i-1}) = v(\delta_j^i, \alpha_{j+1}^{2i}) = j + 1.5$, and
- $v(g,h) = -M - 1$ for all $g, h \in N$ such that the utility is not defined, yet, where M is the maximum utility any agents could receive by the previous utilities.

Let $\pi^* = \{\{a^i_j, b^i_j, c^i_j\} \colon j \in [0, k], i \in [2^j]\} \cup \{\{\alpha^i_j, \beta^i_j\}, \{\gamma^i_j, \delta^i_j\} \colon j \in [0, k-1], i \in [2^j]\} \cup \{\{y_1, y_2\}\} \cup \{\{y^s_1, y^s_2\} \colon s \in S\}$ and $x = c^1_0$.

Now consider a partition $\pi \neq \pi^*$.

We will prove the following claim by induction over $j = k, \ldots, 0$. For every $i \in [2^j]$ holds:

1. If $\{b^i_j, a^i_j\} \cap \pi(c^i_j) = \emptyset$, then $\phi_{T^i_j}(\pi^*, \pi) \geq 1$ and $\phi_{T^i_j}(\pi^*, \pi) \geq 3$ or $\{b^i_k \colon i \in [2^k]\} \cap T^i_j \subseteq N(\pi, \pi^*)$.
2. If $\alpha^i_j \notin N(\pi, \pi^*)$ and there exists an agent $z \in T^i_j$ with $\pi(z) \neq \pi^*(z)$. Then $\phi_{T^i_j}(\pi^*, \pi) \geq 1$.

We will start by arguing, how the first part of the lemma follows from the induction claim.

First, note that $y_1 \notin N(\pi, \pi^*)$ and if $y_2 \in N(\pi, \pi^*)$, then $y_1 \in N(\pi^*, \pi)$. Similarly, for all $s \in S$, $y^s_1 \notin N(\pi, \pi^*)$ and if $y^s_2 \in N(\pi, \pi^*)$, then $y^s_1 \in N(\pi^*, \pi)$. We can therefore focus on T^1_0 and have $\phi(\pi^*, \pi) \geq \phi_{T^1_0}(\pi^*, \pi)$. Define $\rho = \{C \cap T^1_0 \colon C \in \pi\}$ and $\rho^* = \{C \cap T^1_0 \colon C \in \pi^*\}$, which are the partitions π and π^* restricted to agents in T^1_0. If $\rho = \rho^*$, then $\pi \neq \pi^*$ can only happen if some agent outside T^1_0 forms a coalition with a former coalition of π^* in T^1_0. Note that the only agents in T^1_0 that can improve by that are the agents of the type b^i_k. In every case, this will lead to $\phi_{T^1_0}(\pi^*, \pi) \geq 1$. As we have argued above, this implies $\phi(\pi^*, \pi) \geq 1$.

If $\rho \neq \rho^*$, we use the claim for the case $j = 0$ and observe that $\alpha^i_0 \notin N(\pi, \pi^*)$. Hence, $\phi(\pi^*, \pi) \geq 1$ also holds in this case.

It needs still to be shown that if $\pi(x) \cap \pi^*(x) = \{x\}$, then $\phi(\pi^*, \pi) \geq 3$ or (R, S) is a Yes-instance. Assume therefore that $\pi(x) \cap \pi^*(x) = \{x\}$. By the first part of the induction claim, we conclude that $\phi_{T^1_0}(\pi^*, \pi) \geq 3$ or $\{b^i_k \colon i \in [2^k]\} \subseteq N(\pi, \pi^*)$. Since we are done in the former case, we assume that $\{b^i_k \colon i \in [2^k]\} \subseteq N(\pi, \pi^*)$. This can only happen if, for every $i \in 1, \ldots, |R|$, there exists an $s_i \in S$ with $y^{s_i}_2 \in \pi(b^i_k)$. Define $S' = \{s \in S \colon \pi(y^s_2) \cap \{b^i_k \colon i \in [2^k]\} \neq \emptyset\}$. Now fix $s \in S'$. Then, it holds that $y^s_1 \notin \pi(y^s_2)$, because otherwise agents $b^i_k \in \pi(y^s_1)$ are worse off than in π^*. In particular, $y^s_1 \in N(\pi^*, \pi)$. Now, if at most two of the agents b^i_k corresponding two elements $i \in s$ are in the coalition of y^s_2, then $y^s_2 \in N(\pi^*, \pi)$. Together, $\phi(\pi^*, \pi) \geq \phi_{\{y_1, y_2\}}(\pi^*, \pi) + \phi_{\{y^s_1, y^s_2\}}(\pi^*, \pi) + \sum_{s' \in S \setminus \{s\}} + \phi_{\{y^{s'}_1, y^{s'}_2\}}(\pi^*, \pi) + \phi_{T^1_0}(\pi^*, \pi) \geq 0 + 2 + 0 + 1 = 3$. It remains the case that $\pi(y^s_2) = \{y^s_2, b^i_k, b^j_k, b^w_k\}$ for every $s \in S'$ with $s = \{i, j, w\}$. But then, S' is a 3-partition of R by sets in S.

We will now proceed with the proof of the induction claim.

For the base case $j = k$, we observe that if $A^i_k \cap N(\pi, \pi^*) \neq \emptyset$, then clearly $\phi_{A^i_k}(\pi^*, \pi) \geq 1$. In addition, if $\{b^i_k, a^i_k\} \cap \pi(c^i_k) = \emptyset$, then $\{a^i_k, c^i_k\} \subseteq N(\pi^*, \pi)$ and $b^i_k \in N(\pi^*, \pi) \cup N(\pi, \pi^*)$.

For the induction step, let $j \in \{k-1, \ldots, 0\}$ and fix $i \in [2^j]$. Assume first that there exists an agent $z \in T^i_j$ with $\pi(z) \neq \pi^*(z)$ but no such agent in A^i_j. The premise of the first claim is vacuous and this part is therefore true. Since $z \in T^{2i-1}_{j+1} \vee z \in T^{2i}_{j+1}$, we can apply induction for the second claim since the premise of the second claim for T^{2i-1}_{j+1} or T^{2i}_{j+1} is true. Assume therefore that there exists an agent $z \in A^i_j$ with $\pi(z) \neq \pi^*(z)$.

We make the following observations.

- If $\alpha^i_j \in N(\pi, \pi^*)$, then $\beta^i_j \in N(\pi^*, \pi)$.

- If $\beta_j^i \in N(\pi, \pi^*)$, then $\alpha_j^i \in N(\pi^*, \pi)$.
- If $\gamma_j^i \in N(\pi, \pi^*)$, then $\delta_j^i \in N(\pi^*, \pi)$.
- If $\delta_j^i \in N(\pi, \pi^*)$, then $\gamma_j^i \in N(\pi^*, \pi)$.

Now, we consider the case that $\pi(a_j^i) \neq \pi^*(a_j^i)$.

- We consider first the subcase that $b_j^i \in N(\pi, \pi^*)$. Then $c_j^i \in N(\pi^*, \pi)$.
  - If $\pi(b_j^i) \supseteq \{c_{j+1}^{2i-1}, c_{j+1}^{2i}\}$, then $\phi_{A_j^i}(\pi, \pi^*) \leq 1$ (with the above observations), while by induction $\phi_{T_{j+1}^{2i-1} \cup T_{j+1}^{2i}}(\pi^*, \pi) \geq 2$ and $\phi_{T_{j+1}^{2i-1} \cup T_{j+1}^{2i}}(\pi^*, \pi) \geq 4 \vee \{b_k^i \colon i \in [2^k]\} \cap (T_{j+1}^{2i-1} \cup T_{j+1}^{2i}) \subseteq N(\pi, \pi^*)$ and we are done.
  - Otherwise, $c_j^i \in \pi(b_j^i)$. Then $\phi_{A_j^i}(\pi^*, \pi) \geq 1$ or $a_j^i \in N(\pi, \pi^*)$. The second case can only occur for $\pi(a_j^i) = \{a_j^i, \beta_j^i, \gamma_j^i\}$. Hence, $\phi_{A_j^i}(\pi^*, \pi) \geq 1$ or $\pi(\delta_j^i) = \{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\}$. But then $\phi_{A_j^i}(\pi^*, \pi) \geq -1$ and $\phi_{T_{j+1}^{2i-1} \cup T_{j+1}^{2i}}(\pi^*, \pi) \geq 2$ and we are done.
- We can even assume that $b_j^i \in N(\pi^*, \pi)$, since otherwise $a_j^i \in \pi(b_j^i)$ and $a_j^i, c_j^i \in N(\pi^*, \pi)$ and it follows $\phi_{A_j^i}(\pi^*, \pi) \geq 1$.
- If $c_j^i \in N(\pi, \pi^*)$, then $a_j^i, b_j^i \in N(\pi^*, \pi)$ and therefore $\phi_{A_j^i}(\pi^*, \pi) \geq 1$ and we are done.
- Since $\pi(c_j^i) \neq \pi^*(c_j^i)$, we can assume that $c_j^i \in N(\pi^*, \pi)$.
- Next, consider the case that $a_j^i \in N(\pi, \pi^*)$ and, by the previous cases, $c_j^i, b_j^i \in N(\pi^*, \pi)$.
  - If $\pi(a_j^i) = \{a_j^i, \beta_j^i, \gamma_j^i\}$, then $\phi_{A_j^i}(\pi^*, \pi) \geq 3$ or $\pi(\delta_j^i) = \{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\}$. In the latter case, $\phi_{A_j^i}(\pi^*, \pi) \geq 1$ and $\phi_{T_{j+1}^{2i-1} \cup T_{j+1}^{2i}}(\pi^*, \pi) \geq 2$ by induction and we are done.
  - Otherwise, $\beta_j^i \in \pi(a_j^i) \cap N(\pi^*, \pi)$ or $\gamma_j^i \in \pi(a_j^i) \cap N(\pi^*, \pi)$. In the former case, $\alpha_j^i \in N(\pi^*, \pi)$ and in total $\phi_{A_j^i}(\pi^*, \pi) \geq 3$. In the latter case, again, $\phi_{A_j^i}(\pi^*, \pi) \geq 3$ or $\pi(\delta_j^i) = \{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\}$ and the case is similar as before.
- It remains that $a_j^i, b_j^i, c_j^i \in N(\pi^*, \pi)$ in which case $\phi_{A_j^i}(\pi^*, \pi) \geq 3$.

We may therefore assume that $\pi(a_j^i) = \pi^*(a_j^i)$. Only for the remaining cases, we need that $\alpha_j^i \notin N(\pi, \pi^*)$. If $\pi(\alpha_j^i) \neq \pi^*(\alpha_j^i)$, then $\alpha_j^i, \beta_j^i \in N(\pi^*, \pi)$ and consequently $\phi_{A_j^i}(\pi^*, \pi) \geq 2$. If $\pi(\gamma_j^i) \neq \pi^*(\gamma_j^i)$, then $\phi_{A_j^i}(\pi^*, \pi) \geq 2$ or $\phi_{A_j^i}(\pi, \pi^*) \geq 0 \wedge \pi(\delta_j^i) \cap \{\alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\} \neq \emptyset$ and the claim follows by induction.

For the second part of the lemma, assume that S' is a 3-partition of R through sets in S. Define

$$\begin{aligned}\pi' =&\{\{b_k^v, b_k^w, b_k^x, y_2^s\}, \{y_1^s\} \colon \{r^v, r^w, r^x\} = s \in S'\} \cup \{\{y_1^s, y_2^s\} \colon s \in S \setminus S'\} \\ &\cup \{\{b_k^{|R|+1}, \dots, b_k^{2^k}, y_2\}, \{y_1\}\} \cup \{\{\delta_{k-1}^i, a_k^{2i-1}, a_k^{2i}\} \colon i \in [2^{k-1}]\} \\ &\cup \{\{b_j^i, c_{j+1}^{2i-1}, c_{j+1}^{2i}\}, \{a_j^i, \beta_j^i, \gamma_j^i\} \colon j \in [k-1], i \in [2^j]\} \\ &\cup \{\{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\} \colon j \in [k-2], i \in [2^j]\} \cup \{\{\alpha_0^1\}, \{c_0^1\}\}.\end{aligned}$$

It is easily checked that $\phi(\pi', \pi^*) = 1$ and c_0^1 forms a singleton coalition with $c_0^1 \in N(\pi^*, \pi')$. $\square$

Theorem 10. *Checking whether there exists a strongly popular partition in a symmetric ASHG is coNP-hard.*

Proof. The reduction is from X3C. Given an instance (R, S) of X3C, we consider the symmetric ASHG of Lemma 3 on agent set N with utility function v together with the partition π^* and the special agent $x \in N$. Set $M = \max\{\sum_{w \in N \colon v(y,w)>0} v(y, w) \colon y \in N\}$ and $\alpha = \min_{w \in N \colon v(x,w)>0} v(x, w) > 0$. We define a symmetric ASHG on agent set $N' = N \cup \{z\}$ where the utilities are given by $v'(y, w) = v(y, w)$ if $y, w \in N$, $v'(z, x) = \alpha/2$, and $v'(z, y) = -M - 1$ for $y \in N \setminus \{x\}$. Note that by Lemma 3, this reduction is in polynomial time.

Consider the partition $\sigma^* = \pi^* \cup \{\{z\}\}$ and let $\sigma \neq \sigma^*$ be given and define $\pi = (\sigma \setminus \sigma(z)) \cup \{\sigma(z) \setminus \{z\}\}$, that is, the partition of agent set N where z leaves her coalition. We argue first that $\phi_N(\sigma^*, \sigma) \geq \phi(\pi^*, \pi)$ unless $\pi(x) = \pi^*(x)$. Clearly, if z leaves a coalition, only the agent x can be worse. Now recall that x receives her unique top-ranked coalition in π^*, which means that x forms a coalition precisely with all agents that yield her positive utility. By the choice of $v(x, z)$, the only coalition of x that z is part of and that is not worse for x, is $\pi^*(x) \cup \{z\}$. Hence, the only case that the preferences of x over σ^* and σ is affected by z is if $\pi(x) = \pi^*(x)$.

We perform a case distinction over the coalitions of z to investigate the popularity margin between σ^* and σ. First, if $\sigma(z) = \{z\}$, then $\phi(\sigma^*, \sigma) > 0$ by Lemma 3. If $\sigma(z) = \{z, x\}$, then $\phi(\sigma^*, \sigma) \geq -1 + \phi(\pi^*, \pi) \geq 0$ There, we can use the lemma again to see that the latter inequality is strict if (R, S) is a No-instance. Otherwise, $z \in N(\sigma^*, \sigma)$. If $\pi(x) \neq \pi^*(x)$, then $\phi(\sigma^*, \sigma) \geq 1 + \phi(\pi^*, \pi) \geq 1$. We can therefore assume that $\pi(x) = \pi^*(x)$. If $\pi = \pi^*$, then $\phi(\sigma^*, \sigma) = \phi_{\sigma^*(z)}(\sigma^*, \sigma) > 0$. If $\pi \neq \pi^*$, then $\phi(\sigma^*, \sigma) \geq 1 - 1 + \phi(\pi^*, \pi) > 0$, where the -1 accounts for the case that x may be worse off in π compared to σ. Note that it can not be the case that x is both better off in σ and worse off in π, since the only relevant coalition $\sigma(x) = \pi^*(x) \cup \{z\}$. Together, it follows that σ^* is popular and it is a strongly popular partition if (R, S) is a No-instance.

If (R, S) is a Yes-instance, then σ^* is the only candidate that might be strongly popular. Consider the partition π' from Lemma 3 and define $\sigma' = (\pi' \setminus \{\{x\}\}) \cup \{\{x, z\}\}$. Then, $x \in N(\pi^*, \pi') \cap N(\sigma^*, \sigma')$, whereas $z \in N(\sigma', \sigma^*)$. Therefore, $\phi(\sigma', \sigma^*) = 1 + \phi(\pi', \pi^*) = 0$. Hence, π^* is not strongly popular and there exists no strongly popular partition. $\square$

Theorem 11. *Verifying whether a given partition in a symmetric ASHG is strongly popular is coNP-complete.*

Proof. In the proof of Theorem 10, the partition σ^* is strongly popular if, and only if, (R, S) is a No-instance of X3C. □

Theorem 12. *Computing a mixed popular partition in a symmetric ASHG is NP-hard.*

Proof. We give a Turing reduction from X3C. Given an instance (R, S) of X3C, we consider the symmetric ASHG of Lemma 3 on agent set N with utility function v together with the partition π^* and the special agent $x \in N$. Set $M = \max\{\sum_{w\in N\colon v(y,w)>0} v(y,w)\colon y \in N\}$ and $\alpha = \min_{w\in N\colon v(x,w)>0} v(x,w) > 0$. We define a symmetric ASHG on agent set $N' = N \cup \{z_1, z_2\}$ where the utilities are given by $v'(y,w) = v(y,w)$ if $y, w \in N$, $v'(z_1,z_2) = v'(z_1,x) = v'(z_2,x) = \alpha/3 > 0$, and $v'(z_i,y) = -M-1$ for $i \in [2], y \in N \setminus \{x\}$. Note that by Lemma 3, this reduction is in polynomial time.

Consider the partition $\sigma^* = \pi^* \cup \{\{z_1, z_2\}\}$ and let $\sigma \neq \sigma^*$ be given and define $\pi = (\sigma \setminus (\sigma(z_1) \cup \sigma(z_2))) \cup \{\sigma(z_1) \setminus \{z_1,z_2\}, \sigma(z_2) \setminus \{z_1,z_2\}\}$, that is, the partition of agent set N where z_1 and z_2 leave their coalitions. Assume that (R, S) is a No-instance. We will prove that $\phi(\sigma^*, \sigma) > 0$, and therefore that σ^* is strongly popular. We may assume that $\sigma(z_1) = \{z_1, z_2\}$ or $x \in \sigma(z_i)$ for some i, because otherwise it is a Pareto improvement if z_1 and z_2 leave their coalitions and form a coalition of their own.

If $\sigma(z_1) = \{z_1, z_2\}$, then by Lemma 3, $\phi(\sigma^*, \sigma) = \phi(\pi^*, \pi) > 0$, because $\pi \neq \pi^*$. Otherwise, assume without loss of generality that $x \in \sigma(z_1)$. Since x receives her top-ranked coalition in π^* and the utility provided by agents z_i is sufficiently small, $\phi_N(\sigma^*, \sigma) - \phi(\pi^*, \pi) \geq -1$, where equality can only hold for $\pi^*(x) = \pi(x)$. Now, if $\pi(z_1) \subseteq \{x, z_1, z_2\}$, then $\phi(\sigma^*, \sigma) \geq -2 + \phi(\pi^*, \pi) \geq 1$. If there exists $y \in N \setminus \{x\}$ with $y \in \sigma(z_1)$, then $z_1, z_2 \in N(\sigma^*, \sigma)$ and it follows $\phi(\sigma^*, \sigma) \geq 2 - 1 + \phi(\pi^*, \pi) > 0$. In particular, the unique mixed popular partition consists of σ^* with probability 1.

Now assume that (R, S) is a Yes-instance. Consider the partition π' from Lemma 3 and define $\sigma' = (\pi' \setminus \{\{x\}\}) \cup \{\{x, z_1, z_2\}\}$. Then, $x \in N(\pi^*, \pi') \cap N(\sigma^*, \sigma')$, whereas $z_1, z_2 \in N(\sigma', \sigma^*)$. Therefore, $\phi(\sigma', \sigma) = 2 + \phi(\pi', \pi^*) = 1$. Hence, the pure mixed partition $\{\sigma^*\}$ is not mixed popular.

We can solve X3C by computing a partition σ in the support of a mixed popular partition and checking its probability in case $\sigma = \sigma^*$. □

Theorem 13. *Checking whether there exists a popular partition in a symmetric ASHG is coNP-hard.*

Proof. We provide a reduction from X3C. Given an instance (R, S) of X3C, we consider the symmetric ASHG of Lemma 3 on agent set N with utility function v together with the partition π^* and the special agent $x \in N$. Set $M = \max\{\sum_{w\in N\colon v(y,w)>0} v(y,w)\colon y \in N\}$ and $\alpha = \min_{w\in N\colon v(x,w)>0} v(x,w) > 0$. For $i \in [2]$, let $N_i = \{y_i\colon y \in N\}$ be two copies of N. Accordingly, let π_i^* be their respective copies of π^*.

We define a symmetric ASHG on agent set $N' = N_1 \cup N_2 \cup Z$ where $Z = \{z_k^j\colon k \in [2], j \in [3]\}$. Define $Z^j = \{z_1^j, z_2^j\}$. Utilities are as follows.

- $v'(y_i, w_i) = v(y,w)$ if $y, w \in N_i$ for $i \in [2]$,
- $v'(z_k^j, x_1) = \alpha/7$, $v'(z_k^j, x_2) = \alpha/8$ for $k \in [2], j \in [3]$,

- $v'(z_1^j, z_2^j) = \alpha$ for $j \in [3]$, and
- $v'(u, y) = -M - 1$ for every pair of agents $u, y \in N'$ such that their utility is not defined, yet.

Note that by Lemma 3, this reduction is in polynomial time.

First assume that (R, S) is a No-instance. Then, $\sigma^* = \pi_1^* \cup \pi_2^* \cup \{Z^j : j \in [3]\}$ is popular. To prove this, let σ be an arbitrary partition and define $\pi_i = \{\sigma(y) \cap N_i : y \in N_i\}$ be the coalitions restricted to N_i. For each $j \in [3]$, we can assume that $\sigma(z_k^j) = Z^j$ or there exists a $i \in [2]$ with $Z^j \cap \sigma(x_i) \neq \emptyset$. Otherwise, one can obtain a Pareto-improvement σ' over σ and it suffices to prove that $\phi(\sigma^*, \sigma') \geq 0$. Indeed, if $\sigma(z_k^j) = \{z_k^j\}$ for $k \in [2]$, then creating Z^j is a Pareto-improvement. On the other hand, if $\{z_{3-k}, x_1, x_2\} \cap \sigma(z_k^j) = \emptyset$ and $|\sigma(z_k^j)| \geq 2$, then leaving her coalition with z_k^j yields a Pareto-improvement over σ. Hence, if $x_1, x_2 \notin \sigma(z_k^j)$, then $z_{3-k}^j \in \sigma(z_k^j)$ and putting all potential further agents in the coalition into a singleton coalition would yield a Pareto improvement. Hence, we have already substantially restricted the coalitions of agents in a Z^j.

Next, we argue that we may assume that it does not happen that $\sigma(z_k^j) = \{z_k^j\}$. In this case, there exists an $i \in [2]$ with $z_{3-k}^j \in \sigma(x_i)$. We form a partition σ' by adding z_k^j to $\sigma(z_{3-k}^j) = \sigma(x_i)$. This yields a Partition with $N(\sigma^*, \sigma) \subseteq N(\sigma^*, \sigma')$ and $N(\sigma', \sigma^*) \subseteq N(\sigma, \sigma^*)$, hence $\phi(\sigma^*, \sigma') \geq \phi(\sigma^*, \sigma)$, and it suffices to consider the popularity margin between σ^* and σ'.

By a similar argument, we can assume that $\sigma(x_i) \subseteq Z \cup N_i$ (putting all agents outside $Z \cup N_i$ into singleton coalitions has the same effect).

We can therefore partition the agent set N' into sets of the type Z^j such that $\sigma(z_1^j) = Z^j$, of the type N_i such that $Z \cap \sigma(x_i) = \emptyset$, and of the type $N_i \cup \sigma\{x_i\}$ such that $Z \cap \sigma(x_i) \neq \emptyset$. For the first type, $\phi_{Z^j}(\sigma^*, \sigma) = 0$ and by Lemma 3, $\phi_{N_i}(\sigma^*, \sigma) \geq 0$ for the second type of sets. We prove that $\phi_{N_i \cup \sigma\{x_i\}}(\sigma^*, \sigma) \geq 0$ if $Z \cap \sigma(x_i) \neq \emptyset$.

If $\sigma(x_i) \subseteq Z \cup \{x_i\}$, then $x_i \in N(\sigma^*, \sigma)$ and $\phi_{\sigma(x_i) \setminus \{x_i\}}(\sigma^*, \sigma) \geq -2$. As a consequence, $\phi_{N_i \cup \sigma(x_i)}(\sigma^*, \sigma) \geq -2 + \phi(\pi_i^*, \pi_i) \geq 0$ by Lemma 3.

Otherwise, $Z \cap \sigma(x_i) \subseteq N(\sigma^*, \sigma)$ and the only agent in N_i that can be worse off in π_i compared to σ is x_i. Note that the utilities are designed so that $x_i \notin N(\sigma, \sigma^*) \cap N(\pi^*, \pi)$. It follows $\phi_{N_i \cup \sigma(x_i)}(\sigma^*, \sigma) = \phi_{N_i}(\sigma^*, \sigma) + \phi_{\sigma(x_i) \cap Z}(\sigma^*, \sigma) \geq \phi_{N_i}(\sigma^*, \sigma) + 1 \geq -1 + \phi(\pi_i^*, \pi_i) + 1 \geq 0$.

Together, it is shown that σ^* is popular.

Conversely, assume that (R, S) is a Yes-instance and assume for contradiction that σ is popular and define $\pi_i = \{\sigma(y) \cap N_i : y \in N_i\}$ as above. The Pareto-improvements of the first implication show that for all j, $Z^j \in \sigma$ or $\sigma(x_i) \cap Z^j \neq \emptyset$. Define $I = \{i \in [2] : Z \cap \sigma(x_i) \neq \emptyset\}$. The first crucial step is to prove that for all $i \in I$, it holds that there exists a $j \in [3]$ with $\sigma(x_i) = \{x_i\} \cup Z^j$.

Let therefore $i \in I$. First, $\sigma(x_i) \cap N_i = \{x_i\}$ since otherwise splitting $\sigma(x_i)$ into singleton coalitions is more popular. In addition, $x_{3-i} \notin \sigma(x_i)$. If this happens and $|\sigma(x_i) \cap Z| \neq 2$, then splitting into singleton coalitions is more popular. On the other hand, if $|\sigma(x_i) \cap Z| = 2$, there exists $j^* \in [3]$ with $Z^{j^*} \in \sigma$. We form the partition σ' by leaving her coalition with x_1 and forming $\{x_1, z_1^{j^*}, z_2^{j^*}\}$. Then, $\{x_1, x_2, z_1^{j^*}, z_2^{j^*}\} \subseteq N(\sigma', \sigma)$ while $N(\sigma, \sigma') \subseteq \sigma(x_i) \cap Z$. Hence, σ' is more popular.

Hence, $\sigma(x_i) \subseteq Z \cup \{x_i\}$. If for $j \neq j'$, $Z^j \cap \sigma(x_i) \neq \emptyset$ and $Z^{j'} \cap \sigma(x_i) \neq \emptyset$, then dissolving $\sigma(x_i)$ is again more popular. Finally, if $|\sigma(x_i) \cap Z| = 1$, we find again a $j^* \in [3]$ with $Z^{j^*} \in \sigma$. We form the partition σ' by forming $\pi(x_i) \cap Z$ and $\{x_i, z_1^{j^*}, z_2^{j^*}\}$ which is more popular.

The next step is to show that $I = \{1, 2\}$. Assume for contradiction that $Z \cap \sigma(x_i) = \emptyset$. Then we can assume that for all $y \in N_i$, $\sigma(y) \subseteq N_i$. If $\pi_i \neq \pi_i^*$, then replacing π_i by π_i^* is more popular (by Lemma 3). Otherwise $\pi_i = \pi_i^*$ and we consider the partition π_i' of the last part of Lemma 3 for N_i. By the pigeon hole principle, there exists a $j^* \in [3]$ with $Z^{j^*} \in \sigma$. We obtain $\sigma' = (\sigma \setminus (\pi_i \cup \{Z^{j^*}\})) \cup ((\pi_i' \setminus \{\{x_i\}\}) \cup \{\{x_i, z_1^{j^*}, z_2^{j^*}\}\})$. Then, $\phi(\sigma', \sigma) = \phi_{N_i \cup Z^{j^*}}(\sigma', \sigma) = -1 + 2 = 1$ and σ' is more popular.

Together, we can assume that there exist $j_1, j_2 \in [3]$ with $\sigma(x_i) = \{x_i, z_1^{j_i}, z_2^{j_i}\}$, for $i \in [2]$. Let $j_3 \in [3] \setminus \{j_1, j_2\}$ be the third index. Note that $Z^{j_3} \in \sigma$. Define $\sigma' = (\sigma \setminus \{\sigma(z_1^j) \colon j \in [3]\}) \cup \{\{x_1, z_1^{j_2}, z_2^{j_2}\}, \{x_2, z_1^{j_3}, z_2^{j_3}\}, Z^{j_1}\}$. Then, $N(\sigma', \sigma) = Z^{j_2} \cup Z^{j_3}$ while $N(\sigma, \sigma') = Z^{j_1}$. Hence, σ' is more popular.

All in all, it is shown that there exists no popular partition if (R, S) is a Yes-instance. This concludes the proof of the theorem. □

A.3 Fractional Hedonic Games

Theorem 14. *Checking whether there exists a popular partition in a symmetric FHG is NP-hard, even if all utilities are non-negative.*

Proof. The reduction is from X3C to deciding whether there exists a popular partition.

Let (R, S) be an instance of X3C. We transform it into an FHG $(N, \succsim)$ defined by the graph $G = (N, E)$ that is given as follows:

$N = \{c^r, l_j^r \colon r \in R, j \in [6]\} \cup \{y^s, z_j^s \colon s \in S, j \in [2]\}$ and $E = E^R \cup E^C \cup E_6 \cup E^S$ where $E^R = \{\{c^r, l_j^r\} \colon r \in R, j \in [6]\}, E^C = \{\{l_6^r, y^s\} \colon s \in S, r \in s\}, E_6 = \{\{l_6^r, l_6^t\} \colon r \neq t, r, t \in s \text{ for } s \in S\}, E^S = \{\{y^s, z_j^s\}, \{z_1^s, z_2^s\} \colon s \in S, j \in [2]\}$. The edge set E^C connects the gadgets for the ground set and the subsets for the X3C instance.

The weights are 1, except $v(e) = \frac{1}{2}$ for $e \in E^C$ and $v(e) = \frac{1}{4}$ for $e \in E_6$. A schematic of the reduction for a certain set $s = \{i, j, k\} \in S$ is depicted in Figure 6.

We show that there exists a popular partition of $(N, \succsim)$ if and only if (R, S) is a Yes-instance of X3C.

Assume (R, S) is a Yes-instance of X3C. Then, there exists $S' \subseteq S$ such that S' is a partition of R. Consider the partition $\pi = \{\{c^r, l_1^r, l_2^r, l_3^r\} \colon r \in R\} \cup \{\{l_j^r\} \colon r \in R, j = 4, 5\} \cup \{\{y^s, l_6^i, l_6^j, l_6^k\} \colon s = \{i, j, k\} \in S'\} \cup \{\{z_1^s, z_2^s\} \colon s \in S'\} \cup \{\{y^s, z_1^s, z_2^s\} \colon s \in S \setminus S'\}$. We claim that π is popular.

Assume for contradiction that π' is more popular than π and let π' be with $\phi(\pi', \pi)$ maximal. We will prove that $\phi(\pi, \pi') \geq 0$, deriving a contradiction.

We introduce some notation for the proof. Let $V^r = \{c^r, l_j^r \colon j \in [6]\}$, where $r \in R$, and $W^s = \{y^s, z_1^s, z_2^s\}$, where $s \in S$. Also denote $V^R = \cup_{r \in R} V^r, W^S = \cup_{s \in S} W^s$ and $A_6 = \{l_6^r \colon r \in R\}$ and $Y^c = \{s \in S \colon \exists a \in A_6 \text{ with } a \in \pi'(y^s)\}$.

To derive a contradiction, we prove several claims.

1. Let $r \in R$ such that, for all $s \in S$ with $r \in s$, it holds that $y^s \notin \pi'(l_6^r)$. Then, $\phi_{V^r}(\pi, \pi') \geq 1$.

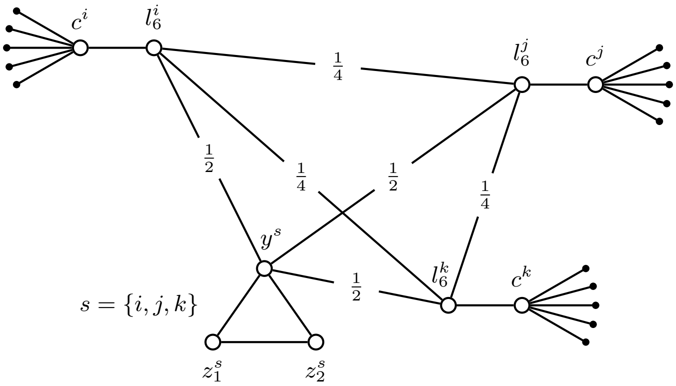


Figure 6: Reduction for existence problem of popular partitions in FHGs. The schematic displays the part of the network corresponding to one specific set $s = \{i, j, k\}$.

2. There exist no $r \in R, s, s' \in S$ with $s \neq s'$ and $\{y^s, y^{s'}\} \subseteq \pi'(l_6^r)$.

3. For all $s \in S$, it holds that $\pi'(y^s) \cap W^s = \{y^s\}$ or $\pi'(y^s) \subseteq W^s$.

4. For all $r \in R$, $\phi_{V^r}(\pi, \pi') \geq 0$.

5. It holds that $\phi_{W^S}(\pi', \pi) \leq |R| - 3|Y^c|$.

The first claim says that we need sufficient external influence for V^r to be 'locally' popular. The second and third claim give insight on the structure of potential more popular partitions. The forth claim shows that we locally do best for every V^r. The final claim calculates the tradeoff between forming a coalition W^s and joining the agents in V^r.

In order to complete the proof from the claims, we apply Claims 1 and 4 to obtain $\phi_{V^R}(\pi, \pi') \geq \max\{0, |R| - 3|Y^c|\} \geq |R| - 3|Y^c|$. Combining this inequality with the one of Claim 5 yields $\phi(\pi, \pi') \geq 0$.

The first claim is a straightforward case distinction considering $\pi'(c^r)$. Observe that by construction of her neighboring agents, $l_6^r \in N(\pi, \pi')$ or $l_6^r \in \pi'(c^r)$. This property makes l_6^r play an equivalent role compared to the agents l_5^r and l_4^r in the analysis.

We proceed with the second claim. Therefore, assume for contradiction that $r \in R, s, s' \in S$ with $s \neq s'$ and $\{y^s, y^{s'}\} \subseteq \pi'(l_6^r)$. We denote $C = \pi'(l_6^r)$ for this part. We claim that we can change π' while strictly increasing $\phi(\pi', \pi)$. This is done by forming a partition π'' that consists of coalitions W^t whenever $y^t \in C$. The agents outside W^S in C form a coalition of their own. Other coalitions are not changed.

- Let $t \in S$ with $y^t \in C$. If $\pi(y^t) = W^t$, then $W^t \subseteq N(\pi, \pi')$. This is immediate for the z_j^t. In addition, by assumption on C, at least 3 agents are present, and the utility is estimated as $v_{y^t}(\pi') \leq \max\{\frac{\frac{1}{2}}{3}, \frac{\frac{3}{2}}{4}, \frac{\frac{5}{2}}{5}, \frac{\frac{6}{2}}{6}, \frac{\frac{7}{2}}{7}\} = \frac{1}{2} < \frac{2}{3} = v_{y^t}(\pi)$
- If $\pi(y^t) \neq W^t$, then $z_1^t, z_2^t \notin N(\pi', \pi)$ and $y^t \notin N(\pi', \pi) \vee (\exists i : z_i^t \in N(\pi, \pi'))$.

Define $Y = |\{t \in S : y^t \in C\}|$. These first two insights yield, that $\phi_{\{W^s : s \in Y\}}(\pi'', \pi) \geq 3|Y| + \phi_{\{W^s : s \in Y\}}(\pi', \pi)$. There is an increase of at least 6 by the assumption that $\{y^s, y^{s'}\} \subseteq C$.

- The only agents that can decrease (π'', π) compared to (π', π) are in A_6. Note that if $a \in A_6 \cap C$ has at most one neighbor in Y, then for some p (the number of neighbors in A_6), $v_a(\pi') = \frac{\frac{1}{2}+\frac{p}{4}}{3+p} < \frac{1}{4} = v_a(\pi)$. Define the improving agents in A_6 via $\mathcal{I} = C \cap A_6 \cap N(\pi', \pi)$ and the non-worsened agents as $\mathcal{I}' = C \cap A_6 \setminus (\mathcal{I} \cup N(\pi, \pi'))$.
  - If $|\mathcal{I}| \leq 2$, then $\phi_{C \cap A_6}(\pi'', \pi) \geq \phi_{C \cap A_6}(\pi', \pi) - 4$ (the agents in $\mathcal{I}$ each counted twice for being worse instead of better off).
  - If $|\mathcal{I}| \geq 3$, we know that $|Y| \geq 3$ (otherwise, three agents in $\mathcal{I}$ are incident to the same two y^t, but then in the instance of X3C, we had two identical 3-elementary sets). This means for any $a \in A_6 \cap C$ that has exactly two neighbors in Y that for some p, $v_a(\pi') \leq \frac{1+\frac{p}{4}}{4+p} = \frac{1}{4}$. Hence, $a \notin N(\pi', \pi)$.

    Agents in $\mathcal{I}$ need therefore three neighbors in Y and agents in $\mathcal{I}'$ two. Since every agent in Y has at most three neighbors, this accumulates to $|Y| \geq |\mathcal{I}| + \frac{2}{3}|\mathcal{I}'|$. Consequently, for $M = C \cap (A_6 \cup W^S)$,

$$\begin{aligned}\phi(\pi'', \pi) &= \phi_{N \setminus M}(\pi'', \pi) + \phi_{C \cap A_6}(\pi'', \pi) + \phi_{W^S \cap C}(\pi'', \pi) \\ &\geq \phi_{N \setminus M}(\pi', \pi) + \phi_{C \cap A_6}(\pi', \pi) - 2|\mathcal{I}| - |\mathcal{I}'| + \phi_{W^S \cap C}(\pi', \pi) + 3|Y| \\ &> \phi(\pi', \pi).\end{aligned}$$

In both cases, we contradict the maximality of $\phi(\pi', \pi)$.

The third claim is proven similarly, but we have to refine some calculation of the previous claim, since we do not get the same lower bounds for the denominators of the utilities.

Assume for contradiction that $s \in S$ with $\pi'(y^s) \cap A_6 \neq \emptyset$ and $\pi'(y^s) \cap W^s \neq \emptyset$. We set $C = \pi'(y^s)$.

- First, we argue that we may assume that $A_6 \cap C \cap N(\pi', \pi) = \emptyset$. Otherwise, by the previous claim, if $l_6^r \in A_6 \cap C \cap N(\pi', \pi)$, then $c^r \in C$. Consequently, $l_j^r \in N(\pi, \pi')$ for $j \in [3]$ and $c^r \in N(\pi, \pi')$. The latter is due to $v_{c^r}(\pi') \leq \frac{6}{9} < \frac{3}{4} = v_{c^r}(\pi)$. Also, there exists $j \in \{4, 5\} : l_j^r \notin C$ or $l_6^r \notin N(\pi', \pi)$. Indeed, if the first is wrong, then for some p, $v_{l_6^r}(\pi') \leq \frac{1+\frac{1}{2}+\frac{p}{4}}{6+p} = \frac{1}{4} = v_{l_6^r}(\pi)$. Hence resetting the coalition within V^r to π yields a coalition contradicting the maximality of $\phi(\pi', \pi)$.

- Now, we consider two cases. First assume that $\pi(y^s) \neq W^s$. We claim that rearranging π' by means of removing agents of W^s from $\pi'(y^s)$ improves $\phi(\pi', \pi)$. Indeed, $z_j^s \notin N(\pi', \pi)$, but they will be after the rearrangement, and $y^s \in N(\pi', \pi)$ afterwards. Also, for all $a \in A_6 \cap C$, $v_a(\pi') \leq \frac{\frac{1}{2}+\frac{p}{4}}{p+3} < \frac{1}{4}$ and these agents are already worse off in the original π'.

- If $\pi(y^s) = W^s$, the same holds for agents in $A_6 \cap C$. Since $W^s \subseteq N(\pi, \pi')$, the same rearrangement improves $\phi(\pi', \pi)$.

We proceed with the next claim and fix $r \in R$. We may assume that for some s, $y^s \in \pi'(l_6^r)$ (since the other case is already covered in the first claim). In addition, if $c^r \notin \pi'(l_6^r)$, then $l_6^r \notin N(\pi', \pi)$ (by the previous claims). In this case, the coalition π restricted to $V^r \setminus \{l_6^r\}$ is popular and the claim is true.

Denote $C = \pi'(l_6^r)$ and assume therefore $c^r \in C$. We also know that $\{l_1^r, l_2^r, l_3^r\} \cap N(\pi', \pi) = \emptyset$ and $|\{l_1^r, l_2^r, l_3^r\} \cap N(\pi, \pi')| \geq 2$. Consequently, if $\{l_4^r, l_5^r\} \cap C = \emptyset$, we are done. If $\{l_4^r, l_5^r\} \cap C \neq \emptyset$, $\{l_1^r, l_2^r, l_3^r\} \subseteq N(\pi, \pi')$. Hence, in the final case, $|N(\pi', \pi)| \leq 3$ while $|N(\pi, \pi')| \geq 3$ and the claim is true.

For the fifth claim, we consider the coalitions in π for different y^s:

- If $W^s = \pi(y^s)$, then $W^s \cap N(\pi', \pi) = \emptyset$ (by Claim 3) and if $s \in Y^c$, then $W^s \subseteq N(\pi, \pi')$. This gives $|N(\pi, \pi') \cap W^S| \geq 3|\{s \in Y^c \colon \pi(y^s) = W^s\}|$.
- If $W^s \neq \pi'(y^s)$ and $s \in Y^c$, then $W^s \cap N(\pi', \pi) = \emptyset$ (again using Claim 3). Consequently, $|N(\pi', \pi) \cap W^S| \leq 3|\{s \notin Y^c \colon \pi(y^s) \neq W^s\}|$.

  Combining the inequalities yields

$$
\begin{aligned}
&|N(\pi', \pi) \cap W^S| - |N(\pi, \pi') \cap W^S| \\
&\leq 3(|\{s \notin Y^c \colon \pi(y^s) \neq W^s\}| - |\{s \in Y^c \colon \pi(y^s) = W^s\}|) \\
&= 3(|\{s \notin Y^c \colon \pi(y^s) \neq W^s\}| + |\{s \in Y^c \colon \pi(y^s) \neq W^s\}| \\
&- |\{s \in Y^c \colon \pi(y^s) \neq W^s\}| - |\{s \in Y^c \colon \pi(y^s) = W^s\}|) \\
&= 3|S'| - 3|Y^c| = |R| - 3|Y^c|.
\end{aligned}
$$

This proves the final claim and we have proved that Yes-instances of X3C map to popular partitions of the FHG.

For the reverse implication, assume that π is a popular partition. We exhibit the coalitions of the agents in A_6.

1. For all $r \in R$, there exists a unique $s \in S$ with $y^s \in \pi(l_6^r)$. For this s holds that $r \in s$.
2. For all $r \in R$, $|A_6 \cap \pi(l_6^r)| = 3$.

If the claims are true, $S' = \{s \in S : A_6 \cap \pi(y^s) \neq \emptyset\}$ covers R due to existence and is a partition due to uniqueness and the fact, that uniqueness and the second claim imply that the coalition of the unique y^s must contain precisely l_6^i for $i \in s$.

We start with the first claim. Existence is clear because otherwise the subpartition of π on V^r (possibly restricted to V^r) is popular on V^r, contradicting Proposition 11.

For uniqueness, assume for contradiction that there is $r \in R$ and $s \neq s' \in S$ with $\{y^s, y^{s'}\} \subseteq \pi(l_6^r)$. We obtain a more popular coalition π' as follows: remove the agents in W^s from their partitions in π and let them form a coalition. Then $W^s \cup \{y^{s'}\} \subseteq N(\pi', \pi)$ and $N(\pi, \pi') \subseteq \{l_6^r \colon r \in s\}$. Hence, π' is more popular.

For the second claim, we know due to uniqueness in the first claim that $|A_6 \cap \pi(l_6^r)| \leq 3$. Assume for contradiction that $|A_6 \cap \pi(l_6^r)| < 3$ and let $y^s \in \pi(l_6^r)$. Then, the same coalition π' as in the proof of the previous claim is more popular. This time, $W^s \subseteq N(\pi', \pi)$ and $N(\pi, \pi') \subseteq \{l_6^r \colon r \in s\}$, hence by assumption $|N(\pi, \pi')| \leq 2$.

$\square$

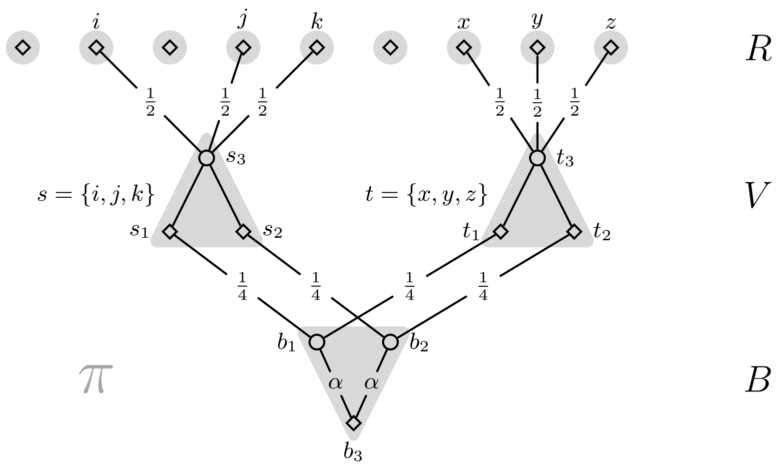


Figure 7: Schematic of the reduction for the verification problem of popular partitions on bipartite FHGs. The bipartition is indicated by the shapes of the agents. The partition π under consideration is marked in gray.

Theorem 15. *Checking whether a given partition in a symmetric FHG is popular is coNP-complete, even if all utilities are non-negative and the underlying graph is bipartite.*

Proof. First of all, the verification problem is in coNP, because a more popular partition serves as a polynomial-time certificate for a No-instance.

For hardness, we reduce again from X3C. Given an instance (R, S) of X3C, we assume without loss of generality that $|R| \geq 6$. We define an FHG $(N, \succsim)$ given by the underlying graph $G = (N, E)$ depicted in Figure 7 and defined as:

$N = R \cup \{s_1, s_2, s_3 \colon s \in S\} \cup \{b_1, b_2, b_3\}$, $E = \{\{s_3, r\} \colon r \in R \cap s\} \cup \{\{s_1, s_3\}, \{s_2, s_3\} \colon s \in S\} \cup \{\{s_j, b_j\} \colon s \in S, j \in [2]\} \cup \{\{b_1, b_3\}, \{b_2, b_3\}\}$.

The symmetric weights v are given as

- $v(i, s_3) = \frac{1}{2}$ if $i \in s$,
- $v(s_1, s_3) = v(s_2, s_3) = 1$ for $s \in S$,
- $v(s_j, b_j) = \frac{1}{4}$ for $s \in S, j \in [2]$, and
- $v(b_1, b_3) = v(b_2, b_3) = \alpha$ for $\frac{3(|R|-3)}{4|R|} < \alpha < \frac{3|R|}{4(|R|+3)}$.

One can choose α with a size bounded polynomially in the input size. For the reduction, only the above bounds matter. We introduce the same notation as in the proof for ASHGs. Denote $V^s = \{s_1, s_2, s_3\}$ for $s \in S$, $B = \{b_1, b_2, b_3\}$, and $V = \cup_{s \in S} V^s$.

G is bipartite with bipartition $(R \cup \{s_1, s_2 \colon s \in S\} \cup \{b_3\}, \{s_3 \colon s \in S\} \cup \{b_1, b_2\})$ and all weights on present edges are positive.

The verification problem is asked for the partition $\pi = \{V^s \colon s \in S\} \cup \{\{r\} \colon r \in R\} \cup \{B\}$. We claim that (R, S) is a Yes-instance of X3C if and only if π is not popular for the FHG given by G.

If (R, S) is a Yes-instance, there exists a subset $S' \subseteq S$ that partitions R. In particular $|R| = 3|S'|$.

Consider the partition given by $\pi' = \{V^s \colon s \in S \setminus S'\} \cup \{\{s_3, i, j, k\} \colon \{i, j, k\} = s \in S'\} \cup \{\{b_j, s_j \colon s \in S'\} \colon j \in [2]\} \cup \{\{b_3\}\}$.

Then, for $j \in [2]$, $v_{b_j}(\pi') = \frac{\frac{1}{4}|S'|}{|S'|+1} = \frac{|R|}{4(|R|+3)} > \frac{\alpha}{3} = v_{b_j}(\pi)$. Since all agents in R have clearly improved their utility, $R \cup \{b_1, b_2\} \subseteq N(\pi', \pi)$ (and in fact equality holds here). Moreover, the utilities of agents in V^s for $s \in S \setminus S'$ have not changed. Consequently, $N(\pi, \pi') \subseteq \cup_{s \in S'} V^s \cup \{b_3\}$. Hence, π' is more popular than π.

Conversely, assume that there exists a more popular partition π' and fix one that maximizes $\phi(\pi', \pi)$. We have to prove that there exists a subset $S' \subseteq S$ that yields a partition of R.

First, we make the observation that if $b_j \in N(\pi', \pi)$ for $j \in [2]$, then $b_3 \in N(\pi, \pi')$. Hence, $\phi_B(\pi', \pi) \leq 1$.

Second, we claim that for all $s \in S$, $N(\pi', \pi) \cap V^s = \emptyset$. Clearly, $s_3 \notin N(\pi', \pi)$ (by construction, since she receives a top coalition with respect to the given utilities). Assume for $j \in [2]$, $s_j \in N(\pi', \pi)$. Then, $\pi'(s_j) = \{s_j, s_3, b_j\}$. Note that both neighbors of s_j are needed to improve utility, but no other agent may be present since for $|\pi'(s_j)| \geq 4$ follows $v_{s_j}(\pi') \leq \frac{\frac{5}{4}}{4} < \frac{1}{3} = v_{s_j}(\pi)$. In addition, $s_{3-j}, b_3 \in N(\pi, \pi')$.

We form a new coalition π'' from π' by having the coalitions V^s and B and all other coalitions remain the same. The exact same case distinction for b_{3-j} as in the case of ASHGs yields a contradiction to the maximality condition on π'.

The remainder of the proof follows a similar strategy as the one for ASHGs, but some arguments are more tedious.

To make this more formal, we introduce the sets $R_I = R \cap N(\pi', \pi)$ of agents in R that form a coalition with a neighbor in π' and $S_C = \{s \in S \colon \pi'(s_3) \cap R \neq \emptyset\}$. The latter is the set of critical sets in S whose corresponding agents s_3 form a coalition with agents in R. We split it into $S_{C,1} = \{s \in S \colon |\pi'(s_3) \cap R| = 1\}$ and $S_{C,2} = S_C \setminus S_{C,1}$.

We have the following facts:

- For $s \in S_C$, $s_3 \in N(\pi, \pi')$.
- For $s \in S_{C,1}$, $s_1 \in N(\pi, \pi') \vee s_2 \in N(\pi, \pi')$.
- For $s \in S_{C,2}$, $s_1 \in N(\pi, \pi') \wedge s_2 \in N(\pi, \pi')$.

Consequently, $|N(\pi, \pi') \cap V| \geq 2|S_{C,1}| + 3|S_{C,2}|$. In addition, $|N(\pi', \pi) \cap R| = |R_I| \leq |S_{C,1}| + 3|S_{C,2}|$.

If $S_{C,1} \neq \emptyset$, then $\phi(\pi, \pi') = \phi_B(\pi, \pi') + \phi_V(\pi, \pi') + \phi_R(\pi, \pi') \geq -1 + 2|S_{C,1}| + 3|S_{C,2}| - (|S_{C,1}| + 3|S_{C,2}|) = |S_{C,1}| - 1 \geq 0$ and π' is not more popular. We conclude that $S_{C,1} = \emptyset$ or equivalently $S_C = S_{C,2}$.

A similar calculation excludes the case $|R_I| < 3|S_{C,2}|$ which means $|R_I| = 3|S_{C,2}|$.

We claim that in fact $|R| = 3|S_C| = 3|S_{C,2}|$. Before we prove this, we show the same two auxiliary claims as for ASHGs.

1. If $B \subseteq \pi'(b_3)$ then $b_1 \notin N(\pi', \pi) \vee b_2 \notin N(\pi', \pi)$.
2. For $j \in [2]$, if $b_j \in N(\pi', \pi)$, then $b_j \in \pi'(b_3) \vee |\{s \in S \colon s_j \in \pi'(b_j)\} \cap \pi'(b_j)| \geq \frac{|R|}{3}$.

For the first claim, assume that $B \subseteq \pi'(b_3)$ and $b_1, b_2 \in N(\pi', \pi)$. Denote $p_j = |\{s \in S\colon s_j \in \pi'(b_3)\}|$. We know that $p_j \geq 1$, since otherwise $b_j \notin N(\pi', \pi)$.

The function $x \mapsto \frac{3(x-3)}{4x}$ is monotonically increasing for $x > 0$. Thus, by the lower bound on α, we know that $\alpha > \frac{3}{8}$ (using $|R| \geq 6$).

Let $j \in [2]$ with $p_j = \min\{p_j, p_{3-j}\}$. Then $|\pi'(b_3)| \geq 3 + 2p_j$. We compute $v_{b_j}(\pi) - v_{b_j}(\pi') = \frac{\alpha}{3} - \frac{\alpha + \frac{p_j}{4}}{3+2p_j} = \frac{p_j}{3(3+2p_j)}(2\alpha - \frac{3}{4}) > 0$. Hence, $b_j \notin N(\pi', \pi)$, a contradiction.

For the second claim, let $j \in [2]$ with $b_j \in N(\pi', \pi)$ and assume $b_j \notin \pi'(b_3)$. Similarly as before, let $p = |\{s \in S\colon s_j \in \pi'(b_j)\}|$. Note that $v_{b_j}(\pi) = \frac{\alpha}{3} > \frac{|R|-3}{4|R|} = \frac{1}{4}\frac{\frac{|R|}{3}-1}{\left(\frac{|R|}{3}-1\right)+1}$. Therefore, $v_{b_j}(\pi) < v_{b_j}(\pi') \leq \frac{1}{4}\frac{p}{p+1}$ only if $p > \frac{|R|}{3} - 1$ and since p is an integer, this implies $p \geq \frac{|R|}{3}$.

The remainder of the proof is identical to the one for ASHGs (Theorem 9). $\square$

Lemma 4. *The class of symmetric FHGs with non-negative utility functions satisfies property PP.*

Proof. Let (R, S) be an instance of X3C. We construct the following game. Let $k = \min\{k \in \mathbb{N}\colon 2^k \geq |R|\}$ define the smallest power of 2 that is larger than the cardinality of R. We define a symmetric FHG with non-negative utility functions on vertex set $N = \{y_1^s, y_2^s\colon s \in S\} \cup \{y_1, y_2\} \cup \bigcup_{j=0}^{k} N_j$, where $N_j = \bigcup_{i=1}^{2^j} A_j^i$ consists of 2^j sets of agents A_j^i.

We define the sets of agents as

- $A_k^i = \{a_k^i, b_k^i, c_k^i\}$ for $i \in [2^k]$, and
- $A_j^i = \{a_j^i, b_j^i, c_j^i, \alpha_j^i, \beta_j^i, \gamma_j^i, \delta_j^i\}$ for $j \in [0, k-1], i \in [2^j]$.

We order the set R in an arbitrary but fixed way, say $R = \{r^1, \dots, r^{|R|}\}$ and for a better understanding of the proof and the preferences, we label the agents $b_k^i = r^i$ for $i \in [|R|]$. If we view the set of agents N as $k+1$ levels of agents, then the ground set R of the instance of X3C is identified with some specific agents in the top level k. We are ready to define the preferences.

- $v(y_1^s, y_2^s) = \frac{21}{10}(k+1)$ for all $s \in S$,
- $v(y_2^s, b_k^i) = \frac{3}{2}(k+1)$ if there exists $s \in S$ with $r^i \in s$,
- $v(y_1, y_2) = 1$,
- $v(y_2, b_k^i) = 2^{k+2}(k+1)$, $i \in [|R|+1, 2^k]$,
- $v(b_k^i, b_k^{i'}) = 0$, $i, i' \in [|R|+1, 2^k]$,
- $v(b_k^i, b_k^{i'}) = \frac{2}{3}(k+1)$, $i, i' \in [|R|]$,
- $v(a_k^i, b_k^i) = v(a_k^i, c_k^i) = v(b_k^i, c_k^i) = k+1$, $i \in [2^k]$,
- For $j \in [0, k-1]$, $i \in [2^k]$,

- $v(a^i_j, b^i_j) = v(a^i_j, c^i_j) = j + 1$, $v(b^i_j, c^i_j) = j + 1.5$,
- $v(b^i_j, c^{2i-1}_{j+1}) = v(b^i_j, c^{2i}_{j+1}) = j + 1.5$,
- $v(\alpha^i_j, \beta^i_j) = j + 1$, $v(\beta^i_j, \gamma^i_j) = \frac{j}{2}$,
- $v(\beta^i_j, a^i_j) = j + 1.75$, $v(\gamma^i_j, a^i_j) = j + 1.25$,
- $v(\gamma^i_j, \delta^i_j) = j + 2$, $v(\delta^i_j, \alpha^{2i-1}_{j+1}) = v(\delta^i_j, \alpha^{2i}_{j+1}) = j + 1.6$, and

- $v(g, h) = 0$ for all $g, h \in N$ such that the utility is not defined, yet.

Let $\pi^* = \{\{a^i_j, b^i_j, c^i_j\} \colon j \in [0, k], i \in [2^j]\} \cup \{\{\alpha^i_j, \beta^i_j\}, \{\gamma^i_j, \delta^i_j\} \colon j \in [0, k-1], i \in [2^j]\} \cup \{\{y_1, y_2\}\} \cup \{\{y^s_1, y^s_2\} \colon s \in S\}$ and $x = c^1_0$.

Now consider a partition $\pi \neq \pi^*$.

We will prove the following claim by induction over $j = k, \dots, 0$. For every $i \in [2^j]$ holds:

1. If $\{b^i_j, a^i_j\} \cap \pi(c^i_j) = \emptyset$, then $\phi_{T^i_j}(\pi^*, \pi) \geq 1$ and $\phi_{T^i_j}(\pi^*, \pi) \geq 3$ or $\{b^i_k \colon i \in [2^k]\} \cap T^i_j \subseteq N(\pi, \pi^*)$.

2. If $\alpha^i_j \notin N(\pi, \pi^*)$ and there exists an agent $z \in T^i_j$ with $\pi(z) \neq \pi^*(z)$. Then $\phi_{T^i_j}(\pi^*, \pi) \geq 1$.

We will start by arguing, how the first part of the lemma follows from the induction claim.

First, note that $y_1 \notin N(\pi, \pi^*)$ and if $y_2 \in N(\pi, \pi^*)$, then $y_1 \in N(\pi^*, \pi)$. Similarly, for all $s \in S$, $y^s_1 \notin N(\pi, \pi^*)$ and if $y^s_2 \in N(\pi, \pi^*)$, then $y^s_1 \in N(\pi^*, \pi)$. We can therefore focus on T^1_0 and have $\phi(\pi^*, \pi) \geq \phi_{T^1_0}(\pi^*, \pi)$. Define $\rho = \{C \cap T^1_0 \colon C \in \pi\}$ and $\rho^* = \{C \cap T^1_0 \colon C \in \pi^*\}$, which are the partitions π and π^* restricted to agents in T^1_0. If $\rho = \rho^*$, then $\pi \neq \pi^*$ can only happen if some agent outside T^1_0 forms a coalition with a former coalition of π^* in T^1_0. Note that the only agents in T^1_0 that can improve by that are the agents of the type b^i_k. In every case, this will lead to $\phi_{T^1_0}(\pi^*, \pi) \geq 1$. As we have argued above, this implies $\phi(\pi^*, \pi) \geq 1$.

If $\rho \neq \rho^*$, we use the claim for the case $j = 0$ and observe that $\alpha^i_0 \notin N(\pi, \pi^*)$. Hence, $\phi(\pi^*, \pi) \geq 1$ also holds in this case.

It needs still to be shown that if $\pi(x) \cap \pi^*(x) = \{x\}$, then $\phi(\pi^*, \pi) \geq 3$ or (R, S) is a Yes-instance. Assume therefore that $\pi(x) \cap \pi^*(x) = \{x\}$. By the first part of the induction claim, we conclude that $\phi_{T^1_0}(\pi^*, \pi) \geq 3$ or $\{b^i_k \colon i \in [2^k]\} \subseteq N(\pi, \pi^*)$. Since we are done in the former case, we assume that $\{b^i_k \colon i \in [2^k]\} \subseteq N(\pi, \pi^*)$. This can only happen if, for every $i \in 1, \dots, |R|$, there exists an $s_i \in S$ with $y^{s_i}_2 \in \pi(b^i_k)$. Indeed, if this is not the case, then the utility of b^i_k is bounded by $\frac{2(k+1) + \frac{2\lambda}{3}(k+1)}{3+\lambda} = \frac{2}{3}(k+1) = v_{b^i_k}(\pi^*)$, where $\lambda = |\{b^j_k \colon j \in [|R|]\} \cap (\pi(b^i_k) \setminus \{b^i_k\})|$. Note that the equality is true for every $\lambda \geq 0$. Hence, $b^i_k \notin N(\pi, \pi^*)$.

Define $S' = \{s \in S \colon \pi(y^s_2) \cap \{b^i_k \colon i \in [|R|]\} \neq \emptyset\}$. Now fix $s \in S'$ and define $C = \pi(y^s_2)$. We deal first with the case that $y^s_1 \in C$ and let $r^i \in R$ with $b^i_k \in C$. We claim that $a^i_k, c^i_k \in C$. Otherwise, for some $\lambda \geq 0$, $v_{b^i_k}(\pi) \leq \frac{\frac{3}{2}(k+1) + (k+1) + \frac{2\lambda}{3}(k+1)}{4+\lambda} < \frac{2}{3}(k+1) = v_{b^i_k}(\pi^*)$, and $b^i_k \notin N(\pi, \pi^*)$, which is a contradiction. Hence, $a^i_k, c^i_k \in C$. If $y^s_2 \in N(\pi^*, \pi)$, we are done, because then $\phi(\pi^*,\pi) \geq \phi_{\{y_1,y_2\}}(\pi^*,\pi)+\phi_{\{y_1^s,y_2^s\}}(\pi^*,\pi)+\sum_{s'\in S\setminus\{s\}} +\phi_{\{y_1^{s'},y_2^{s'}\}}(\pi^*,\pi)+\phi_{T_0^1}(\pi^*,\pi) \geq 0+2+0+1 = 3$. Now, if $C \cap \{b_k^j \colon j \in [|R|]\} = \{b_k^i\}$, then $v_{y_2^s}(\pi) \leq \frac{\frac{21}{10}(k+1)+\frac{3}{2}(k+1)}{5} < \frac{21}{20}(k+1) = v_{y_2^s}(\pi^*)$, but we already excluded that. Thus, there is $i' \neq i$ with $b_k^{i'} \in C$. It is easy to see that $b_k^{i'} \in N(\pi^*,\pi)$, which is contradicting our assumption that $\{b_k^i \colon i \in [2^k]\} \subseteq N(\pi,\pi^*)$. This concludes the case that $y_1^s \in C$ and we assume henceforth that, for all $s \in S'$, $y_1^s \notin C$.

Let $I = s \cap \{r^i \in R \colon b_k^i \in C\}$ the set of members of s whose corresponding agents are in the coalition C. If $|I| \leq 2$, then $v_{y_2^s}(\pi) \leq \frac{\frac{6}{2}(k+1)}{3} = k+1 < \frac{21}{20}(k+1) = v_{y_2^s}(\pi^*)$. However, it is already excluded that $y_2^s \in N(\pi^*,\pi)$. Hence, $|I| = 3$. In other words, $\pi(y_2^s) = \{y_2^s, b_k^i, b_k^j, b_k^w\}$ with $s = \{i,j,w\}$. We conclude that S' is a 3-partition of R by sets in S.

We will now proceed with the proof of the induction claim.

For the base case $j = k$, fix $i \in [2^k]$ and assume that $A_k^i \notin \pi$. We observe that if $A_k^i \cap N(\pi,\pi^*) \neq \emptyset$, then clearly $\phi_{A_k^i}(\pi^*,\pi) \geq 1$. If $A_k^i \cap N(\pi,\pi^*) = \emptyset$, then $\{a_k^i, c_k^i\} \subseteq N(\pi^*,\pi)$ and $\phi_{A_k^i}(\pi^*,\pi) \geq 1$. If in addition $\{b_k^i, a_k^i\} \cap \pi(c_k^i) = \emptyset$, then $b_k^i \in N(\pi^*,\pi) \cup N(\pi,\pi^*)$ and the first part of the claim follows.

For the induction step, let $j \in \{k-1,\dots,0\}$ and fix $i \in [2^j]$. Assume first that there exists an agent $z \in T_j^i$ with $\pi(z) \neq \pi^*(z)$ but no such agent in A_j^i. The premise of the first claim is vacuous and this part is therefore true. Since $z \in T_{j+1}^{2i-1} \vee z \in T_{j+1}^{2i}$, we can apply induction for the second claim since the premise of the second claim for T_{j+1}^{2i-1} or T_{j+1}^{2i} is true. Assume therefore that there exists an agent $z \in A_j^i$ with $\pi(z) \neq \pi^*(z)$.

We make the following observations.

- If $\alpha_j^i \in N(\pi,\pi^*)$, then $\beta_j^i \in N(\pi^*,\pi)$.
- If $\beta_j^i \in N(\pi,\pi^*)$, then $\alpha_j^i \in N(\pi^*,\pi)$.
- If $\gamma_j^i \in N(\pi,\pi^*)$, then $\delta_j^i \in N(\pi^*,\pi)$.
- If $\delta_j^i \in N(\pi,\pi^*)$, then $\gamma_j^i \in N(\pi^*,\pi)$.

Now, we consider the case that $\pi(a_j^i) \neq \pi^*(a_j^i)$.

- We consider first the subcase that $b_j^i \in N(\pi,\pi^*)$. Then $c_j^i \in N(\pi^*,\pi)$.
  - If $\pi(b_j^i) \supseteq \{c_{j+1}^{2i-1}, c_{j+1}^{2i}\}$, then $\phi_{A_j^i}(\pi,\pi^*) \leq 1$ (with the above observations), while by induction $\phi_{T_{j+1}^{2i-1}\cup T_{j+1}^{2i}}(\pi^*,\pi) \geq 2$ and $\phi_{T_{j+1}^{2i-1}\cup T_{j+1}^{2i}}(\pi^*,\pi) \geq 4 \vee \{b_k^i \colon i \in [2^k]\} \cap (T_{j+1}^{2i-1} \cup T_{j+1}^{2i}) \subseteq N(\pi,\pi^*)$ and we are done.
  - Otherwise, $c_j^i \in \pi(b_j^i)$ and $\pi(b_j^i) \cap \{c_{j+1}^{2i-1}, c_{j+1}^{2i}\} \neq \emptyset$. Then $\phi_{A_j^i}(\pi^*,\pi) \geq 1$ or $a_j^i \in N(\pi,\pi^*)$. We only need to consider the second case. Assume for contradiction that $a_j^i \in \pi(b_j^i)$. Then, $\pi(b_j^i) \cap \{\beta_j^i, \gamma_j^i\} \neq \emptyset$ (otherwise, $a_j^i \in N(\pi^*,\pi)$). Then, $v_{b_j^i}(\pi) \leq \frac{3j+4}{5} < \frac{2j+2.5}{3} = v_{b_j^i}(\pi^*)$, contradicting our assumption on b_j^i (note that we used that $\pi(b_j^i) \not\supseteq \{c_{j+1}^{2i-1}, c_{j+1}^{2i}\}$). Therefore, $a_j^i \notin \pi(b_j^i)$ and therefore $\pi(a_j^i) = \{a_j^i, \beta_j^i, \gamma_j^i\}$. Hence, $\phi_{A_j^i}(\pi^*,\pi) \geq 1$ or $\pi(\delta_j^i) = \{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\}$. But then $\phi_{A_j^i}(\pi^*,\pi) \geq -1$ and $\phi_{T_{j+1}^{2i-1}\cup T_{j+1}^{2i}}(\pi^*,\pi) \geq 2$ and we are done.

- We can even assume that $b_j^i \in N(\pi^*, \pi)$, since otherwise $a_j^i \in \pi(b_j^i)$ and $a_j^i, c_j^i \in N(\pi^*, \pi)$ and it follows $\phi_{A_j^i}(\pi^*, \pi) \geq 1$.
- If $c_j^i \in N(\pi, \pi^*)$, then $a_j^i, b_j^i \in N(\pi^*, \pi)$ and therefore $\phi_{A_j^i}(\pi^*, \pi) \geq 1$ and we are done.
- Since $\pi(c_j^i) \neq \pi^*(c_j^i)$, we can assume $c_j^i \in N(\pi^*, \pi)$
- Next, consider the case that $a_j^i \in N(\pi, \pi^*)$ and, by the previous cases, $c_j^i, b_j^i \in N(\pi^*, \pi)$.
  - If $\pi(a_j^i) = \{a_j^i, \beta_j^i, \gamma_j^i\}$, then $\phi_{A_j^i}(\pi^*, \pi) \geq 3$ or $\pi(\delta_j^i) = \{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\}$. In the latter case, $\phi_{A_j^i}(\pi^*, \pi) \geq 1$ and $\phi_{T_{j+1}^{2i-1} \cup T_{j+1}^{2i}}(\pi^*, \pi) \geq 2$ by induction and we are done.
  - Otherwise, $\beta_j^i \in \pi(a_j^i) \cap N(\pi^*, \pi)$ or $\gamma_j^i \in \pi(a_j^i) \cap N(\pi^*, \pi)$. In the former case, $\alpha_j^i \in N(\pi^*, \pi)$ and in total $\phi_{A_j^i}(\pi^*, \pi) \geq 3$. In the latter case, again, $\phi_{A_j^i}(\pi^*, \pi) \geq 3$ or $\pi(\delta_j^i) = \{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\}$ and the case is similar as before.
- Note that a_j^i is not indifferent between $\pi(a_j^i)$ and $\pi^*(a_j^i)$, because $\pi(a_j^i) \neq \pi^*(a_j^i)$. It remains that $a_j^i, b_j^i, c_j^i \in N(\pi^*, \pi)$, in which case $\phi_{A_j^i}(\pi^*, \pi) \geq 3$.

We may therefore assume that $\pi(a_j^i) = \pi^*(a_j^i)$. Only for the remaining cases, we need that $\alpha_j^i \notin N(\pi, \pi^*)$. If $\pi(\alpha_j^i) \neq \pi^*(\alpha_j^i)$, then $\alpha_j^i, \beta_j^i \in N(\pi^*, \pi)$ and consequently $\phi_{A_j^i}(\pi^*, \pi) \geq 2$. If $\pi(\gamma_j^i) \neq \pi^*(\gamma_j^i)$, then $\phi_{A_j^i}(\pi^*, \pi) \geq 2$ or $\phi_{A_j^i}(\pi, \pi^*) \geq 0 \wedge \pi(\delta_j^i) \cap \{\alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\} \neq \emptyset$ and the claim follows by induction.

For the second part of the lemma, assume that S' is a 3-partition of R through sets in S. Define

$$\begin{aligned}
\pi' = &\{\{b_k^v, b_k^w, b_k^x, y_2^s\}, \{y_1^s\} \colon \{r^v, r^w, r^x\} = s \in S'\} \cup \{\{y_1^s, y_2^s\} \colon s \in S \setminus S'\} \\
&\cup \{\{b_k^{|R|+1}, \dots, b_k^{2^k}, y_2\}, \{y_1\}\} \cup \{\{\delta_{k-1}^i, a_k^{2i-1}, a_k^{2i}\} \colon i \in [2^{k-1}]\} \\
&\cup \{\{b_j^i, c_{j+1}^{2i-1}, c_{j+1}^{2i}\}, \{a_j^i, \beta_j^i, \gamma_j^i\} \colon j \in [k-1], i \in [2^j]\} \\
&\cup \{\{\delta_j^i, \alpha_{j+1}^{2i-1}, \alpha_{j+1}^{2i}\} \colon j \in [k-2], i \in [2^j]\} \cup \{\{\alpha_0^1\}, \{c_0^1\}\}.
\end{aligned}$$

It is easily checked that $\phi(\pi', \pi^*) = 1$ and that c_0^1 forms a singleton coalition with $c_0^1 \in N(\pi^*, \pi')$. $\square$

Theorem 16. *Checking whether there exists a strongly popular partition in a symmetric FHG is coNP-hard, even if all utilities are non-negative.*

Proof. The reduction is from X3C. Given an instance (R, S) of X3C, we consider the symmetric, non-negative FHG of Lemma 4 on agent set N with utility function v together with the partition π^* and the special agent $x \in N$. We define a symmetric, non-negative FHG on agent set $N' = N \cup \{z\}$ where the utilities are given by $v'(y, w) = v(y, w)$ if $y, w \in N$, $v'(z, x) = v_x(\pi^*)/2$, and $v'(z, y) = 0$ for $y \in N \setminus \{x\}$. Note that by Lemma 4, this reduction is in polynomial time.

Consider the partition $\sigma^* = \pi^* \cup \{\{z\}\}$ and let a partition $\sigma \neq \sigma^*$ of N' be given. Define $\pi = (\sigma \setminus \sigma(z)) \cup \{\sigma(z) \setminus \{z\}\}$. Note that every agent $y \in N \setminus \{x\}$ can only improve her utility if z leaves her coalition. In addition, the utility $v(x,z)$ is designed so that x still receives her unique top-ranked coalition in σ^* (apply Proposition 10). Hence, $\phi_N(\sigma^*, \sigma) \geq \phi(\pi^*, \pi)$.

We consider the popularity margin between σ^* and σ by a case distinction. If $\pi \neq \pi^*$, then $\phi(\sigma^*, \sigma) \geq -1 + \phi(\pi^*, \pi) \geq 0$ and $\phi(\sigma^*, \sigma) > 0$ if (R,S) is a No-instance. On the other hand, if $\pi = \pi^*$, then $\sigma(z) \neq \{z\}$ (since $\sigma \neq \sigma^*$). As $v_y(\pi^*) > 0$ for all $y \in N$, we know that $|\sigma(z) \setminus \{z\}| \geq 2$ and $y \in N(\sigma^*, \sigma)$ for all $y \in \sigma(z) \setminus \{z\}$ (by design of the utilities, this holds in particular for agent x). Hence, $\phi(\sigma^*, \sigma) = \phi_{\sigma(z)}(\sigma^*, \sigma) \geq -1 + |\sigma(z) \setminus \{z\}| > 0$

It follows that σ^* is popular and it is a strongly popular partition if (R,S) is a No-instance.

If (R,S) is a Yes-instance, then σ^* is the only candidate that might be strongly popular. Consider the partition π' from Lemma 4 and define $\sigma' = (\pi' \setminus \{\{x\}\}) \cup \{\{x,z\}\}$. Then, $x \in N(\pi^*, \pi') \cap N(\sigma^*, \sigma')$, whereas $z \in N(\sigma', \sigma^*)$. Therefore, $\phi(\sigma', \sigma) = 1 + \phi(\pi', \pi^*) = 0$. Hence, π^* is not strongly popular and there exists no strongly popular partition. □

Theorem 17. *Verifying whether a given partition in a symmetric FHG is strongly popular is coNP-complete, even if all utilities are non-negative.*

Proof. In the proof of Theorem 10, the partition σ^* is strongly popular if, and only if, (R,S) is a No-instance of X3C. □

Theorem 18. *Computing a mixed popular partition in a symmetric FHG is NP-hard, even if all utilities are non-negative.*

Proof. We give a Turing reduction from X3C. Given an instance (R,S) of X3C, we consider the symmetric FHG of Lemma 4 on agent set N with utility function v together with the partition π^* and the special agent $x \in N$. We define a symmetric, non-negative FHG on agent set $N' = N \cup \{z_1, z_2\}$ where the utilities are given by $v'(y,w) = v(y,w)$ if $y, w \in N$, $v'(z_1, z_2) = v_x(\pi^*)/2$, $v'(z_1, x) = v'(z_2, x) = v_x(\pi^*)/3 > 0$, and $v'(z_i, y) = 0$ for $i \in [2], y \in N \setminus \{x\}$. Note that by Lemma 4, this reduction is in polynomial time.

Consider the partition $\sigma^* = \pi^* \cup \{\{z_1, z_2\}\}$ and let $\sigma \neq \sigma^*$ be given. Define $\pi = (\sigma \setminus (\sigma(z_1) \cup \sigma(z_2))) \cup \{\sigma(z_1) \setminus \{z_1, z_2\}, \sigma(z_2) \setminus \{z_1, z_2\}\}$, that is, the partition of agent set N where z_1 and z_2 leave their coalitions. Assume that (R,S) is a No-instance. We will prove that $\phi(\sigma^*, \sigma) > 0$, and therefore that σ^* is strongly popular. We may assume that $\sigma(z_1) = \{z_1, z_2\}$ or $x \in \sigma(z_i)$ for some i, because otherwise it is a Pareto improvement if z_1 and z_2 leave their coalitions and form a coalition of their own.

Note that as in the proof of Theorem 16, it holds that $\phi_N(\sigma^*, \sigma) \geq \phi(\pi^*, \pi)$. Now, for $i \in [2]$ holds that $z_i \in N(\sigma^*, \sigma)$ unless $\sigma(z_i) \in \{\{z_1, z_2, x\}, \{z_1, z_2\}\}$. If $\sigma(z_i) = \{z_1, z_2\}$, then $\phi(\sigma^*, \sigma) = \phi(\pi^*, \pi) \geq 1$, because $\pi \neq \pi^*$. On the other hand, $\sigma(z_i) = \{z_1, z_2, x\}$, then $\pi(x) \cap \pi^*(x) = \{x\}$ and it follows that $\phi(\sigma^*, \sigma) \geq -2 + \phi(\pi^*, \pi) \geq 1$ (where the last inequality uses Lemma 4). It remains the case that $z_1, z_2 \in N(\sigma^*, \sigma)$ and we obtain $\phi(\sigma^*, \sigma) \geq 2 + \phi(\pi^*, \pi) \geq 2$. Together, the partition σ^* is strongly popular and therefore, the unique mixed popular partition consists of σ^* with probability 1.

Now assume that (R,S) is a Yes-instance. Consider the partition π' from Lemma 4 and define $\sigma' = (\pi' \setminus \{\{x\}\}) \cup \{\{x, z_1, z_2\}\}$. Then, $x \in N(\pi^*, \pi') \cap N(\sigma^*, \sigma')$, whereas

$z_1, z_2 \in N(\sigma', \sigma^*)$. Therefore, $\phi(\sigma', \sigma) = 2 + \phi(\pi', \pi^*) = 1$. Hence, the pure mixed partition $\{\sigma^*\}$ is not mixed popular.

We can solve X3C by computing a partition σ in the support of a mixed popular partition and checking its probability in case that $\sigma = \sigma^*$. $\square$

Theorem 19. *Checking whether there exists a popular partition in a symmetric FHG is coNP-hard, even if all utilities are non-negative.*

Proof. We provide a reduction from X3C. Given an instance (R, S) of X3C, we consider the symmetric FHG with non-negative utility functions of Lemma 4 on agent set N with utility function v together with the partition π^* and the special agent $x \in N$. Set $\alpha = v_x(\pi^*)$. For $i \in [2]$, let $N_i = \{y_i : y \in N\}$ be two copies of N. Accordingly, let π_i^* be their respective copies of π^*.

We define a symmetric ASHG on agent set $N' = N_1 \cup N_2 \cup Z$ where $Z = \{z_k^j : k \in [2], j \in [3]\}$. Define $Z^j = \{z_1^j, z_2^j\}$. Utilities are as follows.

- $v'(y_i, w_i) = v(y, w)$ if $y, w \in N_i$ for $i \in [2]$,
- $v'(z_k^j, x_1) = 2\alpha/5$, $v'(z_k^j, x_2) = \alpha/3$ for $k \in [2], j \in [3]$,
- $v'(z_1^j, z_2^j) = \alpha/2$ for $j \in [3]$, and
- $v'(u, y) = 0$ for every pair of agents $u, y \in N'$ such that their utility is not yet defined.

By Lemma 4, this reduction is in polynomial time.

First assume that (R, S) is a No-instance. We claim that $\sigma^* = \pi_1^* \cup \pi_2^* \cup \{Z^j : j \in [3]\}$ is popular. To prove this, let $\sigma \neq \sigma^*$ be an arbitrary partition and define $\pi_i = \{\sigma(y) \cap N_i : y \in N_i\}$ be the coalitions restricted to N_i. Let $k \in [2]$ and $j \in [3]$. The first key insight is that if there exists $y \in \sigma(z_k^j) \setminus (Z^j \cup \{x_1, x_2\})$, then $z_k^j \in N(\sigma^*, \sigma)$. Assume that such an agent y exists. Observe that the only agents that provide positive utility to z_k^j are z_{3-k}^j, x_1, and x_2. The maximum utility that under these circumstances can be obtained for z_k^j is if $\sigma(z_k^j) = \{z_k^j, z_{3-k}^j, x_1, y\}$ and even in this case $v_{z_k^j}(\sigma) = \frac{\frac{\alpha}{2} + \frac{2\alpha}{5}}{4} = \frac{9\alpha}{40} < \frac{\alpha}{4} = v_{z_k^i}(\sigma^*)$.

We will use this insight to show that we can assume for every $k \in [2]$, $j \in [3]$ that $\sigma(z_k^j) \subseteq Z^j \cup \{x_1, x_2\}$. Fix again $k \in [2], j \in [3]$ and assume otherwise. Then, $\sigma(z_k^j) \cap (Z^j \cup \{x_1, x_2\}) \subseteq N(\sigma^*, \sigma)$. This follows for agents in Z^j from what we have just shown before, and for agents x_i by the design of their utilities and the fact that they received a top-ranked coalition in π_i^* and by Proposition 10 in σ^*. We modify σ by leaving the coalition with the agents in Z^j, that is, we define $\sigma' = (\sigma \setminus \sigma(z_k^j)) \cup \{\sigma(z_k^j) \setminus Z^j, \sigma(z_k^j) \cap Z^j\}$. Then, $N(\sigma^*, \sigma') \subseteq N(\sigma^*, \sigma)$ and $N(\sigma, \sigma^*) \subseteq N(\sigma', \sigma^*)$, which implies that $\phi(\sigma^*, \sigma) \geq \phi(\sigma^*, \sigma')$ and it suffices to consider σ' and show a non-negative popularity margin for that partition.

We are ready to compute the popularity margin. Therefore, define $I = \{i \in [2] : \sigma(x_i) \cap Z \neq \emptyset\}$. Note that for $i \in [2]$, $\phi_{N_i}(\sigma^*, \sigma) \geq \phi(\pi_i^*, \pi_i)$. Furthermore, if $i \in I$, then $\pi_i(x_i) \cap N_i = \{x_i\}$ and $|Z \cap \sigma(x_i)| \leq 2$. It follows that $\phi(\sigma^*, \sigma) = \phi_{N_1}(\sigma^*, \sigma) + \phi_{N_2}(\sigma^*, \sigma) + \phi_Z(\sigma^*, \sigma) \geq \sum_{i \in I} \phi_{N_i}(\pi_i^*, \pi_i) + \sum_{i \notin I} \phi_{N_i}(\pi_i^*, \pi_i) + \phi_Z(\sigma^*, \sigma) \geq 3|I| - |\{z \in Z : \sigma(z) \cap \{x_1, x_2\} \neq \emptyset\}| = 3|I| - 2|I| \geq 0$. Hence, σ^* is popular.

Conversely, assume that (R, S) is a Yes-instance and assume for contradiction that σ is popular and define $\pi_i = \{\sigma(y) \cap N_i : y \in N_i\}$ as above.

The overall proof strategy is as follows. First, we show that for $k \in [2]$ and $j \in [3]$, $\sigma(z_k^j) \in \{Z^j, Z^j \cup \{x_1\}, Z_j \cup \{x_2\}\}$. Then we show, that for $i \in [2]$, there exists $j \in [3]$ with $Z^j \cup \{x_i\} \in \sigma$. Finally, we perform a cyclic exchange of such coalitions.

Let $k \in [2]$ and $j \in [3]$ and define $C = \sigma(z_k^j)$. The first crucial step is to show that $C \subseteq \{x_1, x_2\} \cup Z^j$. To see this, assume for contradiction that there exists an agent $y \in C \setminus (\{x_1, x_2\} \cup Z^j)$. We may assume that $v_y(\sigma) > 0$, since otherwise leaving the coalition with y yields a Pareto-improvement. Recall, that we have shown in the first part of the proof that, under these circumstances, $v_{z_k^j}(Z^j) > v_{z_k^j}(\sigma)$. The same holds for z_{3-k}^j in both the case that $z_{3-k}^j \in C$ and $z_{3-k}^j \notin C$. Define $\sigma' = (\sigma \setminus \{\sigma(z_1^j), \sigma(z_2^j)\}) \cup \{\sigma(z_1^j) \setminus \{z_1^j\}, \sigma(z_2^j) \setminus \{z_2^j\}, Z^j\}$. Then $\{z_1^j, z_2^j, y\} \subseteq N(\sigma', \sigma)$, while $N(\sigma, \sigma') \subseteq \{x_1, x_2\}$. Hence, σ' is more popular, which is a contradiction. It follows that $C \subseteq \{x_1, x_2\} \cup Z^j$.

Next, we claim that $z_{3-k}^j \in \sigma(z_k^j)$. Assume otherwise. If one of z_k^j and z_{3-k}^j is in a singleton coalition, it is a Pareto improvement to form $\sigma(z_k^j) \cup \sigma(z_{3-k}^j)$. Otherwise, there exists $i \in [2]$ with $\sigma(z_k^j) = \{x_i, z_k^j\}$ and if $\sigma(z_{3-k}^j) = \{z_{3-k}^j, x_{3-i}\}$. Hence, if z_{3-k}^j leaves her coalition and joins $\sigma(z_k^j)$, we obtain a more popular partition.

Define $I = \{i \in [2] : Z \cap \sigma(x_i) \neq \emptyset\}$ and let $i \in I$. We claim that there exists $j \in [3]$ with $\sigma(x_i) = \{x_i\} \cup Z^j$. Let $k \in [2], j \in [3]$ with $z_k^j \in \sigma(x_i)$. We already know that then $Z^j \subseteq \sigma(x_i) \subseteq Z^j \cup \{x_1, x_2\}$. Furthermore, by the pigeon hole principle, for some $j' \in [3] \setminus \{j\}$ holds $Z^{j'} \in \sigma$. Assume for contradiction that $x_{3-i} \in \sigma(x_i)$. Then, $\sigma' = (\sigma \setminus \{\sigma(x_i), Z^{j'}\}) \cup \{Z^j \cup \{x_1\}, Z^{j'} \cup \{x_2\}\}$ is more popular. Indeed, $N(\sigma', \sigma) = \{x_1, x_2, z_1^{j'}, z_2^{j'}\}$, while $N(\sigma, \sigma') = Z^j$.

The remainder of the proof is identical to the proof for ASHGs, namely we show that $I = \{1, 2\}$ and find a more popular partition even in this case.

All in all, it is shown that there exists no popular partition if (R, S) is a Yes-instance. This concludes the proof of the theorem. $\square$

References

Abraham, D. J., Leravi, A., Manlove, D. F., & O'Malley, G. (2008). The stable roommates problem with globally ranked pairs. *Internet Mathematics*, *5*(4), 493–515.

Abraham, D. K., Irving, R. W., Kavitha, T., & Mehlhorn, K. (2007). Popular matchings. *SIAM Journal on Computing*, *37*(4), 1030–1034.

Aziz, H., Brandl, F., Brandt, F., & Brill, M. (2018). On the tradeoff between efficiency and strategyproofness. *Games and Economic Behavior*, *110*, 1–18.

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

Aziz, H., Brandt, F., & Harrenstein, P. (2013a). Pareto optimality in coalition formation. *Games and Economic Behavior*, *82*, 562–581.

Aziz, H., Brandt, F., & Seedig, H. G. (2013b). Computing desirable partitions in additively separable hedonic games. *Artificial Intelligence*, *195*, 316–334.

Aziz, H., Brandt, F., & Stursberg, P. (2013c). On popular random assignments. In *Proceedings of the 6th International Symposium on Algorithmic Game Theory (SAGT)*, Vol. 8146 of *Lecture Notes in Computer Science (LNCS)*, pp. 183–194. Springer-Verlag.

Aziz, H., & Savani, R. (2016). Hedonic games. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (Eds.), *Handbook of Computational Social Choice*, chap. 15. Cambridge University Press.

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

Biró, P., Irving, R. W., & Manlove, D. F. (2010). Popular matchings in the marriage and roommates problems. In *Proceedings of the 7th Italian Conference on Algorithms and Complexity (CIAC)*, pp. 97–108.

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

Brandl, F., & Brandt, F. (2020). Arrovian aggregation of convex preferences. *Econometrica*, *88*(2), 799–844.

Brandl, F., Brandt, F., & Hofbauer, J. (2017). Random assignment with optional participation. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 326–334.

Brandl, F., Brandt, F., & Seedig, H. G. (2016). Consistent probabilistic social choice. *Econometrica*, *84*(5), 1839–1880.

Brandl, F., Brandt, F., & Stricker, C. (2022). An analytical and experimental comparison of maximal lottery schemes. *Social Choice and Welfare*, *58*(1), 5–38.

Brandl, F., & Kavitha, T. (2018). Two problems in max-size popular matchings. *Algorithmica*, *81*(7), 2738–2764.

Brandt, F., Hofbauer, J., & Suderland, M. (2017). Majority graphs of assignment problems and properties of popular random assignments. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 335–343.

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

Cseh, Á. (2017). Popular matchings. In Endriss, U. (Ed.), *Trends in Computational Social Choice*, chap. 6. AI Access.

Cseh, Á., Huang, C.-C., & Kavitha, T. (2015). Popular matchings with two-sided preferences and one-sided ties. In *Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming (ICALP)*, Vol. 9134 of *Lecture Notes in Computer Science (LNCS)*, pp. 367–379. Springer-Verlag.

Cseh, Á., & Kavitha, T. (2018). Popular matchings in complete graphs. In *Proceedings of the 37th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*.

Cseh, Á., & Peters, J. (2022). Three-dimensional popular matching with cyclic preferences. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. Forthcoming.

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

Edmonds, J. (1965). Maximum matching and a polyhedron with 0,1-vertices. *Journal of Research of the National Bureau of Standards B*, *69*, 125–130.

Faenza, Y., Kavitha, T., Power, V., & Zhang, X. (2019). Popular matchings and limits to tractability. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2790–2809.

Fishburn, P. C. (1984). Probabilistic social choice based on simple voting comparisons. *Review of Economic Studies*, *51*(4), 683–692.

Gärdenfors, P. (1975). Match making: Assignments based on bilateral preferences. *Behavioral Science*, *20*(3), 166–173.

Grötschel, M., Lovász, L., & Schrijver, A. (1981). The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, *1*, 169–197.

Gupta, S., Misra, P., Saurabh, S., & Zehavi, M. (2019). Popular matching in roommates setting is np-hard. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2810–2822.

Huang, C.-C., & Kavitha, T. (2011). Popular matchings in the stable marriage problem. In *Proceedings of the 38th International Colloquium on Automata, Languages, and Programming (ICALP)*, Vol. 6755 of *Lecture Notes in Computer Science (LNCS)*, pp. 666–677. Springer-Verlag.

Huang, C.-C., & Kavitha, T. (2017). Popularity, mixed matchings, and self-duality. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2294–2310.

Huang, C.-C., & Kavitha, T. (2021). Popularity, mixed matchings, and self-duality. *Mathematics of Operations Research*, *46*(2), 405–427.

Karp, R. M. (1972). Reducibility among combinatorial problems. In Miller, R. E., & Thatcher, J. W. (Eds.), *Complexity of Computer Computations*, pp. 85–103. Plenum Press.

Kavitha, T. (2014). A size-popularity tradeoff in the stable marriage problem. *SIAM Journal on Computing*, *43*(1), 52–71.

Kavitha, T., Mestre, J., & Nasre, M. (2011). Popular mixed matchings. *Theoretical Computer Science*, *412*(24), 2679–2690.

Kavitha, T., & Nasre, M. (2009). Optimal popular matchings. *Discrete Applied Mathematics*, *157*, 3181–3186.

Khachiyan, L. (1979). A polynomial algorithm in linear programming. *Soviet Mathematics Doklady*, *20*, 191–194.

Király, T., & Mészáros-Karkus, Z. (2017). Finding strongly popular b-matchings in bipartite graphs. *Electronic Notes in Discrete Mathematics*, *61*, 735–741.

Lam, C.-K., & Plaxton, C. G. (2019). On the existence of three-dimensional stable matchings with cyclic preferences. In *Proceedings of the 12th International Symposium on Algorithmic Game Theory (SAGT)*, Vol. 11801 of *Lecture Notes in Computer Science (LNCS)*, pp. 329–342. Springer-Verlag.

Mahdian, M. (2006). Random popular matchings. In *Proceedings of the 7th ACM Conference on Electronic Commerce (ACM-EC)*, pp. 238–242.

Manlove, D. F. (2013). *Algorithmics of Matching Under Preferences*. World Scientific Publishing Company.

McCutchen, R. M. (2008). The least-unpopularity-factor and least-unpopularity-margin criteria for matching problems with one-sided preferences. In *Proceedings of the 8th Latin American Conference on Theoretical Informatics (LATIN)*, Vol. 4957 of *Lecture Notes in Computer Science (LNCS)*, pp. 593–604.

Padberg, M. W., & Wolsey, L. A. (1984). Fractional covers for forests and matchings. *Mathematical Programming*, *29*, 1–14.

von Neumann, J. (1928). Zur Theorie der Gesellschaftspiele. *Mathematische Annalen*, *100*(1), 295–320.

von Neumann, J., & Morgenstern, O. (1947). *Theory of Games and Economic Behavior* (2nd edition). Princeton University Press.

8

CORE PUBLICATION 3: BOUNDARIES TO SINGLE-AGENT STABILITY IN ADDITIVELY SEPARABLE HEDONIC GAMES

SUMMARY

Coalition formation considers the question of how to partition a set of agents into coalitions with respect to their preferences. Additively separable hedonic games are a dominant model where cardinal single-agent values are aggregated into utilities for coalitions by taking sums of the single-agent values of all coalition members.

Output partitions in hedonic games are typically measured by notions of stability, which is defined by the absence of beneficial deviations. We follow this approach by considering stability based on single-agent deviations, where some agent abandons their current coalition to join another existing coalition or to form a singleton coalition. Naturally, permissible deviations should always lead to an improvement in utility for the deviator. However, deviations may also be constrained by demanding the consent of agents involved in the deviations, i.e., by agents in the abandoned or joined coalition. Most of the existing research focuses on the unanimous consent of one or both of these coalitions, but more recent research relaxes this to majority-based consent.

Our contribution is twofold. First, we settle the computational complexity of the existence of contractually Nash stable partitions. Contractual Nash stability requires the absence of deviations that are constrained by the unanimous consent of the abandoned coalition. This resolves the complexity of the last classical stability notion for additively separable hedonic games.

Second, we identify clear boundaries to the tractability of stable partitions with respect to majority-based stability concepts. These results even hold in severely restricted classes of additively separable hedonic games where the utility values are restricted to one positive and one negative weight. In these classes, agents can be viewed as either friends or enemies, dependent on whether they yield positive or negative utility.

A key challenge in proving hardness results for these restricted classes of additively separable hedonic games is to construct first No-instances, i.e., instances in which no stable partition exists. The obtained hardness results are opposed by efficient algorithms under slight further restrictions.

REFERENCE

M. Bullinger. **Boundaries to single-agent stability in additively separable hedonic games**. In *Proceedings of the 47th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 26:1–26:15, 2022.
DOI: https://doi.org/10.4230/LIPIcs.MFCS.2022.26

INDIVIDUAL CONTRIBUTION

This is a single-authored publication and I, Martin Bullinger, am responsible for all parts of this publication.

COPYRIGHT AGREEMENT

LIPIcs – Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Publishing
`http://www.dagstuhl.de/lipics`

Event: **47th International Symposium on Mathematical Foundations of Computer Science (MFCS 2022)**
Editors: **Stefan Szeider, Robert Ganian, and Alexandra Silva**

Title of contribution (named "contribution" in the following)

___

___

Authors (named "author" in the following)

___

___

Corresponding author's name, address, affiliation and e-mail

Name: ___

Address: ___

Affiliation: ___

E-Mail: ___

___

Date, City *Signature of the corresponding author*

[1]For details, see `https://creativecommons.org/licenses/by/4.0/`.
[2]For details, see `https://creativecommons.org/publicdomain/zero/1.0/`.

Boundaries to Single-Agent Stability in Additively Separable Hedonic Games

Martin Bullinger ✉
Technical University of Munich, Germany

Abstract

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

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

Keywords and phrases Coalition Formation, Hedonic Games, Stability

Digital Object Identifier 10.4230/LIPIcs.MFCS.2022.26

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

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

1 Introduction

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

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

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

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

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

1.1 Contribution

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

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

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

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

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

1.2 Related Work

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

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

2 Preliminaries

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

2.1 Hedonic Games

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

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

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

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

2.2 Single-Agent Stability

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

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

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

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

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

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

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

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

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

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

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

3 Contractual Nash Stability

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

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

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

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

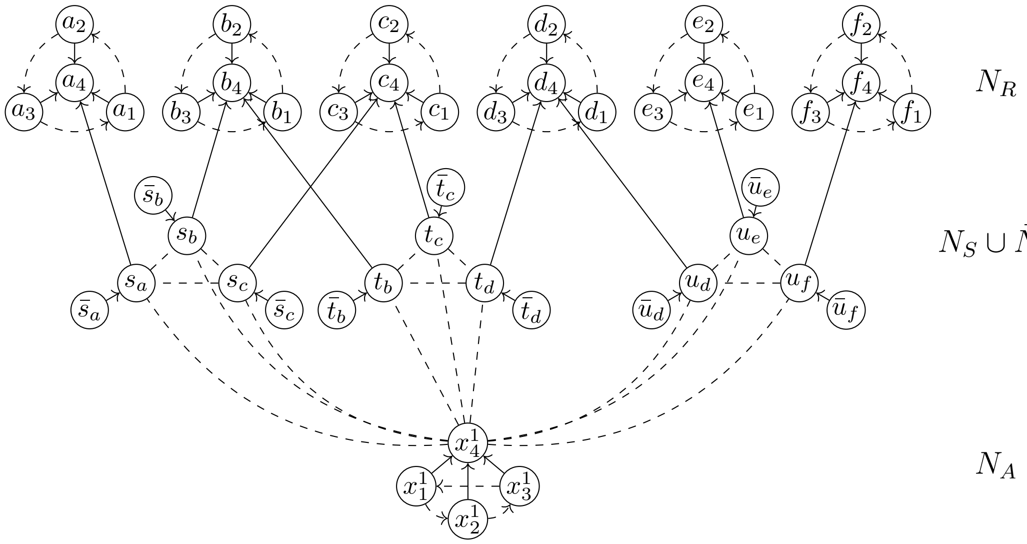


Figure 2 Schematic of the reduction from the proof of Theorem 1. We depict the reduced instance for the instance (R, S) of E3C where $R = \{a, b, c, d, e, f\}$, and $S = \{s, t, u\}$, with $s = \{a, b, c\}$, $t = \{b, c, d\}$, and $u = \{d, e, f\}$. Fully drawn edges mean a positive utility, which is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

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

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

We define valuations v as follows:

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

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

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

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

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

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

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

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

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

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

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

4 Appreciation-of-Friends Games

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

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

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

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

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

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

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

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

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

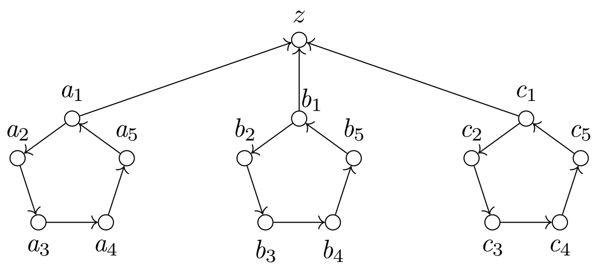


Figure 3 AFG without an MOS partition. The depicted (directed) edges represent friends, i.e., a utility of n, whereas missing edges represent a utility of -1.

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

Case 2. $|\pi(z)| > 1$.

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

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

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

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

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

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

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

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

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

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

5 Friends-and-Enemies Games

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

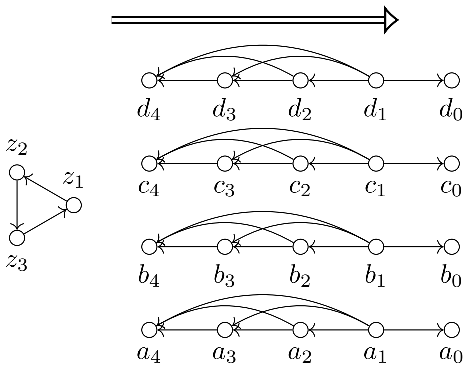


Figure 4 FEG without an MOS partition. The depicted (directed) edges represent friends. The double arrow means that every agent to the left of the tail of the arrow has every agent below the arrow as a friend.

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

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

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

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

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

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

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

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

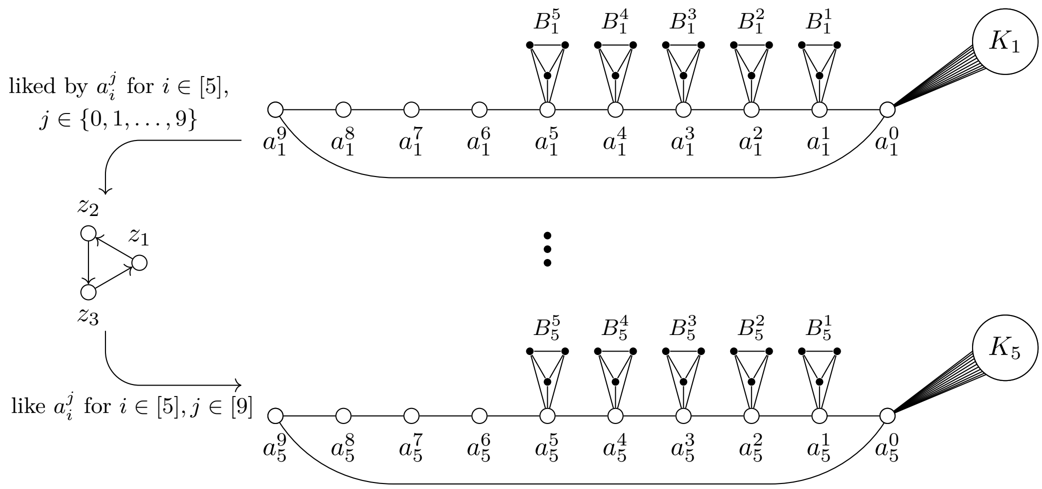


Figure 5 FEG without an MIS partition. The depicted edges represent friends. Undirected edges represent mutual friendship. For $i \in [5]$, some of the edges of agents in A_i are omitted. In fact, these agents form cliques. Also, each K_i represents a clique of 11 agents.

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

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

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

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

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

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

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

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

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

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

6 Discussion and Conclusion

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

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

9 CORE PUBLICATION 4: REACHING INDIVIDUALLY STABLE COALITION STRUCTURES IN HEDONIC GAMES

SUMMARY

The formal study of coalition formation in multi-agent systems is typically realized in the framework of so-called hedonic games, which originate from economic theory. The main focus of this branch of research has been on the existence and the computational complexity of deciding about the existence of coalition structures that satisfy various stability criteria. The actual process of forming coalitions based on individual behavior has received little attention.

In this paper, we study the convergence of simple dynamics leading to stable partitions. The basic idea is simple. Consider an arbitrary partition. Either this partition is stable or we have a cause of instability due to an agent who would like to perform a deviation. If we let this agent perform their deviation, we end up at a new partition and can once again ask for instabilities. Iterating this process we obtain a sequence of partitions induced by a sequence of deviations. We call this process a dynamics. The dynamics we consider is based on individual stability: an agent will join another coalition if she is better off and no member of the welcoming coalition is worse off.

We study dynamics in a variety of classes of hedonic games, including anonymous, fractional, and dichotomous hedonic games as well as hedonic diversity games. In this course, we identify conditions for the convergence of dynamics, provide elaborate counterexamples to the existence of individually stable partitions, and study the computational complexity of problems related to the coalition formation dynamics. The constructed counterexamples are interesting because they entail that dynamics may inevitably cycle. For instance, we find a small anonymous hedonic game and a first symmetric fractional hedonic game, in which no individually stable partition exists. Moreover, we show that dynamics may cycle in hedonic diversity games. These results settle open problems suggested by Bogomolnaia and Jackson (2002), Brandl, Brandt, and Strobel (2015), and Boehmer and Elkind (2020).

REFERENCE

F. Brandt, M. Bullinger, and A. Wilczynski. **Reaching individually stable coalition structures in hedonic games**. *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, pages 5211–5218, 2021.
DOI: https://doi.org/10.1609/aaai.v35i6.16658

INDIVIDUAL CONTRIBUTION

I, Martin Bullinger, am the main author of this publication. In particular, I am responsible for the joint development and conceptual design of the research project, many of the proofs and their write-up (in particular, all of the results about fractional hedonic games as well as Proposition 4), and the write-up of the introduction and conclusion.

COPYRIGHT AGREEMENT

The right to present this paper in a doctoral thesis has been granted by the publisher, the Association for the Advancement of Artificial Intelligence (AAAI), in the copyright form presented below. There, AAAI grants permission to "*Personal reuse of all or portions of the above article/paper in other works of their own authorship.*" This form can also be found at
https://proceedings.aaai.org/docs/AAAI-21-copyrightform.pdf
(accessed November 10, 2022).

Association for the Advancement of Artificial Intelligence

1900 Embarcadero Road, Suite 101, Palo Alto, California 94303

Palo Alto, California 94303 USA

AAAI COPYRIGHT FORM

Title of Article/Paper: ______________________________

Publication in Which Article/Paper Is to Appear: ______________________________

Author's Name(s): ______________________________

Please type or print your name(s) as you wish it (them) to appear in print

PART A – COPYRIGHT TRANSFER FORM

The undersigned, desiring to publish the above article/paper in a publication of the Association for the Advancement of Artificial Intelligence, (AAAI), hereby transfer their copyrights in the above article/paper to the Association for the Advancement of Artificial Intelligence (AAAI), in order to deal with future requests for reprints, translations, anthologies, reproductions, excerpts, and other publications.

This grant will include, without limitation, the entire copyright in the article/paper in all countries of the world, including all renewals, extensions, and reversions thereof, whether such rights current exist or hereafter come into effect, and also the exclusive right to create electronic versions of the article/paper, to the extent that such right is not subsumed under copyright.

The undersigned warrants that they are the sole author and owner of the copyright in the above article/paper, except for those portions shown to be in quotations; that the article/paper is original throughout; and that the undersigned right to make the grants set forth above is complete and unencumbered.

If anyone brings any claim or action alleging facts that, if true, constitute a breach of any of the foregoing warranties, the undersigned will hold harmless and indemnify AAAI, their grantees, their licensees, and their distributors against any liability, whether under judgment, decree, or compromise, and any legal fees and expenses arising out of that claim or actions, and the undersigned will cooperate fully in any defense AAAI may make to such claim or action. Moreover, the undersigned agrees to cooperate in any claim or other action seeking to protect or enforce any right the undersigned has granted to AAAI in the article/paper. If any such claim or action fails because of facts that constitute a breach of any of the foregoing warranties, the undersigned agrees to reimburse whomever brings such claim or action for expenses and attorneys' fees incurred therein.

Returned Rights

In return for these rights, AAAI hereby grants to the above author(s), and the employer(s) for whom the work was performed, royalty-free permission to:

1. Retain all proprietary rights other than copyright (such as patent rights).
2. Personal reuse of all or portions of the above article/paper in other works of their own authorship. This does not include granting third-party requests for reprinting, republishing, or other types of reuse. AAAI must handle all such third-party requests.
3. Reproduce, or have reproduced, the above article/paper for the author's personal use, or for company use provided that AAAI copyright and the source are indicated, and that the copies are not used in a way that implies AAAI endorsement of a product or service of an employer, and that the copies per se are not offered for sale. The foregoing right shall not permit the posting of the article/paper in electronic or digital form on any computer network, except by the author or the author's employer, and then only on the author's or the employer's own web page or ftp site. Such web page or ftp site, in addition to the aforementioned requirements of this Paragraph, shall not post other AAAI copyrighted materials not of the author's or the employer's creation (including tables of contents with links to other papers) without AAAI's written permission.
4. Make limited distribution of all or portions of the above article/paper prior to publication.
5. In the case of work performed under a U.S. Government contract or grant, AAAI recognized that the U.S. Government has royalty-free permission to reproduce all or portions of the above Work, and to authorize others to do so, for official U.S. Government purposes only, if the contract or grant so requires.

In the event the above article/paper is not accepted and published by AAAI, or is withdrawn by the author(s) before acceptance by AAAI, this agreement becomes null and void.

(1)

______________________________	______________________________
Author/Authorized Agent for Joint Author's Signature	Date (MM/DD/YYYY)
______________________________	______________________________
Employer for whom work was performed	Title (if not author)

(For jointly authored Works, all joint authors should sign unless one of the authors has been duly authorized to act as agent for the others.)

Reaching Individually Stable Coalition Structures in Hedonic Games

Felix Brandt[1], **Martin Bullinger**[1], **Anaëlle Wilczynski**[2]
[1] Institut für Informatik, Technische Universität München
[2] MICS, CentraleSupélec, Université Paris-Saclay, France
brandtf@in.tum.de, bullinge@in.tum.de, anaelle.wilczynski@centralesupelec.fr

Abstract

The formal study of coalition formation in multiagent systems is typically realized using so-called hedonic games, which originate from economic theory. The main focus of this branch of research has been on the existence and the computational complexity of deciding the existence of coalition structures that satisfy various stability criteria. The actual process of forming coalitions based on individual behavior has received little attention. In this paper, we study the convergence of simple dynamics leading to stable partitions in a variety of classes of hedonic games, including anonymous, dichotomous, fractional, and hedonic diversity games. The dynamics we consider is based on individual stability: an agent will join another coalition if she is better off and no member of the welcoming coalition is worse off. We identify conditions for convergence, provide elaborate counterexamples of existence of individually stable partitions, and study the computational complexity of problems related to the coalition formation dynamics. In particular, we settle open problems suggested by Bogomolnaia and Jackson (2002), Brandl, Brandt, and Strobel (2015), and Boehmer and Elkind (2020).

Introduction

Coalitions and coalition formation are central concerns in the study of multiagent systems as well as cooperative game theory. Typical real-world examples include individuals joining clubs or societies such as orchestras, choirs, or sport teams, countries organizing themselves in international bodies like the European Union (EU) or the North Atlantic Treaty Organization (NATO), students living together in shared flats, or employees forming unions. The formal study of coalition formation is often realized using so-called hedonic games, which originate from economic theory and focus on coalition structures (henceforth partitions) that satisfy various stability criteria based on the agents' preferences over coalitions. A partition is defined to be stable if single agents or groups of agents cannot gain by deviating from the current partition by means of leaving their current coalition and joining another coalition or forming a new one. Which kinds of deviations are permitted depends on the underlying notion of stability. Two important and well-studied questions in this context concern the existence of stable partitions in restricted classes of hedonic games and the computational complexity of finding a stable partition. However, stability is only concerned with the end-state of the coalition formation process and ignores how these desirable partitions can actually be reached. Essentially, an underlying assumption in most of the existing work is that there is a central authority that receives the preferences of all agents, computes a stable partition, and has the means to enforce this partition on the agents. By contrast, our work focuses on simple dynamics, where starting with some partition (e.g., the partition of singletons), agents deliberately decide to join and leave coalitions based on their individual preferences. We study the convergence of such a process and the stable partitions that can arise from it. For example, in some cases the only partition satisfying a certain stability criterion is the grand coalition consisting of all agents, while the dynamics based on the agents' individual decisions can never reach this partition and is doomed to cycle.

The dynamics we consider is based on *individual stability*, a natural notion of stability going back to Drèze and Greenberg (1980): an agent will join another coalition if she is better off and no member of the welcoming coalition is worse off. Individual stability is suitable to model the situations mentioned above. For instance, by Article 49 of the Treaty on European Union, admitting new members to the EU requires the unanimous approval of the current members. Similarly, by Article 10 of their founding treaty, unanimous agreement of all parties is necessary to become a member of the NATO. Also, for joining a choir or orchestra it is often necessary to audition successfully, and joining a shared flat requires the consent of all current residents. This distinguishes individual stability from Nash stability, which ignores the consent of members of the welcoming coalition.

The analysis of coalition formation processes provides more insight in the natural behavior of agents and the conditions that are required to guarantee that desirable social outcomes can be reached without a central authority. Similar dynamic processes have been studied in the special domain of matching, which only allows coalitions of size 2 (e.g., Roth and Vande Vate 1990; Abeledo and Rothblum 1995; Brandt and Wilczynski 2019). More recently, the dynamics of coalition formation have also come under scrutiny in the context of hedonic games (Bilò et al. 2018; Hoefer, Vaz, and Wagner 2018; Carosi, Monaco, and Moscardelli 2019).

While coalition formation dynamics are an object of study worthy for itself, they can also be used as a means to design algorithms that compute stable outcomes, and have been implicitly used for this purpose before. For example, the algorithm by Boehmer and Elkind (2020) for finding an individually stable partition in hedonic diversity games predefines a promising partition and then reaches an individually stable partition by running the dynamics from there. Similarly, the algorithm by Bogomolnaia and Jackson (2002) for finding an individually stable partition on games with ordered characteristics, a generalization of anonymous hedonic games, runs the dynamics using a specific sequence of deviations starting from the singleton partition.

In many cases, the convergence of the dynamics of deviations follows from the existence of potential functions, whose local optima form individually stable states. Generalizing a result by Bogomolnaia and Jackson (2002), Suksompong (2015) has shown via a potential function argument that an individually stable—and even a Nash stable—partition always exists in subset-neutral hedonic games, a generalization of symmetric additively-separable hedonic games. Using the same potential function, it can straightforwardly be shown that the dynamics converge.[1]

Another example are hedonic games with the common ranking property, a class of hedonic games where preferences are induced by a common global order (Farrell and Scotchmer 1988). The dynamics associated with core-stable deviations is known to converge to a core-stable partition that is also Pareto-optimal, thanks to a potential function argument (Caskurlu and Kizilkaya 2019). The same potential function implies convergence of the dynamics based on individual stability.

In this paper, we study the coalition formation dynamics based on individual stability for a variety of classes of hedonic games, including anonymous hedonic games (AHGs), hedonic diversity games (HDGs), fractional hedonic games (FHGs), and dichotomous hedonic games (DHGs). Whether we obtain positive or negative results often depends on the initial partition and on restrictions imposed on the agents' preferences. Computational questions related to the dynamics are investigated in two ways: the existence of a *path to stability*, that is the existence of a sequence of deviations that leads to a stable state, and the *guarantee of convergence* where every sequence of deviations should lead to a stable state. The former gives an optimistic view on the behavior of the dynamics and may be used to motivate the choice of reachable stable partitions (we can exclude "artificial" stable partitions that may never naturally form). If such a sequence can be computed efficiently, it enables a central authority to coordinate the deviations towards a stable partition. However, since this approach does not give any guarantee on the outcome of the dynamics, we also study the latter, more pessimistic, problem. Our main results are as follows.

[1]By inclusion, convergence also holds for symmetric additively-separable hedonic games. Symmetry is essential for this result to hold since an individually stable partition may not exist in additively-separable hedonic games, even under additional restrictions (Bogomolnaia and Jackson 2002).

- In AHGs, the dynamics converges for (naturally) single-peaked strict preferences. We provide a 15-agent example showing the non-existence of individually stable partitions in general AHGs. The previous known counterexample by Bogomolnaia and Jackson (2002) requires 63 agents and the existence of smaller examples was an acknowledged open problem (see Ballester 2004; Boehmer and Elkind 2020).
- In HDGs, the dynamics converges for strict and naturally singled-peaked preferences when starting from the singleton partition. In contrast to empirical evidence reported by Boehmer and Elkind (2020), we show that these preference restrictions are not sufficient to guarantee convergence from an arbitrary initial partition.
- In FHGs, the dynamics converges for simple symmetric preferences when starting from the singleton partition or when preferences form an acyclic digraph. We show that individually stable partitions need not exist in general symmetric FHGs, which was left as an open problem by Brandl, Brandt, and Strobel (2015).
- For each of these four classes, including DHGs, we show that deciding whether there is a sequence of deviations leading to an individually stable partition is NP-hard while deciding whether all sequences of deviations lead to an individually stable partition is co-NP-hard. Some of these results hold under preference restrictions and even when starting from the singleton partition.

Preliminaries

Let $N = [n] = \{1, \dots, n\}$ be a set of n agents. The goal of a coalition formation problem is to partition the agents into different disjoint coalitions according to their preferences. A solution is then a partition $\pi : N \to 2^N$ such that $i \in \pi(i)$ for every agent $i \in N$ and either $\pi(i) = \pi(j)$ or $\pi(i) \cap \pi(j) = \emptyset$ holds for every agents i and j, where $\pi(i)$ denotes the coalition to which agent i belongs. Two prominent partitions are the *singleton partition* π given by $\pi(i) = \{i\}$ for every agent $i \in N$, and the *grand coalition* π given by $\pi = \{N\}$.

Since we focus on dynamics of deviations, we assume that there exists an initial partition π_0, which could be a natural initial state (such as the singleton partition) or the outcome of a previous coalition formation process.

Classes of Hedonic Games

In a hedonic game, the agents only express preferences over the coalitions to which they belong, i.e., there are no externalities. Let $\mathcal{N}_i$ denote all possible coalitions containing agent i, i.e., $\mathcal{N}_i = \{C \subseteq N : i \in C\}$. A hedonic game is defined by a tuple $(N, (\succsim_i)_{i \in N})$ where $\succsim_i$ is a weak order over $\mathcal{N}_i$ which represents the preferences of agent i. Since $|\mathcal{N}_i| = 2^{n-1}$, the preferences are rarely given explicitly, but rather in some concise representation. These representations give rise to several classes of hedonic games:

- *Anonymous hedonic games (AHGs)* (Bogomolnaia and Jackson 2002): The agents only care about the size of the coalition they belong to, i.e., for each agent $i \in N$, there exists a weak order $\succsim_i$ over integers in $[n]$ such that $\pi(i) \succsim_i \pi'(i)$ iff $|\pi(i)| \succsim_i |\pi'(i)|$.

- *Hedonic diversity games (HDGs)* (Bredereck, Elkind, and Igarashi 2019): The agents are divided into two different types, red and blue agents, represented by the subsets R and B, respectively, such that $N = R \cup B$ and $R \cap B = \emptyset$. Each agent only cares about the proportion of red agents present in her own coalition, i.e., for each agent $i \in N$, there exists a weak order $\succsim_i$ over $\{\frac{p}{q} : p \in [|R|] \cup \{0\}, q \in [n]\}$ such that $\pi(i) \succsim_i \pi'(i)$ iff $\frac{|R \cap \pi(i)|}{|\pi(i)|} \succsim_i \frac{|R \cap \pi'(i)|}{|\pi'(i)|}$.
- *Fractional Hedonic Games (FHGs)* (Aziz et al. 2019): The agents evaluate a coalition according to how much they like each of its members on average, i.e., for each agent i, there exists a utility function $v_i : N \to \mathbb{R}$ where $v_i(i) = 0$ such that $\pi(i) \succsim_i \pi'(i)$ iff $\frac{\sum_{j \in \pi(i)} v_i(j)}{|\pi(i)|} \geq \frac{\sum_{j \in \pi'(i)} v_i(j)}{|\pi'(i)|}$. An FHG can be represented by a weighted complete directed graph $G = (N, E)$ where the weight of arc (i, j) is equal to $v_i(j)$. An FHG is *symmetric* if $v_i(j) = v_j(i)$ for every pair of agents i and j, i.e., it can be represented by a weighted complete undirected graph with weights $v(i, j)$ on each edge $\{i, j\}$. An FHG is simple if $v_i : N \to \{0, 1\}$ for every agent i, i.e., it can be represented by an unweighted directed graph where $(i, j) \in E$ iff $v_i(j) = 1$. We say that a simple FHG is *asymmetric* if, for every pair of agents i and j, $v_i(j) = 1$ implies $v_j(i) = 0$, i.e., it can be represented by an asymmetric directed graph.
- *Dichotomous hedonic games (DHGs)*: The agents only approve or disapprove coalitions, i.e., for each agent i there exists a utility function $v_i : \mathcal{N}_i \to \{0, 1\}$ such that $\pi(i) \succsim_i \pi'(i)$ iff $v_i(\pi(i)) \geq v_i(\pi'(i))$. When the preferences are represented by a propositional formula, such games are called *Boolean hedonic games* (Aziz et al. 2016).

An anonymous game (resp., hedonic diversity game) is *generally single-peaked* if there exists a linear order $>$ over integers in $[n]$ (resp., over ratios in $\{\frac{p}{q} : p \in [|R|] \cup \{0\}, q \in [n]\}$) such that for each agent $i \in N$ and each triple of integers $x, y, z \in [n]$ (resp., $x, y, z \in \{\frac{p}{q} : p \in |R| \cup \{0\}, q \in [n]\}$) with $x > y > z$ or $z > y > x$, $x \succsim_i y$ implies $y \succsim_i z$. The obvious linear order $>$ that comes to mind is, of course, the natural order over integers (resp., over rational numbers). We refer to such games as *naturally single-peaked*. Clearly, a naturally single-peaked preference profile is generally single-peaked but the converse is not true.

Dynamics of Individually Stable Deviations

Starting from the initial partition, agents can leave and join coalitions in order to improve their well-being. We focus on unilateral deviations, which occur when a single agent decides to move from one coalition to another. A *unilateral deviation* performed by agent i transforms a partition π into a partition π' where $\pi(i) \neq \pi'(i)$ and, for all agents $j \neq i$,

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

Since agents are assumed to be rational, agents only engage in a unilateral deviation if it makes them better off, i.e., $\pi'(i) \succ_i \pi(i)$. Any partition in which no such deviation is possible is called *Nash stable (NS)*.

This type of deviation can be refined by additionally requiring that no agent in the welcoming coalition is worse off when agent i joins. A partition in which no such deviation is possible is called *individually stable (IS)*. Formally, a unilateral deviation performed by agent i who moves from coalition $\pi(i)$ to $\pi'(i)$ is an IS-deviation if $\pi'(i) \succ_i \pi(i)$ and $\pi'(i) \succsim_j \pi(j)$ for all agents $j \in \pi'(i)$. Clearly, an NS partition is also IS.[2] In this article, we focus on dynamics based on IS-deviations. By definition, all terminal states of the dynamics have to be IS partitions.

We are mainly concerned with whether sequences of IS-deviations can reach or always reach an IS partition. If there exists a sequence of IS-deviations leading to an IS partition, i.e., a path to stability, then agents can coordinate (or can be coordinated) to reach a stable partition. The corresponding decision problem is described as follows.

$\exists$-IS-SEQUENCE-[HG]	
Input:	Instance of a particular class of hedonic games [HG], initial partition π_0
Question:	Does there exist a sequence of IS-deviations starting from π_0 leading to an IS partition?

In order to provide some guarantee, we also examine whether *all* sequences of IS-deviations terminate. Whenever this is the case, we say that the dynamics *converges*. The corresponding decision problem is described below.

$\forall$-IS-SEQUENCE-[HG]	
Input:	Instance of a particular class of hedonic games [HG], initial partition π_0
Question:	Does every sequence of IS-deviations starting from π_0 reach an IS partition?

We mainly investigate this problem via the study of its complement: given a hedonic game and an initial partition, does there exist a sequence of IS-deviations that cycles?

A common idea behind hardness reductions concerning these two problems is to exploit prohibitive subconfigurations that evolve from instances without an IS partition or instances which allow for cycling starting from a certain partition.

Anonymous Hedonic Games (AHGs)

Bogomolnaia and Jackson (2002) showed that IS partitions always exist in AHGs under naturally single-peaked preferences, and proved that this does not hold under general preferences, by means of a 63-agent counterexample. Here, we provide a counterexample that only requires 15 agents and additionally satisfies general single-peakedness.

Due to space restrictions, we omit some of the proofs or provide only proof sketches.

[2]It is possible to weaken the notion of individual stability even further by also requiring that no member of the *former* coalition of agent i is worse off. The resulting stability notion is called contractual individual stability and guarantees convergence of our dynamics.

Proposition 1. *There may not exist an IS partition in AHGs even when $n = 15$ and the agents have strict and generally single-peaked preferences.*

Sketch of proof. Let us consider an AHG with 15 agents with the following (incompletely specified) preferences.

$$\begin{array}{rccccccccccccc}
1: & 2 & \succ & 3 & \succ & 13 & \succ & 12 & \succ & 1 & \succ & [\dots] & & \\
2: & 13 & \succ & 3 & \succ & 2 & \succ & 1 & \succ & 12 & \succ & [\dots] & & \\
3,4: & 3 & \succ & 2 & \succ & 1 & \succ & [\dots] & & & & & & \\
5,\dots,15: & 13 & \succ & 12 & \succ & 15 & \succ & 14 & \succ & 11 & \succ & [\dots] & \succ & 1
\end{array}$$

They can be completed to be generally single-peaked w.r.t. axis $1 < 2 < 3 < 13 < 12 < 15 < 14 < 11 < \cdots < 4$.

One can prove that in an IS partition,

(i) agents 3 and 4 are in a coalition of size at most 3;
(ii) agents 5 to 15 are in the same coalition;
(iii) agents 3 and 4 are in the same coalition;
(iv) agents 1 and 2 cannot be both alone.

Therefore, agents 3 and 4 must be together, as well as agents 5 to 15, but not in the same coalition. It remains to identify the coalitions of agents 1 and 2. By *(i)*, they cannot be both with agents 3 and 4. If one agent among them is alone and the other one with agents 5 to 15, then the alone agent can deviate to join them, a contradiction. The remaining possible partitions are present in the cycle of IS-deviations below (the deviating agent is written on top of the arrows).

$$\begin{tikzcd}
\{\{1\},\{2,3,4\},\{5,\dots,15\}\} \arrow[r, "1"] & \{\{2,3,4\},\{1,5,\dots,15\}\} \arrow[r, "2"] & \{\{3,4\},\{1,2,5,\dots,15\}\} \arrow[d, "1"] \\
\{\{1,2\},\{3,4\},\{5,\dots,15\}\} \arrow[u, "2"] & \{\{2\},\{1,3,4\},\{5,\dots,15\}\} \arrow[l, "1"'] & \{\{1,3,4\},\{2,5,\dots,15\}\} \arrow[l, "2"']
\end{tikzcd}$$

Hence, there is no IS partition in this instance. □

However, even in smaller examples where IS partitions do exist, there may still be cycles in the dynamics.

Proposition 2. *The dynamics of IS-deviations may cycle in AHGs even when starting from the singleton partition or grand coalition, for strict generally single-peaked preferences, and for $n < 15$.*

Proof. Let us consider an AHG with 7 agents with the following (incompletely specified) preferences.

$$\begin{array}{rccccccccccc}
1: & 2 & \succ & 3 & \succ & 5 & \succ & 4 & \succ & 1 & \succ & [\dots] \\
2: & 5 & \succ & 3 & \succ & 2 & \succ & 1 & \succ & 4 & \succ & [\dots] \\
3,4: & 3 & \succ & 2 & \succ & 1 & \succ & [\dots] & & & & \\
5,6,7: & 5 & \succ & 4 & \succ & 3 & \succ & 2 & \succ & 1 & \succ & [\dots]
\end{array}$$

They can be completed to be generally single-peaked w.r.t. axis $1 < 2 < 3 < 5 < 4 < 6 < 7$. Note that $\{\{1\},\{3,5,6\},\{2,4,7\}\}$ is an IS partition. We represent below a cycle in IS-deviations that can be reached from the singleton partition or the grand coalition.

$$\begin{tikzcd}
\{1,2\},\{3,4\},\{5,6,7\} \arrow[r, "2"] & \{1\},\{2,3,4\},\{5,6,7\} \arrow[r, "1"] & \{2,3,4\},\{1,5,6,7\} \arrow[d, "2"] \\
\{2\},\{1,3,4\},\{5,6,7\} \arrow[u, "1"] & \{1,3,4\},\{2,5,6,7\} \arrow[l, "2"'] & \{3,4\},\{1,2,5,6,7\} \arrow[l, "1"']
\end{tikzcd}$$

□

We know that it is NP-complete to recognize instances for which an IS partition exists in AHGs, even for strict preferences (Ballester 2004). We prove that both checking the existence of a sequence of IS-deviations ending in an IS partition and checking convergence are hard.

Theorem 3. $\exists$-IS-SEQUENCE-AHG *is* NP*-hard and* $\forall$-IS-SEQUENCE-AHG *is* co-NP*-hard, even for strict preferences.*

However, this hardness result does not hold under strict naturally single-peaked preferences, since we show in the next proposition that *every* sequence of IS-deviations is finite under such a restriction.

Proposition 4. *The dynamics of IS-deviations always converges to an IS partition in AHGs for strict naturally single-peaked preferences.*

Proof. Assume for contradiction that there exists a cycle of IS-deviations. The key idea is to construct an infinite sequence of agents $(a_k)_{k\geq 1}$ that perform deviations from coalitions $(C_k)_{k\geq 1}$, which are strictly increasing in size. Let a_1 be an agent that deviates within this cycle towards a larger coalition by an IS-deviation. This transforms, say, partition π_1 into partition π_1^1. Set $C_1 = \pi_1(a_1)$ and $\hat{C}_1 = \pi_1^1(a_1)$. One can for instance take an agent that performs a deviation from a coalition of minimum size amongst all coalitions from which any deviation is performed. We will now observe how the coalition $\hat{C}_1$ evolves during the cycle. After possibly some agents outside $\hat{C}_1$ joined it or some left it, some agent b originally in $\hat{C}_1$ must deviate from the coalition evolved from $\hat{C}_1$. Otherwise, we cannot reach partition π_1 again in the cycle. If $b \neq a_1$, we assume that the deviation transforms partition π_2 into partition π_2^1 and we set $a_2 = b$, $C_2 = \pi_2(b)$, and $\hat{C}_2 = \pi_2^1(b)$. Note that $|\hat{C}_2| > |C_2| \geq |\hat{C}_1|$, by single-peakedness and the fact that $|\hat{C}_2| \succ_b |C_2| \succ_b |C_2| - 1 \succ_b \cdots \succ_b |\hat{C}_1| \succ_b |\hat{C}_1| - 1$ (where all preferences but the first follow from the assumption of strictness when some other agent joined the coalition of b). In particular, $|C_2| > |C_1|$.

If $b = a_1$, assume that the deviation transforms partition π_1^2 into π_1^3, where possibly $\pi_1^2 = \pi_1^1$. We update $\hat{C}_1 = \pi_1^3(a_1)$. Note that still $|\hat{C}_1| > |C_1|$ by single-peakedness, because the original deviation of a_1 performed in partition π_1 was towards a larger coalition and $|\pi_1^2(a_1)| \succeq_{a_1} |\pi_1^1(a_1)|$ (equality if the partitions are the same). We consider again the next deviation from $\hat{C}_1$ until it is from an agent $b \neq a_1$, in which case we proceed as in the first case. This must eventually happen, because every time the deviation is again performed by agent a_1 she gets closer to her peak. We proceed in the same manner. In step k, we are given a coalition $\hat{C}_k$ with $|\hat{C}_k| > |C_k|$ which was just joined by an agent. When the next agent originally in $\hat{C}_k$ deviates from the coalition evolved from $\hat{C}_k$, it is either an agent different from a_k and we call it a_{k+1}, and find coalitions C_{k+1} and $\hat{C}_{k+1}$ with $|\hat{C}_{k+1}| > |C_{k+1}| \geq |\hat{C}_k|$; or this agent is a_k, she moves towards an updated coalition $\hat{C}_k$ which maintains $|\hat{C}_k| > |C_k|$.

We have thus constructed an infinite sequence of coalitions $(C_k)_{k\geq 1}$ occurring in the cycle with $|C_{k+1}| > |C_k|$ for all $k \geq 1$, a contradiction. □

An interesting open question is whether this convergence result still holds under naturally single-peaked preferences with indifference. However, convergence is also guaranteed under other constrained anonymous games, called *neutral anonymous games*, which are subset-neutral, as defined by Suksompong (2015), thanks to the use of the same potential function argument.

Hedonic Diversity Games (HDGs)

Hedonic diversity games take into account more information about the identity of the agents, changing the focus from coalition sizes to proportions of given types of agents. We obtain more positive results regarding the existence of IS partitions. Indeed, there always exists an IS partition in a hedonic diversity game, even with preferences that are not single-peaked (Boehmer and Elkind 2020). However, we prove that there may exist cycles in IS-deviations, even under some strong restrictions. This stands in contrast to empirical evidence for convergence based on extensive computer simulations by Boehmer and Elkind (2020).

Proposition 5. *The dynamics of IS-deviations may cycle in HDGs even*

1. *when preferences are strict and naturally single-peaked,*
2. *when preferences are strict and the initial partition is the singleton partition or the grand coalition, or*
3. *when preferences are naturally single-peaked and the initial partition is the singleton partition.*

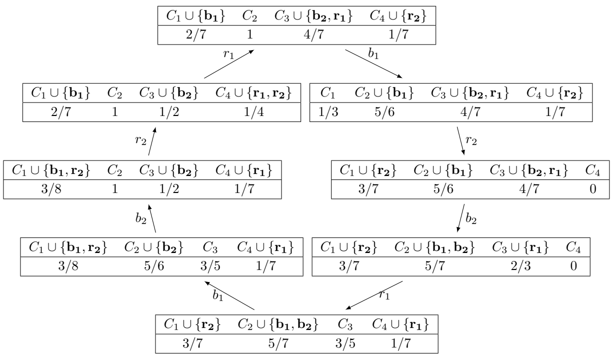
Sketch of proof. We only provide the counterexample for an HDG with strict and naturally single-peaked preferences (restriction 1). Let us consider an HDG with 26 agents: 12 red agents and 14 blue agents. There are four deviating agents: red agents r_1 and r_2 and blue agents b_1 and b_2, and four fixed coalitions C_1, C_2, C_3 and C_4 such that:

- C_1 contains 2 red agents and 4 blue agents;
- C_2 contains 5 red agents;
- C_3 contains 3 red agents and 2 blue agents;
- C_4 contains 6 blue agents.

The relevant part of the preferences is given below.

$$
\begin{array}{llll}
b_1: & \frac{3}{8} \succ \frac{5}{7} \succ \frac{5}{6} \succ \frac{2}{7} & C_1: & \frac{3}{8} \succ \frac{3}{7} \succ \frac{1}{3} \\
b_2: & \frac{5}{7} \succ \frac{4}{7} \succ \frac{1}{2} \succ \frac{5}{6} & C_2: & \frac{5}{7} \succ \frac{5}{6} \succ 1 \\
r_1: & \frac{4}{7} \succ \frac{1}{4} \succ \frac{1}{7} \succ \frac{2}{3} & C_3: & \frac{4}{7} \succ \frac{1}{2} \succ \frac{3}{5} \\
r_2: & \frac{1}{4} \succ \frac{3}{8} \succ \frac{3}{7} \succ \frac{1}{7} & C_4: & \frac{1}{4} \succ \frac{1}{7} \succ 0
\end{array}
$$

Consider the following sequence of IS-deviations that describe a cycle in the dynamics. The four deviating agents of the cycle b_1, b_2, r_1 and r_2 are marked in bold and the specific deviating agent between two states is indicated next to the arrows.



□

This example does not show the impossibility to reach an IS partition since the IS partition $\{C_1 \cup \{b_1, r_2\}, C_2, C_3 \cup \{r_1, b_2\}, C_4\}$ can be reached via IS-deviations from some partitions in the cycle. Thus, starting in these partitions, a path to stability may still exist. Nevertheless, it may be possible that every sequence of IS-deviations cycles, even for strict or naturally single-peaked preferences (with indifference), as the next proposition shows. An interesting open question is whether strict and single-peaked preferences allow for the existence of a path to stability.

Proposition 6. *The dynamics of IS-deviations may never reach an IS partition in HDGs, whatever the chosen path of deviations, even for (1) strict preferences or (2) naturally single-peaked preferences with indifference.*

However, convergence is guaranteed by combining all previous restrictions, as stated in the next proposition.

Proposition 7. *The dynamics of IS-deviations starting from the singleton partition always converges to an IS partition in HDGs for strict naturally single-peaked preferences.*

Sketch of proof. One can easily prove that at any step of the dynamics, a coalition is necessarily of the form $\{r_1, b_1, \ldots, b_k\}$ or $\{b_1, r_1, \ldots, r_{k'}\}$ or $\{b_1\}$ or $\{r_1\}$ where $r_i \in R$ and $b_j \in B$ for every $i \in [k']$, $j \in [k]$ and $k \leq |B|$ and $k' \leq |R|$. Therefore, the ratio of a coalition can only be equal to $\frac{1}{k+1}$, $\frac{k'}{k'+1}$, 0 or 1. Let us define as $\rho(C)$ the modified ratio of a valid coalition C formed by the dynamics where $\rho(C) = \begin{cases} \frac{|R \cap C|}{|C|} & \text{if } C = \{b_1, r_1, \ldots, r_{k'}\} \text{ for } k' \geq 1 \\ 1 - \frac{|R \cap C|}{|C|} & \text{if } C = \{r_1, b_1, \ldots, b_k\} \text{ for } k \geq 2 \\ 0 & \text{otherwise, i.e., } C = \{r_1\} \text{ or } C = \{b_1\} \end{cases}$.
For each partition in a sequence of IS-deviations, we consider the vector composed of the modified ratios $\rho(C)$ for all coalitions C in the partition. One can prove that for each sequence of IS-deviations, either this vector strictly increases lexicographically at each deviation or there is an equivalent sequence of IS-deviations where it does. □

Under strict preferences, checking the existence of a path to stability and convergence are hard.

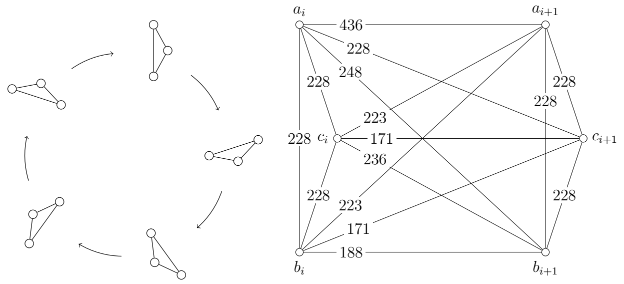
Theorem 8. $\exists$-IS-SEQUENCE-HDG *is* NP*-hard and* $\forall$-IS-SEQUENCE-HDG *is* co-NP*-hard, even for strict preferences.*

Fractional Hedonic Games (FHGs)

Next, we study fractional hedonic games, which are closely related to hedonic diversity games, but instead of agent types, utilities rely on a cardinal valuation function of the other agents. The first part of the section deals with symmetric games, the second part with simple games.

An open problem for symmetric FHGs was whether they always admit an IS partition (Brandl, Brandt, and Strobel 2015). Here, we provide a counterexample using 15 agents.

Theorem 9. *There exists a symmetric FHG without an IS partition.*



(a) Five triangles ordered in a cycle. There is a tendency of agents in N_i to deviate to coalitions in N_{i+1}.

(b) The transition weights between the triangles allow for infinite loops of deviations.

Figure 1: Description of the graph associated with the constructed symmetric FHG without an IS partition.

Sketch of proof. Define the sets of agents $N_i = \{a_i, b_i, c_i\}$ for $i \in \{1, \dots, 5\}$ and consider the FHG on the agent set $N = \bigcup_{i=1}^{5} N_i$ where symmetric weights are given as in Figure 1b. All weights not specified in this figure are set to -2251. The FHG consists of five triangles that form a cycle. Its structure is illustrated in Figure 1a.

There is an infinite cycle of deviations starting with the partition $\{N_5 \cup N_1, N_2, N_3, N_4\}$. First, a_1 deviates by joining N_2. Then, b_1 joins this new coalition, then c_1. After this step, we are in an isomorphic state as in the initial partition. It can be shown that there exists no IS partition in this instance. ☐

Employing this counterexample, the methods of Brandl, Brandt, and Strobel (2015), which originate from hardness constructions of Sung and Dimitrov (2010), can be used to show that it is NP-hard to decide about the existence of IS partitions in symmetric FHGs.

Corollary 10. *Deciding whether there exists an individually stable partition in symmetric FHGs is* NP*-hard.*

If we consider symmetric, non-negative utilities, the grand coalition forms an NS, and therefore IS, partition of the agents. However, deciding about the convergence of the IS dynamics starting with the singleton partition is NP-hard. The reduction is similar to the one in the previous statement and avoids negative weights by the fact that, due to symmetry of the weights, in a dynamics starting with the singleton partition, all coalitions that can be obtained in the process must have strictly positive mutual utility for all pairs of agents in the coalition.

Theorem 11. $\exists$-IS-SEQUENCE-FHG *is* NP*-hard and* $\forall$-IS-SEQUENCE-FHG *is* co-NP*-hard, even in symmetric FHGs with non-negative weights. The former is even true if the initial partition is the singleton partition.*

From now on, we consider simple FHGs. We start with the additional assumption of symmetry.

Proposition 12. *The dynamics of IS-deviations starting from the singleton partition always converges to an IS partition in simple symmetric FHGs in at most $\mathcal{O}(n^2)$ steps. The dynamics may take $\Omega(n\sqrt{n})$ steps.*

Sketch of proof. We only prove the upper bound. Note that all coalitions formed through the deviation dynamics are cliques. Hence, every deviation step will increase the total number of edges in all coalitions. More precisely, the dynamics will increase the potential $\Lambda(\pi) = \sum_{C\in\pi} |C|(|C|-1)/2$ in every step by at least 1. Since the total number of edges is quadratic, this proves the upper bound. ☐

Note that there is a simple way to converge in a linear number of steps starting with the singleton partition by forming largest cliques and removing them from consideration.

If we allow for asymmetries, the dynamics is not guaranteed to converge anymore. For instance, the IS dynamics on an FHG induced by a directed triangle will not converge for any initial partition except for the grand coalition. We can, however, characterize convergence on asymmetric FHGs. Tractability highly depends on the initial partition. First, we assume that we start from the singleton partition.

The key insight is that throughout the dynamic process on an asymmetric FHG starting from the singleton partition, the subgraphs induced by coalitions are always transitive and complete. Convergence is then shown by a potential function argument.

Proposition 13. *The dynamics of IS-deviations starting from the singleton partition converges in asymmetric FHGs if and only if the underlying graph is acyclic. Moreover, under acyclicity, it converges in $\mathcal{O}(n^4)$ steps.*

The previous statement shows convergence of the dynamics for asymmetric, acyclic FHGs. In addition, it is easy to see that there is always a sequence converging after n steps, starting with the singleton partition. One can use a topological order of the agents and allow agents to deviate in decreasing topological order towards a best possible coalition.

There are two interesting further directions. One can weaken either the restriction on the initial partition or on asymmetry. If we allow for general initial partitions, we immediately obtain hardness results that apply in particular to the broader class of simple FHGs.

Theorem 14. $\exists$-IS-SEQUENCE-FHG *is* NP*-hard and* $\forall$-IS-SEQUENCE-FHG *is* co-NP*-hard, even in asymmetric FHGs.*

On the other hand, if we transition to simple FHGs while maintaining the initial partition, the problem of deciding whether a path to stability exists becomes hard.

Theorem 15. $\exists$-IS-SEQUENCE-FHG *is* NP*-hard even in simple FHGs when starting from the singleton partition.*

Dichotomous Hedonic Games (DHGs)

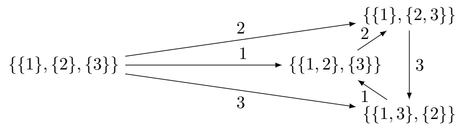
By taking into account the identity of other agents in the preferences of agents over coalitions, it can be more complicated to get positive results regarding individual stability (see, e.g., Theorem 9). However, by restricting the evaluation of coalitions to dichotomous preferences, the existence of an IS partition is guaranteed (Peters 2016), as well as convergence of the dynamics of IS-deviations, when starting from the grand coalition (Boehmer and Elkind 2020).

Nevertheless, the convergence of the dynamics is not guaranteed for an arbitrary initial partition and no sequence of IS-deviations may ever reach an IS partition.

Proposition 16. *The dynamics of IS-deviations may never reach an IS partition in DHGs, whatever the chosen path of deviations, even when starting from the singleton partition.*

Proof. Let us consider an instance of a DHG with three agents where the only coalition approved by agent i is $\{i, i+1\}$ for $i \in \{1, 2\}$ and $\{1, 3\}$ for agent 3.

There is a unique IS partition which consists of the grand coalition $\{1, 2, 3\}$. We represent below all possible IS-deviations between all the other possible partitions. An IS-deviation between two partitions is indicated by an arrow mentioning the name of the deviating agent.



One can check that the described deviations are IS-deviations. A cycle is necessarily reached when starting from a partition different from the unique IS partition, which can be reached only if it is the initial partition. ☐

Moreover, it is hard to decide on the existence of a sequence of IS-deviations ending in an IS partition, even when starting from the singleton partition, as well as checking convergence.

Theorem 17. $\exists$-IS-SEQUENCE-DHG *is* NP*-hard even when starting from the singleton partition, and* $\forall$-IS-SEQUENCE-DHG *is* co-NP*-hard.*

Note that the counterexample provided in the proof of Proposition 16 exhibits a global cycle in the preferences of the agents: $\{1, 2\} \rhd \{1, 3\} \rhd \{2, 3\} \rhd \{1, 2\}$. However, by considering dichotomous preferences with *common ranking property*, that is, each agent has a threshold for acceptance in a given global order, we obtain convergence thanks to the same potential function argument used by Caskurlu and Kizilkaya (2019), for proving the existence of a core-stable partition in hedonic games with common ranking property.

Note that when assuming that if a coalition is approved by one agent, then it must be approved by all the members of the coalition (so-called *symmetric dichotomous preferences*), we obtain a special case of preferences with common ranking property where all the approved coalitions are at the top of the global order. Therefore, convergence is also guaranteed under symmetric dichotomous preferences.

Conclusion

We have investigated dynamics of deviations based on individual stability in hedonic games. The two main questions we considered were whether there exists *some* sequence of deviations terminating in an IS partition, and whether *all* sequences of deviations terminate in an IS partition, i.e., the dynamics converges. Our results are mostly negative with examples of cycles in dynamics or even non-existence of IS partitions under rather strong preference restrictions. In particular, we have answered a number of open problems proposed in the literature. On the other hand, we have identified natural conditions for convergence that are mostly based on preferences relying on a common scale for the agents, like the common ranking property, single-peakedness or symmetry. An overview of our results can be found in Table 1.

		Convergence		Hardness
AHGs	✓	strict & nat. SP (single-peaked) (Prop. 4)	$\exists$	strict (Th. 3)
	○	strict & gen. SP; singletons / grand coalition (Prop. 2)	$\forall$	strict (Th. 3)
HDGs	✓	strict & nat. SP; singletons (Prop. 7)	$\exists$	strict (Th. 8)
	○	[strict & nat. SP] or [strict; singletons/grand coalition] or [nat. SP; singletons] (Prop. 5)	$\forall$	strict (Th. 8)
FHGs	✓	simple & sym.; singletons (Prop. 12)	$\exists$	sym. (Th. 11)
	✓	acyclic digraph (Th. 13)	$\exists$	simple; singletons (Th. 15)
	○	sym. (Th. 9)	$\exists$	asym. (Th. 14)
			$\forall$	sym. (Th. 11)
			$\forall$	asym. (Th. 14)
DHGs	✓	grand coalition (Boehmer and Elkind 2020)	$\exists$	singletons (Th. 17)
	○	singletons (Prop. 16)	$\forall$	general (Th. 17)

Table 1: Convergence and hardness results for the dynamics of IS-deviations in various classes of hedonic games. Symbol ✓ marks convergence under the given preference restrictions and initial partition (if applicable) while ○ marks non-convergence, i.e., cycling dynamics. Symbol $\exists$ (resp., $\forall$) denotes that problem $\exists$-IS-SEQUENCE-HG (resp., $\forall$-IS-SEQUENCE-HG) is NP-hard (resp., co-NP-hard).

For all hedonic games under study, it turned out that the existence of cycles for IS-deviations is sufficient to prove the hardness of recognizing instances for which there exists a finite sequence of deviations or whether all sequences of deviations are finite, i.e., the dynamics converges. While our results cover a broad range of hedonic games considered in the literature, there are still promising directions for further research. First, even though our hardness results hold under strong restrictions, the complexity of these questions remains open for interesting preference restrictions, some of which do not guarantee convergence. Following our work, the most intriguing cases are AHGs under single-peaked weak preferences, simple symmetric FHGs with arbitrary initial partitions, and HDGs under single-peaked preferences. Secondly, one could investigate more specific rules of IS-deviations that quickly terminate in IS partitions. For instance, for simple symmetric FHGs, there is the possibility of very fast convergence, but the selection of the deviating agents in this approach requires to solve a maximum clique problem (cf. the discussion after Proposition 12). Finally, the dynamics we consider only concerns individual stability. One could also aim at reaching outcomes that satisfy Pareto optimality or other desirable properties in addition.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/12-1. We thank the anonymous reviewers for helpful comments.

References

Abeledo, H.; and Rothblum, U. G. 1995. Paths to marriage stability. *Discrete Applied Mathematics* 63(1): 1–12.

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

Aziz, H.; Harrenstein, P.; Lang, J.; and Wooldridge, M. 2016. Boolean Hedonic Games. In *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 166–175.

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

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

Boehmer, N.; and Elkind, E. 2020. Individual-Based Stability in Hedonic Diversity Games. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, 1822–1829.

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

Brandl, F.; Brandt, F.; and Strobel, M. 2015. Fractional Hedonic Games: Individual and Group Stability. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 1219–1227.

Brandt, F.; and Wilczynski, A. 2019. On the Convergence of Swap Dynamics to Pareto-Optimal Matchings. In *Proceedings of the 15th International Conference on Web and Internet Economics (WINE)*, 100–113.

Bredereck, R.; Elkind, E.; and Igarashi, A. 2019. Hedonic Diversity Games. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 565–573.

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

Caskurlu, B.; and Kizilkaya, F. E. 2019. On Hedonic Games with Common Ranking Property. In *Proceedings of the 11th International Conference on Algorithms and Complexity (CIAC)*, 137–148. Springer International Publishing.

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

Farrell, J.; and Scotchmer, S. 1988. Partnerships. *Quarterly Journal of Economics* 103: 279–297.

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

Peters, D. 2016. Complexity of Hedonic Games with Dichotomous Preferences. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 579–585.

Roth, A. E.; and Vande Vate, J. H. 1990. Random Paths to Stability in Two-Sided Matching. *Econometrica* 58(6): 1475–1480.

Suksompong, W. 2015. Individual and Group Stability in Neutral Restrictions of Hedonic Games. *Mathematical Social Sciences* 78: 1–5.

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

10 CORE PUBLICATION 5: SINGLE-AGENT DYNAMICS IN ADDITIVELY SEPARABLE HEDONIC GAMES

SUMMARY

The formation of stable coalitions is a central goal in multi-agent systems. A considerable stream of research defines stability via the absence of beneficial deviations by single agents. The simplest type of such a deviation is a Nash deviation, where it is required that an agent improves their utility by performing the deviation. However, requiring that a coalition structure allows for no Nash deviation is a strong demand. Therefore, further restrictions for conducting deviations have been imposed in the literature, which are based on the consent of the agents in the abandoned as well as the welcoming coalition. If all agents in the abandoned (or welcoming) coalition have to approve a deviation, we speak of a contractual (or individual) deviation.

While Nash deviations are not constrained by agents other than the deviating agent, contractual and individual deviations require the unanimous consent of agents. We strive for a compromise between these two extremes, and also study consent decided by majority votes. In particular, we introduce two new stability notions that can be seen as local variants of popularity.

We investigate all resulting stability notions in additively separable hedonic games. There, we pinpoint boundaries to computational complexity depending on the type of consent and restrictions on the utility functions. The latter restrictions shed new light on well-studied classes of games based on the appreciation of friends or the aversion to enemies. While majority-based stability leads to intractabilities in general additively separable hedonic games, we find more positive results in the obtained restricted classes of games. In particular, involving both the abandoned and welcoming coalition in the majority vote leads to tractability.

Many of our positive results follow from the *Deviation Lemma*, a general combinatorial observation, which can be leveraged to prove the convergence of simple and natural single-agent dynamics under fairly general conditions.

REFERENCE

F. Brandt, M. Bullinger, and L. Tappe. **Single-agent dynamics in additively separable hedonic games**. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, pages 4867–4874, 2022.
DOI: https://doi.org/10.1609/aaai.v36i5.20415

INDIVIDUAL CONTRIBUTION

I, Martin Bullinger, am the main author of this publication. In particular, I am responsible for the development and conceptual design of the research project, proofs of some results (Proposition 1 and 2), advice and corrective changes to all other results, and the write-up of the manuscript.

COPYRIGHT AGREEMENT

The right to present this paper in a doctoral thesis has been granted by the publisher, the Association for the Advancement of Artificial Intelligence (AAAI), in the copyright form presented below. There, AAAI grants permission to "*Personal reuse of all or portions of the above article/paper in other works of their own authorship*." This form can also be found at
https://proceedings.aaai.org/docs/AAAI-22-copyrightform.pdf
(accessed November 10, 2022).

Association for the Advancement of Artificial Intelligence

1900 Embarcadero Road, Suite 101, Palo Alto, California 94303
Palo Alto, California 94303 USA

AAAI COPYRIGHT FORM

Title of Article/Paper: ______________________________

Publication in Which Article/Paper Is to Appear: ______________________________

Author's Name(s): ______________________________

Please type or print your name(s) as you wish it (them) to appear in print

PART A – COPYRIGHT TRANSFER FORM

The undersigned, desiring to publish the above article/paper in a publication of the Association for the Advancement of Artificial Intelligence, (AAAI), hereby transfer their copyrights in the above article/paper to the Association for the Advancement of Artificial Intelligence (AAAI), in order to deal with future requests for reprints, translations, anthologies, reproductions, excerpts, and other publications.

This grant will include, without limitation, the entire copyright in the article/paper in all countries of the world, including all renewals, extensions, and reversions thereof, whether such rights current exist or hereafter come into effect, and also the exclusive right to create electronic versions of the article/paper, to the extent that such right is not subsumed under copyright.

The undersigned warrants that they are the sole author and owner of the copyright in the above article/paper, except for those portions shown to be in quotations; that the article/paper is original throughout; and that the undersigned right to make the grants set forth above is complete and unencumbered.

If anyone brings any claim or action alleging facts that, if true, constitute a breach of any of the foregoing warranties, the undersigned will hold harmless and indemnify AAAI, their grantees, their licensees, and their distributors against any liability, whether under judgment, decree, or compromise, and any legal fees and expenses arising out of that claim or actions, and the undersigned will cooperate fully in any defense AAAI may make to such claim or action. Moreover, the undersigned agrees to cooperate in any claim or other action seeking to protect or enforce any right the undersigned has granted to AAAI in the article/paper. If any such claim or action fails because of facts that constitute a breach of any of the foregoing warranties, the undersigned agrees to reimburse whomever brings such claim or action for expenses and attorneys' fees incurred therein.

Returned Rights

In return for these rights, AAAI hereby grants to the above author(s), and the employer(s) for whom the work was performed, royalty-free permission to:

1. Retain all proprietary rights other than copyright (such as patent rights).
2. Personal reuse of all or portions of the above article/paper in other works of their own authorship. This does not include granting third-party requests for reprinting, republishing, or other types of reuse. AAAI must handle all such third-party requests.
3. Reproduce, or have reproduced, the above article/paper for the author's personal use, or for company use provided that AAAI copyright and the source are indicated, and that the copies are not used in a way that implies AAAI endorsement of a product or service of an employer, and that the copies per se are not offered for sale. The foregoing right shall not permit the posting of the article/paper in electronic or digital form on any computer network, except by the author or the author's employer, and then only on the author's or the employer's own web page or ftp site. Such web page or ftp site, in addition to the aforementioned requirements of this Paragraph, shall not post other AAAI copyrighted materials not of the author's or the employer's creation (including tables of contents with links to other papers) without AAAI's written permission.
4. Make limited distribution of all or portions of the above article/paper prior to publication.
5. In the case of work performed under a U.S. Government contract or grant, AAAI recognized that the U.S. Government has royalty-free permission to reproduce all or portions of the above Work, and to authorize others to do so, for official U.S. Government purposes only, if the contract or grant so requires.

In the event the above article/paper is not accepted and published by AAAI, or is withdrawn by the author(s) before acceptance by AAAI, this agreement becomes null and void.

(1)

______________________________ ______________________________

Author/Authorized Agent for Joint Author's Signature Date (MM/DD/YYYY)

______________________________ ______________________________

Employer for whom work was performed Title (if not author)

(For jointly authored Works, all joint authors should sign unless one of the authors has been duly authorized to act as agent for the others.)

Single-Agent Dynamics in Additively Separable Hedonic Games

Felix Brandt, Martin Bullinger, and Leo Tappe

Institut für Informatik, Technische Universität München
brandtf@in.tum.de, bullinge@in.tum.de, leo.tappe@tum.de

Abstract

The formation of stable coalitions is a central concern in multiagent systems. A considerable stream of research defines stability via the absence of beneficial deviations by single agents. Such deviations require an agent to improve her utility by joining another coalition while possibly imposing further restrictions on the consent of the agents in the welcoming as well as the abandoned coalition. While most of the literature focuses on unanimous consent, we also study consent decided by majority vote, and introduce two new stability notions that can be seen as local variants of popularity. We investigate these notions in additively separable hedonic games by pinpointing boundaries to computational complexity depending on the type of consent and restrictions on the utility functions. The latter restrictions shed new light on well-studied classes of games based on the appreciation of friends or the aversion to enemies. Many of our positive results follow from the *Deviation Lemma*, a general combinatorial observation, which can be leveraged to prove the convergence of simple and natural single-agent dynamics under fairly general conditions.

Introduction

Coalition formation is a central concern in multi-agent systems and considers the question of grouping a set of agents, e.g., humans or machines, into coalitions such as teams, clubs, or societies. A prominent framework for studying coalition formation is that of hedonic games, where agents' utilities are solely based on the coalition they are part of, and which thus disregards inter-coalitional relationships (Drèze and Greenberg 1980). Hedonic games have been successfully used to model various scenarios evolving from operations research or the mathematical social sciences such as research team formation (Alcalde and Revilla 2004), task allocation (Saad et al. 2011), or community detection (Newman 2004; Aziz et al. 2019). Identifying desirable coalition structures is often based on the prospect of coalitions to stay together. To this end, various notions of stability have been introduced and studied. A coalition structure (henceforth partition) is stable when no individual or group of agents benefits by joining another coalition or by forming a new coalition.

In this paper, we focus on deviations by single agents. The simplest example is a Nash deviation where some agent unilaterally decides to leave her current coalition in order to join another coalition. While such a deviation clearly captures the incentive of single agents to perform deviations, it completely ignores the other agents' opinions about the deviation. To overcome this shortcoming, various restrictions of Nash deviations have been proposed. This has motivated stability notions such as individual stability or contractual Nash stability, which consider the unanimous consent of some or all of the coalitions directly affected by the deviation. While unanimous consent is in fact used in the formation process of international bodies like the EU or the NATO, it might be impractical and even undesirable in small- or medium-scale coalition formation scenarios. As a compromise, we also study intermediate notions of stability based on majority votes among the involved coalitions. This setting has received little attention so far (Gairing and Savani 2019), and we will also define some new majority-based stability notions.

The study of hedonic games was initiated by Drèze and Greenberg (1980), and later popularized by Banerjee, Konishi, and Sönmez (2001), Cechlárová and Romero-Medina (2001), and Bogomolnaia and Jackson (2002). Since then, a large body of research has been devoted to defining suitable game representations and solution concepts. An overview of many important aspects is provided in the survey by Aziz and Savani (2016). One prominent, natural, and arguably simple type of hedonic games is given by additively separable hedonic games (Bogomolnaia and Jackson 2002). In these games, agents entertain cardinal utilities for other agents and the utility for a coalition is defined by taking the sum of the individual utility values. This game representation allows, for instance, the modeling of settings where agents have friends and enemies, and their goal is to simultaneously maximize the number of friends and minimize the number of enemies, while one of these two goals can have higher priority than the other one (Dimitrov et al. 2006). Our work provides exact boundaries for the computational tractability of stability concepts based on single-agent deviations in additively separable hedonic games, showing a clear cut between Nash stability and stability notions under consent. We give simple and precise conditions for restricted classes of utility functions that pinpoint the boundaries of computational tractability. This includes well-studied classes of games where agents only distinguish between friends and enemies.

A more recent line of research on stability notions focuses on the dynamical aspects leading to the formation of stable outcomes (e.g., Bilò et al. 2018; Hoefer, Vaz, and Wagner 2018; Carosi, Monaco, and Moscardelli 2019; Brandt, Bullinger, and Wilczynski 2021). This yields an important distributed perspective on the coalition formation process. The value of some positive computational results in the context of hedonic games is diminished by the fact that they implicitly assume that a central authority has the means to collect all individual preferences, compute a stable partition, and enforce this partition on the agents. In contrast, simple dynamics based on single-agent deviations provide a much more plausible explanation for the formation of stable partitions. A versatile tool to prove the convergence of dynamics are potential functions, which guide the dynamics towards stable states (e.g., Bogomolnaia and Jackson 2002; Suksompong 2015). We extend the applicability of this approach by considering *non-monotonic* potential functions, i.e., potential functions that might decrease in some rounds of the dynamic process. This is possible because the total number of rounds can be bounded by observing the potential function from a global perspective using a new general combinatorial insight that we call the Deviation Lemma. The Deviation Lemma is not restricted to additively separable utilities or the specific type of single-agent deviations. For instance, the combinatorial relationship of the lemma also arises naturally in the analysis of deviation dynamics in classes of games beyond the scope of this paper, such as anonymous hedonic games (Bogomolnaia and Jackson 2002). In fact, the lemma holds for every sequence of partitions such that each partition evolves from its predecessor by having one element move to another partition class. It establishes a relationship between the development of the sizes of coalitions involved in deviations to information solely based on the starting partition and the terminal partition of the sequence.

For the special case of symmetric utility functions, additively separable hedonic games are well understood: the standard notion of utilitarian social welfare represents an increasing potential function for the dynamics induced by Nash stability (Bogomolnaia and Jackson 2002), but finding stable states (even under unanimous consent of the welcoming coalition) leads to PLS-complete problems (Gairing and Savani 2019). As we will see, this implies worst-case exponential running time of the dynamics. By contrast, our results hold for restricted sets of non-symmetric utility functions and our computational boundaries lie between polynomial-time computability and NP-completeness. In fact, whenever we identify a potential function guaranteeing the existence of stable outcomes, we are also able to prove that, from any starting partition, the corresponding simple dynamics of single-agent deviations converges to a stable partition in a polynomial number of rounds.

Preliminaries and Model

In this section we introduce hedonic games and our stability concepts. We use the notation $[k] = \{1, \dots, k\}$ for any positive integer k.

Hedonic Games

Throughout the paper, we consider settings with a set $N = [n]$ of n agents. The goal of coalition formation is to find a partition of the agents into different disjoint coalitions according to their preferences. A *partition* of N is a subset $\pi \subseteq 2^N$ such that $\bigcup_{C\in\pi} C = N$, and for every pair $C, D \in \pi$, it holds that $C = D$ or $C \cap D = \emptyset$. An element of a partition is called *coalition*, and given a partition π, we denote by $\pi(i)$ the coalition containing agent i. We refer to the partition π given by $\pi(i) = \{i\}$ for every agent $i \in N$ as the *singleton partition*, and to $\pi = \{N\}$ as the *grand coalition*.

Let $\mathcal{N}_i$ denote all possible coalitions containing agent i, i.e., $\mathcal{N}_i = \{C \subseteq N : i \in C\}$. A *hedonic game* is defined by a tuple $(N, \succsim)$, where N is an agent set and $\succsim = (\succsim_i)_{i\in N}$ is a tuple of weak orders $\succsim_i$ over $\mathcal{N}_i$ which represent the preferences of the respective agent i. Hence, agents express preferences only over the coalitions which they are part of without considering externalities. The generality of the definition of hedonic games gives rise to many interesting subclasses of games that have been proposed in the literature. Many of these classes rely on cardinal utility functions $v_i\colon N \to \mathbb{R}$ for every agent i, which are aggregated in various ways (Aziz et al. 2019; Bogomolnaia and Jackson 2002; Olsen 2012). One particularly natural and prominent such model considers aggregation by taking the sum of individual utilities. Formally, following Bogomolnaia and Jackson (2002), an *additively separable hedonic game* (ASHG) (N, v) consists of an agent set N and a tuple $v = (v_i)_{i\in N}$ of utility functions $v_i\colon N \to \mathbb{R}$ such that $\pi(i) \succsim_i \pi'(i)$ iff $\sum_{j\in\pi(i)} v_i(j) \geq \sum_{j\in\pi'(i)} v_i(j)$. Clearly, ASHGs are a subclass of hedonic games, and we can assume without loss of generality that $v_i(i) = 0$ (or set the utility of an agent for herself to an arbitrary constant). ASHGs have a natural representation by a complete directed graph $G = (N, E)$ with weight $v_i(j)$ on arc (i, j). An ASHG is called *symmetric* if $v_i(j) = v_j(i)$ for every pair of agents i and j, and it can then be represented by a complete undirected graph with weight $v_i(j)$ on edge $\{i, j\}$. There are various classes of ASHGs with certain restrictions for the utility functions that allow a natural interpretation in terms of friends and enemies. An agent j is called *friend* (respectively, *enemy*) of agent i if $v_i(j) > 0$ (respectively, $v_i(j) < 0$). An ASHG is called *friends-and-enemies game* (FEG) if $v_i(j) \in \{-1, 1\}$ for every pair of agents i and j. Further, following Dimitrov et al. (2006), an ASHG is called an *appreciation of friends game* (AFG) (respectively, an *aversion to enemies game* (AEG)) if $v_i(j) \in \{-1, n\}$ (respectively, $v_i(j) \in \{-n, 1\}$). In all of these games, agents pursue the objective to maximize their number of friends while minimizing their number of enemies. In the case of an FEG, these two goals have equal priority, while there is a strict priority for one of the goals in AFGs and AEGs.

Stability Based on Single-Agent Deviations

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

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

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

We consider the case of myopic rational agents who only engage in a deviation if it immediately makes them better off. Formally, a *Nash deviation* is a single-agent deviation performed by agent i making agent i better off, i.e., $\pi'(i) \succ_i \pi(i)$. Any partition in which no Nash deviation is possible is called *Nash stable* (NS).

This concept of stability is very strong and comes with the drawback that only the preferences of the deviating agent are considered. Therefore, various refinements have been proposed which additionally require the consent of the abandoned and the welcoming coalition. For a compact representation, we introduce them via the notion of favor sets.

Let $C \subseteq N$ be a coalition and $i \in N$ be an agent. The *favor-in set* of C with respect to i is the set of agents in C (excluding i) that strictly favor having i inside of C than outside, i.e., $F_{\text{in}}(C,i) = \{j \in C \setminus \{i\} \colon C \cup \{i\} \succ_j C \setminus \{i\}\}$. Similarly, the *favor-out set* of C with respect to i is the set of agents in C (excluding i) that strictly favor having i outside of C than inside, i.e., $F_{\text{out}}(C,i) = \{j \in C \setminus \{i\} \colon C \setminus \{i\} \succ_j C \cup \{i\}\}$.

Following Bogomolnaia and Jackson (2002) and Dimitrov and Sung (2007), an *individual deviation* (respectively, *contractual deviation*) is a Nash deviation $\pi \xrightarrow{i} \pi'$ such that $F_{\text{out}}(\pi'(i), i) = \emptyset$ (respectively, $F_{\text{in}}(\pi(i), i) = \emptyset$). A single-agent deviation that is both an individual and a contractual deviation is called *contractual individual deviation*. All of these deviation concepts give rise to a respective stability concept. A partition is called *individually stable* (IS), *contractually Nash stable* (CNS), or *contractually individually stable* (CIS) if it allows for no individual, contractual, or contractual individual deviations, respectively.

While these stability concepts include agents affected by the deviation, they require unanimous consent, which might be unnecessarily strong in some settings. Based on this observation, we define several hybrid stability concepts where the possibility of a deviation by some agent is decided via majority votes of the involved agents.

A Nash deviation $\pi \xrightarrow{i} \pi'$ is called *majority-in deviation* (respectively, *majority-out deviation*) if $|F_{\text{in}}(\pi'(i), i)| \geq |F_{\text{out}}(\pi'(i), i)|$ (respectively, $|F_{\text{out}}(\pi(i), i)| \geq |F_{\text{in}}(\pi(i), i)|$). A single-agent deviation that is both a majority-in deviation and a majority-out deviation is called *separate-majorities deviation*. As before, a partition is called *majority-in stable* (MIS), *majority-out stable* (MOS), or *separate-majorities stable* (SMS) if it allows for no majority-in, majority-out, or separate-majorities deviations, respectively. The concepts MIS and MOS are a special case of voting-based stability notions by Gairing and Savani (2019) for a threshold of $1/2$.

Finally, it is possible to relax SMS by performing one joint vote instead of two separate votes. A Nash deviation $\pi \xrightarrow{i} \pi'$ is called a *joint-majority deviation* if $|F_{\text{out}}(\pi(i), i)| + |F_{\text{in}}(\pi'(i), i)| \geq |F_{\text{in}}(\pi(i), i)| + |F_{\text{out}}(\pi'(i), i)|$. A partition is then called *joint-majority stable* (JMS) if it allows for no joint-majority deviations. JMS is particularly interesting as it is a natural local version of popularity (Pop), an axiom recently studied in the context of hedonic games (Gärdenfors 1975; Cseh 2017; Brandt and Bullinger 2020).[1]

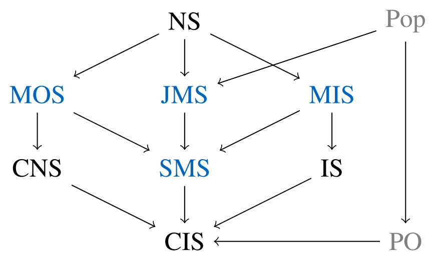


Figure 1: Logical relationships between stability notions and other solutions concepts. An arrow from concept α to concept β indicates that if a partition satisfies α, it also satisfies β. Majority-based stability notions are highlighted in blue, other single-agent based stability notions in black.

Also note that while CIS is a refinement of Pareto optimality (PO), there is no logical relationship between other (majority-based) stability concepts and PO. In particular, we denote the stability concepts based on single-agent deviations by $\mathcal{S}$, i.e., $\mathcal{S} = \{\text{NS}, \text{IS}, \text{CNS}, \text{CIS}, \text{MIS}, \text{MOS}, \text{SMS}, \text{JMS}\}$. A taxonomy of our related solution concepts is provided in Figure 1. For a more concise notation, we refer to deviations with respect to stability concept $\alpha \in \mathcal{S}$ as α-*deviations*, e.g., IS-deviations for $\alpha = \text{IS}$.

All these stability concepts naturally induce dynamics where we choose some starting partition and obtain a successor partition by having some agent perform a deviation from the current partition. More precisely, given a stability concept $\alpha \in \mathcal{S}$, an execution of α-*dynamics* is an infinite or finite sequence $(\pi_j)_{j \geq 0}$ of partitions and a corresponding sequence $(i_j)_{j \geq 1}$ of (deviating) agents such that $\pi_{j-1} \xrightarrow{i_j} \pi_j$ is an α-deviation for every j. The partition π_0 is then called the *starting partition*. Given a hedonic game G, and a stability concept $\alpha \in \mathcal{S}$, we say that the dynamics *converges for starting partition* π_0 if every execution of the α-dynamics on G with starting partition π_0 is finite. Additionally, the α-dynamics *converges* on G if it converges for every starting partition.

Proving convergence of dynamics is a very natural way to prove the existence of stable states and underlines the robustness of the stability concept. It complements a static solution concept with a decentralized process to reach a solution.

Results

In this section, we present our results.

Computational Boundaries for Nash Stability

First, we consider the notion of Nash stability. In the absence of negative utility values, the partition consisting solely of the grand coalition is Nash stable. Conversely, in the absence of

[1] Informally speaking, a partition is popular if there is no other partition preferred by a majority of the agents. JMS partitions can only be challenged by partitions evolving through Nash deviations.

positive utility values, the singleton partition is Nash stable. It is therefore necessary for an ASHG to have both positive and negative utility values in order to admit a non-trivial Nash stable partition (see also Gairing and Savani 2019).

Sung and Dimitrov (2010) showed that deciding whether an ASHG has an NS partition is NP-hard by a reduction from EXACT3COVER. This reduction produces an ASHG with four distinct positive utility values and one negative utility value. We improve upon this result by showing that a reduction is possible with only one positive and one negative utility value. Moreover, it is possible for *any* choice of these two utility values, as long as the absolute value of the negative utility value is at least as large as the positive utility value. We state the theorem in a general way allowing the positive and negative utility value to be dependent on the number of agents of the particular instance. In this way, we simultaneously cover several important cases. For instance, the hardness holds for fixed constant positive and negative utility values as in FEGs, or for AFGs and AEGs. Note that for all of our stability notions, a stable partition is a polynomial-time verifiable certificate: one can simply check whether any agent can perform a deviation, and if no one can, the partition is stable. Therefore, we omit the proof of membership in NP in all of our reductions. The proof of the next and some subsequent results are omitted due to space restrictions.

Theorem 1. *Let $f^+ : \mathbb{N} \to \mathbb{Q}_{>0}$ and $f^- : \mathbb{N} \to \mathbb{Q}_{<0}$ be two polynomial-time computable functions satisfying $|f^-(m)| \geq f^+(m)$ for all $m \in \mathbb{N}$. Then, the problem of deciding whether an ASHG with utility values restricted to $\{f^-(n), f^+(n)\}$ has an NS partition is* NP*-complete.*

Theorem 1 requires the negative utility value to be at least as large in absolute value as the positive utility value. While we leave open the computational complexity for completely arbitrary pairs of negative and positive values, we can show that the problem is also hard when the positive utility value is significantly larger than the absolute value of the negative utility value. The reduction is a variant of the reduction in Theorem 1.

Theorem 2. *Deciding whether an AFG has an NS partition is* NP*-complete.*

Deviation Lemma and Applications

By contrast, restricting the utility values to one positive and one negative value leads to positive results for other notions of stability. These results can be shown in a unified manner using a potential function argument that crucially hinges on the following general observation.

Lemma 1 (Deviation Lemma)**.** *Let $\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k$ be a sequence of k single-agent deviations. Then, the following identity holds:*

$$\sum_{j\in[k]} |\pi_j(i_j)| - |\pi_{j-1}(i_j)| = \frac{1}{2}\sum_{i\in N}|\pi_k(i)| - |\pi_0(i)|. \quad (1)$$

Proof. Let $\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k$ be a sequence of k single-agent deviations and fix some $j \in [k]$. Then, the following facts hold:

$$|\pi_j(i_j)| = \left(\sum_{i\in\pi_j(i_j)\setminus\{i_j\}} |\pi_j(i)| - |\pi_{j-1}(i)|\right) + 1,$$
$$|\pi_{j-1}(i_j)| = \left(\sum_{i\in\pi_{j-1}(i_j)\setminus\{i_j\}} |\pi_{j-1}(i)| - |\pi_j(i)|\right) + 1,$$
$$\pi_j(i) = \pi_{j-1}(i) \quad \forall i \in N \setminus (\pi_j(i_j) \cup \pi_{j-1}(i_j)).$$

Combining these facts allows us to express the difference of the deviator's coalition sizes as follows:

$$\begin{aligned}|\pi_j(i_j)| - |\pi_{j-1}(i_j)| &= \left(\sum_{i\in\pi_j(i_j)\setminus\{i_j\}} |\pi_j(i)| - |\pi_{j-1}(i)|\right)\\ &- \left(\sum_{i\in\pi_{j-1}(i_j)\setminus\{i_j\}} |\pi_{j-1}(i)| - |\pi_j(i)|\right)\\ &+ \sum_{i\in N\setminus(\pi_j(i_j)\cup\pi_{j-1}(i_j))} |\pi_j(i)| - |\pi_{j-1}(i)|\\ &= \sum_{i\in N\setminus\{i_j\}} |\pi_j(i)| - |\pi_{j-1}(i)|.\end{aligned}$$

Adding $|\pi_j(i_j)| - |\pi_{j-1}(i_j)|$ to both sides yields

$$2\left(|\pi_j(i_j)| - |\pi_{j-1}(i_j)|\right) = \sum_{i\in N}|\pi_j(i)| - |\pi_{j-1}(i)|.$$

Summing these terms for all $j \in [k]$, interchanging summation order, and telescoping gives

$$\begin{aligned}\sum_{j\in[k]} 2\left(|\pi_j(i_j)| - |\pi_{j-1}(i_j)|\right) &= \sum_{j\in[k]}\sum_{i\in N}|\pi_j(i)| - |\pi_{j-1}(i)|\\ 2\sum_{j\in[k]}|\pi_j(i_j)| - |\pi_{j-1}(i_j)| &= \sum_{i\in N}\sum_{j\in[k]}|\pi_j(i)| - |\pi_{j-1}(i)|\\ 2\sum_{j\in[k]}|\pi_j(i_j)| - |\pi_{j-1}(i_j)| &= \sum_{i\in N}|\pi_k(i)| - |\pi_0(i)|.\end{aligned}$$

Dividing both sides by 2 completes the proof. □

The Deviation Lemma is especially useful as the right-hand side of Equation (1) does not depend on k, and we can therefore also find bounds for its left-hand side solely depending on the number of players n.

Lemma 2. *Consider a sequence of k successive single-agent deviations*

$$\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k.$$

Then, the following bounds hold:

$$-\frac{n(n-1)}{2} \leq \sum_{j\in[k]}|\pi_j(i_j)| - |\pi_{j-1}(i_j)| \leq \frac{n(n-1)}{2}.$$

Proof. Observe that for all $i \in N$ and all partitions π, we have

$$1 \leq |\pi(i)| \leq n.$$

Thus, we can find the bounds

$$-n(n-1) \leq \sum_{i \in N} |\pi_k(i)| - |\pi_0(i)| \leq n(n-1).$$

Applying Lemma 1 yields the desired result. $\square$

We demonstrate the power of the Deviation Lemma by proving convergence of the dynamics for a variety of deviation types and classes of ASHGs.

Theorem 3. *The dynamics of IS-deviations always converges in ASHGs with at most one nonnegative utility value.*

Proof. Let (N, v) be an ASHG such that the v_i take on at most one nonnegative value. If there are no nonnegative valuations, all IS-deviations are singleton formations, so after at most n deviations, we reach a stable partition. Now, suppose that there is exactly one nonnegative utility value $x \geq 0$. If there are no negative valuations, then in case $x = 0$ we terminate immediately, and in case $x > 0$ the grand coalition will form after at most n^2 deviations. The latter holds because every deviation increases the number of pairs of agents which are part of the same coalition. Thus, we will now assume that in addition to the single nonnegative utility value x, there is at least one negative utility value, and we denote the largest absolute value of a negative utility value by y. Further, define $\Delta = \min\{v_i(C) - v_i(C') \colon i \in N, C, C' \in \mathcal{N}_i, v_i(C) > v_i(C')\}$. Intuitively, $\Delta > 0$ is the minimum improvement any agent is guaranteed to have when making a NS-deviation. Further, consider the potential function Φ defined by the social welfare of a partition as $\Phi(\pi) = \sum_{i \in N} v_i(\pi)$.

Let us investigate how this potential changes for a single IS-deviation $\pi \xrightarrow{i} \pi'$.

$$\begin{aligned}
&\Phi(\pi') - \Phi(\pi) = \underbrace{v_i(\pi') - v_i(\pi)}_{\text{deviator}} \\
&+ \underbrace{\sum_{j \in \pi'(i) \setminus \{i\}} v_j(\pi') - v_j(\pi)}_{\text{welcoming coalition}} + \underbrace{\sum_{j \in \pi(i) \setminus \{i\}} v_j(\pi') - v_j(\pi)}_{\text{abandoned coalition}} \\
&= v_i(\pi') - v_i(\pi) + \sum_{j \in \pi'(i) \setminus \{i\}} v_j(i) - \sum_{j \in \pi(i) \setminus \{i\}} v_j(i) \\
&= v_i(\pi') - v_i(\pi) + x\left(|\pi'(i)| - 1\right) - \sum_{j \in \pi(i) \setminus \{i\}} v_j(i) \\
&\geq \Delta + x\left(|\pi'(i)| - 1\right) - x\left(|\pi(i)| - 1\right) \\
&= \Delta + x\left(|\pi'(i)| - |\pi(i)|\right).
\end{aligned}$$

The third equality comes from the fact that i performs an IS-deviation, so all agents $j \in \pi'(i) \setminus \{i\}$ must accept i, which means they must have $v_j(i) = x$. Now, let π_0 be any initial partition and consider any sequence of k successive IS-deviations

$$\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k.$$

Telescoping and termwise application of the above inequality yields $\Phi(\pi_k) - \Phi(\pi_0) = \sum_{j \in [k]} \Phi(\pi_j) - \Phi(\pi_{j-1}) \geq \sum_{j \in [k]} \Delta + x\left(|\pi_j(i_j)| - |\pi_{j-1}(i_j)|\right) = k\Delta + x \sum_{j \in [k]} |\pi_j(i_j)| - |\pi_{j-1}(i_j)|$. We recognize the sum from the Deviation Lemma, which can be bounded from below using Lemma 2:

$$\Phi(\pi_k) - \Phi(\pi_0) \geq k\Delta - x\frac{n(n-1)}{2}. \tag{2}$$

As the right hand side is unbounded in k, the sequence must be finite. To be precise, we can bound the potentials of the initial and final partitions by

$$\Phi(\pi_0) \geq -n(n-1)y, \quad \Phi(\pi_k) \leq n(n-1)x.$$

Substituting in these bounds and rearranging for k gives

$$k \leq \frac{(2y + 3x)n(n-1)}{2\Delta}. \tag{3}$$

$\square$

There are a few important insights gained by the previous proof. First, the bound obtained via the Deviation Lemma does not mean that the potential function Φ is increasing in every round. In fact, since utilities are not necessarily symmetric, the deviating agent might move from a rather large coalition to a smaller coalition only improving her utility by Δ whereas the utility of all agents in the abandoned coalition are decreased by x. In fact, the Deviation Lemma does not give us control of the potential function in a single round. Also, it does not control the utility changes caused by the deviator. We apply it to control the utility changes of agents involved in deviations except for the deviator to obtain Equation (2). Hence, we can bound their utility changes by a global constant solely depending on input data. The utility changes caused by the deviator will then eventually lead to the potential reaching a local maximum.

Second, we can easily obtain polynomial bounds on the running time of the dynamics. If x and y are polynomially bounded in n and all valuations are integer, polynomial running time is directly obtained from Equation (3). In particular, this is the case for FEGs, AFGs, and AEGs, so individually stable partitions can be found in polynomial time for these games. After showing two more applications of the Deviation Lemma for other types of deviations, we will capture this observation in Corollary 1.

Third, the previous theorem is tight in the sense that the dynamics can cycle if we have two nonnegative utility values. Indeed, in an instance with agent set $N = [3]$ and utility values $v_i(j) = 1, v_j(i) = 0$ for $(i,j) \in \{(1,2),(2,3),(3,1)\}$, the dynamics can infinitely cycle among the partitions $\{\{1,2\},\{3\}\}$, $\{\{1\},\{2,3\}\}$, and $\{\{1,3\},\{2\}\}$. However, the partition consisting of the grand coalition is individually stable and can be reached through the dynamics.

Our next application of the Deviation Lemma considers contractual Nash stability, where we obtain a similar result if we allow at most one nonpositive value. The proof is completely analogous and is therefore omitted. Note that this result also breaks down if we simultaneously allow the utility values -1 and 0 by constructing a similar cycle as in the previous example.

Theorem 4. *The dynamics of CNS-deviations always converges in ASHGs with at most one nonpositive utility value.*

Theorems 3 and 4 use the Deviation Lemma to derive positive results for the single-sided unanimity-based stability notions IS and CNS. In a third application of the deviation lemma, we show that this technique is also applicable to majority-based stability notions, at least when we involve both the welcoming and the abandoned coalition in the vote. The key idea is a suitable arrangement of the terms occurring in the difference of the potential with respect to the agents affected by a deviation.

Theorem 5. *The dynamics of JMS-deviations always converges in ASHGs with at most two distinct utility values.*

Note that since every JMS-deviation is also an SMS-deviation, the previous result holds for SMS as well. As in the discussion after Theorem 3, we obtain a polynomial running time of the dynamics for appropriate restrictions of the cases. We collect important consequences in the following corollary. In particular, we extend results by Dimitrov et al. (2006) and Aziz and Brandl (2012) who proved the existence of IS partitions for AFGs and AEGs, respectively.[2]

Corollary 1. *The dynamics of IS-, CNS-, and JMS-deviations always converges in polynomial time in AFGs, AEGs, and FEGs.*

We would like to stress that convergence of the dynamics does not guarantee a polynomial running time in general. An example is the case of symmetric utility values in ASHGs. For NS this can be directly inferred from the PLS-reduction by (Gairing and Savani 2019), which satisfies *tightness*, a property of reductions defined by Schäffer and Yannakakis (1991).

Proposition 1. *The dynamics of NS-deviations in symmetric ASHGs may require exponentially many rounds before converging to an NS partition.*

Proof. It is easy to verify that the PLS-reduction from PartyAffiliation under the Flip neighborhood by Gairing and Savani (2019, Observation 2) is tight. Schäffer and Yannakakis (1991, Lemma 3.3) showed that tight reductions preserve the existence of exponentially long running times of the standard local search algorithm, i.e., the NS-dynamics in our case. Note that the standard local search algorithm of the source problem can have an exponential running time, because PartyAffiliation is a generalization of MaxCut whose standard local search algorithm can run in exponential time with respect to the flip neighborhood (Schäffer and Yannakakis 1991, Theorem 5.15).[3] □

While the previous proposition uses a nonconstructive argument avoiding to construct an explicit example with an exponential running time, it is possible to construct such an example even in the more restricted case of IS-dynamics. To this end, it is possible to modify an example for MaxCut provided by Monien and Tscheuschner (2010) by essentially reverting the sequence of flips for MaxCut to obtain an execution of the IS-dynamics. Thus, we generalize the previous proposition via a constructive proof.

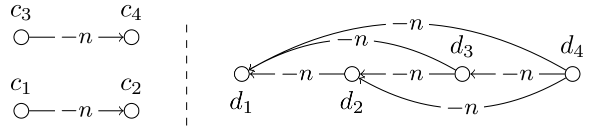


Figure 2: The *aversion to enemies* games without MIS partition (left) and MOS partition (right) from Proposition 3. Omitted edges have weight 1.

Proposition 2. *The dynamics of IS-deviations in symmetric ASHGs may require exponentially many rounds before converging to an IS partition.*

Stability under Majority Consent

In this section, we study stability under majority consent. First, the existential results of Theorem 3 and Theorem 4 are contrasted with the non-existence of stable partitions in AEGs under the majority-based relaxations of the respective stability concepts.

Proposition 3. *There exists an AEG which contains no MIS (respectively, MOS) partition.*

Proof. First, we provide an AEG with no MIS partition. Let $N = \{c_1, c_2, c_3, c_4\}$, i.e., there are $n = 4$ agents, and valuations defined as $v_{c_1}(c_2) = v_{c_3}(c_4) = -n$ and all other valuations set to 1. The AEG is illustrated in Figure 2 (left).

Assume for contradiction that there exists an MIS partition π. Then, $c_1 \notin \pi(c_2)$ and $c_3 \notin \pi(c_4)$. Also, $|\pi(c_1)| \leq 1$ (respectively, $|\pi(c_3)| \leq 1$), because otherwise, c_2 (respectively, c_4) would join via an MIS-deviation). But then $\pi(c_1) = \{c_1\}$ and $\pi(c_3) = \{c_3\}$, and c_1 could deviate to join $\pi(c_3)$, a contradiction.

Second, we provide an AEG without MOS partition. Let $N = \{d_1, d_2, d_3, d_4\}$, and define valuations for all $i, j \in [4]$ with $i < j$ as $v_{d_i}(d_j) = 1$ and $v_{d_j}(d_i) = -4$. An illustration is provided in Figure 2 (right).

Assume for contradiction that there exists an MOS partition π. Then, every coalition $C \in \pi$ must fulfill $|C| \leq 2$. Otherwise, the agent of C with the second smallest index would form a singleton via an MOS-deviation. In addition, there cannot be a singleton, because if some agent is in a singleton, there must be a second such agent, and then the one with the smaller index would join the other one. Hence, π consists of two pairs. But then d_1 would deviate to the pair not containing her, a contradiction. □

We can leverage the AEGs provided in the previous proposition as gadgets in reductions to show hardness of the associated decision problems. This can be interpreted as a more exact boundary (compared to Theorem 1) of the tractabilities encountered in Theorem 3 and Theorem 4 for the special case of AEGs.

Theorem 6. *It is* NP*-complete to decide if there exists an MIS (respectively, MOS) partition in AEGs.*

[2] These contributions actually show existence of partitions satisfying properties stronger than IS.

[3] We refer to the respective references for formal definitions of the involved combinatorial problems.

The utility restrictions in Theorem 6 are not as flexible as in the negative result for Nash stability in Theorem 1 or the positive results for unanimity-based dynamics in Theorem 3 and Theorem 4. In fact, the picture for majority-based notions is more diverse, because we obtain another positive result in the class of AFGs.

Theorem 7. *When starting from the grand coalition, the dynamics of MIS-deviations converges after at most n rounds in AFGs.*

Proof. The key insight is that there can only be deviations to form a new singleton coalition yielding no more than n deviations. Let $\pi_0 = \{N\}$ be the initial partition, and consider a sequence of k MIS-deviations

$$\pi_0 \xrightarrow{i_1} \pi_1 \xrightarrow{i_2} \dots \xrightarrow{i_k} \pi_k.$$

We inductively define coalitions evolving from the grand coalition if removing the deviator as $G_0 = N$, and $G_j = G_{j-1} \setminus \{i_j\}$ for $j > 0$.

Now, we proceed to simultaneously prove the following claims by induction:

1. $\forall j \in [k] : \pi_{j-1}(i_j) = G_{j-1}$.
2. $\forall j \in [k] : \pi_j(i_j) = \{i_j\}$.
3. $\forall j \in [k] : \{i \in \pi_{j-1}(i_j) \colon v_{i_j}(i) = n\} = \emptyset$.

The base case $j = 1$ is immediate. For the induction step, let $2 \leq j \leq k$ and suppose the claims are true for all $1 \leq l < j$. We start with the first claim. By the induction hypothesis, $\pi_{j-1} = \{G_{j-1}\} \cup \{\{i_l\} \colon 1 \leq l < j\}$. This means that if $\pi_{j-1}(i_j) \neq G_{j-1}$, we must have $\pi_{j-1}(i_j) = \{i_j\}$, indicating $i_j = i_l$ for some $l < j$. Then, the welcoming coalition cannot be G_{j-1}, as i_j, by induction hypothesis, abandoned G_{l-1} due to not having any friends in G_{l-1}, and thus has, by $G_{j-1} \subseteq G_{l-1}$, no friends in G_{j-1}, either. The alternative is that i_j joins another singleton coalition $\{i_m\}$ to form a pair. However, if i_m abandoned G_m at some point $m < l$, then she dislikes i_j, and won't allow her to join. If i_m abandoned G_m at some point $m > l$, then i_j dislikes i_m, and has no incentive to join. Hence, $\pi_{j-1}(i_j) = G_{j-1}$. For the second claim, note that i_j cannot join another singleton $\{i_m\}$, because i_m abandoned G_{m-1} at some point $m < j$ and thus dislikes i_j. Hence, i_j must form a singleton $\pi_j(i_j) = \{i_j\}$, which she only wants to do if $\{i \in \pi_{j-1}(i_j) \colon v_{i_j}(i) = n\} = \emptyset$. This accomplishes the third claim, and completes the induction proof.

Finally, as there can be at most n singletons, the dynamics must terminate after at most n rounds. $\square$

The computational boundaries in this section encountered so far only hold for one-sided stability notions where either the welcoming or the abandoned coalition takes a vote. On the other hand, Theorem 5 shows that these are opposed by tractabilities under two-sided majority consent.

For general utilities, existence of SMS (and therefore JMS) partitions is not guaranteed anymore, and we show that the tractabilities break down.

Theorem 8. *Deciding whether an ASHG contains an SMS (respectively, JMS) partition is* NP*-complete.*

Conclusion and Discussion

We studied stability based on single-agent deviations in additively separable hedonic games with a particular focus on games with restricted utility functions that can be naturally interpreted in terms of friends and enemies. We identified a computational boundary between Nash stability and stability with unanimous consent. The picture is less clear when deviations are governed by majority consent. While stable partitions always exist when considering both the abandoned and the welcoming coalition of the deviating agent, we obtain both positive and negative results if only one of these coalitions is considered. Table 1 summarizes our results and compares them with related results from the literature. Notably, we obtain all of our positive results through the convergence of simple and natural dynamics. This also extends previously known results about IS. Aziz and Brandl (2012) obtain a polynomial algorithm essentially by running a dynamics from the singleton partition, whereas Dimitrov et al. (2006) take a different, graph-theoretical approach considering strongly connected components. The construction of CIS partitions by Aziz, Brandt, and Seedig (2013) is done by iteratively identifying specific coalitions, and it is not known whether CIS-dynamics converge in polynomial time for natural starting partitions such as the singleton partition or grand coalition. An important tool in establishing our results concerning convergence of dynamics is the Deviation Lemma, a general combinatorial insight that allows us to study dynamics from a global perspective.

	General	$FEGs$	$AEGs$	$AFGs$
NS	NP-c[d]	NP-c (Th. 1)	NP-c (Th. 1)	NP-c (Th. 2)
IS	NP-c[d]	FP (Th. 3)	FP[a] (Th. 3)	FP[c] (Th. 3)
CNS	NP	FP (Th. 4)	FP (Th. 4)	FP (Th. 4)
CIS	FP[b]	FP[b]	FP[b]	FP[b]
MIS	NP-c (Th. 6)	?	NP-c (Th. 6)	FP (Th. 7)
MOS	NP-c (Th. 6)	?	NP-c (Th. 6)	?
JMS	NP-c (Th. 8)	FP (Th. 5)	FP (Th. 5)	FP (Th. 5)
SMS	NP-c (Th. 8)	FP (Th. 5)	FP (Th. 5)	FP (Th. 5)

Table 1: Overview of our computational results. The NP-completeness results concern deciding on the existence of a stable partition. The positive results mean that a stable partition can be constructed in polynomial time (Function-P) by executing a dynamics. Question marks indicate that it is even unknown whether a stable partition always exists.
[a]: Aziz and Brandl (2012), [b]: Aziz, Brandt, and Seedig (2013), [c]: Dimitrov et al. (2006), [d]: Sung and Dimitrov (2010)

Our work offers a wide range of interesting follow-up questions. First, Table 1 contains some problems left open in our analysis. Specifically, despite the existence of partitions without CNS partitions, the complexity of the existence problem of CNS partitions remains open for general utilities. Also, the voting-based stability notions deserve further investigation, and might even lead to interesting discoveries in other classes of hedonic games. Lastly, an intriguing further direction is to study further applications of the Deviation Lemma, particularly in domains other than coalition formation.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/12-1.

References

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

Aziz, H.; and Brandl, F. 2012. Existence of Stability in Hedonic Coalition Formation Games. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 763–770.

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

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

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

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

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

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

Brandt, F.; and Bullinger, M. 2020. Finding and Recognizing Popular Coalition Structures. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 195–203.

Brandt, F.; Bullinger, M.; and Wilczynski, A. 2021. Reaching Individually Stable Coalition Structures in Hedonic Games. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 5211–5218.

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

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

Cseh, Á. 2017. Popular Matchings. In Endriss, U., ed., *Trends in Computational Social Choice*, chapter 6. AI Access.

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

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

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

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

Gärdenfors, P. 1975. Match Making: Assignments based on bilateral preferences. *Behavioral Science*, 20(3): 166–173.

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

Monien, B.; and Tscheuschner, T. 2010. On the power of nodes of degree four in the local max-cut problem. In *Proceedings of the 7th International Conference on Algorithms and Complexity (CIAC)*, number 6078 in Lecture Notes in Computer Science (LNCS), 264–275. Springer-Verlag.

Newman, M. E. J. 2004. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2): 321–330.

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

Saad, W.; Han, Z.; Basar, T.; Debbah, M.; and Hjorungnes, A. 2011. Hedonic Coalition Formation for Distributed Task Allocation among Wireless Agents. *IEEE Transactions on Mobile Computing*, 10(9): 1327–1344.

Schäffer, A. A.; and Yannakakis, M. 1991. Simple Local Search Problems that are Hard to Solve. *SIAM Journal on Computing*, 20(1): 56–87.

Suksompong, W. 2015. Individual and Group Stability in Neutral Restrictions of Hedonic Games. *Mathematical Social Sciences*, 78: 1–5.

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

11

CORE PUBLICATION 6: LOYALTY IN CARDINAL HEDONIC GAMES

SUMMARY

A common theme of decision making in multi-agent systems is to assume selfish behavior of agents following Bentham's utilitarianism. Based on this paradigm, agents simply assign values to alternatives (or output options), which they seek to maximize. This rationale is questionable in coalition formation where agents are affected by other members of their coalition. We propose the concept of loyalty in hedonic games, a binary relation dependent on agents' utilities. This concept is based on the assumption that agents are benevolent towards other agents they like to form coalitions with.

Given a hedonic game, we can define an associated loyal variant where agents' utilities are defined by taking the minimum of their utility and the utilities of agents towards which they are loyal. The goal is to analyze the loyal variant in comparison to the original benchmark game and to carve out the influence of the modified incentives.

Since the loyal variant of a hedonic game is a hedonic game itself, taking the loyal variant can be iterated to obtain various degrees of loyalty. Interestingly, this process terminates after a finite number of steps in a game with a high degree of loyalty where the incentives of agents lead to a locally egalitarian behavior with respect to the benchmark game.

Surprisingly, loyalty leads to an increase of complexity regarding the description of the game. Even if the preferences in the benchmark game follow a simple structure based on the consideration of friends and enemies, computing best coalitions in the loyal variants leads to computational hardness.

Moreover, we determine the desirability of coalition structures in the obtained game variants by means of group stability and efficiency. Specifically, we consider the problem of finding coalition structures in the core as well as coalition structures satisfying Pareto optimality. We obtain strong existential results, opposed by computational intractabilities. In particular, the limit game possesses Pareto-optimal coalition structures in the core, while it is generally hard to compute solutions in the core.

REFERENCE

M. Bullinger and S. Kober. **Loyalty in cardinal hedonic games**. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 66–72, 2021.
DOI: https://doi.org/10.24963/ijcai.2021/10

INDIVIDUAL CONTRIBUTION

I, Martin Bullinger, am the main author of this publication. In particular, I am responsible for the development and conceptual design of the research project, joint development and write-up of proofs and results, and the write-up of the manuscript.

COPYRIGHT AGREEMENT

The right to present this paper in a doctoral thesis has been granted by the publisher, the International Joint Conference on Artificial Intelligence (IJCAI), in the copyright form presented below. There, IJCAI grants permission to "*personally reuse all or portions of the paper in other works of their own authorship.*" This form can also be found at
https://proceedings.ijcai.org/docs/IJCAI-Copyright_Transfer_Agreement.pdf
(accessed November 10, 2022).

INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE

TRANSFER OF COPYRIGHT AGREEMENT

Title of Article/Paper: ______________________________

Publication in Which Article Is to Appear: ______________________________

Author's Name(s): ______________________________

Please type or print your name as you wish it to appear in print

(Please read and sign Part A only, unless you are a government employee and created your article/paper as part of your employment. If your work was performed under Government contract, but you are not a Government employee, sign Part A and see item 6 under returned rights.)

PART A–Copyright Transfer Form

The undersigned, desiring to publish the above article/paper in a publication of the International Joint Conferences on Artificial Intelligence, Inc., hereby transfer their copyrights in the above paper to the International Joint Conferences on Artificial Intelligence, Inc. (IJCAI), in order to deal with future requests for reprints, translations, anthologies, reproductions, excerpts, and other publications.

This grant will include, without limitation, the entire copyright in the paper in all countries of the world, including all renewals, extensions, and reversions thereof, whether such rights currently exist or hereafter come into effect, and also the exclusive right to create electronic versions of the paper, to the extent that such right is not subsumed under copyright. The undersigned warrants that he/she is the sole author and owner of the copyright in the above paper, except for those portions shown to be in quotations; that the paper is original throughout; and that the undersigned's right to make the grants set forth above is complete and unencumbered.

If anyone brings any claim or action alleging facts that, if true, constitute a breach of any of the foregoing warranties, the undersigned will hold harmless and indemnify IJCAI, their grantees, their licensees, and their distributors against any liability, whether under judgment, decree, or compromise, and any legal fees and expenses arising out of that claim or actions, and the undersigned will cooperate fully in any defense IJCAI may make to such claim or action. Moreover, the undersigned agrees to cooperate in any claim or other action seeking to protect or enforce any right the undersigned has granted to IJCAI in the paper. If any such claim or action fails because of facts that constitute a breach of any of the foregoing warranties, the undersigned agrees to reimburse whomever brings such claim or action for expenses and attorney's fees incurred therein.

Returned Rights

In return for these rights, IJCAI hereby grants to the above authors, and the employers for whom the work was performed, royalty-free permission to:

1. retain all proprietary rights (such as patent rights) other than copyright and the publication rights transferred to IJCAI;
2. personally reuse all or portions of the paper in other works of their own authorship;
3. make oral presentation of the material in any forum;
4. reproduce, or have reproduced, the above paper for the author's personal use, or for company use provided that IJCAI copyright and the source are indicated, and that the copies are not used in a way that implies IJCAI endorsement of a product or service of an employer, and that the copies per se are not offered for sale. The foregoing right shall not permit the posting of the paper in electronic or digital form on any computer network, except by the author or the author's employer, and then only on the author's or the employer's own World Wide Web page or ftp site. Such Web page or ftp site, in addition to the aforementioned requirements of this Paragraph, must provide an electronic reference or link back to the IJCAI electronic server (http://www.ijcai.org), and shall not post other IJCAI copyrighted materials not of the author's or the employer's creation (including tables of contents with links to other papers) without IJCAI's written permission;
5. make limited distribution of all or portions of the above paper prior to publication.
6. In the case of work performed under U.S. Government contract, IJCAI grants the U.S. Government royalty-free permission to reproduce all or portions of the above paper, and to authorize others to do so, for U.S. Government purposes. In the event the above paper is not accepted and published by IJCAI, or is withdrawn by the author(s) before acceptance by IJCAI, this agreement becomes null and void.

______________________________ ______________________________

Author's (or Employer's Representative) Signature Date

______________________________ ______________________________

Employer for whom work was performed Title (if not author)

Loyalty in Cardinal Hedonic Games

Martin Bullinger and **Stefan Kober**
Technical University of Munich
bullinge@in.tum.de, stefan.kober@tum.de

Abstract

A common theme of decision making in multi-agent systems is to assign utilities to alternatives, which individuals seek to maximize. This rationale is questionable in coalition formation where agents are affected by other members of their coalition. Based on the assumption that agents are benevolent towards other agents they like to form coalitions with, we propose loyalty in hedonic games, a binary relation dependent on agents' utilities. Given a hedonic game, we define a loyal variant where agents' utilities are defined by taking the minimum of their utility and the utilities of agents towards which they are loyal. This process can be iterated to obtain various degrees of loyalty, terminating in a locally egalitarian variant of the original game.

We investigate axioms of group stability and efficiency for different degrees of loyalty. Specifically, we consider the problem of finding coalition structures in the core and of computing best coalitions, obtaining both positive and intractability results. In particular, the limit game possesses Pareto optimal coalition structures in the core.

1 Introduction

Decision making in multi-agent systems is highly driven by the idea of the *homo economicus*, a rational decision taker that seeks to maximize her individual well-being. Following the classical *Theory of Games and Economic Behavior* by von Neumann and Morgenstern, agents assign utilities to alternatives and aim for an outcome that maximizes individual utility. Such behavior entails many delicate situations in non-cooperative game theory such as the prisoner's dilemma or the tragedy of the commons [Hardin, 1968], where agents take decisions in their individual interest without regarding other agents. This leads to outcomes that are bad for the society as a whole and often, as it is the case in the prisoner's dilemma, agents have an incentive to coordinate to improve their respective situation.

From the theoretical point of view, one can either accept the existence of such dilemmata and study their social impact, for instance, by means of the price of anarchy [Koutsoupias and Papadimitriou, 1999], or one can ask for the degree of individual dependency on the social outcome necessary to escape a situation of inferior welfare. The latter idea is implemented by adapting the utility function of players as a weighted sum of individual and joint utility, an idea repeatedly developed in network design [Elias *et al.*, 2010], artificial intelligence [Apt and Schäfer, 2014], or public choice [Mueller, 1986]. Specifically, the *selfishness level* by Apt and Schäfer is the lowest weight on the joint utility such that a Nash equilibrium becomes a social optimum.

On the other hand, empirical evidence does not only question whether agents behave according to the utility model by von Neumann and Morgenstern [Kahneman and Tversky, 1979], but even supports the hypothesis that human behavior is steered by the well-being of the whole society [Colman *et al.*, 2008]. However, in scenarios of high competition, agents might also act spiteful towards other agents, i.e., there is an incentive to harm other agents [Levine, 1998].

In cooperative game theory, it seems to be an even more reasonable assumption to include other agents into the own valuation. We follow this line of thought in the setting of coalition formation, where we propose loyalty, a possibility to modify utilities by taking into account other agents' utilities towards which loyalty is perceived. Loyalty is a binary relation directly extracted from the agents' utilities over partnership, i.e., coalitions of size 2. Loyalty is sensed towards the agents within the own coalition that yield positive utility in a partnership. Following the paradigm of a chain that is only as strong as its weakest link, loyal utilities are obtained by taking the minimum of the own utility and the utilities of agents receiving our loyalty. As such, we obtain a loyal variant of the original game, which is itself a coalition formation game, and we can iterate towards various degrees of loyalty. As we will see, this process terminates in a game which satisfies a high degree of egalitarianism. We consider common solution concepts concerning group stability and efficiency for different degrees of loyalty and the limit game, and provide both existential and computational results.

We study coalition formation in the framework of hedonic games [Drèze and Greenberg, 1980; Banerjee *et al.*, 2001; Bogomolnaia and Jackson, 2002]. Our contribution lies in studying aspects of empathy in hedonic games [Brânzei and Larson, 2011; Monaco *et al.*, 2018; Nguyen *et al.*, 2016]. Previous work considers empathy between agents through various alternative utility functions based on friendship relations among the agents extracted from utility functions or a social network. Closest to our work are altruistic hedonic games introduced by Nguyen *et al.* [2016] and subsequently studied by Wiechers and Rothe [2020], Kerkmann and Rothe [2020], and Schlueter and Goldsmith [2020]. Our first degree of loyalty in symmetric friend-oriented hedonic games coincides with minimum-equal-treatment altruistic hedonic games as defined by Wiechers and Rothe [2020]. We significantly extend their model, but since most of our hardness results work for the restricted class of symmetric friend-oriented hedonic games, they have immediate consequences for this type of altruistic hedonic games. Also, loyal variants of hedonic games fit into the framework of super altruistic hedonic games by Schlueter and Goldsmith [2020] if their aggregation is modified by taking the average instead of the minimum of other agents' utilities.

2 Preliminaries and Model

We start with some notation. Define $[i] = \{1, \dots, i\}$ and $[i, j] = \{i, \dots, j\}$ for $i, j \in \mathbb{Z}, i \leq j$.

Also, we use standard notions from graph theory. Let $G = (V, E)$ be an undirected graph. For a subset of agents $W \subseteq V$, denote by $G[W]$ the subgraph of G *induced by* W. Given two vertices $v, w \in V$, we denote by $d_G(v, w)$ their *distance* in G, i.e., the length of a shortest path connecting them. The graph G is called *regular* if there exists a non-negative integer r such that every vertex of G has degree r.

In the following subsections, we introduce hedonic games, our concept of loyalty, and desirable properties of coalition structures.

2.1 Cardinal Hedonic Games

Let $N = \{1, \dots, n\}$ be a finite set of agents. A *coalition* is a non-empty subset of N. By $\mathcal{N}_i$ we denote the set of coalitions agent i belongs to, i.e., $\mathcal{N}_i = \{S \subseteq N : i \in S\}$. A *coalition structure*, or simply a *partition*, is a partition π of the agents N into disjoint coalitions, where $\pi(i)$ denotes the coalition agent i belongs to. A *hedonic game* is a pair $(N, \succsim)$, where $\succsim = (\succsim_i)_{i \in N}$ is a preference profile specifying the preferences of each agent i as a complete and transitive preference relation $\succsim_i$ over $\mathcal{N}_i$. In hedonic games, agents are only concerned about their own coalition. Accordingly, preferences over coalitions naturally extend to preferences over partitions as follows: $\pi \succsim_i \pi'$ if and only if $\pi(i) \succsim_i \pi'(i)$.

Throughout the paper, we assume that rankings over the coalitions in $\mathcal{N}_i$ are given by utility functions $u_i : \mathcal{N}_i \to \mathbb{R}$, which are extended to evaluate partitions in the hedonic way by setting $u_i(\pi) = u_i(\pi(i))$. A hedonic game together with a representation by utility functions is called *cardinal hedonic game*. Because the sets $\mathcal{N}_i$ are finite, preferences could in principle always be represented by cardinal values. This is impractical due to two reasons. First, such utility functions require exponential space to represent. Therefore it would be desirable to consider classes of hedonic games with succinct representations. Second, we would like to compare different agents' utility functions such that a certain cardinal value expresses the same intensity of a preference for all agents. This cannot be guaranteed by arbitrary utility representations of ordinal preferences. Our model of loyalty is therefore particularly meaningful in succinctly representable classes of cardinal hedonic games. These include the following classes of hedonic games, which aggregate utility functions over single agents of the form $u_i : N \to \mathbb{R}$ where $u_i(i) = 0$, which can be represented by a complete weighted digraph.

- *Additively separable hedonic games (ASHGs)* [Bogomolnaia and Jackson, 2002]: utilities are aggregated by taking the sum of single utilities, i.e., $u_i(\pi) = \sum_{j \in \pi(i)} u_i(j)$.
- *Friend-oriented hedonic games (FOHGs)* [Dimitrov *et al.*, 2006]: the restriction of ASHGs where utilities for other agents are either n (the agent is a *friend*) or -1 (the agent is an *enemy*), i.e., for all $i, j \in N$ with $i \neq j$, $u_i(j) \in \{n, -1\}$. Given an FOHG, the set $F_i = \{j \in N : u_i(j) = n\}$ is called *friend set* of agent i. The unweighted digraph $G_F = (N, A)$ where $(i, j) \in A$ if and only if $j \in F_i$ is called *friendship graph*. An FOHG can be represented by specifying the friend set for every agent or by its friendship graph.
- *Modified fractional hedonic games (MFHGs)* [Olsen, 2012]: utilities are aggregated by dividing the sum of single utilities by the size of the coalition minus 1, i.e., $u_i(\pi) = 0$ if $\pi(i) = \{i\}$, and $u_i(\pi) = \frac{\sum_{j \in \pi(i)} u_i(j)}{|\pi(i)| - 1}$, otherwise. In other words, the utility of a coalition structure is the expected utility achieved through another agent in the own coalition selected uniformly at random.

A cardinal hedonic game is called *mutual* if, for all pairs of agents $i, j \in N$, $u_i(j) > 0$ implies $u_j(i) > 0$. It is called *symmetric* if, for all pairs of agents $i, j \in N$, $u_i(j) = u_j(i)$. Clearly, symmetric games are mutual. Throughout most of the paper, we will consider at least mutual variants of the classes of hedonic games, which we just introduced.

2.2 Loyalty in Hedonic Games

We are ready to define our concept of loyalty. Given a cardinal hedonic game, its loyal variant needs to specify two key features. First, for every agent, we need to identify a loyalty set, which contains the agents towards which loyalty is expressed. Second, we need to specify how loyalty is expressed, i.e., how to obtain new, loyal utility functions.

Formally, given a cardinal hedonic game and an agent $i \in N$, we define her *loyalty set* as $L_i = \{j \in N \setminus \{i\} : u_i(\{i, j\}) > 0\}$. In other words, agents are affected by agents that influence them positively when being in a joint coalition. Note that for all hedonic games considered in this paper, the loyalty set is equivalently given by $L_i = \{j \in N \setminus \{i\} : u_i(\{i, j\}) > u_i(i)\}$, i.e., it contains the agents with which i would rather form a coalition of size 2 than staying on her own. The *loyalty graph* is the directed graph $G_L = (N, A)$ where $(i, j) \in A$ if and only if $j \in L_i$.

It remains to specify how agents aggregate utilities in a loyal way. Given a cardinal hedonic game, its *loyal variant* is defined on agent set N by the utility function $u_i^L(\pi) = \min_{j \in \pi(i) \cap (L_i \cup \{i\})} u_j(\pi(i))$. Interestingly, the loyal variant is itself a hedonic game, and we can consider its own loyal variant. Following this reasoning, we recursively define the *k-fold loyal variant* by setting the 1-fold loyal variant to the loyal variant and the $(k+1)$-fold loyal variant to the loyal variant of the k-fold loyal variant. Also, we denote by u_i^k and L_i^k the utility function and the loyalty set of an agent i, and by G_L^k the loyalty graph of the k-fold loyal variant.

In fact, we will see that this process terminates after at most n steps in a limit game that satisfies egalitarianism at the level of coalitions. For simplicity, we restrict attention to mutual cardinal hedonic games, where the loyalty sets defines a symmetric binary relation and the loyalty graph can be represented by an undirected graph.[1] For an agent $i \in N$, let $G_L^\pi(i)$ be the agents in the connected component of the subgraph of G_L induced by $\pi(i)$ containing i. Now, define the *locally egalitarian variant* of a cardinal hedonic game as the game on agent set N with utilities given by $u_i^E(\pi) = \min_{j\in G_L^\pi(i)} u_j(\pi)$. In other words, an agent receives the minimum utility among all agents reachable within her coalition in the loyalty graph.

Finally, we introduce a technical assumption. A mutual cardinal hedonic game is called *loyalty-connected* if, for all agents $i \in N$ and coalition structures π, $u_i(G_L^\pi(i)) \geq u_i(\pi)$. This property precludes negative influence through agents outside the reach of loyalty, and is satisfied by reasonable classes of cardinal hedonic games like ASHGs, MFHGs, or fractional hedonic games [Aziz *et al.*, 2019].

2.3 Solution Concepts

We evaluate the quality of coalition structures by measures of stability and efficiency.

A common concept of group stability is the core. Given a coalition structure π, a coalition $C \subseteq N$ is *blocking* π (respectively, *weakly blocking* π) if for all agents $i \in C$, $u_i(C) > u_i(\pi)$ (respectively, for all agents $i \in C$, $u_i(C) \geq u_i(\pi)$, where the inequality is strict for some agent in C). A coalition structure π is in the *core* (respectively, *strict core*) if there exists no non-empty coalition blocking (respectively, weakly blocking) π.

A fundamental concept of efficiency is Pareto optimality. A coalition structure π' *Pareto dominates* a coalition structure π if, for all $i \in N$, $u_i(\pi'(i)) \geq u_i(\pi(i))$, where the inequality is strict for some agent in N. A coalition structure π is called *Pareto optimal* if it is not Pareto dominated by another coalition structure. In other words, given a Pareto optimal coalition structure, every other coalition structure that is better for some agent, is also worse for another agent.

Another concept of efficiency concerns the welfare of a coalition structure. There are many notions of welfare dependent on how to aggregate single agents' utilities for a social evaluation. In the context of loyalty, egalitarianism seems to be especially appropriate. It aims to maximize the well-being of the agent that is worst off. Formally, the *egalitarian welfare* of a partition π is defined as $\mathcal{E}(\pi) = \min_{i\in N} u_i(\pi(i))$. Also, let $\mathcal{E}^k(\pi)$ denote the egalitarian welfare of the k-fold loyal variant. Following this definition, coalition structures maximizing egalitarian welfare are not necessarily Pareto optimal. However, there exists always a Pareto optimal coalition structure maximizing egalitarian welfare. Specifically, a coalition structure maximizes *leximin welfare* if its utility vector, sorted in non-decreasing order, is lexicographically largest. A coalition structure maximizing leximin welfare is Pareto optimal and maximizes egalitarian welfare.

Apart from finding efficient coalition structures, an individual goal of an agent i is to be in a *best coalition*, i.e., in a coalition in $\mathcal{N}_i$ maximizing her utility. Formally, the problem of, given a cardinal hedonic game, an agent $i^* \in N$, and a rational number $q \in \mathbb{Q}$, deciding if there exists a subset $C \subseteq N$ with $i^* \in C$ and $u_{i^*}(C) \geq q$, is called `BestCoalition`.

3 Loyalty Propagation and Best Coalitions

Our first proposition collects some initial observations. It states, how loyalty propagates through the loyalty graph for higher degree loyal variants, terminating with the locally egalitarian variant, and considers egalitarian welfare.

Proposition 1. *Let a mutual cardinal hedonic game on agent set N with $|N| = n$ be given. Let $k \geq 1$, $i \in N$, and π a coalition structure. Then, the following statements hold.*

1. *The loyalty graph and loyalty sets are the same for all loyal variants, i.e., $G_L^k = G_L^1$ and $L_i^k = L_i^1$.*
2. *Loyalty extends to agents at distance k, i.e., $u_i^k(\pi) = \min\{u_j(\pi) \colon j \in \pi(i) \text{ with } d_{G_L[\pi(i)]}(i,j) \leq k\}$.*
3. *Utilities converge to the utilities of the locally egalitarian variant, i.e., $u_i^l = u_i^E$ for all $l \geq n$.*
4. *Egalitarian welfare is preserved, i.e., $\mathcal{E}^k(\pi) = \mathcal{E}(\pi)$.*

Proof. The first statements follow immediately from mutuality. We prove the second statement by induction over k. For $k = 1$, the assertion follows directly from the definition of the loyal variant.

Now, let $k \geq 2$ be an integer. Let $C = \pi(i)$, $C_L = \pi(i) \cap (L_i \cup \{i\})$, $H = G_L[\pi(i)]$, and for $p \geq 1$, let $C_p(j) = \{m \in C \text{ with } d_H(j,m) \leq p\}$. Then,

$$\begin{aligned} u_i^k(\pi) &= \min_{j\in C_L} u_j^{k-1}(C) \\ &= \min_{j\in C_L} \min\{u_m(C) \colon m \in C_{k-1}(j)\} \\ &= \min_{j\in C\colon d_H(i,j)\leq 1} \min\{u_m(C) \colon m \in C_{k-1}(j)\} \\ &= \min\{u_j(C) \colon j \in C \text{ with } d_H(i,j) \leq k\}. \end{aligned}$$

There, the second equality follows by induction, the third equality by definition of the loyalty graph, and the last equality by observing that the vertices with a distance of at most k from i are precisely the vertices with a distance of at most $k-1$ from an arbitrary neighbor.

The third statement follows from the second one, and the final statement follows from the observation that the minimum utility among agents in a coalition structure is preserved when transitioning to a loyal variant. $\square$

Example 1. *We provide an example showing that part 4 of Proposition 1 does not extend to leximin welfare.*

[1] This restriction is in accordance with our results, but it can be lifted with some technical effort.

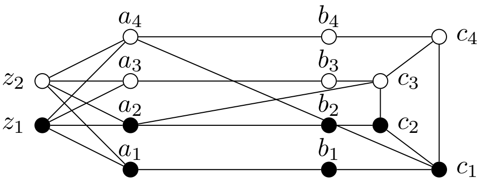


Figure 1: Friendship graph of Example 1. The black and white coalitions constitute a coalition structure minimizing leximin welfare for the 2-fold loyal variant, which is not Pareto optimal under the original utilities.

Consider a symmetric FOHG with agent set $N = \{a_i, b_i, c_i \colon 1 \leq i \leq 4\} \cup \{z_1, z_2\}$*, and the friendship graph in Figure 1. It can be shown that the coalition structure* $\pi = \{\{z_i, a_{2i-1}, a_{2i}, b_{2i-1}, b_{2i}, c_{2i-1}, c_{2i}\} \colon i = 1, 2\}$ *maximizes leximin welfare for its* 2*-fold loyal variant (consider agents of type* b_i*). However,* π *is not even Pareto optimal under the original utilities. Indeed,* $\pi' = \{\{z_1, a_1, a_4, b_1, b_4, c_1, c_4\}, \{z_2, a_2, a_3, b_2, b_3, c_2, c_3\}\}$ *is a Pareto improvement. All agents receive at least the same utility, and* a_2, a_4, c_1*, and* c_3 *are better off.*

Our next goal is to reason about finding best coalitions for an agent. Note that this problem can usually be solved in polynomial time. For instance, in ASHGs, given an agent i, every coalition that contains i together with all agents that give positive utility to i and no agent that gives negative utility to i is a best coalition for i. By contrast, we obtain hardness results for loyalty even in symmetric FOHGs. While it is possible to determine the number of friends of the unhappiest friend in a best coalition in polynomial time [Wiechers and Rothe, 2020], the problem becomes hard if the number of enemies is to be minimized at the same time. We omit some proof details and proofs due to space restrictions, but they can all be found in the extended version of the paper.

Theorem 2. *Let* $k \geq 1$*. Then,* `BestCoalition` *is NP-complete for the* k*-fold loyal variant of symmetric FOHGs.*

Proof sketch. Membership in NP is clear. For hardness, we provide a reduction from the NP-complete problem `SetCover` [Karp, 1972]. An instance of `SetCover` consists of a triple (A, S, κ), where A is some finite ground set, $S \subseteq 2^A$ is a set of subsets of A, and κ is an integer. The instance (A, S, κ) is a Yes-instance if there exists $S' \subseteq S$ with $\bigcup_{B \in S'} B = A$ and $|S'| \leq \kappa$. The reduction is illustrated in Figure 2.

Let $k \in \mathbb{N}$. Define $M = \left\lfloor \frac{k-1}{2} \right\rfloor$. Given an instance (A, S, κ) of `SetCover`, define $a = |A|$. We define an instance $((N, (F_i)_{i \in N}), i^*, q)$ of `BestCoalition` based on an FOHG $(N, (F_i)_{i \in N})$ represented via friend sets by specifying each individual component. The agent set is defined as $N = \{w_i \colon i \in [0, a+2]\} \cup \{v_i \colon i \in [0, a-1]\} \cup \{\alpha_i^j, \beta_i^j \colon i \in [a], j \in [M]\} \cup A \cup S$, and consists of representatives of the elements of A and S, and auxiliary agents. If k is even, set $i^* = w_0$ and if k is odd, $i^* = v_0$. The friend sets are given as

- $F_{w_0} = \{w_1, v_0, \ldots, v_{a-1}\}$,
- $F_{w_1} = \{w_0, w_2, w_3, \ldots, w_{a+2}\}$,

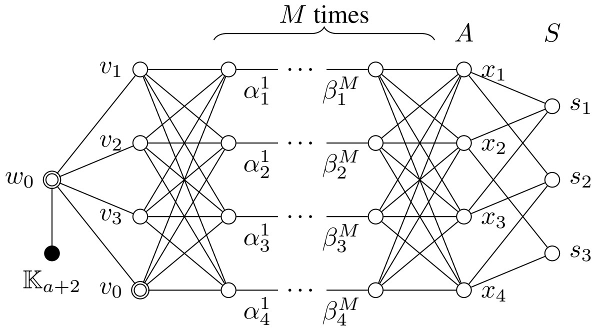


Figure 2: Schematic of the hardness reduction in Theorem 2 for $k \geq 2$. The figure shows the friendship graph for the instance of `SetCover` given by $A = \{x_1, x_2, x_3, x_4\}$ and $S = \{s_1 = \{x_1, x_2, x_3\}, s_2 = \{x_1, x_3, x_4\}, s_3 = \{x_2, x_4\}\}$. The black vertex indicates a complete subgraph on $a + 2$ vertices. We ask `BestCoalition` for the agents v_0 and w_0, respectively, indicated by double circles.

- $F_{w_i} = \{w_j \colon j \in [a+2], j \neq i\}$ for $i \in [2, a+2]$,
- $F_{v_i} = \{w_0, \alpha_1^1, \ldots, \alpha_a^1\}$ for $i \in [0, a-1]$ if $k > 2$,
- $F_{v_i} = \{w_0\} \cup A$ for $i \in [0, a-1]$ if $k \leq 2$,
- $F_{\alpha_i^1} = \{v_0, \ldots, v_{a-1}, \beta_i^1\}$ for $i \in [a]$,
- $F_{\alpha_i^j} = \{\beta_1^{j-1}, \ldots, \beta_a^{j-1}, \beta_i^j\}$ for $i \in [a], j \in [2, M]$,
- $F_{\beta_i^j} = \{\alpha_i^j, \alpha_1^{j+1}, \ldots, \alpha_a^{j+1}\}$ for $i \in [a], j \in [M-1]$,
- $F_{\beta_i^M} = \{\alpha_i^M\} \cup A$ for $i \in [a]$,
- $F_x = \{\beta_i^M \colon i \in [a]\} \cup \{s \in S \colon x \in s\}$ for $x \in A$ if $k > 2$,
- $F_x = \{v_0, \ldots, v_{a-1}\} \cup \{s \in S \colon x \in s\}$ for $x \in A$ if $k \leq 2$, and
- $F_s = \{x \in A \colon x \in s\}$ for $s \in S$ (in other words, $F_s = s$).

Finally, with $n = |N|$, specify the threshold utility $q = n(a+1) - (a+\kappa)$ for $k = 1$ and $q = n(a+1) - (1 + \kappa + 2(M+1)a)$, otherwise. Note that the distance between i^* and the x_i in the loyalty graph is exactly k.

If (A, S, κ) is a Yes-instance, let $S' \subseteq S$ be a set cover of A with at most κ sets. For $k = 1$, consider the coalition $C = A \cup S' \cup \{v_0, \ldots, v_{a-1}, w_0, w_1\}$. For $k \geq 2$, consider the coalition $C = (N \setminus S) \cup S'$. It is quickly checked that in each case $u_{v_0}^k(C) \geq q$.

Conversely, assume that C is a coalition with $i^* \in C$ and $u_{i^*}^k(C) \geq q$. Then, all agents that have a distance of at most k in the loyalty graph have to be included due to the degrees of vertices at a distance of at most k. In particular, $A \subseteq C$ for any k. Let $S' = C \cap S$.

First, consider the case $k = 1$. Then, $u_{v_0}(C) = n(a+1) - a - |S'|$. Hence $u_{v_0}^1(C) \geq q$ implies that $|S'| \leq \kappa$. In addition, every agent $x \in A$ must have at least $a+1$ friends present in C. In other words, for every $x \in A$ there exists $s \in S'$ with $x \in s$. Hence, S' is a cover of A with at most κ elements.

For arbitrary $k \geq 2$, it holds that $u_{i^*}(C) = n(a+1) - 1 - |S'| - (M+2)a$. Hence $u^1_{v_0}(C) \geq q$ implies that $|S'| \leq \kappa$. The remainder follows analogous to the case $k = 1$. □

Since the instances in the previous reduction contain agents with an arbitrarily large distance (parametrized by k), we cannot deduce direct consequences for the locally egalitarian variant. However, it is possible to bound the diameter in the reduced instances globally to obtain a similar result.

Theorem 3. *`BestCoalition` is NP-complete for the locally egalitarian variant of symmetric FOHGs.*

If we change the underlying class of hedonic games, we can circumvent the hardness results of the last two theorems.

Theorem 4. *Let $k \geq 1$. Then, `BestCoalition` can be solved in polynomial time for the k-fold loyal variant and the locally egalitarian variant of symmetric MFHGs.*

4 Coalition Structures in the Core

In this section we consider group stability in the locally egalitarian variant and the loyal variants.

4.1 Core in the Locally Egalitarian Variant

We start with a general lemma yielding a sufficient condition for existence of Pareto optimal coalition structures in the core.

Lemma 5. *Consider a class of hedonic games with the following two properties:*

1. *Restrictions of the game to subsets of agents are in the class.*
2. *For every coalition in any game of the class, the value of the coalition is the same for every player in the coalition.*

Then, for every game in the class, there exists a coalition structure in the core which is Pareto optimal.

Weakening the second condition of the lemma to the existence of some coalition that is best for all of its members is sufficient to find a coalition structure in the core. We discuss this in the extended version of the paper. Interestingly, the lemma can be applied to the locally egalitarian variant of cardinal hedonic games under fairly weak assumptions.

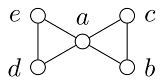
Theorem 6. *Let a loyalty-connected, mutual cardinal hedonic game be given. Then, there exists a Pareto optimal coalition structure in the core of its locally egalitarian variant.*

Proof. Let a loyalty-connected, mutual cardinal hedonic game be given and consider its locally egalitarian variant. We modify the utility functions such that $u^E_i(C)$ stays the same if C is connected in the loyalty graph, and set it to 0, otherwise. It suffices to find a Pareto optimal member of the core under this modification, because, by loyalty-connectivity, splitting coalitions into their connected components in the loyalty graph is weakly better for every agent, even under u^E. Consider the class of hedonic games given by this modified n-fold loyal variant together with all of its restrictions, in which we apply the same modifications towards the utility values for non-connected coalitions.

By Proposition 1, the utility for a coalition is the same for every player in the coalition. Hence, all requirements of Lemma 5 are satisfied and we find the desired coalition structure. □

In the extended version of the paper, we provide an example for the necessity of loyalty-connectivity in the previous theorem.

Example 2. *We extend an example by Wiechers and Rothe [2020] that shows that the previous result cannot be strengthened to find a coalition structure in the strict core. Consider the symmetric FOHG on agent set $\{a, b, c, d, e\}$ with loyalty graph depicted below.*



Consider its locally egalitarian variant. Then, $\{a, b, c\}$ is the unique best coalition for agents b and c and among the best coalitions for agent a. Hence, it has to be contained in every coalition structure in the strict core. Similarly, $\{a, d, e\}$ has to be a coalition in the strict core. As these conditions cannot be satisfied simultaneously, the strict core is empty.

Note that both the coalition structure $\{\{a, b, c\}, \{d, e\}\}$ and $\{\{a, d, e\}, \{b, c\}\}$ are in the core and Pareto optimal.

The construction in Lemma 5 gives rise to a simple recursive algorithm that computes Pareto optimal coalition structures in the core. Still, the computational complexity highly depends on the underlying cardinal hedonic game. While a modified version of the algorithm by Bullinger [2020] for computing Pareto optimal coalition structures in symmetric MFHGs finds a coalition structure in the core of their locally egalitarian variants, a version of our reduction on best coalitions shows an intractability for FOHGs.

Theorem 7. *The following statements hold.*

1. *Computing a coalition structure in the core can be done in polynomial time for the locally egalitarian variant of symmetric MFHGs.*
2. *Computing a coalition structure in the core is NP-hard for the locally egalitarian variant, even in the class of symmetric FOHGs with non-empty core.*

4.2 Core in the Loyal Variants

In contrast to the locally egalitarian variant, the k-fold loyal variant may have an empty core for arbitrary k. This is even true in a rather restricted class of symmetric ASHGs with individual values restricted to $\{n, n+1, -1\}$.

Proposition 8. *For every $k \geq 1$, there exists a symmetric ASHG with $\mathcal{O}(k)$ agents such that the core of its k-fold loyal variant is empty.*

Proof sketch. We only describe the instance. Let $k \in \mathbb{N}$. We define an ASHG $(N, (u_i)_{i \in N})$. Set $m = k$ if k is an even number and $m = k+1$ if k is odd. Let $A_i = \{a_i, b^i_1, \dots, b^i_m, c^i_1, \dots, c^i_m\}$ for $i \in [3]$. Define $N = \bigcup_{i=1}^3 A_i$ as the set of agents and let $n = |N|$. Reading indices i modulo 3, we define symmetric utilities according to

- $u(a_i, b^i_1) = u(a_i, c^i_1) = n + 1$ for $i \in [3]$,
- $u(b^i_m, a_{i+1}) = u(c^i_m, a_{i+1}) = n$ for $i \in [3]$,

- $u(b^i_j, b^i_{j+1}) = u(c^i_j, c^i_{j+1}) = n + 1$ for $i \in [3], j \in [m-1]$, and
- $u(v, w) = -1$ for all other utilities.

Note that $|N| = 3(2m + 1) = \mathcal{O}(k)$. □

We can use the previous counterexample as a gadget in a sophisticated reduction to obtain computational hardness.

Theorem 9. *Let $k \geq 1$. Deciding whether the core is non-empty is NP-hard for the k-fold loyal variant of symmetric ASHGs.*

Naturally, the question arises whether the core is always non-empty for loyal variants of FOHGs. While we leave the ultimate answer to this question as an open problem, we give evidence into both directions. First, we determine certain graph topologies that allow for coalition structures in the core. By contrast, we provide an intractability result for the computation of coalition structures in the core, and in the extended version of the paper we show that the dynamics related to blocking coalitions can cycle.

Proposition 10. *Let a symmetric FOHG with connected, regular friendship graph be given. Then the coalition structure consisting of the grand coalition is in the strict core for the k-fold loyal variant for every $k \geq 1$.*

Proof. Assume that the friendship graph is regular with every vertex having degree r. Singleton coalitions are clearly not weakly blocking, so we may assume that $r \geq 2$. In addition, we may assume that a weakly blocking coalition induces a connected subgraph of G. In a weakly blocking coalition $C \subsetneq N$, some agent would have less than r friends, strictly decreasing her utility. Hence, the grand coalition is in the strict core. □

Albeit the previous proposition may look rather innocent, regular substructures in the loyalty graph have been very useful in dealing with core (non-)existence (see, e.g., the many cycles in the games of Proposition 8 and Theorem 9).

For symmetric FOHGs with a tree as loyalty graph, it is easy to see that a coalition structure is in the core if and only if its coalitions form an inclusion-maximal matching. In the case of ASHGs, we can apply a greedy matching algorithm to compute coalition structures in the core.

Proposition 11. *Let $k \geq 1$. A coalition structure in the core of the k-fold loyal variant can be computed in polynomial time for symmetric ASHGs with a tree as loyalty graph.*

On the negative side, even under the existence of core partitions, it may be hard to compute them. Interestingly, the next theorem does not cover the case $k = 1$.

Theorem 12. *Let $k \geq 2$. Computing a coalition structure in the core is NP-hard for the k-fold loyal variant of symmetric FOHGs with non-empty core.*

On the other hand, if the games originate from symmetric MFHGs, we obtain a polynomial-time algorithm by a modification of the algorithm in Theorem 7.

Theorem 13. *Let $k \geq 1$. Computing a coalition structure in the core can be done in polynomial time for the k-fold loyal variant of symmetric MFHGs.*

Symmetric k-fold loyal variant		Best Coalition	Core Solution
FOHGs	orig.	poly.	poly.+ [Dimitrov *et al.*, 2006]
	$k = 1$	NP-h.[Thm. 2]	open?
	$k \geq 2$	NP-h.[Thm. 2]	NP-h.? [Thm. 12]
	limit	NP-h.[Thm. 3]	NP-h.+ [Thm. 7]
ASHGs	orig.	poly.	NP-h.− [Aziz *et al.*, 2013]
	$k \geq 1$	NP-h.[Thm. 2]	NP-h.− [Thm. 9]
	limit	NP-h.[Thm. 3]	NP-h.+ [Thm. 7]
MFHGs	all	poly.[Thm. 4]	poly.+ [Thms. 7,13]

Table 1: Computational complexity of computing best coalitions and core partitions. The circled $+$, $-$, and ? indicate whether elements in the core always exist, may not exist, or whether this is unknown.

5 Conclusion and Open Problems

We have introduced loyalty in hedonic games as a possible way to integrate relationships of players in a coalition into the coalition formation process. Given a hedonic game, players can modify their utilities to obtain a new hedonic game which regards loyalty among coalition partners. Applying loyalty multiple times yields a sequence of hedonic games with increasing loyalty, eventually terminating in a hedonic game with utilities that represent a local form of egalitarianism. The limit game usually contains Pareto optimal coalition structures in the core, but their efficient computability is dependent on the initial input game. We show that computing best coalitions is hard if the input is an FOHG, a reduction that can also be applied to the computation of coalition structures in the core, revealing a close relationship of the two problems. An overview of our results is given in Table 1.

Our work offers plenty directions for further investigation. First, similarly to altruistic hedonic games, one can make the aggregation mechanism for loyal utilities dependent on a priority amongst the agents, or take averages instead of sums. This yields new notions of loyalty that are worth to investigate and compare. Second, it would be interesting to approach loyalty for other underlying classes of hedonic games such as fractional hedonic games. This includes also to find a reasonable way to define loyalty for purely ordinal input. Note that our (equivalent) definition of the loyalty set is also applicable in this case. Finally, an intriguing open problem concerns the existence of coalition structures in the core for loyal variants of FOHGs, in particular for the 1-fold variant, where we could not show hardness of the computational problem.

Acknowledgements

This work was supported by the German Research Foundation under grants BR 2312/12-1 and 277991500/GRK2201. We would like to thank Ina Seidel, Johannes Bullinger, Anna Maria Kerkmann, and Jörg Rothe for valuable discussions, and thank the anonymous reviewers.

References

[Apt and Schäfer, 2014] K. R. Apt and G. Schäfer. Selfishness level of strategic games. *Journal of Artificial Intelligence Research*, 49:207–240, 2014.

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

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

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

[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.

[Brânzei and Larson, 2011] S. Brânzei and K. Larson. Social distance games. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 273–279, 2011.

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

[Colman *et al.*, 2008] A. M. Colman, B. D. Pulford, and J. Rose. Collective rationality in interactive decisions: Evidence for team reasoning. *Acta psychologica*, 128(2):387–397, 2008.

[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.

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

[Edmonds, 1965] J. Edmonds. Paths, trees and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.

[Elias *et al.*, 2010] J. Elias, F. Martignon, K. Avrachenkov, and G. Neglia. Socially-aware network design games. In *Proceedings of the 29th IEEE Conference on Computer Communications (INFOCOM)*, pages 1–5. IEEE, 2010.

[Hardin, 1968] G. Hardin. The tragedy of the commons. *Science*, 1632:1243–1248, 1968.

[Kahneman and Tversky, 1979] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

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

[Kerkmann and Rothe, 2020] A. M. Kerkmann and J. Rothe. Altruism in coalition formation games. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 461–467, 2020.

[Koutsoupias and Papadimitriou, 1999] E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 404–413, 1999.

[Levine, 1998] D. K. Levine. Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, 1(3):593–622, 1998.

[Monaco *et al.*, 2018] G. Monaco, L. Moscardelli, and Y. Velaj. Hedonic games with social context. In *Proceedings of the 19th Italian Conference on Theoretical Computer Science*, pages 24–35, 2018.

[Mueller, 1986] D. C. Mueller. Rational egoism versus adaptive egoism as fundamental postulate for a descriptive theory of human behavior. *Public Choice*, 51(1):3–23, 1986.

[Nguyen *et al.*, 2016] N-T. Nguyen, A. Rey, L. Rey, J. Rothe, and L. Schend. Altruistic hedonic games. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 251–259, 2016.

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

[Schlueter and Goldsmith, 2020] J. Schlueter and J. Goldsmith. Super altruistic hedonic games. In *Proceedings of the 33rd International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, pages 160–165, 2020.

[von Neumann and Morgenstern, 1947] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 2nd edition, 1947.

[Wiechers and Rothe, 2020] A. Wiechers and J. Rothe. Stability in minimization-based altruistic hedonic games. In *Proceedings of the 9th European Starting AI Researcher Symposium (STAIRS)*, 2020.

12 CORE PUBLICATION 7: WELFARE GUARANTEES IN SCHELLING SEGREGATION

SUMMARY

The emergence of segregation is a long-term object of study in social sciences. About 50 years ago, Thomas Schelling proposed a simple, yet highly influential model that reveals how individual perceptions and incentives can lead to residential segregation. Even a small degree of homophily at the individual level can cause segregation at the global level. Schelling's model has inspired a constant stream of research during the last decades.

While the early research on Schelling's model mostly encompasses results obtained from simulations, there is a recent stream of work studying a game-theoretic variant of Schelling's model. In this variant, agents are assigned to the nodes of a topology graph, and act strategically to obtain a good position. We contribute to this research and study welfare guarantees and complexity with respect to several welfare measures.

First, we show that while maximizing the social welfare is NP-hard, computing an assignment of the agents to the nodes of any topology graph with approximately half of the maximum welfare can be done in polynomial time. We then consider Pareto optimality, and introduce two novel optimality notions that are a compromise between Pareto optimality and welfare maximization. We establish mostly tight bounds on the worst-case welfare loss for assignments satisfying any of the introduced optimality notions as well as the complexity of computing such assignments.

In addition, we study the problem of finding an assignment where every agent is the neighbor of at least one similar agent. This can be interpreted as a weak notion of homophily. We show that for tree topologies, it is possible to decide whether there exists an assignment that gives every agent a positive utility in polynomial time. Moreover, when every node in the topology graph has a degree of at least 2, such an assignment always exists and can be found efficiently.

REFERENCE

M. Bullinger, W. Suksompong, and A. A. Voudouris. **Welfare guarantees in Schelling segregation**. *Journal of Artificial Intelligence Research*, 71:143–174, 2021.
DOI: https://doi.org/10.1613/jair.1.12771

INDIVIDUAL CONTRIBUTION

I, Martin Bullinger, am the main author of this publication. In particular, I am responsible for the joint development and conceptual design of the research project, most of the proofs and their write-up (in particular, Theorems 3.2, 3.3, 4.9, and A.2, Propositions 6.4, A.3, and A.4, and all results in Section 5), the joint development of further results, and corrective changes in other parts of the paper.

COPYRIGHT AGREEMENT

The right to present this paper in a doctoral thesis has been granted by the publisher, AI Access Foundation, in the copyright form presented below. There, it is stated that "*The author retains the right to use this material in future works of his or her own authorship.*" This form can also be found at http://jair.org/public/resources/copyright.pdf (accessed November 10, 2022).

AI Access Foundation

Transfer of Copyright

Title of Contribution:

Author:

The undersigned (Author), desiring to publish the article named above, in the Journal of Artificial Intelligence Research, hereby grants and assigns exclusively to the AI Access Foundation (Publisher) all rights of copyright in this article, and the exclusive right to copy and distribute the article throughout the world, and the authority to exercise or to dispose of all subsidiary rights in all countries and in all languages.

The author retains the right to use this material in future works of his or her own authorship. The author retains the right to post an electronic copy of the article on his or her web site, on web sites operated by the author's institution, and on scholarly repositories, provided there is no commercial purpose involved, and with appropriate bibliographic citation to the JAIR article.

The Author agrees that the contribution does not contain any material created by others, or from other copyrighted works, unless the written consent of the owner of such material is attached.

The undersigned warrants that s/he has full power to make this agreement, and has not previously granted, assigned, or encumbered any of the rights granted and assigned herein; that the material submitted for publication is the work of the author, and is original, unpublished, and not previously the subject of any application for copyright registration (except material for which the Publisher has written grants of permission to include, as described above).

Author	Date

Journal of Artificial Intelligence Research 71 (2021) 143-174 Submitted 02/2021; published 05/2021

Welfare Guarantees in Schelling Segregation

Martin Bullinger MARTIN.BULLINGER@IN.TUM.DE
Institut für Informatik
Technische Universität München
Boltzmannstr. 3, 85748 Garching, Germany

Warut Suksompong WARUT@COMP.NUS.EDU.SG
School of Computing
National University of Singapore
13 Computing Drive, Singapore 117417, Singapore

Alexandros A. Voudouris ALEXANDROS.VOUDOURIS@ESSEX.AC.UK
School of Computer Science and Electronic Engineering
University of Essex
Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

Abstract

Schelling's model is an influential model that reveals how individual perceptions and incentives can lead to residential segregation. Inspired by a recent stream of work, we study welfare guarantees and complexity in this model with respect to several welfare measures. First, we show that while maximizing the social welfare is NP-hard, computing an assignment of agents to the nodes of any topology graph with approximately half of the maximum welfare can be done in polynomial time. We then consider Pareto optimality, introduce two new optimality notions based on it, and establish mostly tight bounds on the worst-case welfare loss for assignments satisfying these notions as well as the complexity of computing such assignments. In addition, we show that for tree topologies, it is possible to decide whether there exists an assignment that gives every agent a positive utility in polynomial time; moreover, when every node in the topology has degree at least 2, such an assignment always exists and can be found efficiently.

1. Introduction

Schelling's model was proposed half a century ago to illustrate how individual perceptions and incentives can lead to racial segregation, and has been used to study this phenomenon in residential metropolitan areas in particular (Schelling, 1969, 1971). The model is rather simple to describe. There are a number of agents, each of whom belongs to one of two predetermined types and occupies a location; in his original work, Schelling assumed that the locations are cells of a rectangular board, which can be represented as a grid graph. Every agent would like to occupy a node on the graph such that the fraction of other agents of the same type in the neighborhood of that node is at least a predefined tolerance threshold $\tau \in [0, 1]$. If this condition is not met for an agent, then the agent can relocate to a randomly chosen empty node on the grid. One of the most surprising findings of Schelling is that, starting from a random initial assignment of the agents to the nodes of the grid, the dynamics may converge to segregated assignments even when $\tau \approx 1/3$, contrasting the intuition that segregation should happen only when $\tau \geq 1/2$.

Throughout the years, hundreds of researchers in sociology and economics reconfirmed Schelling's observations and made similar ones for numerous variants of the model using computer simulations—see, for example, the work of Clark and Fossett (2008). More recent work, mainly in computer science, performed rigorous analyses of such variants, some of which are quite close to the original model, and showed that the dynamics according to which the agents relocate converges to assignments in which the agents form large monochromatic regions (that is, subgraphs consisting only of agents of the same type); in addition, this line of work established bounds on the size of these regions. We refer to the papers by Pollicott and Weiss (2001), Young (2001), Zhang (2004), Pancs and Vriend (2007), Brandt et al. (2012), Barmpalias et al. (2014, 2015), Bhakta et al. (2014), and Immorlica et al. (2017) for results of this flavor.

While most of the literature on Schelling's model has focused on properties related to segregation between the two types, segregation itself is only one side of the story, especially when we allow different, possibly more complex location graphs. Given that the agents are willing to relocate to be close to other agents of the same type, another natural question is whether the resulting assignments satisfy some sort of *efficiency*. This has been considered in part by a recent array of papers (Chauhan et al., 2018; Echzell et al., 2019; Elkind et al., 2019; Agarwal et al., 2020; Bilò et al., 2020; Chan et al., 2020; Kanellopoulos et al., 2020), which have studied Schelling's model from a game-theoretic perspective. In particular, instead of randomly relocating, the agents are assumed to be strategic and each of them aims to select a location that maximizes her *utility*, defined as the fraction of same-type agents in her neighborhood.

Besides questions related to the existence and computation of equilibria (i.e., assignments in which no agent has an incentive to relocate in order to increase her utility), the authors of some of the aforementioned papers have also studied the efficiency of assignments in terms of *social welfare*, defined as the total utility of the agents. For this objective, these authors have shown that computing assignments (not necessarily equilibria) maximizing the social welfare is NP-hard under specific assumptions about the graph and the behavior of the agents. Furthermore, they established several bounds on the worst-case ratio between the maximum social welfare (achieved by any possible assignment) and the social welfare of the best or worst equilbrium assignment, also known as the *price of stability* (Anshelevich et al., 2008) and the *price of anarchy* (Koutsoupias & Papadimitriou, 1999), respectively. These ratios quantify the welfare that is lost due to the agents aiming to maximize their individual utilities rather than their collective welfare.

Inspired by this active stream of work, we study welfare guarantees and complexity in Schelling's model, not only with respect to the social welfare, but also to different notions of efficiency, such as Pareto optimality and natural variants of it.

1.1 Our Contribution

Our setting consists of n agents partitioned into two types, and a location graph known as the *topology*; agents of the same type are "friends", and agents of different types are "enemies". Each agent is assigned to a single node of the graph, and the utility of the agent is defined as the fraction of her friends among the agents in her neighborhood.

Welfare notion P		Price of P Lower bound	Upper bound	Welfare guarantee
Maximum welfare		1	1	$\frac{n}{2}-1$ [Thm. 3.1]
GWO		$\Omega(n)$ [Prop. 4.6]	$O(n)$ [Thm. 4.8]	$\frac{n}{n-1}$ [Thm. 4.8]
UVO		$\Omega(n)$ [Thm. 4.7]	$O(n)$ [Thm. 4.7]	1 [Thm. A.1]
PO	general	$\Omega(n)$ [Prop. 4.6]	$O(n\sqrt{n})$ [Thm. 4.9]	$\frac{1}{\sqrt{n}}$ [Thm. 4.9]
	trees	$\Omega(n)$ [Cor. 4.11]	$O(n)$ [Cor. 4.11]	$\frac{n}{n-1}$ [Thm. 4.10]
	approx. balanced	$\Omega(n)$ [Cor. 4.13]	$O(n)$ [Cor. 4.13]	$\Omega(1)$ [Prop. 4.12]

Table 1: An overview of our results on the price and welfare guarantees of the optimality notions we consider. The 'approximately balanced' results for Pareto optimality hold when the numbers of agents of the two types are within a constant factor of each other (and the number of agents is equal to the number of nodes). Combining the lower and upper bounds, we obtain a price of $\Theta(n)$ for all welfare notions except for maximum welfare (whose price is trivially 1) and PO on general topologies.

We start by considering the social welfare. We show that for any topology and any distribution of the agents into types, there always exists an assignment with social welfare at least $n/2-1$, and we provide a polynomial-time algorithm for computing such an assignment. Since the social welfare never exceeds n, our algorithm produces an assignment with at least approximately half of the maximum social welfare. We complement this result by showing that maximizing the social welfare is NP-hard, even when the topology is a graph such that the number of nodes is equal to the number of agents. This improves upon previous hardness results of Elkind et al. (2019) and Agarwal et al. (2020) whose reductions use instances with "stubborn agents" (who are assigned to fixed nodes in advance and cannot move), and either a topology with the number of nodes larger than the number of agents, or at least three types of agents instead of just two. These results are presented in Section 3.

Even if an assignment does not maximize the social welfare, it can still be optimal in other senses. With this in mind, in Section 4, we turn our attention to different notions of optimality. In particular, we consider the well-known notion of *Pareto optimality* (PO), according to which it should not be possible to improve the utility of an agent without decreasing that of another agent. We also introduce two variants of PO, called *utility-vector optimality* (UVO) and *group-welfare optimality* (GWO), which are particularly appropriate for Schelling's model and may be of interest in other settings as well. Informally, an assignment is UVO if we cannot improve the sorted utility vector of the agents, and GWO if it is not possible to increase the total utility of one type of agents without decreasing that of the other type. We prove several results on these three notions of optimality. First, while UVO and GWO imply PO by definition, we show that they are not implied by each other or by PO. Then, for each $P \in \{\mathrm{PO}, \mathrm{UVO}, \mathrm{GWO}\}$, we establish mostly tight bounds on the *price of* P, which is an analogue of the price of anarchy: the price of P is defined as the worst-case ratio between the maximum social welfare (among all assignments) and the minimum social welfare among all assignments satisfying P.[1] Several of our results in Sections 3 and 4 are summarized in Table 1.

Next, in Section 5, we address the complexity of computing assignments satisfying different optimality notions. Our NP-hardness reduction for social welfare maximization in Section 3 also yields a corresponding hardness for GWO. We then consider *perfect* assignments, in which every agent receives the maximum utility of 1, and show that deciding whether such an assignment exists is NP-complete. As consequences, we obtain hardness results for computing a UVO or PO assignment, as well as for maximizing the *egalitarian* welfare (defined as the minimum utility among all agents) and the *Nash* welfare (defined as the product of the agents' utilities). When perfect assignments do not exist, a reasonable relaxation is to require that every agent receives the maximum utility that she can receive in any assignment for that instance; we call assignments satisfying this requirement *individually optimal*. While deciding whether an individually optimal assignment exists is again NP-complete in general, when the number of agents is equal to the number of nodes in the topology, we present a characterization of instances admitting such an assignment—this characterization allows us to solve the decision problem in polynomial time for the special case.

Finally, another important measure of efficiency is the number of agents who receive a positive utility in the assignment. Even though only requiring the utility to be nonzero seems minimal, there exist simple instances in which not all of the agents can obtain a positive utility simultaneously. We show that for trees, it is possible to decide in polynomial time whether there exists an assignment such that all agents receive a positive utility. We then observe that it is always possible to guarantee a positive utility for at least half of the agents. Moreover, when every node in the topology has degree at least 2, an assignment in which all agents receive a positive utility is guaranteed to exist, and such an assignment can be computed in polynomial time. These results are presented in Section 6.

1.2 Further Related Work

As already mentioned, Schelling's model and its variants have been studied extensively from many different perspectives in several disciplines. For an overview of early work on the model, we refer the reader to the work of Immorlica et al. (2017).

Most related to our present work are the papers by Elkind et al. (2019), Agarwal et al. (2020), Bilò et al. (2020), and Kanellopoulos et al. (2020), which studied game-theoretic and complexity questions related to the social welfare in Schelling games. In particular, Elkind et al. (2019) considered *jump Schelling games* in which there are $k \geq 2$ types of agents, and the topology is a graph with more nodes than agents so that there are empty nodes to which unhappy agents can jump. They showed that equilibrium assignments do not always exist, proved that computing equilibrium assignments and assignments with social welfare close to n (the maximum possible) is NP-hard, and bounded the price of anarchy and stability for both general and restricted games.

1. Note that an analogue of the price of *stability*, where we consider the worst-case ratio between the maximum social welfare and the *maximum* social welfare among assignments satisfying the optimality notion, is uninteresting: for all of the optimality notions we consider, this price is simply 1.

Later on, Agarwal et al. (2020) considered the complement case of *swap Schelling games* in which the number of nodes in the topology is equal to the number of agents; since there are no empty nodes to which the agents can jump, the agents can increase their utility only by swapping positions pairwise. For this setting, the authors showed results similar to those of Elkind et al. (2019). Bilò et al. (2020) improved some of the price of anarchy bounds of Agarwal et al. (2020), and also studied a variation of the model in which the agents have a restricted view of the topology and can only swap with their neighbors. Kanellopoulos et al. (2020) investigated the price of anarchy and stability in jump Schelling games, but with a slightly different utility function according to which an agent considers herself as part of her set of neighbors.

Schelling games are closely related to variants of hedonic games, most notably unweighted *fractional hedonic games* (Aziz et al., 2019), in which there is a set of agents and an unweighted graph that indicates friendship relations among them. In such games, the agents split into disjoint coalitions and, similarly to Schelling games, the utility of each agent is equal to the fraction of her friends in her coalition. The main difference between the two models is that, in Schelling games, the agents occupy the nodes of a topology graph, thereby leading to overlapping coalitions.

The price of Pareto optimality was first considered in the context of fractional hedonic games by Elkind et al. (2020), and was also implicitly studied by Bullinger (2020). Since Pareto optimality is a fundamental notion in various settings, its price has also been studied in the context of social distance games (Balliu et al., 2017) and fair division (Bei et al., 2019). To the best of our knowledge, this is the first time that Pareto optimality is studied in Schelling's model.

2. Preliminaries

Let $N = \{1, \dots, n\}$ be a set of $n \geq 2$ *agents*. The agents are partitioned into two different *types* (or *colors*), red and blue. Denote by r and b the number of red and blue agents, respectively; we have $r + b = n$. The distribution of agents into types is called *balanced* if $|r - b| \leq 1$. We say that two agents $i, j \in N$ such that $i \neq j$ are *friends* if i and j are of the same type; otherwise we say that they are *enemies*. For each $i \in N$, we denote the set of all friends of agent i by $F(i)$.

A *topology* is a simple connected undirected graph $G = (V, E)$, where $V = \{v_1, \dots, v_t\}$. Each agent in N has to select a node of this graph so that there are no collisions. A tuple $I = (N, G)$ is called a *Schelling instance*. Given a set of agents N and a topology $G = (V, E)$ with $|V| \geq n$, an *assignment* is an n-tuple $\mathbf{v} = (v(1), \dots, v(n)) \in V^n$ such that $v(i) \neq v(j)$ for all $i, j \in N$ with $i \neq j$; here, $v(i)$ is the node of the topology where agent i is positioned. A node $v \in V$ is *occupied* by agent i if $v = v(i)$. For a given assignment $\mathbf{v}$ and an agent $i \in N$, let $N_i(\mathbf{v}) = \{j \in N : \{v(i), v(j)\} \in E\}$ be the set of neighbors of agent i. Let $f_i(\mathbf{v}) = |N_i(\mathbf{v}) \cap F(i)|$ be the number of neighbors of i in $\mathbf{v}$ who are her friends. Similarly, let $e_i(\mathbf{v}) = |N_i(\mathbf{v})| - f_i(\mathbf{v})$ be the number of neighbors of i in $\mathbf{v}$ who are her enemies. Following prior work, we define the utility $u_i(\mathbf{v})$ of an agent i in $\mathbf{v}$ to be 0 if $|N_i(\mathbf{v})| = 0$; otherwise, her utility is defined as the fraction of her friends among the agents in her neighborhood:

$$u_i(\mathbf{v}) = \frac{f_i(\mathbf{v})}{|N_i(\mathbf{v})|} = \frac{f_i(\mathbf{v})}{f_i(\mathbf{v}) + e_i(\mathbf{v})}.$$

The *social welfare* of an assignment $\mathbf{v}$ is defined as the total utility of all agents:

$$\mathrm{SW}(\mathbf{v}) = \sum_{i \in N} u_i(\mathbf{v}).$$

Let $\mathbf{v}^*(I)$ be an assignment that maximizes the social welfare for a given instance I; we refer to it as a *maximum-welfare* assignment. Note that for any assignment $\mathbf{v}$, we have $u_i(\mathbf{v}) \leq 1$, and so $\mathrm{SW}(\mathbf{v}^*) \leq n$. Denote by $\mathrm{SW}_R(\mathbf{v})$ and $\mathrm{SW}_B(\mathbf{v})$ the sum of the utilities of the red and blue agents, respectively; we have $\mathrm{SW}_R(\mathbf{v}) + \mathrm{SW}_B(\mathbf{v}) = \mathrm{SW}(\mathbf{v})$.

3. Social Welfare

The first question we address is whether a high social welfare can always be achieved in any Schelling instance. Even though it may seem that we can obtain high welfare simply by grouping the agents of each type together, given the possibly complex topology in combination with the distribution of agents into types, it is unclear how this idea can be executed in general or what guarantee it results in. Nevertheless, we show that high welfare is indeed always achievable. Moreover, we provide a tight lower bound on the maximum welfare for each number of agents.

For any positive integer n, define

$$g(n) = \begin{cases} \frac{n(n-2)}{2(n-1)} & \text{if } n \text{ is even;} \\ \frac{n-1}{2} & \text{if } n \text{ is odd.} \end{cases}$$

Note that $g(n) \geq n/2 - 1$ for all n. Our approach is to choose an assignment uniformly at random among all possible assignments. Equivalently, we place agents in the following iterative manner: for an arbitrary unoccupied node, assign a uniformly random agent who is unassigned thus far. We show that the expected welfare of the assignment resulting from this simple randomized algorithm is at least $g(n)$, which implies the existence of an assignment with this welfare guarantee.

Theorem 3.1. *For any Schelling instance with n agents, there exists an assignment with social welfare at least $g(n)$. Moreover, the bound $g(n)$ cannot be improved.*

Proof. First, note that we may assume that the number of agents is equal to the number of nodes by restricting our attention to an arbitrary connected subgraph of G with the desired size. For $v_i \in V$, let $N_{v_i} = \{v_j \in V \mid \{v_i, v_j\} \in E\}$ be the neighborhood of node v_i in G, and $n_{v_i} = |N_{v_i}|$ be its size.

Consider an assignment of the agents to the nodes of G chosen uniformly at random. Let W be a random variable denoting the social welfare of this assignment, U_i a random variable denoting the expected utility of the agent placed at node v_i, and X_i a binary random variable describing the color of this agent, where $X_i = 1$ if node v_i is occupied by a blue agent and $X_i = 0$ if it is occupied by a red agent. By the linearity of expectation and the law of total expectation, we have

$$\begin{aligned}\mathbb{E}[W] &= \sum_{i=1}^{n} \mathbb{E}[U_i] \\ &= \sum_{i=1}^{n} (\Pr(X_i = 1) \cdot \mathbb{E}[U_i \mid X_i = 1] + \Pr(X_i = 0) \cdot \mathbb{E}[U_i \mid X_i = 0]).\end{aligned}$$

Now, for a fixed $v_i \in V$, it holds that

$$\begin{aligned}\mathbb{E}[U_i \mid X_i = 1] &= \frac{1}{n_{v_i}} \sum_{v_j \in N_{v_i}} \mathbb{E}[X_j \mid X_i = 1] \\ &= \frac{1}{n_{v_i}} \sum_{v_j \in N_{v_i}} \Pr(v_j \text{ blue} \mid v_i \text{ blue}) = \frac{1}{n_{v_i}} \sum_{v_j \in N_{v_i}} \frac{b-1}{n-1} = \frac{b-1}{n-1},\end{aligned}$$

where the first equality is again due to linearity of expectation. Similarly, we have $\mathbb{E}[U_i \mid X_i = 0] = \frac{r-1}{n-1}$. Hence,

$$\begin{aligned}\mathbb{E}[W] &= \sum_{i=1}^{n} \left(\frac{b}{n} \cdot \frac{b-1}{n-1} + \frac{r}{n} \cdot \frac{r-1}{n-1} \right) \\ &= b \cdot \frac{b-1}{n-1} + r \cdot \frac{r-1}{n-1} \\ &= \frac{1}{n-1}(b(b-1) + (n-b)(n-b-1)) = \frac{1}{n-1}(n^2 - n + 2b(b-n)).\end{aligned}$$

Observe that the function $b(b-n)$ is decreasing in the range $b \in [0, n/2]$ and increasing in the range $b \in [n/2, n]$. This means that for even n, we have

$$\mathbb{E}[W] \geq \frac{1}{n-1}\left(n^2 - n + 2 \cdot \frac{n}{2} \cdot \left(-\frac{n}{2}\right)\right) = \frac{n(n-2)}{2(n-1)} = g(n).$$

For n odd, since b is an integer, it holds that

$$\mathbb{E}[W] \geq \frac{1}{n-1}\left(n^2 - n + 2 \cdot \frac{n-1}{2} \cdot \left(-\frac{n+1}{2}\right)\right) = \frac{n-1}{2} = g(n),$$

implying that $\mathbb{E}[W] \geq g(n)$ in both cases. Hence, there exists an assignment with social welfare at least $g(n)$.

Finally, it can be verified that when G is a complete graph with n nodes and the distribution of agents into types is balanced, every assignment has social welfare exactly $g(n)$. □

Next, we derandomize the algorithm in Theorem 3.1 to produce an efficient deterministic algorithm that computes an assignment with welfare at least $g(n)$. The pseudocode of the algorithm, which shares the notation of Theorem 3.1, can be found in Algorithm 1. The main idea is that when we choose an agent to be assigned to an unassigned node, we pick a type such that the expected welfare is maximized, where the expectation is taken with respect to the uniform distribution of the remaining agents to the remaining nodes.

Algorithm 1 Assignment with high social welfare

Input: Schelling instance $I = (N, G)$ with $G = (V, E)$
Output: Assignment with social welfare at least $g(n)$

```
for i = 1, ..., n do
    if there is a unique assignment v consistent with X_1 = a_1, ..., X_{i-1} = a_{i-1} (up to
    permuting agents of the same color) then
        return v
    W_0 = E[W | X_1 = a_1, ..., X_{i-1} = a_{i-1}, X_i = 0]
    W_1 = E[W | X_1 = a_1, ..., X_{i-1} = a_{i-1}, X_i = 1]
    if W_1 ≥ W_0 then
        a_i = 1 /*assign a blue agent to v_i*/
    else
        a_i = 0 /*assign a red agent to v_i*/
return Assignment corresponding to (a_1, ..., a_n)
```

Theorem 3.2. *Algorithm 1 returns an assignment with social welfare at least $g(n)$ in polynomial time.*

Proof. We use the same notation as in the proof of Theorem 3.1.

First, we prove that the welfare of the returned assignment is at least $g(n)$. For $i = 0, \dots, n$, denote by A_i the event $X_1 = a_1 \wedge X_2 = a_2 \wedge \cdots \wedge X_i = a_i$. In particular, A_0 is the entire sample space. We will show by induction that for each i, $\mathbb{E}[W \mid A_i] \geq \mathbb{E}[W]$. The base case $i = 0$ holds trivially. For $i \in \{1, \dots, n\}$, if there is a unique assignment consistent with $X_1 = a_1 \wedge \cdots \wedge X_{i-1} = a_{i-1}$, then the social welfare of the returned assignment is $\mathbb{E}[W \mid A_{i-1}] \geq \mathbb{E}[W] \geq g(n)$, where the first inequality follows from the induction hypothesis and the second inequality from Theorem 3.1. Otherwise, we have

$$\begin{aligned}
\mathbb{E}[W] &\leq \mathbb{E}[W \mid A_{i-1}] \\
&= \Pr(X_i = 0 \mid A_{i-1}) \cdot \mathbb{E}[W \mid A_{i-1} \wedge X_i = 0] + \Pr(X_i = 1 \mid A_{i-1}) \cdot \mathbb{E}[W \mid A_{i-1} \wedge X_i = 1] \\
&\leq \Pr(X_i = 0 \mid A_{i-1}) \cdot \mathbb{E}[W \mid A_i] + \Pr(X_i = 1 \mid A_{i-1}) \cdot \mathbb{E}[W \mid A_i] \\
&= \mathbb{E}[W \mid A_i],
\end{aligned}$$

where we use the law of total expectation for the first equality and the choice of a_i in the algorithm for the second inequality. This completes the induction. Hence, if the algorithm terminates in the jth iteration, the welfare of the returned assignment is $\mathbb{E}[W \mid A_j] \geq \mathbb{E}[W] \geq g(n)$.

We next show that the algorithm can be implemented in polynomial time. To this end, it suffices to show that the quantities W_0 and W_1 can be computed efficiently for each fixed $i \in \{1, \dots, n\}$. If there is only one type of agents left after having assigned the first i agents, this is straightforward, so assume that both types of agents still remain. By the linearity of expectation, for each $x \in \{0, 1\}$,

$$\mathbb{E}[W \mid A_{i-1} \wedge X_i = x] = \sum_{j=1}^{n} \mathbb{E}[U_j \mid A_{i-1} \wedge X_i = x].$$

By the law of total expectation,

$$\mathbb{E}[U_j \mid A_{i-1} \wedge X_i = x] = \Pr(X_j = 0 \mid A_{i-1} \wedge X_i = x) \cdot \mathbb{E}[U_j \mid A_{i-1} \wedge X_i = x \wedge X_j = 0] + \Pr(X_j = 1 \mid A_{i-1} \wedge X_i = x) \cdot \mathbb{E}[U_j \mid A_{i-1} \wedge X_i = x \wedge X_j = 1],$$

where a probability can be 0 if v_j has already been assigned an agent (i.e., if $j \le i$). When $j > i$, we have

$$\Pr(X_j = 1 \mid A_{i-1} \wedge X_i = x) = \frac{b - \sum_{k=1}^{i-1} a_k - x}{n - i}.$$

Also, by the linearity of expectation,

$$\mathbb{E}[U_j \mid A_{i-1} \wedge X_i = x \wedge X_j = 1] = \frac{1}{n_{v_j}} \sum_{v_k \in N_{v_j}} \mathbb{E}[X_k \mid A_{i-1} \wedge X_i = x \wedge X_j = 1].$$

Finally,

$$\mathbb{E}[X_k \mid A_{i-1} \wedge X_i = x \wedge X_j = 1] = \begin{cases} a_k & \text{if } k \le i-1; \\ x & \text{if } k = i; \\ \frac{b - \sum_{\ell=1}^{i-1} a_\ell - x - 1}{n-i-1} & \text{if } k > i. \end{cases}$$

The computations for $X_j = 0$ as well as for $j \le i$ can be done similarly. □

Since the social welfare of any assignment is at most n, Algorithm 1 always produces an assignment with at least roughly half of the optimal welfare. This raises the question of whether it is possible to compute a maximum-welfare assignment for any given instance in polynomial time. Unfortunately, Elkind et al. (2019) proved that maximizing the social welfare is NP-hard. However, their proof relies on the existence of a "stubborn agent", who is assigned to a fixed node in advance and cannot move, and uses a topology with more nodes than agents.[2] We show that the hardness remains even when both of these assumptions are removed and the topology is a regular graph, i.e., a graph in which all nodes have the same degree.

Theorem 3.3. *The following problem is NP-complete: Given a Schelling instance and a rational number s, decide whether there exists an assignment with social welfare at least s. The hardness holds even for the class of instances where the number of agents is equal to the number of nodes and the topology is a regular graph.*

Proof. The problem belongs to NP since computing the social welfare of a given assignment can be done efficiently. For the hardness, we reduce from the Maximum Clique problem for regular graphs, i.e., given a regular graph G and an integer k, is there a clique of size at least k? Note that this problem is NP-hard: Indeed, the Independent Set problem is NP-hard for regular graphs (Garey & Johnson, 1979, pp. 194–195), and a set of vertices

2. Agarwal et al. (2020) showed that the hardness holds when the numbers of agents and nodes are equal, but still required stubborn agents and moreover assumed at least three types of agents.

forms a clique in a given graph exactly when these vertices form an independent set in the complement graph.[3]

Let (G, k) be an instance of Maximum Clique, where $G = (V, E)$ is a ρ-regular graph on n vertices, and k an integer. Define a Schelling instance on topology G with k red and $n-k$ blue agents. For any assignment $\mathbf{v}$, the social welfare $\mathrm{SW}(\mathbf{v})$ is equal to $n - 2\delta(\mathbf{v})/\rho$, where $\delta(\mathbf{v})$ denotes the number of edges connecting a red agent and a blue agent in $\mathbf{v}$. If G has a clique of size k, then by assigning all red agents to nodes in this clique, we have $\delta(\mathbf{v}) = k\rho - 2\binom{k}{2}$; indeed, the sum of degrees of the red agents is $k\rho$, from which we have to subtract twice the number of red-red edges. Similarly, if G does not have a clique of size k, then for any assignment $\mathbf{v}$, we have $\delta(\mathbf{v}) > k\rho - 2\binom{k}{2}$. Hence, there exists an assignment with welfare at least $n - 2k + 4\binom{k}{2}/\rho$ if and only if G has a clique of size k, so we may set $s = n - 2k + 4\binom{k}{2}/\rho$ to complete the reduction. $\square$

Note that Theorem 3.3 also yields the hardness of computing a maximum-welfare assignment. Indeed, any algorithm that computes such an assignment can also be used to decide whether there exists an assignment with a certain social welfare.

4. Optimality Notions

Even when an assignment does not achieve maximum social welfare, there can still be other ways in which it is "optimal". In this section, we consider some optimality notions and quantify them in relation to social welfare. We begin with a classic notion, Pareto optimality.

Definition 4.1. An assignment $\mathbf{v}$ is said to be *Pareto dominated* by an assignment $\mathbf{v}'$ if $u_i(\mathbf{v}) \leq u_i(\mathbf{v}')$ for all $i \in N$, with the inequality being strict for at least one agent. An assignment $\mathbf{v}$ is *Pareto optimal* (PO) if it is not Pareto dominated by any other assignment.

Given two vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ of the same length k, we say that $\mathbf{w}_1$ *weakly dominates* $\mathbf{w}_2$ if for each $i \in \{1, \ldots, k\}$, the ith element of $\mathbf{w}_1$ is at least that of $\mathbf{w}_2$. We say that $\mathbf{w}_1$ *strictly dominates* $\mathbf{w}_2$ if at least one of the inequalities is strict.

For an assignment $\mathbf{v}$, denote by $\mathbf{u}(\mathbf{v})$ the vector of length n consisting of the agents' utilities $u_i(\mathbf{v})$, sorted in non-increasing order. Similarly, denote by $\mathbf{u}_R(\mathbf{v})$ and $\mathbf{u}_B(\mathbf{v})$ the corresponding sorted vectors of length r and b for the red and blue agents, respectively. Note that an assignment $\mathbf{v}$ is Pareto optimal if and only if there is no other assignment $\mathbf{v}'$ such that $\mathbf{u}_X(\mathbf{v}')$ weakly dominates $\mathbf{u}_X(\mathbf{v})$ for $X \in \{R, B\}$ and at least one of the dominations is strict. Motivated by this observation, we define two new optimality notions appropriate for Schelling instances.

Definition 4.2. An assignment $\mathbf{v}$ is said to be

- *group-welfare dominated* by an assignment $\mathbf{v}'$ if $\mathrm{SW}_X(\mathbf{v}') \geq \mathrm{SW}_X(\mathbf{v})$ for $X \in \{R, B\}$ and at least one of the inequalities is strict;

3. Given a graph $G = (V, E)$, its *complement graph* is the graph $\overline{G} = (V, \overline{E})$ with $\overline{E} = \{e \subseteq V : |e| = 2, e \notin E\}$, i.e., there is an edge between vertices $v_1, v_2 \in V$ in $\overline{G}$ exactly when there is no edge between them in G.

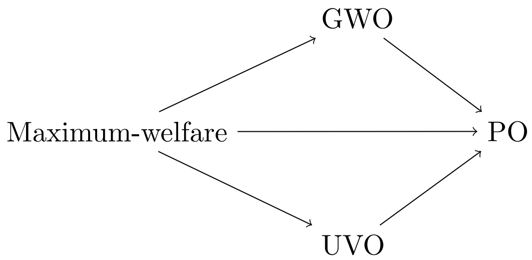


Figure 1: Implication relations among optimality notions.

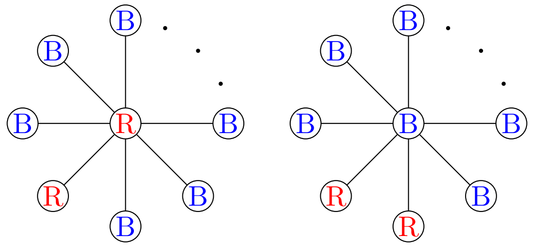


Figure 2: Example showing that GWO does not imply UVO.

- *utility-vector dominated* by an assignment $\mathbf{v}'$ if $\mathbf{u}(\mathbf{v}')$ strictly dominates $\mathbf{u}(\mathbf{v})$.

An assignment $\mathbf{v}$ is *group-welfare optimal* (GWO) if it is not group-welfare dominated by any other assignment. Similarly, an assignment $\mathbf{v}$ is *utility-vector optimal* (UVO) if it is not utility-vector dominated by any other assignment.

The implication relations in Figure 1 follow immediately from the definitions; in particular, both of the new notions lie between welfare maximality and Pareto optimality. We claim that no other implications exist between these notions. To establish this claim, it suffices to show that GWO and UVO do not imply each other.

Proposition 4.3. *GWO does not imply UVO.*

Proof. Assume that the topology is a star as in Figure 2, and there are two red and $n-2$ blue agents, where $n \geq 5$. The left assignment $\mathbf{v}$ is GWO, since putting a blue agent at the center as in the right assignment $\mathbf{v}'$ leaves both red agents with utility 0. However, $\mathbf{v}$ is not UVO, as

$$\mathbf{u}(\mathbf{v}) = (1, 1/(n-1), 0, \dots, 0)$$

is strictly dominated by

$$\mathbf{u}(\mathbf{v}') = (1, \dots, 1, (n-3)/(n-1), 0, 0).$$ □

Proposition 4.4. *UVO does not imply GWO.*

Proof. Let n be a multiple of 4. Suppose that the topology is a complete bipartite graph with $n/2$ nodes on each side, and there are $n/2$ red and $n/2$ blue agents (Figure 3). The left

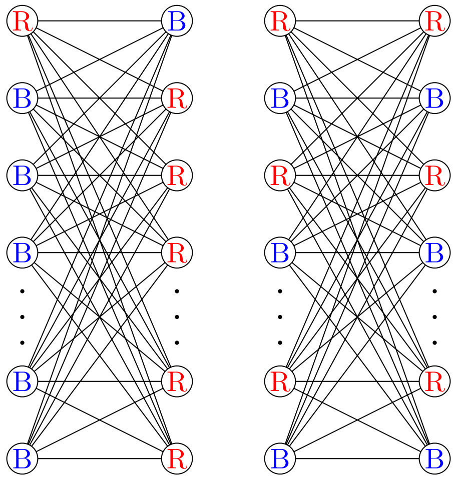


Figure 3: Example showing that UVO does not imply GWO. The topology is a complete bipartite graph.

assignment $\mathbf{v}$, which assigns one red agent to the left side and one blue agent to the right side, is UVO. Indeed, the red agent assigned to the left side receives utility $(n/2-1)/(n/2)$, and any assignment in which an agent receives equal or higher utility must have the same sorted utility vector as $\mathbf{v}$. We have

$$\mathrm{SW}(\mathbf{v}) = 2 \cdot \frac{n/2-1}{n/2} + 2(n/2-1) \cdot \frac{1}{n/2} = 4 - \frac{8}{n},$$

with each group receiving half of the welfare, i.e., $2 - 4/n$. On the other hand, in the right assignment $\mathbf{v}'$, which assigns half of the agents of each color to each side, every agent receives utility $1/2$. Hence $\mathrm{SW}(\mathbf{v}') = n/2$, and each group receives a total utility of $n/4$. It follows that when $n \geq 8$, $\mathbf{v}$ is UVO but not GWO. □

In order to quantify the welfare guarantee that each optimality notion provides, we define the price of a notion as follows.

Definition 4.5. Given a property P of assignments and a Schelling instance, the *price of* P for that instance is defined as the ratio between the maximum social welfare (of any assignment) and the minimum social welfare of an assignment satisfying P:

$$\text{Price of } P \text{ for instance } I = \frac{\mathrm{SW}(\mathbf{v}^*(I))}{\min_{\mathbf{v} \in P(I)} \mathrm{SW}(\mathbf{v})},$$

where $P(I)$ is the set of all assignments satisfying P in instance I.[4] The *price of* P for a class of instances is then defined as the supremum price of P over all instances in that class.

For $P \in \{\text{PO, GWO, UVO}\}$, we have $\mathbf{v}^*(I) \in P(I)$, so the price of P is always well-defined and at least 1. Note also that $\max_{\mathbf{v} \in P(I)} \mathrm{SW}(\mathbf{v}) = \mathrm{SW}(\mathbf{v}^*(I))$.

4. We interpret the ratio $\frac{0}{0}$ in this context to be equal to 1.

In Figure 2, the left assignment is GWO and PO and has social welfare $n/(n-1)$, whereas the maximum-welfare assignment on the right has social welfare $n(n-3)/(n-1)$. We therefore have the following bound (for $n \leq 4$, the bound holds trivially).

Proposition 4.6. *For every n, both the price of GWO and the price of PO are at least $n-3$.*

The following result shows that the welfare of a UVO assignment can also be a linear factor away from the maximum welfare, but not more.

Theorem 4.7. *The price of UVO is $\Theta(n)$.*

Proof. We prove the lower and the upper bound separately.

Lower bound: Consider the topology in Figure 3. As in the proof of Proposition 4.4, the left assignment $\mathbf{v}$ is UVO and has social welfare $4-8/n$. On the other hand, the right assignment $\mathbf{v}'$ has social welfare $n/2$, meaning that the ratio $\mathrm{SW}(\mathbf{v}')/\mathrm{SW}(\mathbf{v})$ is greater than $n/8$.

Upper bound: We claim that if $n \geq 3$, any UVO assignment has social welfare at least[5] $1/2$; since the maximum social welfare is at most n, this yields the desired bound.

Assume first that the number of agents is equal to the number of nodes. Let $\mathbf{v}$ be a UVO assignment. If there is a red agent and a blue agent both receiving utility 0, then since no node is empty and $n \geq 3$, swapping them yields an improvement with respect to the utility vector. So we may assume that all agents of one type, say blue, receive a positive utility. If at least $n/2$ agents receive a positive utility, then $\mathrm{SW}(\mathbf{v}) \geq n/(2n-2) > n/(2n) = 1/2$. Assume therefore that more than $n/2$ agents receive utility 0; these agents must all be red. Swap b of these red agents receiving utility 0 with all b blue agents to obtain an assignment $\mathbf{v}'$. Notice that the utility in $\mathbf{v}'$ of each of these b red agents is at least as high as the utility of the blue agent in $\mathbf{v}$ with whom she was swapped, while all blue agents receive utility 0 in $\mathbf{v}'$. In addition, every other (red) agent is not worse off, and at least one of them is better off (in particular, one who receives utility 0 in $\mathbf{v}$, which must exist since $n/2 > b$). Hence $\mathbf{v}$ is utility-vector dominated by $\mathbf{v}'$, a contradiction.

Now, assume that the number of agents is less than the number of nodes. Since $n \geq 3$, any UVO assignment $\mathbf{v}$ must have $\mathrm{SW}(\mathbf{v}) > 0$, so there exists a connected component (of the topology restricted to the nodes occupied according to $\mathbf{v}$) with a positive social welfare. Let n' be the size of this component. If $n' = 2$, then $\mathrm{SW}(\mathbf{v}) \geq 2$. Else, the assignment restricted to this component is also UVO, and by our earlier arguments has social welfare at least $1/2$. □

Next, we show that the price of GWO is also $\Theta(n)$. The lower bound follows from Proposition 4.6, while the upper bound follows from the fact that the social welfare never exceeds n along with the following theorem, which establishes a lower bound of (at least) 1 on the social welfare of GWO assignments.

Theorem 4.8. *Any GWO assignment has social welfare at least $n/(n-1)$ for $n \geq 4$, and 1 for $n = 3$. Moreover, these bounds cannot be improved.*

5. In Theorem A.1, we improve this bound to 1 via a longer proof.

Proof. To see that the bounds cannot be improved, consider the left assignment in Figure 2 for $n \geq 4$, and a triangle topology with two red and one blue agents for $n = 3$.

Assume first that the number of agents is equal to the number of nodes. The case $n = 3$ can be verified directly, since the only two possible topologies are a triangle and a path. Let $n \geq 4$, and assume for contradiction that there exists a GWO assignment $\mathbf{v}$ with social welfare less than $n/(n-1)$. Since the least possible positive utility of an agent is $1/(n-1)$, this means that some agent receives utility 0. Take such an agent i, and assume without loss of generality that i is red. Since the numbers of agents and nodes are equal, i is connected to a set $A \neq \emptyset$ of blue agents. Consider the following cases.

<u>Case 1</u>: There is a blue agent j outside A. We swap i and j to obtain an assignment $\mathbf{v}'$. After the swap, j has utility 1, and each blue agent in A has utility at least $1/(n-1)$, so $\mathrm{SW}_B(\mathbf{v}') \geq n/(n-1) > \mathrm{SW}(\mathbf{v}) \geq \mathrm{SW}_B(\mathbf{v})$. In addition, no red agent receives a lower utility in $\mathbf{v}'$ than in $\mathbf{v}$. Hence $\mathbf{v}$ is not GWO, a contradiction.

<u>Case 2</u>: There are no blue agents outside A, but at least one red agent besides i. Since no node is empty, there is a blue agent $j \in A$ who is adjacent to another red agent. We swap i and j to obtain an assignment $\mathbf{v}'$. No agent receives a lower utility in $\mathbf{v}'$ than in $\mathbf{v}$. Moreover, i's utility strictly increases. Hence $\mathbf{v}$ is not GWO, a contradiction.

<u>Case 3</u>: There are no blue agents outside A, and no red agent besides i. This means that i receives utility 0 in any assignment. Consider an assignment $\mathbf{v}'$ such that the blue agents form a connected component. Each blue agent receives utility at least $1/2$, so $\mathrm{SW}_B(\mathbf{v}') \geq b/2 \geq 3/2 > n/(n-1)$. It follows that $\mathbf{v}$ is group-welfare dominated by $\mathbf{v}'$, a contradiction.

Now, assume that the number of agents is less than the number of nodes. Since $n \geq 3$, any GWO assignment $\mathbf{v}$ must have $\mathrm{SW}(\mathbf{v}) > 0$, so there exists a connected component (of the topology restricted to the nodes occupied according to $\mathbf{v}$) with a positive social welfare. Since the assignment restricted to this component is GWO, we are done if the component is of size at least $n' \geq 4$, because then $\mathrm{SW}(\mathbf{v}) \geq n'/(n'-1) \geq n/(n-1)$. If there is a component of size 2 with positive welfare, the social welfare is at least 2 and we are also done. Since any component of positive welfare has a welfare of at least 1, we are again done if there are at least two components of positive welfare. Thus, we can assume that there is a single component of positive welfare of size 3, which we can moreover assume has the topology of a triangle (the only other possibility being a path, which guarantees a welfare of $3/2$), and that there are exactly two agents of one type, say blue, and one agent of the other type in this component.

If there is another blue agent outside this component, $\mathbf{v}$ is group-welfare dominated by the assignment that swaps the red agent of the triangle and this blue agent. Finally, consider the case where there is another red agent and no blue agent outside the triangle. The red agent must have an empty neighboring node. We obtain a group-welfare improvement by moving the red agent of the triangle to this empty node. Hence, the remaining case is that all agents are part of the triangle, meaning that $n = 3$; in this case, the social welfare is 1, as desired. □

We now turn to Pareto optimality, for which we prove a weaker lower bound on the social welfare.

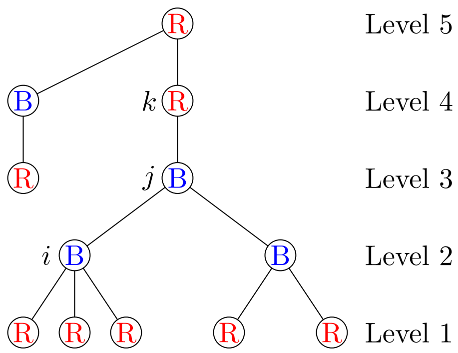


Figure 4: Illustration for the proof of Theorem 4.10.

Theorem 4.9. *When $n \geq 3$, any PO assignment has social welfare at least $1/\sqrt{n}$.*

Proof. By an argument similar to that in the upper bound part of Theorem 4.7, and since the function $1/\sqrt{n}$ is decreasing, it suffices to consider the case where the number of agents is equal to the number of nodes. Let $\mathbf{v}$ be a PO assignment. If at least $\sqrt{n}$ agents receive a positive utility, then $\mathrm{SW}(\mathbf{v}) \geq \sqrt{n}/(n-1) \geq 1/\sqrt{n}$, so assume that fewer than $\sqrt{n}$ agents receive a positive utility. Similarly, we may assume that every agent receives utility less than $1/\sqrt{n}$. If there is a red agent and a blue agent both receiving utility 0, then since no node is empty and $n \geq 3$, swapping them yields a Pareto improvement. So we may assume that all agents of one type, say blue, receive a positive utility. This means in particular that $b < \sqrt{n}$, and more than $n - \sqrt{n}$ red agents only have blue neighbors. Hence, there exists a blue agent i with at least $(n - \sqrt{n})/\sqrt{n} = \sqrt{n} - 1 \geq b - 1$ red neighbors. Let A be a set containing $b-1$ of these red neighbors.

Swap the $b-1$ blue agents other than i with the red agents in A to obtain an assignment $\mathbf{v}'$. Since these red agents are not adjacent to any red agent in $\mathbf{v}$, no red agent is worse off in $\mathbf{v}'$. Each of the $b-1$ blue agents receives utility at least $1/b > 1/\sqrt{n}$ in $\mathbf{v}'$, so all of them are strictly better off. Furthermore, i is adjacent to all of these $b-1$ blue agents in $\mathbf{v}'$, and therefore cannot be worse off. Hence $\mathbf{v}$ is not PO, a contradiction. $\square$

Combined with Proposition 4.6, Theorem 4.9 implies that when $n \geq 3$, the price of PO is at least $n-3$ and at most $n\sqrt{n}$. We conjecture that the welfare guarantee in Theorem 4.9 can be improved to $n/(n-1)$ for $n \geq 4$, which would be tight due to the left assignment in Figure 2. Next, we confirm this conjecture when the topology is a tree.

Theorem 4.10. *When $n \geq 3$ and the topology is a tree, any PO assignment has social welfare at least $n/(n-1)$. Moreover, this bound cannot be improved.*

Proof. The bound cannot be improved due to the left assignment in Figure 2. To establish the bound, note first that by an argument similar to that in the upper bound part of Theorem 4.7, and since the function $n/(n-1)$ is decreasing, it suffices to consider the case where the number of agents is equal to the number of nodes. Let $\mathbf{v}$ be a PO assignment. As in the proof of Theorem 4.9, we may assume that all agents of one type, say blue, receive a positive utility. If an agent occupying a leaf node is adjacent to another agent of the same type, then $\mathrm{SW}(\mathbf{v}) \geq 1 + 1/(n-1)$ and we are done. Hence, assume that all leaf nodes are occupied by red agents receiving utility 0.

Root the tree at an arbitrary node. The deepest level of the tree consists only of leaf nodes; call this "level 1", and call the other levels accordingly (see Figure 4). Consider an arbitrary parent i of a leaf at level 1—this parent must be blue. Since all blue agents receive a positive utility, i must have a blue parent j. Any branch originating from j has to go down to level 1; otherwise, the branch stops at level 2 with a red leaf, and we can swap this red leaf with i to obtain a Pareto improvement. In particular, in j's subtree, all nodes on level 2 are blue. If j does not have a parent or has a blue parent, then she has utility 1 and $\mathrm{SW}(\mathbf{v}) \geq n/(n-1)$. Assume therefore that j has a red parent k. Note that j receives utility at least $1/2$, and any of its child at least $1/(n-1)$.

Suppose that k has another child. If k has a red child, this child cannot be a leaf (because all leaves have utility 0), so it must in turn have a blue child (otherwise it receives utility 1 and we are done), but then this blue child receives utility 0, a contradiction. Hence k can only have blue children. Suppose that k has a blue child $\ell \neq j$, which cannot be a leaf. If ℓ only has blue children, it receives utility at least $1/2$, and we are done since $\mathrm{SW}(\mathbf{v}) \geq 1/2+1/(n-1)+1/2 = n/(n-1)$. So assume that ℓ has a red child m, which must also be a leaf. Since all blue agents receive a positive utility, ℓ must also have a blue child o, which must in turn have only red children. Now, swapping m and o leads to a Pareto improvement, a contradiction. Hence we may assume that k's only child is j.

Finally, if k does not have a parent or has a blue parent, swapping k with i yields a Pareto improvement. So assume that k has a red parent. This means that k receives utility $1/2$. Combining this with the utility of j and her children, we again have $\mathrm{SW}(\mathbf{v}) \geq n/(n-1)$. □

Together with Proposition 4.6, which holds for trees, Theorem 4.10 gives a tight bound on the price of PO for trees.

Corollary 4.11. *When the topology is a tree, the price of PO is* $\Theta(n)$.

To finish this section, we show that if $b/r \in \Theta(1)$, i.e., the fraction of agents of each type is at least a certain constant, then a constant welfare can again be guaranteed.

Proposition 4.12. *Suppose that the number of agents is equal to the number of nodes. Then any PO assignment has social welfare at least* $\min\left\{\frac{b}{r+1}, \frac{r}{b+1}\right\}$.

Proof. Let $\mathbf{v}$ be a PO assignment. Since the number of agents is equal to the number of nodes, as in the proof of Theorem 4.9, we may assume that all agents of one type, say blue, receive a positive utility. Since a blue agent is adjacent to at least one other blue agent and at most r red agents, each blue agent receives utility at least $1/(r+1)$. This means that $\mathrm{SW}(\mathbf{v}) \geq b/(r+1)$. Similarly, if all red agents receive a positive utility, then $\mathrm{SW}(\mathbf{v}) \geq r/(b+1)$. The conclusion follows. □

When the ratio b/r is upper and lower bounded by constants, the welfare guarantee provided by Proposition 4.12 is also constant, which is at most a linear factor away from the maximum welfare. Since Figure 3 shows an instance with $b = r$ and a PO assignment whose welfare is a linear factor away from the maximum welfare, we obtain the following:

Corollary 4.13. *When* $b/r \in \Theta(1)$ *and the number of agents is equal to the number of nodes, the price of PO is* $\Theta(n)$.

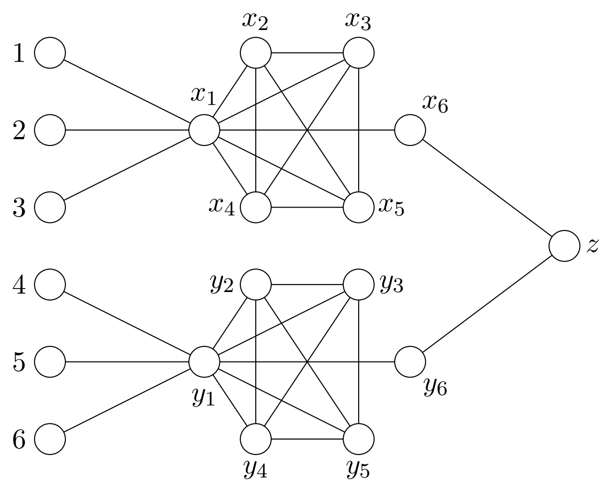


Figure 5: Illustration for the proof of Theorem 5.3 when $R = \{1, 2, 3, 4, 5, 6\}$ and $S = \{x, y\}$ where $x = \{1, 2, 3\}$ and $y = \{4, 5, 6\}$.

5. Computing Optimal Assignments

In the previous section, we introduced two new concepts of optimality and studied the welfare guarantees that they provide along with Pareto optimality. In this section, we continue our investigation of these optimality notions by examining the complexity of computing assignments satisfying them. Furthermore, we consider other common welfare notions such as egalitarian welfare and Nash welfare.

First, we observe that in the reduced instances of Theorem 3.3, an assignment is GWO if and only if it maximizes the social welfare.[6] This immediately yields the following intractability.

Theorem 5.1. *Computing a GWO assignment is NP-hard, even for the class of Schelling instances where the number of agents is equal to the number of nodes and the topology is a regular graph.*

For the other two optimality notions, we will establish a close relationship to the problem of computing a "perfect assignment", wherein every agent receives utility 1.

Definition 5.2. An assignment $\mathbf{v}$ is called *perfect* if $u_i(\mathbf{v}) = 1$ for all $i \in N$.

We start by showing the hardness of this problem.

Theorem 5.3. *Deciding whether there exists a perfect assignment is NP-complete.*

Proof. Membership in NP is clear: a perfect assignment can be verified in polynomial time.

The hardness reduction is from Exact 3-Cover (X3C). An instance of X3C consists of a tuple (R, S), where R is a ground set whose size is divisible by 3, and S is a collection

6. To see this, note that for any assignment $\mathbf{v}$, the sum of utilities of the red and blue agents is equal to $\mathrm{SW}_R(\mathbf{v}) = r - \delta(\mathbf{v})/\rho$ and $\mathrm{SW}_B(\mathbf{v}) = b - \delta(\mathbf{v})/\rho$, respectively, where $\delta(\mathbf{v})$ denotes the number of edges connecting a red agent and a blue agent in $\mathbf{v}$. Hence, both the GWO and maximum-welfare assignments are precisely the assignments minimizing the number of these interconnection edges.

of 3-element subsets of R, where $|S| \geq |R|/3$. A Yes-instance is an instance in which there exists a subcollection $S' \subseteq S$ of size $|R|/3$ that exactly partitions R. It is well-known that X3C is NP-hard (Garey & Johnson, 1979, p. 221).

Let an instance (R, S) of X3C be given. We define a Schelling instance with $1 + |S| - |R|/3$ blue and $|R| + 2|S|^2 + |R|/3$ red agents. Define the topology graph $G = (V, E)$ with $V = R \cup \{s_i \colon 1 \leq i \leq 2|S| + 2, s \in S\} \cup \{z\}$ and edges given by

- $\{r, s_1\} \in E$ if $r \in R$ and $s \in S$ with $r \in s$;
- $\{s_i, s_j\} \in E$ for $s \in S$, $1 \leq i, j \leq 2|S| + 1$;
- $\{s_1, s_{2|S|+2}\}, \{s_{2|S|+2}, z\} \in E$ for $s \in S$; and
- no further edges are in E.

See Figure 5 for an illustration. Note that the number of nodes is $|R| + 2|S|^2 + 2|S| + 1$ and the number of agents is $|R| + 2|S|^2 + |S| + 1$, so exactly $|S|$ nodes are left empty in any assignment. We claim that (R, S) is a Yes-instance if and only if there exists a perfect assignment in the Schelling instance.

Assume first that (R, S) is a Yes-instance, and let $S' \subseteq S$ be a partition of R using sets in S. We assign the blue agents to the nodes in the set $A = \{z\} \cup \{s_{2|S|+2} \colon s \in S \setminus S'\}$, and the red agents to the vertices in the set $B = R \cup \{s_i \colon 2 \leq i \leq 2|S|+1, s \in S\} \cup \{s_1 \colon s \in S'\}$. Note that $|A| = 1 + |S| - |R|/3$, $B = |R| + 2|S|^2 + |R|/3$, and the nodes in A induce a connected subgraph of G that has no neighbor in B. This means that all blue agents receive utility 1. Since S' covers R, all red agents also receive utility 1, meaning that the assignment is perfect.

Conversely, assume that there is a perfect assignment. Since every assignment leaves exactly $|S|$ nodes empty, no blue agent can be assigned to a vertex in $\{s_i \colon 1 \leq i \leq 2|S|+1, s \in S\}$, because she would then have a red neighbor. Additionally, no blue agent can be assigned to a vertex in R, because then some of her only neighbors in the set $\{s_1 \colon s \in S\}$ would also have to be blue, which is impossible by the previous sentence. We conclude that the blue agents are assigned to the nodes in $\{z\} \cup \{s_{2|S|+2} \colon s \in S\}$, which means in particular that some blue agent is assigned to z. Define $S' = \{s \in S \colon s_1 \text{ is red}\}$. Then, the empty nodes are precisely $\{s_{2|S|+2} \colon s \in S'\} \cup \{s_1 \colon s \in S \setminus S'\}$, and we have $|S'| = |R|/3$. In particular, all nodes in R must have red agents. Now, such an agent can receive a positive utility only if at least one of her neighbors is red. Hence, S' covers R. Since $|S'| = |R|/3$, it follows that S' forms a partition of R. □

Theorem 5.3 turns out to be particularly useful for deriving hardness results with respect to other optimality and welfare notions.

Corollary 5.4. *Computing a UVO assignment (resp., PO assignment) is NP-hard.*

Proof. Observe that if there exists a perfect assignment in an instance, then every UVO (resp., PO) assignment in that instance must be perfect. Hence, an algorithm for computing a UVO (resp., PO) assignment can be used to decide whether a perfect assignment exists. The conclusion follows from Theorem 5.3. □

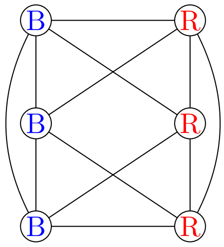


Figure 6: Example of an individually optimal assignment.

In addition, we obtain the hardness of computing assignments with maximum egalitarian or Nash welfare. Recall that the *egalitarian welfare* of an assignment is the minimum among the agents' utilities in that assignment, and the *Nash welfare* is the product of the agents' utilities in that assignment.

Corollary 5.5. *The following problem is NP-complete: Given a Schelling instance and a rational number s, decide whether there exists an assignment with egalitarian (resp., Nash) welfare at least s.*

Proof. Membership in NP is trivial. For the hardness, observe that an assignment is perfect if and only if its egalitarian (resp., Nash) welfare is (at least) 1, so the conclusion follows from Theorem 5.3. □

We emphasize that it is crucial for Theorem 5.3 and its corollaries that the number of nodes in the topology is larger than the number of agents. If the two numbers are equal, then perfect assignments do not exist (unless there is only one type of agents), and the corresponding decision problem becomes trivial. Nevertheless, it remains interesting to ask for assignments that are "individually optimal" for all agents.

Definition 5.6. An assignment $\mathbf{v}$ is called *individually optimal for agent i* if $u_i(\mathbf{v}) \geq u_i(\mathbf{v}')$ for all assignments $\mathbf{v}'$. An assignment is called *individually optimal* if it is individually optimal for all agents.

An example of an individually optimal assignment is shown in Figure 6, where every agent receives a utility of $1/2$.

Note that it is easy to compute the utility that an agent receives in an individually optimal assignment for her: Assign this agent to a node of minimum degree; then, assign agents of the same type to as many neighbors as possible and leave other neighbors empty; finally, if needed, assign agents from the other type to the remaining neighbors. An individually optimal assignment for all agents, if it exists, is clearly optimal with respect to all welfare measures that we consider and treats agents of the same type equally.

In the reduction of Theorem 5.3, we can assume that there are at least three agents of each type, and some nodes in the topology have[7] degree 2. Hence, an assignment is

7. More precisely, we can assume without loss of generality that $|S| \geq 2 + |R|/3$ in an instance (R, S) of X3C, since the problem can be solved by brute force otherwise. Then, there are at least three blue agents. Also, there are at least three red agents whenever $|R| \geq 3$. Finally, nodes of the type $s_{2|S|+2}$ have degree 2.

individually optimal if and only if every agent receives utility 1. Consequently, perfect assignments and individually optimal assignments coincide, and we obtain the following:

Corollary 5.7. *Deciding whether there exists an individually optimal assignment is NP-complete.*

However, we show next that if the numbers of agents and nodes are equal, then this decision problem becomes efficiently solvable. More precisely, we provide simple conditions characterizing the instances with an individually optimal assignment. Recall the definition of a complement graph from Footnote 3.

Theorem 5.8. *Given a Schelling instance with an equal number of agents and nodes, there exists an individually optimal assignment if and only if all of the following three conditions are met:*

(1) The numbers of red and blue agents are the same, or the topology graph is a complete graph.

(2) The topology graph is regular.

(3) The complement graph of the topology graph is bipartite.

Deciding whether the three conditions are met can be done in polynomial time. Moreover, if all three conditions are met, we can compute an individually optimal assignment in polynomial time.

Proof. Consider a Schelling instance with an equal number of agents and nodes and a topology graph $G = (V, E)$. Denote the complement graph by $\overline{G}$. Note that if G is a complete graph, then all assignments are individually optimal and conditions (2) and (3) are trivially met. Therefore, assume from now on that G is not a complete graph.

First, suppose that there exists an individually optimal assignment, and consider such an assignment. Recall that there are b blue and r red agents. Let δ denote the minimum degree among the nodes of G. If $b \geq \delta + 1$ or $r \geq \delta + 1$, then the individually optimal utility for one type of agents is 1, and there does not exist an individually optimal assignment. Hence, this cannot be the case, and the individually optimal utility is $(b-1)/\delta$ and $(r-1)/\delta$ for the blue and red agents, respectively. In particular, the topology graph must be δ-regular; otherwise, an agent assigned to a node with degree greater than δ cannot receive her individually optimal utility. Condition (2) is therefore met.

Let B and R be the sets of nodes occupied by the blue and red agents, respectively. For the assignment to be individually optimal, B and R must form cliques in the topology graph. In other words, (B, R) forms a bipartition of the complement graph, meaning that condition (3) is met.

The regularity of the topology graph implies the regularity of its complement graph, which is not the empty graph (because we already excluded a complete topology graph). Since (B, R) forms a bipartition of the complement graph which is $(n-1-\delta)$-regular, the number of edges adjacent to exactly one node in B is $(n-1-\delta)b$, the number of edges adjacent to exactly one node in R is $(n-1-\delta)r$, and these two numbers must be equal. It follows that $b = r$, so condition (1) is met as well.

Conversely, assume that the three conditions are met. Since we have already handled the case of a complete topology graph, we have that the numbers of blue and red agents are the same.

Let (X, Y) be a bipartition of the node set V in $\overline{G}$. Since G is regular, so is $\overline{G}$. Again, by counting the (nonzero) number of edges incident to nodes in X and Y, regularity implies that $|X| = |Y|$. Now, consider the assignment that assigns all blue agents at the nodes in X and all red agents at the nodes in Y; this assignment is feasible because the numbers of blue and red agents are the same. Since (X, Y) forms a bipartition in the complement graph, the set of blue agents as well as the set of red agents form cliques in the topology graph. Regularity therefore implies that every agent is individually optimal, so the assignment is individually optimal.

The proof of the converse also shows that we can compute an individually optimal assignment in polynomial time by computing the complement graph and an arbitrary bipartition. This yields an individually optimal assignment, provided that all three conditions are met. It is clear that checking whether the three conditions are met can also be done in polynomial time. □

The key algorithmic problem in the proof of Theorem 5.8 is to decide whether the nodes of a given regular graph can be partitioned into two equally-sized subsets in such a way that each subset forms a clique. This problem is related to some NP-hard problems, and its tractability may therefore be of broader interest. Indeed, it appears similar not only to the Maximum Clique problem, which we have seen to be NP-hard for regular graphs (cf. Theorem 3.3), but also to the Minimum Bisection problem, which is likewise NP-hard for regular graphs (Bui et al., 1987). The latter problem asks for a partition of the nodes of a given graph into two equally-sized subsets such that the number of edges between these two sets is minimized. Our proof of Theorem 5.8 shows that the variant where we ask for the two subsets of nodes to form cliques is solvable in polynomial time.

6. Number of Positive Agents

In this section, we consider the problem of maximizing the number of agents receiving a positive utility, who we refer to as *positive agents*. This problem is closely related to egalitarian and Nash welfare, because an assignment has nonzero egalitarian (resp., Nash) welfare if and only if it makes every agent positive. Notice that it is not always possible to make every agent positive—for example, in a star, every agent whose type is different from the center agent receives zero utility. We begin by showing that for trees, deciding whether it is possible to make every agent positive can be done efficiently. Our algorithm is based on dynamic programming and shares some similarities with the algorithm of Elkind et al. (2019) for deciding whether an equilibrium exists on a tree.

Theorem 6.1. *There is a polynomial-time algorithm that decides whether there exists an assignment in which every agent receives a positive utility when the topology is a tree.*

Proof. Pick an arbitrary node v_{root} to be the root of G. For each node $v \in V$, let *tree*(v) be the set of descendants of v, including v itself. For each v, we fill out a table τ_v, which contains an entry $\tau_v(C, n_B, n_R, n_E, q)$ for each tuple (C, n_B, n_R, n_E, q), where

- $C \in \{\textit{blue}, \textit{red}, \textit{empty}\}$;
- $n_B, n_R, n_E \in \{0, 1, \ldots, n\}$;
- $q \in \{\textit{yes}, \textit{no}\}$.

The number of entries in each table is $6(n+1)^3$. The value of each entry is either *true* or *false*. Specifically, $\tau_v(C, n_B, n_R, n_E, q) = \textit{true}$ if and only if there exists an assignment of a subset of agents to the nodes in *tree*(v) satisfying the following conditions:

1. If $C = \textit{empty}$, then node v is empty; otherwise, it is assigned to an agent of color C.
2. There are n_B blue agents, n_R red agents, and n_E empty nodes in *tree*(v).
3. If $C \in \{\textit{blue}, \textit{red}\}$, then $q = \textit{yes}$ if and only if the agent in node v has at least one child of the same color.
4. Every node in *tree*(v) different from v has at least one neighbor of the same color.

An assignment in which every agent receives a positive utility exists if and only if in the table $\tau_{v_{\text{root}}}$ of the root node v_{root}, there exists (C, n_B, n_R, n_E, q) such that $\tau_{v_{\text{root}}}(C, n_B, n_R, n_E, q) = \textit{true}$, $n_B = b$, $n_R = r$, and if $C \in \{\textit{blue}, \textit{red}\}$ then $q = \textit{yes}$.

The tables for the leaf nodes can be filled in trivially. We now show how to fill the table τ_v of each $v \in V$ given the tables of its children. If v has L children $w_1, \ldots, w_L$, we construct intermediate tables $\theta_v^0, \theta_v^1, \ldots, \theta_v^L$. Each table θ_v^i takes parameters $(n_B, n_R, n_E, q_B, q_R, \widehat{q_B}, \widehat{q_R})$. The entry of the table is set to *true* if it is possible to place agents in the first i subtrees so that the following conditions hold: there are a total of n_B blue agents, n_R red agents, and n_E empty nodes; q_B (resp., q_R) indicates whether there is at least one blue (resp., red) agent among the first i children of v; $\widehat{q_B}$ (resp., $\widehat{q_R}$) indicates whether there is at least one blue (resp., red) agent among the first i children of v who does not have a blue (resp., red) child. The table θ_v^0 can be filled in trivially: only the entry $\theta_v^0(0, 0, 0, \textit{no}, \textit{no}, \textit{no}, \textit{no})$ is set to *true*. By combining the tables θ_v^{i-1} and τ_{w_i}, we can fill in the table θ_v^i in polynomial time. Specifically, we set the entry $\theta_v^i(n_B, n_R, n_E, q_B, q_R, \widehat{q_B}, \widehat{q_R})$ to *true* if and only if there exist entries $\theta_v^{i-1}(n'_B, n'_R, n'_E, q'_B, q'_R, \widehat{q_B}', \widehat{q_R}')$ and $\tau_{w_i}(C'', n''_B, n''_R, n''_E, q'')$ such that both are *true* and the following three conditions hold:

1. $n'_B + n''_B = n_B$, $n'_R + n''_R = n_R$, and $n'_E + n''_E = n_E$.
2. $q_B = \textit{yes}$ if and only if $q'_B = \textit{yes}$ or $C'' = \textit{blue}$. Analogously, $q_R = \textit{yes}$ if and only if $q'_R = \textit{yes}$ or $C'' = \textit{red}$.
3. $\widehat{q_B} = \textit{yes}$ if and only if (i) $\widehat{q_B}' = \textit{yes}$ or (ii) $C'' = \textit{blue}$ and $q'' = \textit{no}$. Analogously, $\widehat{q_R} = \textit{yes}$ if and only if (i) $\widehat{q_R}' = \textit{yes}$ or (ii) $C'' = \textit{red}$ and $q'' = \textit{no}$.

The table θ_v^L can then be used to fill in τ_v in polynomial time. Specifically, we set the entry $\tau_v(C, n_B, n_R, n_E, q)$ to *true* if and only if there exists an entry $\theta_v^L(n'_B, n'_R, n'_E, q'_B, q'_R, \widehat{q_B}', \widehat{q_R}')$ which has been set to *true* and such that the following conditions hold:

1. If $C = \textit{blue}$, then $n_B = n'_B + 1$, $n_R = n'_R$, and $n_E = n'_E$. Else, if $C = \textit{red}$, then $n_B = n'_B$, $n_R = n'_R + 1$, and $n_E = n'_E$. Finally, if $C = \textit{empty}$, then $n_B = n'_B$, $n_R = n'_R$, and $n_E = n'_E + 1$.

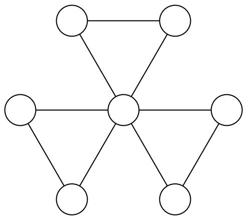


Figure 7: Example showing that Theorem 6.3 does not hold when the number of nodes is greater than the number of agents. There are three red and three blue agents. No matter how the agents are placed, at least one of them will receive utility 0.

2. If $C = blue$, then $q = q'_B$. Else, if $C = red$, then $q = q'_R$.

3. If $\widehat{q_B}' = yes$, then $C = blue$. Similarly, if $\widehat{q_R}' = yes$, then $C = red$.

This concludes the proof. □

Observe that for any topology, an assignment in which at least half of the agents are positive is guaranteed to exist and can be easily found by using depth-first search for the majority type.

Proposition 6.2. *For any $n \geq 3$, there exists a polynomial-time algorithm that computes an assignment in which at least $\lceil n/2 \rceil$ agents receive a positive utility.*

Proof. Assume without loss of generality that there are at least as many red as blue agents, so there are at least $\lceil n/2 \rceil \geq 2$ red agents. Starting from an arbitrary node of the topology, we first assign the red agents to nodes as we perform a depth-first search, and then assign the blue agents to any subset of the remaining nodes. Since the topology is connected, every red agent will have at least one red neighbor, meaning that at least $\lceil n/2 \rceil$ agents receive a positive utility. □

The bound $\lceil n/2 \rceil$ is tight when the topology is a star and there are $\lceil n/2 \rceil$ red and $\lfloor n/2 \rfloor$ blue agents.

Next, we show that when every node has degree at least 2 and the number of agents is equal to the number of nodes, it is possible to give every agent a positive utility. Note that the latter condition is also necessary—for the topology given in Figure 7, if there are three red and three blue agents (so one node is left unoccupied), it is easy to see that no assignment makes every agent positive.

Theorem 6.3. *Suppose that every node in the topology has degree at least 2, the number of agents is equal to the number of nodes, and there are at least two agents of each type. Then there exists an assignment such that every agent receives a positive utility.*

Proof. Consider an arbitrary assignment $\mathbf{v}$. If every agent is already positive, we are done, so assume that there is an agent i with utility 0. Without loss of generality, i is a blue agent. Among all paths from i to another blue agent, consider one with maximum length—suppose that the path goes to agent j. Since there are at least two blue agents, such a path must exist; moreover, since i has utility 0, the path contains at least one red agent. Let k be the last red agent on the path before reaching j. Swap i and k.

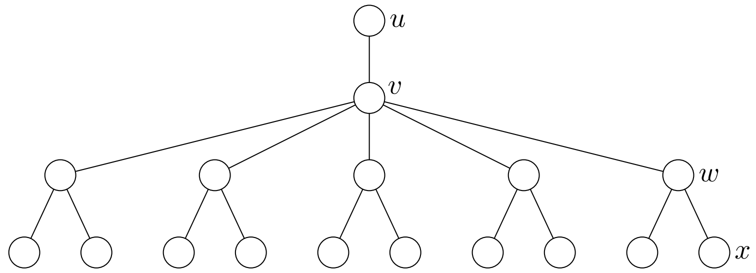


Figure 8: Illustration for the proof of Proposition 6.4.

We claim that in the resulting assignment $\mathbf{v}'$, the number of positive agents increases by at least 1; by applying such swaps repeatedly, we will reach an assignment in which all agents are positive. To establish the claim, it suffices to show that i, k, as well as any agent adjacent to either of them are positive in $\mathbf{v}'$. Since i has utility 0 in $\mathbf{v}$, she has at least two red neighbors in $\mathbf{v}$, so k is positive in $\mathbf{v}'$. Moreover, i is adjacent to j in $\mathbf{v}'$ and therefore becomes positive. Any other red agent on the path remains positive, and all agents adjacent to k in $\mathbf{v}'$ are red (besides possibly i, if k is the only red agent on the path) and are therefore positive. Finally, consider any red agent ℓ adjacent to i in $\mathbf{v}'$ not lying on the path. Since every node has degree at least 2, agent ℓ must have a neighbor $m \neq i$ (possibly $m = j$). If ℓ is adjacent to k in $\mathbf{v}'$, then ℓ is positive since k is red. Else, if m is a blue agent, we obtain a longer path from i to m in $\mathbf{v}$ than the original longest path, a contradiction. Hence m must be red, and ℓ is positive in $\mathbf{v}'$, proving the claim. □

Since the longest path problem is known to be NP-hard (Garey & Johnson, 1979, p. 213), the proof of Theorem 6.3 does not give rise to a polynomial-time algorithm for computing a desired assignment. In Theorem A.2, we present an inductive approach that is more involved but leads to an efficient algorithm.

For our final result of this section, we show that maximizing the social welfare and maximizing the number of positive agents can be conflicting goals. Recall that an assignment has nonzero egalitarian welfare if and only if it makes every agent positive. To avoid confusion, we will refer to our main notion of social welfare (i.e., the sum of the agents' utilities) as *utilitarian welfare.*

Proposition 6.4. *There exists a Schelling instance in which the maximum egalitarian welfare is nonzero but the egalitarian welfare of every assignment that maximizes the utilitarian welfare is zero.*

Proof. Consider the topology in Figure 8, and assume that there are 2 blue and 15 red agents (so no node can be left empty). The only way to achieve nonzero egalitarian welfare is to assign the blue agents to u and v, yielding utilitarian welfare 14.5. However, assigning the blue agents to w and x results in a higher utilitarian welfare of $91/6 \approx 15.17$. □

The utilitarian welfare gap in the example in Proposition 6.4 is rather small. However, one can easily modify the example by adding more three-node gadgets as subtrees of the node v to obtain instances wherein the utilitarian welfare of the unique assignment yielding a positive egalitarian welfare is an additive factor of $\Theta(n)$ lower than the maximum possible utilitarian welfare. Moreover, even assignments that maximize the utilitarian welfare can differ significantly in terms of egalitarian welfare. In Propositions A.3 and A.4, we present examples illustrating the possibility that one such assignment has a positive egalitarian welfare while another one has zero, or that two such assignments have positive egalitarian welfare differing by a multiplicative factor of $\Theta(n)$.

7. Conclusion and Future Work

In this paper, we have studied questions regarding welfare guarantees and complexity in Schelling segregation. Several of our findings are positive: An assignment with high social welfare always exists and can be found efficiently, and the welfare of assignments satisfying most optimality notions are at most a linear factor away from the maximum social welfare in the worst case. Furthermore, even though an assignment yielding a positive utility to every agent may not exist, the existence can be guaranteed when every node in the topology has degree at least 2, a realistic assumption in well-connected metropolitan areas. By contrast, computing an assignment that maximizes the social welfare or satisfies any of the optimality notions is NP-hard, and assignments maximizing the (utilitarian) social welfare can differ significantly in terms of egalitarian welfare.

A number of interesting directions remain from our work. On the technical side, it would be useful to close the gap on the price of Pareto optimality, which we conjecture to be $\Theta(n)$, as well as to characterize the topologies for which an assignment such that every agent receives a positive utility always exists. Another question is whether we can obtain in polynomial time a better approximation of social welfare than the factor of 2 in Theorem 3.2, or whether there is in fact an inapproximability result. From a more conceptual perspective, one could try to extend our results to a model with more than two types of agents or more complex friendship relations (e.g., friendship relations defined by a social network, Elkind et al., 2019) or modified utility functions (Kanellopoulos et al., 2020). Questions concerning the convergence behavior in best-response dynamics also remain open: do such dynamics always converge to an optimal assignment? Finally, studying our new optimality notions from Section 4 in related settings such as hedonic games, especially when agents are partitioned into types, or optimality notions derived from other known welfare measures, may lead to intriguing discoveries as well.

Acknowledgments

This work has been supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/12-1, by the European Research Council (ERC) under grant number 639945 (ACCORD), and by an NUS Start-up Grant. A preliminary version of the paper appeared in Proceedings of the 35th AAAI Conference on Artificial Intelligence (Bullinger et al., 2021). We would like to thank Jiarui Gan, Pascal Lenzner, and Louise Molitor for interesting discussions, and the anonymous reviewers of AAAI 2021 and JAIR for helpful comments.

Appendix A. Additional Results

In the upper bound part of Theorem 4.7, we showed that any UVO assignment has social welfare at least $1/2$. Here we establish an improved bound of 1, albeit with a much longer proof.

Theorem A.1. *When $n \geq 3$, any UVO assignment has social welfare at least* 1.

Proof. By an argument similar to that in the upper bound part of Theorem 4.7, it suffices to consider the case where the number of agents is equal to the number of nodes. Let $\mathbf{v}$ be a UVO assignment. If all agents receive nonzero utility, then since the least possible positive utility is $1/(n-1)$, the social welfare would be at least $n/(n-1) > 1$. Hence we may assume without loss of generality that there is a red agent with utility 0, and that all blue agents receive a positive utility.

Decompose the topology into maximal monochromatic components. We claim that there cannot be a set of red components and a set of blue components such that the two sets have the same total number of nodes and together they do not cover all nodes; call this claim (*). To see why (*) is true, assume for contradiction that two such sets exist. We swap all agents in the set of red components with those in the set of blue components to obtain an assignment $\mathbf{v}'$. The utility of each swapped red agent in $\mathbf{v}'$ is at least that of the blue agent whom she replaces in $\mathbf{v}$; an analogous statement holds for the swapped blue agents in $\mathbf{v}'$. Moreover, an unswapped agent who is adjacent to at least one swapped agent receives a strictly higher utility in $\mathbf{v}'$ than in $\mathbf{v}$. Hence $\mathbf{v}$ is not UVO, a contradiction that establishes (*).

Next, assume for contradiction that $\mathrm{SW}(\mathbf{v}) < 1$, and let s be the number of maximal monochromatic components in the topology. Call an edge "monochromatic" if it connects two agents of the same color. Since a component with t nodes has at least $t-1$ edges, the total number of monochromatic edges is at least $n-s$. A monochromatic edge generates a utility of at least $1/(n-1)$ for each of the two agents adjacent to it, so the generated utility is at least $2/(n-1)$ for each edge. Since $\mathrm{SW}(\mathbf{v}) < 1$, this means that the number of monochromatic edges is less than $(n-1)/2$, and therefore at most $(n-2)/2$. Thus, we have $n - s \leq (n-2)/2$, which implies that $n \leq 2s - 2$.

Let $k \geq 1$ be the number of red components of size 1, i.e., the number of red agents with utility 0. By (*), the size of any blue component is at least k. Suppose the smallest blue component has size at least $k+1$. Then, considering the k singleton red components and the smallest blue component, there are $k+1$ components with a total of at least $2k+1$ nodes. Letting n_0 and s_0 denote the number of nodes and components considered so far, we have $n_0 > 2s_0 - 2$. Since every other component has size at least 2, we also have $n > 2s-2$, a contradiction with $n \leq 2s-2$. So the smallest blue component must have size k. If there is any other component, we are again done by (*). Hence, the topology consists exactly of k red components of size 1 and one blue component of size k. In particular, $k \geq 2$.

We now show that the blue component of size k is a star, i.e., all but one blue agents have exactly one blue neighbor. Assume that at least two blue agents have more than one blue neighbor each. Each of these two agents receives utility at least $\frac{2}{k+2}$, while every other blue agent receives at least $\frac{1}{k+1}$. Hence $\mathrm{SW}(\mathbf{v}) \geq 2 \cdot \frac{2}{k+2} + (k-2) \cdot \frac{1}{k+1} = \frac{k(k+4)}{(k+1)(k+2)}$, which is at least 1 because $k \geq 2$. So the blue component is a star, and $\mathrm{SW}(\mathbf{v}) \geq \frac{k-1}{2k-1} + (k-1) \cdot \frac{1}{k+1} = \frac{3k(k-1)}{(k+1)(2k-1)}$, which is at least 1 whenever $k \geq 4$.

It remains to consider the cases $k = 2$ and $k = 3$. Since $\mathrm{SW}(\mathbf{v}) < 1$, each blue agent must have at least one red neighbor. Moreover, between the two blue agents with one blue neighbor, at least one must have more than one red neighbor. Let i be such a blue agent, j be her blue neighbor, ℓ be a red neighbor of j, and $m \neq \ell$ be a red neighbor of i. We swap i and ℓ to obtain an assignment $\mathbf{v}'$. From $\mathbf{v}$ to $\mathbf{v}'$, j receives the same utility, m receives a strictly higher utility, i receives a higher utility than ℓ's utility in $\mathbf{v}$ (which is 0), and ℓ receives at least as much utility as i does in $\mathbf{v}$. It follows that $\mathbf{v}$ is not UVO, a final contradiction which completes the proof. □

Next, we present an efficient algorithm for computing an assignment that gives every agent a positive utility when every node has degree at least 2; a shorter proof that such an assignment exists can be found in Theorem 6.3.

Theorem A.2. *Suppose that every node in the topology has degree at least* 2*, the number of agents is equal to the number of nodes, and there are at least two agents of each type. Then, it is possible to compute an assignment in which every agent receives a positive utility in polynomial time.*

Proof. We present an inductive approach which inherently gives rise to a polynomial-time algorithm. First, if there is an edge connecting two nodes of degree at least 3, deleting it still leaves a topology in which every node has degree at least 2. The topology may stay connected or break into two connected components. We first show how to deal with a connected topology in which no edge connects two nodes of degree at least 3, and specify later how to proceed when the topology breaks into two components upon the removal of an edge. Assume that we have a connected topology such that no edge connects two nodes of degree at least 3. If the topology is a cycle, a desired assignment can be easily found, so assume otherwise.

Call the nodes of degree at least 3 "primary nodes", and the remaining nodes (which have degree 2) "secondary nodes". Note that no edge connects two primary nodes and there exists at least one primary node due to our previous assumptions. In addition, each secondary node belongs either to a path connecting two primary nodes or to a cycle going from a primary node back to itself. We prove our claim for the class of graphs satisfying these conditions, but more generally without the assumption that primary nodes have degree at least 3 (so they may have degree 1 or 2). Nevertheless, the fact that the primary nodes under our original assumptions can be easily identified by their degree will be useful for constructing an efficient algorithm. From this point on, primary nodes will remain primary throughout the procedure regardless of their degree.

Consider the "meta-graph" H whose nodes correspond to the primary nodes of our topology G, where two nodes in H are connected by an edge if and only if there exists a path connecting the two corresponding primary nodes not going through any other primary node in G. Since G is connected, so is H (note that H may consist of a single node). Let v be a primary node in G such that removing the corresponding node in H leaves H connected—for example, any leaf in a spanning tree of H satisfies this property.

We will color certain nodes in G blue in the following order. Start with a cycle connecting v back to itself (if there is any), and color nodes excluding v in sequence starting from a node adjacent to v; then move on to the next cycle if there is another left. The only exception is if the total amount of nodes to be colored blue only allows us to color one more node when we are about to start coloring a new cycle—in this case, we color v blue instead of the first node of a new cycle. Whenever we run out of nodes to be colored blue, we color all of the remaining nodes red. If we have colored all nodes in cycles adjacent to v and there are still nodes left to be colored blue, we color v blue, followed by secondary nodes on paths adjacent to v (call this set of secondary nodes A); for each such path, we color nodes closer to v before those further away from v. If we run out of blue nodes, color all remaining nodes red. In case we have also colored all nodes in A and removed them, the remaining topology is again connected due to our choice of v. If we have only one node left to be colored blue, color one of the primary nodes adjacent to a secondary node in A blue, and all of the remaining nodes red; any remaining primary node is still adjacent to at least one secondary node. Else, there are at least two nodes left to be colored blue, and we apply induction on the remaining topology. Note that if we only have one primary node v left, G is a union of cycles that only intersect each other at v, and the same procedure still applies.

We now show how to proceed when, upon removing an edge connecting two nodes of degree at least 3, the topology breaks into two connected components C_1 and C_2. We try to allocate an appropriate number of red and blue agents to each component, and recurse on the two smaller problems. Assume without loss of generality that $|C_1| \leq |C_2|$ and $r \leq b$. Since every node in the remaining (disconnected) topology still has degree at least 2, we must have $|C_1| \geq 3$. Assume first that $|C_1| \geq 4$. If $r \geq 4$, we allocate two red and two blue agents to each component, and the remaining agents arbitrarily so that the number of agents is equal to the number of nodes in each subproblem—this ensures that the subproblems satisfy the conditions of the original problem. If $r = 2$, or if $r = 3$ and $|C_2| \geq 5$, we simply allocate all red agents to C_2. The remaining case is that $r = 3$, $|C_1| = |C_2| = 4$, and $b = 5$. In this case, suppose that the removed edge is adjacent to node x in C_1. We assign the blue agents to $C_2 \cup \{x\}$, and the red agents to $C_1 \setminus \{x\}$. Since each node in C_1 is adjacent to at least one other node in C_1 besides x, all agents are positive.

Consider now the case where $|C_1| = 3$. If one of the types has three or at least five agents, we may allocate three agents of that type to C_1. Hence, assume that both types consist of either two or four agents. The only two possibilities are then $(|C_1|, |C_2|, r, b) = (3, 3, 2, 4)$ and $(3, 5, 4, 4)$. Both cases can be handled similarly to the case $(4, 4, 3, 5)$ in the previous paragraph.

Since each step of our procedure removes at least one node or edge, by following the procedure, we obtain a polynomial-time algorithm that computes a desired assignment. □

In fact, the algorithm in Theorem A.2 runs in time linear in the size of the input. Indeed, finding a spanning tree of H can be done by breadth-first search, and all other steps of the algorithm take time $O(m)$, where m denotes the number of edges in G.

Finally, we continue our discussion after Proposition 6.4 by presenting further examples relating egalitarian and utilitarian welfare.

Proposition A.3. *There exists a Schelling instance in which one assignment maximizing utilitarian welfare also maximizes egalitarian welfare, while another such assignment has zero egalitarian welfare.*

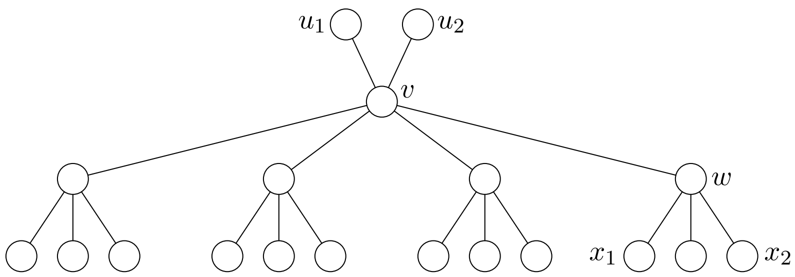


Figure 9: Illustration for the proof of Proposition A.3.

Proof. Consider the topology in Figure 9, and assume there are 3 blue and 16 red agents (so no node can be left empty). The only way to achieve nonzero egalitarian welfare is to assign the blue agents to u_1, u_2, and v, yielding utilitarian welfare 52/3. Assigning the blue agents to w, x_1, and x_2 results in the same utilitarian welfare, while the egalitarian welfare is then zero. It can be verified that both assignments maximize the utilitarian welfare. □

Proposition A.4. *There exists a class of Schelling instances such that the ratio between the maximum and minimum egalitarian welfare among assignments that maximize utilitarian welfare is* $\Theta(n)$*.*

Proof. Let k be a positive integer. We define the topology $G = (V, E)$ of a Schelling instance with node set V given by

$$V = \{a_i \colon 1 \leq i \leq k\} \cup \{x_1, x_2, y_1, y_2\} \cup \{b_i^j \colon 1 \leq j \leq 3, 1 \leq i \leq k\}$$

and edge set E given by

- $\{a_i, a_\ell\} \in E$ for $1 \leq i, \ell \leq k$;
- $\{a_i, x_\ell\} \in E$ for $1 \leq i \leq k$, $1 \leq \ell \leq 2$;
- $\{x_1, y_1\}, \{x_2, y_2\} \in E$;
- $\{y_j, b_i^j\} \in E$ for $1 \leq i \leq k$, $1 \leq j \leq 2$;
- $\{b_i^j, b_\ell^j\} \in E$ for $1 \leq i, \ell \leq k$, $1 \leq j \leq 3$;
- $\{b_i^j, b_i^3\} \in E$ for $1 \leq i \leq k$, $1 \leq j \leq 2$;
- no further edges are in E.

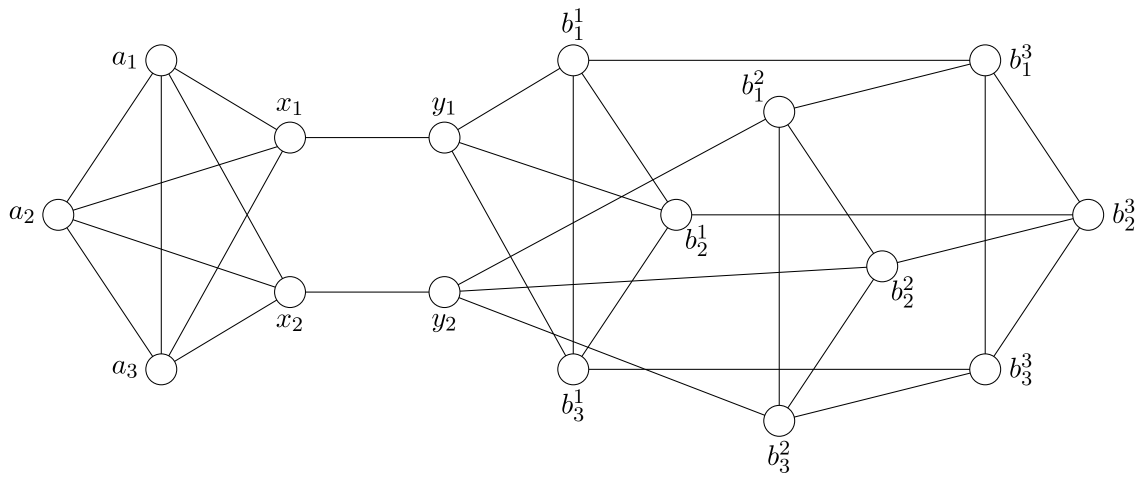


Figure 10: Illustration for the proof of Proposition A.4.

See Figure 10 for an illustration for $k = 3$.

Assume that there are $n = 4k + 4$ agents, composed of k blue and $3k + 4$ red agents, so no node can be left empty. Note that the topology graph is $(k + 1)$-regular and contains cliques of size k. Hence, as in the proof of Theorem 3.3, an assignment maximizes the utilitarian welfare if and only if the blue agents form a clique.

Consider the assignment where the blue agents form the clique $\{a_i : 1 \leq i \leq k\}$. The utility of each neighboring (red) agent x_j is $1/(k+1)$, so the egalitarian welfare is $1/(k+1)$. On the other hand, consider the assignment with the blue agents forming the clique $\{b_i^3 : 1 \leq i \leq k\}$. In this assignment, every agent has utility at least $(k-1)/(k+1)$, where the minimum utility is attained by the blue agents. Hence, the egalitarian welfare is $(k-1)/(k+1)$, which is a multiplicative factor of $\Theta(k) = \Theta(n)$ higher than that of the first assignment. $\square$

References

Agarwal, A., Elkind, E., Gan, J., & Voudouris, A. A. (2020). Swap stability in Schelling games on graphs. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1758–1765.

Anshelevich, E., Dasgupta, A., Kleinberg, J. M., Tardos, É., Wexler, T., & Roughgarden, T. (2008). The price of stability for network design with fair cost allocation. *SIAM Journal on Computing*, *38*(4), 1602–1623.

Aziz, H., Brandl, F., Brandt, F., Harrenstein, P., Olsen, M., & Peters, D. (2019). Fractional hedonic games. *ACM Transactions on Economics and Computation*, *7*(2), 6:1–6:29.

Balliu, A., Flammini, M., & Olivetti, D. (2017). On Pareto optimality in social distance games. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pp. 349–355.

Barmpalias, G., Elwes, R., & Lewis-Pye, A. (2014). Digital morphogenesis via Schelling segregation. In *Proceedings of the 55th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 156–165.

Barmpalias, G., Elwes, R., & Lewis-Pye, A. (2015). Unperturbed Schelling segregation in two or three dimensions. *Journal of Statistical Physics*, *164*, 1460–1487.

Bei, X., Lu, X., Manurangsi, P., & Suksompong, W. (2019). The price of fairness for indivisible goods. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 81–87.

Bhakta, P., Miracle, S., & Randall, D. (2014). Clustering and mixing times for segregation models on $\mathbb{Z}^2$. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 327–340.

Bilò, D., Bilò, V., Lenzner, P., & Molitor, L. (2020). Topological influence and locality in swap Schelling games. In *Proceedings of the 45th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pp. 15:1–15:15.

Brandt, C., Immorlica, N., Kamath, G., & Kleinberg, R. (2012). An analysis of one-dimensional Schelling segregation. In *Proceedings of the 44th Symposium on Theory of Computing Conference (STOC)*, pp. 789–804.

Bui, T. N., Chaudhuri, S., Leighton, F. T., & Sipser, M. (1987). Graph bisection algorithms with good average case behavior. *Combinatorica*, *7*(2), 171–191.

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

Bullinger, M., Suksompong, W., & Voudouris, A. A. (2021). Welfare guarantees in Schelling segregation. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*.

Chan, H., Irfan, M. T., & Than, C. V. (2020). Schelling models with localized social influence: a game-theoretic framework. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 240–248.

Chauhan, A., Lenzner, P., & Molitor, L. (2018). Schelling segregation with strategic agents. In *Proceedings of the 11th International Symposium on Algorithmic Game Theory (SAGT)*, pp. 137–149.

Clark, W., & Fossett, M. (2008). Understanding the social context of the Schelling segregation model. *Proceedings of the National Academy of Sciences*, *105(11)*, 4109–4114.

Echzell, H., Friedrich, T., Lenzner, P., Molitor, L., Pappik, M., Schöne, F., Sommer, F., & Stangl, D. (2019). Convergence and hardness of strategic Schelling segregation. In *Proceedings of the 15th International Conference on Web and Internet Economics (WINE)*, pp. 156–170.

Elkind, E., Fanelli, A., & Flammini, M. (2020). Price of Pareto optimality in hedonic games. *Artificial Intelligence*, *288*, 103357.

Elkind, E., Gan, J., Igarashi, A., Suksompong, W., & Voudouris, A. A. (2019). Schelling games on graphs. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 266–272.

Garey, M. R., & Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman.

Immorlica, N., Kleinberg, R., Lucier, B., & Zadimoghaddam, M. (2017). Exponential segregation in a two-dimensional Schelling model with tolerant individuals. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 984–993.

Kanellopoulos, P., Kyropoulou, M., & Voudouris, A. A. (2020). Modified Schelling games. In *Proceedings of the 13th International Symposium on Algorithmic Game Theory (SAGT)*, pp. 241–256.

Koutsoupias, E., & Papadimitriou, C. H. (1999). Worst-case equilibria. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, pp. 404–413.

Pancs, R., & Vriend, N. J. (2007). Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics*, *91*(1), 1–24.

Pollicott, M., & Weiss, H. (2001). The dynamics of Schelling-type segregation models and a nonlinear graph Laplacian variational problem. *Advances in Applied Mathematics*, *27*(1), 17–40.

Schelling, T. C. (1969). Models of segregation. *American Economic Review*, *59*(2), 488–493.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*(2), 143–186.

Young, H. P. (2001). *Individual Strategy and Social Structure: an Evolutionary Theory of Institutions*. Princeton University Press.

Zhang, J. (2004). A dynamic model of residential segregation. *Journal of Mathematical Sociology*, *28(3)*, 147–170.

Abeledo, H. and U. G. Rothblum (1995). "Paths to marriage stability". In: *Discrete Applied Mathematics* 63.1, pp. 1–12 [p. 21].

Abraham, D. J., R. W. Irving, T. Kavitha, and K. Mehlhorn (2007a). "Popular matchings". In: *SIAM Journal on Computing* 37.4, pp. 1030–1034 [p. 32].

Abraham, D. J., A. Leravi, D. F. Manlove, and G. O'Malley (2007b). "The stable roommates problem with globally-ranked pairs". In: *Proceedings of the 3rd*. Vol. 4858. Lecture Notes in Computer Science (LNCS). Springer-Verlag, pp. 431–444 [p. 10].

Agarwal, A., E. Elkind, J. Gan, A. Igarashi, W. Suksompong, and A. A. Voudouris (2021). "Schelling games on graphs". In: *Artificial Intelligence* 301, p. 103576 [pp. 6, 24, 25, 43].

Alcalde, J. and P. Revilla (2004). "Researching with whom? Stability and manipulation". In: *Journal of Mathematical Economics* 40.8, pp. 869–887 [p. 5].

Apt, K. R. and G. Schäfer (2014). "Selfishness level of strategic games". In: *Journal of Artificial Intelligence Research* 49, pp. 207–240 [p. 22].

Arora, S. and B. Barak (2009). *Computational Complexity: A Modern Approach*. Cambridge University Press [p. 7].

Aumann, R. J. and J. H. Drèze (1975). "Cooperative games with coalition structures". In: *International Journal of Game Theory* 3.4, pp. 217–237 [p. 4].

Aumann, R. J. and M. Maschler (1964). "The Bargaining Set for Cooperative Games". In: *Advances in Game Theory*. Ed. by M. Dresher, L. S. Shapley, and A. W. Tucker. Vol. 52. Annals of Mathematics Studies. Princeton University Press, pp. 443–476 [p. 4].

Aziz, H., F. Brandl, F. Brandt, P. Harrenstein, M. Olsen, and D. Peters (2019). "Fractional Hedonic Games". In: *ACM Transactions on Economics and Computation* 7.2, pp. 1–29 [pp. 5, 13, 49].

Aziz, H., F. Brandt, and P. Harrenstein (2013a). "Pareto Optimality in Coalition Formation". In: *Games and Economic Behavior* 82, pp. 562–581 [p. 32].

Aziz, H., F. Brandt, and H. G. Seedig (2013b). "Computing Desirable Partitions in Additively Separable Hedonic Games". In: *Artificial Intelligence* 195, pp. 316–334 [pp. 19, 29, 33, 55, 66].

Aziz, H., S. Gaspers, J. Gudmundsson, J. Mestre, and H. Täubig (2015). "Welfare Maximization in Fractional Hedonic Games". In: *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 461–467 [p. 66].

Aziz, H. and R. Savani (2016). "Hedonic Games". In: *Handbook of Computational Social Choice*. Ed. by F. Brandt, V. Conitzer, U. Endriss, J.

Lang, and A. D. Procaccia. Cambridge University Press. Chap. 15 [p. 15].

Ballester, C. (2004). "NP-completeness in hedonic games". In: *Games and Economic Behavior* 49.1, pp. 1–30 [pp. 10, 11, 55].

Balliu, A., M. Flammini, G. Melideo, and D. Olivetti (2022). "On Pareto optimality in social distance games". In: *Artificial Intelligence* 312, p. 103768 [p. 26].

Banerjee, S., H. Konishi, and T. Sönmez (2001). "Core in a simple coalition formation game". In: *Social Choice and Welfare* 18, pp. 135–153 [pp. 5, 16].

Barmpalias, G., R. Elwes, and A. Lewis-Pye (2014). "Digital morphogenesis via Schelling segregation". In: *Proceedings of the 55th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 156–165 [p. 6].

Barmpalias, G., R. Elwes, and A. Lewis-Pye (2015). "Unperturbed Schelling segregation in two or three dimensions". In: *Journal of Statistical Physics* 164, pp. 1460–1487 [p. 6].

Bilò, V., A. Fanelli, M. Flammini, G. Monaco, and L. Moscardelli (2018). "Nash Stable Outcomes in Fractional Hedonic Games: Existence, Efficiency and Computation". In: *Journal of Artificial Intelligence Research* 62, pp. 315–371 [p. 21].

Bilò, V., G. Monaco, and L. Moscardelli (2022). "Hedonic Games with Fixed-Size Coalitions". In: *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)* [p. 48].

Biró, P., R. W. Irving, and D. F. Manlove (2010). "Popular Matchings in the Marriage and Roommates Problems". In: *Proceedings of the 7th Italian Conference on Algorithms and Complexity (CIAC)*, pp. 97–108 [pp. 32, 33].

Boehmer, N., M. Bullinger, and A. M. Kerkmann (2023). "Causes of Stability in Dynamic Coalition Formation". In: *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*. Forthcoming [pp. 39, 65].

Boehmer, N. and E. Elkind (2020). "Individual-Based Stability in Hedonic Diversity Games". In: *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1822–1829 [p. 36].

Bogomolnaia, A. and M. O. Jackson (2002). "The Stability of Hedonic Coalition Structures". In: *Games and Economic Behavior* 38.2, pp. 201–230 [pp. 5, 12, 13, 15, 16, 18, 29, 30, 35, 38, 48].

Borel, É. (1921). "La théorie du jeu et les équations intégrales à noyau symétrique". In: *Comptes Rendus de l'Académie des Sciences* 173, pp. 1304–1308 [p. 3].

Brandl, F. and F. Brandt (2020). "Arrovian Aggregation of Convex Preferences". In: *Econometrica* 88.2, pp. 799–844 [p. 20].

Brandl, F., F. Brandt, and M. Strobel (2015). "Fractional Hedonic Games: Individual and Group Stability". In: *Proceedings of the 14th*

International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pp. 1219–1227 [pp. 29, 30].
Brandt, C., N. Immorlica, G. Kamath, and R. Kleinberg (2012). "An analysis of one-dimensional Schelling segregation". In: *Proceedings of the 44th Symposium on Theory of Computing Conference (STOC)*, pp. 789–804 [p. 6].
Brandt, F. and M. Bullinger (2022). "Finding and Recognizing Popular Coalition Structures". In: *Journal of Artificial Intelligence Research* 74, pp. 569–626 [pp. 10, 11, 20, 32–34, 62].
Brandt, F., M. Bullinger, and L. Tappe (2022a). "Single-Agent Dynamics in Additively Separable Hedonic Games". In: *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 4867–4874 [pp. 15, 18, 39].
Brandt, F., M. Bullinger, and A. Wilczynski (2021). "Reaching Individually Stable Coalition Structures in Hedonic Games". In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 5211–5218 [pp. 21, 30, 36, 37, 51, 52, 54].
Brandt, F., M. Bullinger, and A. Wilczynski (2022b). *Reaching Individually Stable Coalition Structures*. Tech. rep. https://arxiv.org/abs/2211.09571 [pp. 22, 36, 38, 55].
Brandt, F. and A. Wilczynski (2019). "On the Convergence of Swap Dynamics to Pareto-Optimal Matchings". In: *Proceedings of the 15th International Workshop on Internet and Network Economics (WINE)*. Lecture Notes in Computer Science (LNCS). Springer-Verlag, pp. 100–113 [p. 21].
Bredereck, R., E. Elkind, and A. Igarashi (2019). "Hedonic Diversity Games". In: *Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 565–573 [p. 12].
Bullinger, M. (2020). "Pareto-Optimality in Cardinal Hedonic Games". In: *Proceedings of the 19th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 213–221 [pp. 30, 31].
Bullinger, M. (2022). "Boundaries to Single-Agent Stability in Additively Separable Hedonic Games". In: *Proceedings of the 47th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, 26:1–26:15 [pp. 29, 39, 56, 57, 59].
Bullinger, M. and S. Kober (2021). "Loyalty in Cardinal Hedonic Games". In: *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 66–72 [pp. 23, 40, 41].
Bullinger, M. and W. Suksompong (2023). "Topological Distance Games". In: *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*. Forthcoming [p. 48].
Bullinger, M., W. Suksompong, and A. Voudouris (2021). "Welfare Guarantees in Schelling Segregation". In: *Journal of Artificial Intelligence Research* 71, pp. 143–174 [pp. 25–27, 42–44].
Carosi, R., G. Monaco, and L. Moscardelli (2019). "Local Core Stability in Simple Symmetric Fractional Hedonic Games". In: *Proceedings*

of the 18th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pp. 574–582 [p. 21].

Cechlárová, K. and A. Romero-Medina (2001). "Stability in Coalition Formation games". In: *International Journal of Game Theory* 29, pp. 487–494 [pp. 5, 11, 13].

Chauhan, A., P. Lenzner, and L. Molitor (2018). "Schelling segregation with strategic agents". In: *Proceedings of the 11th International Symposium on Algorithmic Game Theory (SAGT)*, pp. 137–149 [pp. 6, 24].

Clark, W. and M. Fossett (2008). "Understanding the social context of the Schelling segregation model". In: *Proceedings of the National Academy of Sciences* 105(11), pp. 4109–4114 [p. 6].

Cseh, Á. (2017). "Popular Matchings". In: *Trends in Computational Social Choice*. Ed. by U. Endriss. AI Access. Chap. 6 [p. 19].

Dimitrov, D., P. Borm, R. Hendrickx, and S. C. Sung (2006). "Simple Priorities and Core Stability in Hedonic Games". In: *Social Choice and Welfare* 26.2, pp. 421–433 [pp. 13, 66].

Dimitrov, D. and S. C. Sung (2007). "On top responsiveness and strict core stability". In: *Journal of Mathematical Economics* 43.2, pp. 130–134 [p. 16].

Drèze, J. H. and J. Greenberg (1980). "Hedonic Coalitions: Optimality and Stability". In: *Econometrica* 48.4, pp. 987–1003 [pp. 4, 5, 9, 15, 18].

Easley, D. and J. Kleinberg (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press [p. 6].

Echzell, H., T. Friedrich, P. Lenzner, L. Molitor, M. Pappik, F. Schöne, F. Sommer, and D. Stangl (2019). "Convergence and hardness of strategic Schelling segregation". In: *Proceedings of the 15th International Conference on Web and Internet Economics (WINE)*, pp. 156–170 [p. 6].

Edmonds, J. (1965a). "Maximum Matching and a Polyhedron with 0,1-vertices". In: *Journal of Research of the National Bureau of Standards B* 69, pp. 125–130 [p. 32].

Edmonds, J. (1965b). "Paths, Trees and Flowers". In: *Canadian Journal of Mathematics* 17, pp. 449–467 [p. 32].

Elias, J., F. Martignon, K. Avrachenkov, and G. Neglia (2010). "Socially-aware network design games". In: *Proceedings of the 29th IEEE Conference on Computer Communications (INFOCOM)*. IEEE, pp. 1–5 [p. 22].

Elkind, E., A. Fanelli, and M. Flammini (2020). "Price of Pareto optimality in hedonic games". In: *Artificial Intelligence* 288, p. 103357 [pp. 26, 31].

Elkind, E. and M. Wooldridge (2009). "Hedonic coalition nets". In: *Proceedings of the 8th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 417–424 [p. 12].

Faenza, Y., T. Kavitha, V. Power, and X. Zhang (2019). "Popular Matchings and Limits to Tractability". In: *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2790–2809 [p. 33].

Farrell, J. and S. Scotchmer (1988). "Partnerships". In: *Quarterly Journal of Economics* 103, pp. 279–297 [p. 10].

Fishburn, P. C. (1984). "Probabilistic Social Choice Based on Simple Voting Comparisons". In: *Review of Economic Studies* 51.4, pp. 683–692 [pp. 20, 33].

Freeman, L. C. (1978). "Segregation in social networks". In: *Sociological Methods & Research* 6.4, pp. 411–429 [p. 43].

Gairing, M. and R. Savani (2019). "Computing Stable Outcomes in Symmetric Additively Separable Hedonic Games". In: *Mathematics of Operations Research* 44.3, pp. 1101–1121 [pp. 17, 38].

Gärdenfors, P. (1975). "Match Making: Assignments based on bilateral preferences". In: *Behavioral Science* 20.3, pp. 166–173 [pp. 19, 32].

Grötschel, M., L. Lovász, and A. Schrijver (1981). "The Ellipsoid Method and its Consequences in Combinatorial Optimization". In: *Combinatorica* 1, pp. 169–197 [p. 32].

Grötschel, M., L. Lovász, and A. Schrijver (1993). *Geometric Algorithms and Combinatorial Optimization*. Vol. 2. Algorithms and Combinatorics. Springer [p. 51].

Gupta, S., P. Misra, S. Saurabh, and M. Zehavi (2019). "Popular Matching In Roommates Setting is NP-hard". In: *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2810–2822 [p. 33].

Hajduková, J. (2006). "Coalition formation games: A survey". In: *International Game Theory Review* 8.4, pp. 613–641 [p. 5].

Hell, P. and D. G. Kirkpatrick (1984). "Packings by cliques and by finite families of graphs". In: *Discrete Mathematics* 49.1, pp. 45–59 [p. 30].

Hoefer, M., D. Vaz, and L. Wagner (2018). "Dynamics in matching and coalition formation games with structural constraints". In: *Artificial Intelligence* 262, pp. 222–247 [p. 21].

Immorlica, N., R. Kleinberg, B. Lucier, and M. Zadimoghaddam (2017). "Exponential segregation in a two-dimensional Schelling model with tolerant individuals". In: *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 984–993 [p. 6].

Irving, R. W. (1985). "An Efficient Algorithm for the "Stable Roommates" Problem". In: *Journal of Algorithms* 6.4, pp. 577–595 [p. 11].

Johnson, D. S., C. H. Papadimitriou, and M. Yannakakis (1988). "How Easy is Local Search?" In: *Journal of Computer and System Sciences* 37, pp. 79–100 [p. 47].

Kahn, A. B. (1962). "Topological sorting of large networks". In: *Communications of the ACM* 5.11, pp. 558–562 [pp. 48, 49].

Karp, R. M. (1972). "Reducibility among Combinatorial Problems". In: *Complexity of Computer Computations*. Ed. by R. E. Miller and J. W. Thatcher. Plenum Press, pp. 85–103 [p. 58].

Kavitha, T., J. Mestre, and M. Nasre (2011). "Popular mixed matchings". In: *Theoretical Computer Science* 412.24, pp. 2679–2690 [pp. 20, 32].

Kerkmann, A. M., J. Lang, A. Rey, J. Rothe, H. Schadrack, and L. Schend (2020). "Hedonic Games with Ordinal Preferences and Thresholds". In: *Journal of Artificial Intelligence Research* 67, pp. 705–756 [pp. 19, 33].

Kerkmann, A. M., N.-T. Nguyen, A. Rey, L. Rey, J. Rothe, L. Schend, and A. Wiechers (2022). "Altruistic Hedonic Games". In: *Journal of Artificial Intelligence Research* 75, pp. 129–169 [pp. 22, 23, 66].

Kerkmann, A. M. and J. Rothe (2020). "Altruism in Coalition Formation Games". In: *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 461–467 [p. 33].

Khachiyan, L. (1979). "A Polynomial Algorithm in Linear Programming". In: *Soviet Mathematics Doklady* 20, pp. 191–194 [p. 32].

Koutsoupias, E. and C. H. Papadimitriou (1999). "Worst-case equilibria". In: *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, pp. 404–413 [p. 26].

Kreisel, L., N. Boehmer, V. Froese, and R. Niedermeier (2022). "Equilibria in Schelling Games: Computational Hardness and Robustness". In: *Proceedings of the 21st International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 761–769 [p. 24].

Manlove, D. F. (2013). *Algorithmics of Matching Under Preferences*. World Scientific Publishing Company [p. 33].

McCutchen, R. M. (2008). "The Least-Unpopularity-Factor and Least-Unpopularity-Margin Criteria for Matching Problems with One-Sided Preferences". In: *Proceedings of the 8th Latin American Conference on Theoretical Informatics (LATIN)*. Vol. 4957. Lecture Notes in Computer Science (LNCS), pp. 593–604 [p. 32].

Meade, J. E. (1972). "The theory of labour-managed firms and of profit sharing". In: *The Economic Journal* 82.325, pp. 402–428 [p. 15].

Mitzenmacher, M. and E. Upfal (2005). *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press [p. 42].

Monaco, G., L. Moscardelli, and Y. Velaj (2018). "Stable Outcomes in Modified Fractional Hedonic Games". In: *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 937–945 [p. 66].

Monien, B. and T. Tscheuschner (2010). "On the power of nodes of degree four in the local max-cut problem". In: *Proceedings of the 7th International Conference on Algorithms and Complexity (CIAC)*. Lecture

Notes in Computer Science (LNCS) 6078. Springer-Verlag, pp. 264–275 [p. 39].

Mueller, D. C. (1986). "Rational egoism versus adaptive egoism as fundamental postulate for a descriptive theory of human behavior". In: *Public Choice* 51.1, pp. 3–23 [p. 22].

Nash, J. F. (1950). "Equilibrium Points in n-Person Games". In: *Proceedings of the National Academy of Sciences (PNAS)* 36, pp. 48–49 [p. 15].

Olsen, M. (2012). "On defining and computing communities". In: *Proceedings of the 18th Computing: Australasian Theory Symposium (CATS)*. Vol. 128. Conferences in Research and Practice in Information Technology (CRPIT), pp. 97–102 [p. 14].

Padberg, M. W. and L. A. Wolsey (1984). "Fractional Covers for Forests and Matchings". In: *Mathematical Programming* 29, pp. 1–14 [p. 32].

Peters, D. and E. Elkind (2015). "Simple Causes of Complexity in Hedonic Games". In: *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 617–623 [p. 55].

Roth, A. E. and J. H. Vande Vate (1990). "Random Paths to Stability in Two-Sided Matching". In: *Econometrica* 58.6, pp. 1475–1480 [p. 21].

Saad, W., Z. Han, T. Basar, M. Debbah, and A. Hjorungnes (2011). "Hedonic Coalition Formation for Distributed Task Allocation among Wireless Agents". In: *IEEE Transactions on Mobile Computing* 10.9, pp. 1327–1344 [p. 5].

Schelling, T. C. (1969). "Models of segregation". In: *American Economic Review* 59.2, pp. 488–493 [pp. 6, 7, 24].

Schelling, T. C. (1971). "Dynamic models of segregation". In: *Journal of Mathematical Sociology* 1.2, pp. 143–186 [p. 6].

Schelling, T. C. (2006). *Micromotives and macrobehavior*. WW Norton & Company [p. 6].

Sung, S. C. and D. Dimitrov (2007). "On Myopic Stability Concepts for Hedonic Games". In: *Theory and Decision* 62.1, pp. 31–45 [pp. 29, 56].

Sung, S. C. and D. Dimitrov (2010). "Computational Complexity in Additive Hedonic Games". In: *European Journal of Operational Research* 203.3, pp. 635–639 [pp. 29, 55, 56].

Tappe, L. (2021). "Stability in Coalition Formation Games Based on Single-Agent Deviations". MA thesis. Technical University of Munich [p. 65].

von Neumann, J. (1928). "Zur Theorie der Gesellschaftspiele". In: *Mathematische Annalen* 100.1, pp. 295–320 [p. 3].

von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press [pp. 4, 16].

Williamson, D. P. and D. B. Shmoys (2011). *The design of approximation algorithms*. Cambridge University Press [p. 42].

Woeginger, G. (2013). "A hardness result for core stability in additive hedonic games". In: *Mathematical Social Sciences* 65.2, pp. 101–104 [p. 55].

Young, H. P. (1998). *Individual strategy and social structure." Individual Strategy and Social Structure*. Princeton University Press [p. 6].

Zhang, J. (2004). "A dynamic model of residential segregation". In: *Journal of Mathematical Sociology* 28.3, pp. 147–170 [p. 6].