Dynamics in Matching and Coalition Formation Games with Structural Constraints*

Martin Hoefer† Daniel Vaz‡ Lisa Wagner§

Abstract

Matching and coalition formation are fundamental aspects in the organization of many multi-agent systems. In large populations, the emergence of coalitions is often restricted by structural constraints under which agents can reorganize, e.g., local visibility or externality constraints among the agents. We study this aspect using a novel framework for dynamics with constraints within the popular domain of hedonic coalition formation games. We analyze the effects of structural constraints on the convergence of matching and coalition formation processes to stable states. Our main result are tight characterizations for the constraint structures based on which dynamic coalition formation can stabilize quickly. We show a variety of convergence results for matching and coalition formation games with different forms of locality and externality constraints. In particular, we propose and analyze a new model of graph-based visibility for coalition formation games and tightly characterize the graph structures that allow polynomial-time convergence – it can be achieved if and only if coalition formation is based on complete or star graphs.

Keywords: Stable Matching, Hedonic Games, Path to Stability, Convergence

*This work was supported by DFG Cluster of Excellence MMCI at Saarland University and grant Ho 3831/3-1.

†Max-Planck-Institut für Informatik and Saarland University, Germany. mhoefer@mpi-inf.mpg.de.

‡Max-Planck-Institut für Informatik and Saarland University Graduate School of Computer Science, Germany. ramosvaz@mpi-inf.mpg.de.

§Department of Computer Science, RWTH Aachen University, Germany. lwagner@cs.rwth-aachen.de.

1 Introduction

Stable Matching and Hedonic Games Matching and coalition formation problems form the basis for a variety of assignment and allocation tasks encountered in computer science, operations research, and economics. In multi-agent systems, the formation of coalitions is a natural process to exploit synergies. In this domain, the model of hedonic games [7, 18, 25, 26, 59] represents a natural and versatile approach to analyze aspects of coalition formation processes in multi-agent systems.

Perhaps the most prominent domain of hedonic games are classes of stable matching. In the classic stable marriage problem, there is a set of men and a set of women. Each agent strives to find a partner from the other gender, and it has a preference list over possible partners. Given a matching (i.e., a set of mutually disjoint man-woman pairs), a *blocking pair* is a man-woman pair such that both partners strictly improve over their current match if the pair is formed. A matching without blocking pair is called a *stable matching*. Since its introduction by Gale and Shapley in 1962 [27], stable matching has been successfully applied to capture assignment problems in multi-agent systems in a large variety of applications, e.g., assignment of jobs to workers [5, 41], organs to patients [55], and many others. In addition, stable marriage problems have been successfully used to study distributed resource allocation problems in networks [3, 28, 46]. As such, stable marriage occupies a central position in the analysis of decentralized coalition formation.

There are several natural generalizations of stable marriage that have been intensively studied. Stable roommates instances drop the assumption of two sides – there is simply a single set of n agents, and each agent can match to some (possibly arbitrary) subset of other agents [40]. More generally, hedonic games allow the formation of coalitions of three or more agents. Their main characteristics are that (1) each agent can be part of at most one coalition, and (2) the preference or payoff an agent receives for being in a coalition depends only on the agents in that coalition, but not on how the remaining agents are partitioned. A variety of stability concepts in hedonic games and their aspects have been studied in recent years (see for example [8, 9, 11, 18]). We here concentrate on the natural approach of core-stability, where agents have cardinal preferences for the coalitions they are part of. A state is a set of non-overlapping coalitions, and a blocking coalition is a group of agents that could (all individually) improve by abandoning their current coalitions and forming a new one together. A state is core-stable if there is no blocking coalition.

Structural Constraints and Improvement Dynamics In the standard model of hedonic games, improvement in terms of preferences is the only criterion for deviation to other coalitions. In contrast, there can be a variety of additional constraints that govern the coalition formation process. Our main contribution in this paper is the formulation of a general framework for hedonic coalition formation with additional constraints. We study the consequences of these constraints on the existence of stable states and their reachability by myopic improvement dynamics.

Within our general framework, we put a focus on constraints derived from two natural properties of multi-agent systems – local visibility and (positive) externalities. As a concrete application, consider the prominent special case of stable matching. In large matching markets, it is unrealistic to assume that agents have full information about all other agents and possible partners. Instead, agents are often only aware about a subset of the population. For example, consider agents looking for a partner to do a joint activity (such as, e.g., playing squash or chess). We would not expect to form all possible and profitable matching pairs instantaneously. Instead, a pair of actors first have to get to know (about) each other before they can start a joint project. For the task of finding a partner, one often relies on existing relationships from family or co-workers. As a stylized model, we can capture this idea by extending the standard model of stable matching using a network – possible matching pairs have to be connected in an underlying social network to be available for formation. More interestingly, by engaging in an activity with a new partner, we get to know their friends and partners and possibly discover new and better matches. When we incorporate this aspect, we arrive at what has been termed *locally stable matching* [5,19,23,31,37]. Here the agents of a (local) blocking pair must have a hop-distance of at most 2 in the underlying network of social contacts and existing matched pairs. A variation of this model can be used to study job markets, where we strive to match jobs to workers. It is known that a large portion of jobs are assigned based on social contacts to co-workers in the same firm. Here the social network is present only among workers, but firms can match to multiple workers [5].

More generally, instead of visibility it is also natural to imagine that social contacts have a positive externality (or "altruistic") effect on actors. Beyond anecdotal evidence, there are studies in experimental game theory that support this assumption [43, 44]. In addition to social ties and friendship, such an altruistic or considerate behavior can also arise, e.g., due to the existence of formal contracts or business relations among agents. Recently, such effects have received increased interest in a variety of game-theoretic models, including potential games [21,33,34], matching [4], and hedonic coalition formation [49]. In this case, agents are assumed to consider the (negative) effects on others when forming a new match and thereby, e.g., "stealing" the current partner of one of their friends or family members. A frequently explored model in games with cardinal utility is to express the trade-off between individual improvement and externality using a numerical value $c_{u,v} \geq 0$ for each pair u, v of agents. The perceived utility of an agent u then is a weighted sum of his individual utility and the utilities of all other agents v (weighted with $c_{u,v}$). Here the agents of a (friendship) blocking pair must have improved perceived utilities.

In this paper, our goal is to shed light on the properties of a variety of extensions of stable matching and hedonic coalition formation games, in which coalition formation is restricted by additional constraints. Such constraints can arise due to many aspects, for example, spatial closeness, previous collaborations, or social ties based on family, friendship, or co-worker relations. The general model we study here includes a variety of special cases that have found recent interest in the literature, such as socially and locally stable matchings [5, 6, 19, 23, 31, 37] or friendship stable matching [4]. In addition, we also outline a natural model of *considerate stable matching.*

Our interest is in the properties of these games when the system is governed by dynamic and myopic coalitional deviations. Intuitively, such systems will eventually converge to a stable matching or core-stable state (if it exists). We study improvement dynamics, that is, the type of dynamics arising when we allow the iterative resolution of blocking coalitions until a stable state is reached. A blocking coalition is resolved by deleting all overlapping coalitions from the state and then adding the blocking coalition – possibly leaving some agents single. Thus, some agents improve in terms of their preference while others deteriorate. In consequence, such a sequence of improvement steps can take very long until it reaches a stable state, or it might even run into cycles. Of course, such behavior is undesirable, and our main interest is to identify and characterize conditions under which stable states can be reached (quickly).

A common aspect in all the example domains discussed above is that a blocking pair – in addition to being an improvement in terms of their preference for both incident agents – also has to fulfill additional graph-theoretic properties. A stable matching in these variants is a matching that has no blocking pair that satisfies such additional properties. Consequently, each stable matching remains stable, even when we require additional constraints for deviation. However, a state that is stable under a set of additional constraints might not be a (globally) stable matching. We analyze the effects of such constraints on the reachability of stable states using sequential improvement dynamics.

Since every stable matching remains stable under additional structural constraints, one might be tempted to think that stable matchings with constraints become easier to find and/or reach using distributed dynamics. In contrast, we show that stable matchings with constraints have a rich structure and can behave quite differently than stable matchings in unconstrained settings. More generally, our main results characterize sequences of improvement steps for a broad class of hedonic games with general conditions on the structural constraints.

1.1 Contribution and Outline

We study a broad domain of different matching and coalition formation game settings with structural constraints. In addition, we concentrate on the four matching variants (locally, socially, considerate and friendship stable matching) sketched above with their diverse facets. Our main focus lies on improvement dynamics and on (fast) reachability of stable states in the presence of structural constraints. A central question we study in this paper is the existence of a *path to stability*: Given an initial state, is there a path to stability, i.e., a sequence of resolutions of blocking coalitions that leads into a stable state? The existence of finite paths to stability guarantees that random improvement dynamics, in which blocking coalitions are chosen uniformly at random for resolution, converge to a stable state with probability 1 in the limit. Moreover, we are interested in the length of the paths to stability, and whether we can compute them in polynomial time.

We concentrate on matching and coalition formation games with correlated preferences, in which each coalition is characterized by a single benefit or quality value, and agents prefer coalitions with higher value. This domain has a natural appeal, e.g., it generalizes weighted matching problems that are ubiquitous in many applications. Furthermore, it guarantees existence of and convergence to core-stable states due to the existence of a lexicographical potential function [1, 46, 59]. In contrast, the case of general preferences over coalitions is much harder to tackle, since already deciding the existence of core-stable states is an extremely hard problem in seemingly special cases [51–53, 58, 59]. Characterizing convergence properties in this general domain represents an interesting direction for future work.

In Section 2 we first review a number of stable matching variants that formulate constraints on the matching process and have found recent interest in the literature. These games are special cases of our novel class of games termed coalition formation games with constraints, which we introduce in Section 3. For these games, we represent the improvement dynamics and consistent constraints using two sets of rules that depend on the current state. There is a set of self-generating coalitions $\mathcal{C}^g$ that can always be considered for deviation. In addition, generation rules specify a subset of coalitions that can be considered for deviation, which is based on the current state. Domination rules specify a subset of coalitions that are unavailable for deviation, which is based on the current state. Blocking coalitions for a state are undominated coalitions that can be generated. In principle, the rules can create arbitrary dependencies of blocking coalitions on the entire state.

We put special emphasis on games with *consistency* conditions given as follows. *Consistent* generation rules are of the form (C', C), when availability of coalition C depends only on existence of a *single overlapping* coalition C' with $C' \cap C \neq \emptyset$. For *consistent* domination rules $(\{C_1, C_2, \ldots, C_k\}, C)$ – which implies C is dominated when all coalitions $C_1, \ldots, C_k$ exist – we assume that there is *at least one overlapping* C_i with $C_i \cap C \neq \emptyset$. With consistent constraints the set of blocking coalitions can depend on the current state in a non-trivial way. However, the non-trivial blocking coalitions are in some sense local – a coalition can be generated only when some other overlapping coalition exists in the current state. Moreover, a coalition being dominated can be traced back to existence of some overlapping coalition in the current state.

We show that consistent constraints guarantee the existence of paths to stability with polynomial length. To highlight the broad applicability of these results, we show that all the matching variants discussed in the introduction yield consistent generation and domination rules. A detailed reduction is given in Appendix A. Furthermore, we show that our characterization is tight. The tightness result does not only apply when dropping consistency for generation and/or domination rules, but also with respect to relaxation of their intrinsic conditions described above. If generation rules are consistent and there are two coalitions allowed in the precondition of a generation rule, exponential paths to stability can become necessary. The same holds for consistent generation rules and when the single coalition in the domination rule is allowed to be non-overlapping with the target coalition. As long as domination rules are consistent, however, we can always guarantee that all paths to stability are finite. In contrast, for consistent domination rules and generation rules that are allowed to be non-overlapping with the target coalition, there are instances where no (finite) path to stability exists.

Consistent constraints guarantee paths of polynomial length to *some* stable state. Moreover, we show that they imply that there is a path of polynomial length to *any reachable* state. However, if we ask if there is a path to stability to a *given* stable state, this question becomes NP-hard to decide in many cases. In fact, we prove a general reduction that shows the hardness result for locally, socially, considerate and friendship stable matching with correlated preferences. Moreover, it shows the hardness result also for classic two-sided stable marriage with strict preferences (and without any additional constraints). To our knowledge, this result was not known before – in contrast, it has been known for more than two decades that polynomial paths to some stable marriage always exist [57].

As an additional domain, in Section 4 we consider a class of hedonic games with correlated preferences and visibility based on formation graphs. This represents a generalization of locally stable matching. In these games, we show another tight characterization of paths to stability with polynomial length. They are guaranteed to exist if and only if the formation graphs are either only complete graphs or only star graphs. While games based on star graphs turn out to yield consistent constraints, for complete graphs we show the result using a different set of arguments.

Our results are strongest when agents strive to form coalitions of small size. Let n be the number of agents and m be the number of possible coalitions that can form. In the prominent domain of matching, coalitions have size at most 2 and thus $m \leq \binom{n}{2}$. Our upper bounds on the length of paths to stability are polynomial in n and m. In contrast, the lower bound gadgets yield lengths exponential in n and only use coalitions of constant size. While the results apply in general, they are strongest when $m \in \mathsf{poly}(n)$ (e.g., for matching or when all coalitions that can form have constant size), since in this case we obtain a clear dichotomy of polynomial in n vs. exponential in n. Moreover, if $m \in \mathsf{poly}(n)$, then many involved problems in constructing improvement steps (finding a local blocking coalition, or the one with maximum benefit) can be solved trivially in time $\mathsf{poly}(n)$.

This work has partly appeared as extended abstracts in the proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI 2015) [35] and the 10th International Conference on Web and Internet Economics (WINE 2014) [36].

1.2 Related Work

Coalition formation and matching games are fundamental in the area of cooperative game theory, as they allow to model a variety of assignment, allocation, and formation problems. Applications can be found, for example, in computer science, operations research, and economics (for an overview see [10, 30]). Hedonic games are a central concept for modeling coalition formation in multi-agent systems, game theory, and algorithms. In these games the payoff for an agent is determined solely by the coalition it is part of. Since the initial work of Dreze and Greenberg [25], stability concepts in hedonic coalition formation games, their efficiency, and their computation have received significant attention in game theory [11, 18, 26, 58] and computer science [7, 20, 59]. In general, computing core-stable partitions is a very hard problem, lying at the second level of the polynomial hierarchy. For a recent overview on this problem we refer to Woeginger [59]. More recent developments on characterizing tractable cases can be found in [51–54].

Besides interest in graph theory, matching problems have received great attention in computer science, economics, and psychology due to many applications in assignment and allocation problems. For the two-sided stable marriage problem, where agents are divided into two sets and the potential matching edges form a bipartite graph, there has been a large amount of work on various aspects, for example, ties, incomplete lists, and many-to-many matchings. For an introduction to stable marriage and many of its variants see, e.g. [29, 45, 56].

There is a significant literature on matching dynamics, especially in economics, which is too broad to survey here. These works usually do not address issues like computational complexity or worst-case bounds. Here we focus on a subset of prominent analytical works related to our scenario.

In the stable marriage problem, a stable matching can always be reached in a polynomial number of improvement steps [57]. On the other hand, if in each step the blocking pair is chosen uniformly at random, convergence time can become exponential [2]. Furthermore, there exist instances and sequences of improvement steps such that the dynamics cycle [42]. However, if preferences are randomized, short convergence time can be observed experimentally [15]. For weighted or correlated matching, random dynamics can be shown to converge in expected polynomial time [2, 47]. More recently, convergence time of random dynamics using combinatorial properties of preferences have been studied [38].

In the stable roommates problem, where every pair of agents can be matched, stable matchings might not exist. However, deciding existence and computing a stable matching, if they exist, can be done in polynomial time [24, 40]. Several works have studied the computation of (variants of) stable matchings using iterative entry dynamics [14, 16, 17, 22], or matching problems in scenarios with payments or profit sharing [4, 13].

For socially stable matchings, the problem of computing a maximum-cardinality stable matching has found interest. There is a 1.5-approximation algorithm and approximation hardness within $1.5-\epsilon$ under the unique games conjecture [6].

Locally stable matching games were first studied for the special case of two-sided job-markets, where social links only exist among one partition [5]. Using a potential function argument, existence and reachability of stable states is always guaranteed for correlated preferences, even for networks with arbitrary links in the roommates problem. There always exist paths to stability of polynomial length, but random dynamics might need an exponential number of steps to converge [31]. For maximum locally stable matchings, the problem cannot be approximated within $(21/19 - \epsilon)$ unless $\mathsf{P}=\mathsf{NP}$ [23] and within 3/2 under the unique games conjecture [37]. Local algorithms perform arbitrarily badly [19]. Similar ideas of limited and dynamic visibility have recently been studied in the context of network creation games [12].

Our approach to locality is related to existing graph-based models for limited interaction. Most prominently, there is a large body of works treating the Myerson value in graph-based cooperative games [48, 50]. While the underlying network does not restrict the allowed deviations, the Myerson value only assigns positive payoff to those coalitions which form a connected component regarding the network. In a similar direction, recent work showed strong hardness results for computing stable states in hedonic games with coalitions that are connected components in a static network [39].

The concept of considerate stable matching is related to ideas of ordinal externalities that have been studied in the context of resource selection games [32]. Friendship and other-regarding preferences in stable matching games have been studied in [4]. They analyze the existence of friendship stable matchings and bound prices of anarchy and stability with equal and unequal sharing of matching benefits. A different approach to friendship-based preferences grounded in ordinal and axiomatic properties is analyzed in [49].

2 Preliminaries

2.1 Matching and Coalition Formation Games

In a *coalition formation game* there is a set of rational agents that strive to group into coalitions. Depending on the group(s) an agent is part of, this generates a benefit for the agent. For most of the paper we assume that each agent can be part of at most one coalition. Given a set of coalitions, to establish a new coalition (and thereby potentially destroy existing ones), all agents of the new coalition have to agree to deviate. An agent is willing to deviate from (and thus destroy) an existing coalition only if the new coalition yields strictly larger benefit.

Definition (Coalition Formation Game) *A* (hedonic) coalition formation game $\mathcal{G} = (V, \mathcal{C}, b)$ *(with correlated preferences) consists of*

- *a set* V *of* agents *or* players,
- *a set of* possible coalitions *or* hyperedges $\mathcal{C} \subseteq 2^V \setminus \{\emptyset\}$,
- *a* benefit*-function* $b : V \times \mathcal{C} \to \mathbb{R}_{>0}$.

Let $n = |V|$, $m = |\mathcal{C}|$, *and* $\Delta = \max\{|C| \mid C \in \mathcal{C}\}$ *be the number of agents, the number of coalitions, and the size of the largest coalition in* $\mathcal{C}$*, respectively.*

In this paper, we focus on the case of *correlated preferences* where $b(v, C) = b(C)$ for all $v \in C$. Generalizing our results to more general domains seems an interesting and very challenging open problem.

A set $\mathcal{S} \subseteq \mathcal{C}$ with $|\{C \mid C \in \mathcal{S}, v \in C\}| \leq 1$ for every $v \in V$ is called a *state* or a *coalition structure*. Given a coalition structure $\mathcal{S}$, the benefit of agent $v \in V$ is given by $b(v, \mathcal{S}) = b(C)$ if there is $C \in \mathcal{S}$ with $v \in C$, and $b(v, \mathcal{S}) = 0$ otherwise.

We consider improvement dynamics that replace coalitions by more worthy ones. These are termed *blocking coalitions*.

Definition (Blocking Coalition, Stability) *For a coalition structure* $\mathcal{S}$*, agent* v *is* saturated *if there is* $C_v \in \mathcal{S}$ *with* $v \in C_v$*. A coalition* $C \in \mathcal{C} \setminus \mathcal{S}$ *is a* blocking coalition *if for every saturated agent* $v \in C$ *it holds* $b(C) > b(C_v)$*. If* $\mathcal{S}$ *has no blocking coalition, it is called* (core)-stable[1].

Intuitively, a blocking coalition C is a coalition such that every involved saturated agent would prefer to be in this coalition over being in the coalition it is part of in $\mathcal{S}$. By assumption, if an agent is not saturated in $\mathcal{S}$, then it prefers to be part of a coalition.

An *improvement step* is the resolution of some blocking coalition C: We drop all existing coalitions from the state that overlap with C. Then we add C to the state. Note that the payoff for all agents in the blocking coalition C strictly increases upon resolution.

Due to its many applications, we put a special focus on matching games.

Definition (Matching Game) *A* matching game *is a coalition formation game with* $|C| = 2$ *for all* $C \in \mathcal{C}$*. We refer to a coalition as a* pair*, and two agents involved in a pair are* matching partners*. A coalition structure is termed a* matching *and denoted by* M*.*

We call a blocking coalition in a matching game a *blocking pair*. A matching without blocking pairs is called a *stable matching*.

A prominent class are *two-sided* or *bipartite* matching games, where V is composed of two disjoint sets U and W, and $\mathcal{C} \subseteq U \times W$. This scenario is also referred to as the *stable marriage problem*.

Due to consistency with the literature, we express a matching game by a simple, undirected graph $G = (V, E)$ (with $E = \mathcal{C}$). Every possible coalition is an edge $e \in E$ with edge benefit $b(e)$.

2.2 Constraints

Important aspects of coalition formation such as locality or externalities are not captured by the standard model of hedonic coalition formation. We model these aspects by the addition of structural constraints. We here recapitulate several models that have been proposed in the literature for local visibility and externality constraints in matching markets, and we propose direct extensions to more general coalition formation scenarios. In addition, we present a novel model of considerate stability to study coalition formation with positive externalities.

[1]Our definition of stability represents the classic notion of core-stability if every $C \in 2^V$ is available to the agents. We can easily extend our setting to fulfill this property by assuming $b(C) = 0$ for all $C \in 2^V \setminus \mathcal{C}$.

2.2.1 Social Stability

In large markets, agents have no chance to gather information about all other agents due to its size or diversity. Hence, it is reasonable to assume that agents know only about a certain subset of other agents and can only deviate to coalitions formed with those agents. To model matching markets where agents can only deviate to known partners, we consider *socially stable matchings*. The agents are embedded into a *network* $N = (V, L)$, where L is a set of *links*. A link $\{u, v\} \in L$ indicates that u and v know about each other. Initially, an agent u can be matched to any other agent w with $\{u, w\} \in E$. Subsequently, however, it can only deviate to match to an agent v with $\{u, v\} \in E \cap L$.

More formally, the only change concerns the definition of blocking pair. A pair $e = \{u, v\}$ is called a *social blocking pair* if e is a blocking pair and $e \in L$. Consequently, a *social improvement step* is the resolution of a social blocking pair, and a state M is a *socially stable matching* if it has no social blocking pair. Socially stable matchings were introduced by Askalidis et al [6]. For an example, see Fig. 1.

We can restrict attention to instances with $L \subseteq E$, since links $\ell \notin E$ have no effect on social blocking pairs. Note that edges $e \in E \setminus L$ can be present in the initial state. However, none of these edges is available for subsequent deviation. Hence, as soon as an agent deviates, any incident matching edge will be from L.

The edges $e \in E \setminus L$ can make a significant difference in terms of size and computational complexity of socially stable matchings (see [6]). A central match-making platform might be able to bring together pairs that the agents themselves would not be able to come up with (and not be willing to deviate from) in the subset of their locally known agents. In this way, edges $e \in E \setminus L$ might enable the platform to create much larger (socially) stable matchings than when creation and deviation are based on the same set of edges. For dynamics, the impact of edges $e \in E \setminus L$ is arguably more limited.

We here extend the model to coalitions with more than two agents. By doing so, there is some freedom to define the relationship between social links among agents and the resulting coalitions that can form. Since the formation of a coalition often requires a significant amount of coordination and cooperation among its members, we find it natural to assume that agents of a blocking coalition know each other, i.e., a blocking coalition must form a clique in N. More generally, it might also be sufficient that the coalition has a small diameter (or is just connected in N) to enable communication and cooperation in order to deviate.

We capture the network requirements for coalition formation using a set of formation graphs $\mathcal{H}$. Each of the graphs $H \in \mathcal{H}$ represents a minimal connectivity structure that should be present for a coalition to achieve visibility and organization of a deviation. A coalition can form if its inherent network structure contains at least one formation graph $H \in \mathcal{H}$ as a subgraph. The formation graphs represent a minimum requirement in the sense that additional links cause no harm to the organization of the coalition as a deviation. Examples for formation graphs are, e.g., cliques, stars (a center agent that knows all other agents of the coalition organizes its creation), or graphs with small diameter.

More formally, in a *social coalition formation game* there is a set $\mathcal{H}$ of graphs. A coalition C is a *social blocking coalition* if it is a blocking coalition and there is a graph $H = (V_H, E_H) \in \mathcal{H}$ and some bijective map $\varphi : V_H \to C$ such that $\{u, v\} \in V_H \Rightarrow \{\varphi(u), \varphi(v)\} \in L$. By abuse of notation we will denote the latter condition by $H(C) \in L$. The graphs in $\mathcal{H}$ are termed *formation graphs*. A state $\mathcal{S}$ is a *socially stable state* if it has no social blocking coalition.

2.2.2 Local Stability

The concept of socially stable states is static since there is no aspect of market exploration. In contrast, *locally stable matching* captures the idea that the set of agents that know about each other and the coalitions available for deviation changes depending on the current state. By matching to some agent v, an agent u might get to see additional agents from the population and thereby

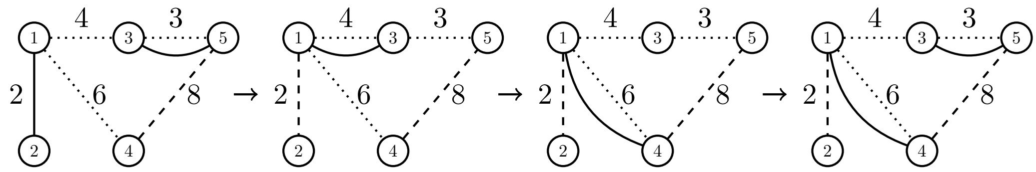

Figure 1: A sequence of social improvement steps. Vertices are agents, thick edges indicate current matching edges, dotted lines symbolize links. The dashed edges show additional pairs in E which can be present in the initial matching, but cannot be created by deviations since they are not part of L. Benefits are indicated by numbers alongside the edges. From left to right: Initial matching $\{\{1,2\},\{3,5\}\}$, edge $\{1,2\}$ exists but not part of L; resolution of social blocking pair $\{1,3\}$ removes $\{1,2\}$ and $\{3,5\}$, where the latter is later re-established and the former is lost; resolution of social blocking pair $\{1,4\}$; resolution of social blocking pair $\{3,5\}$, a socially stable matching is reached. Note that the attractive pair $\{4,5\}$ is not accessible throughout.

obtain additional and possibly more preferred partners to match and deviate. Similarly, if a match is abandoned, an agent might lose information about the population and get more restricted in the way it can match.

There is again a network $N = (V, L)$, in which L is a set of links. Links are static and represent connections independent of the coalition formation problem, e.g., based on family bonds or work relations. In contrast to socially stable matching, these links do not only provide alternative partners but also allow to explore and discover other matching partners. Agents see their 2-hop neighborhood in the graph $(V, M \cup L)$, i.e., the set of possible matching partners depends on L and the current matching M. Thus, the improvement dynamics become a joint matching and exploration process. While the links in L are permanent, note that matching edges cease to exist once they are dropped from M. This can also reduce the set of matching partners that are known to each other.

More formally, we say that a pair $\{u, v\}$ is *accessible* in state M if u and v have hop-distance at most 2 in the graph $(V, L \cup M)$. Put differently, $dist(u, v, (V, L \cup M)) \leq 2$ where $dist$ gives the length of a shortest path from u to v (which is ∞ if u and v lie in different connected components). A pair $e = \{u, v\}$ is called a *local blocking pair* if e is a blocking pair and e is accessible. A *local improvement step* is the resolution of a local blocking pair, and a state M is a *locally stable matching* if it has no social blocking pair. Locally stable matchings were introduced by Arcaute and Vassilvitskii [5] and further studied in [23, 31, 37].

For local blocking pairs, a shortest path that makes them accessible can consist of one link, two links, or of exactly one link and one matching edge. In the latter case, let w.l.o.g. $\{u, w\}$ be a matching edge. As u has at most one incident edge in M, the local improvement step will delete $\{u, w\}$ to create $\{u, v\}$. For simplicity, we will refer to this fact as an "edge moving from $\{u, w\}$ to $\{u, v\}$" or "u's edge moving from w to v". For an example, see Fig. 2.

To extend these ideas to games with larger coalitions, there are again several options to define a notion of accessible coalition. Similar to our approach for social coalition formation games above, we consider a set $\mathcal{H}$ of formation graphs that determines the minimum graph structures that are required among a coalition to be a candidate for deviation. Intuitively, closely connected graph structures of small diameter are natural choices for formation graphs. In this case, we also consider a notion of visibility of edges – a coalition C is *accessible* if there is a graph $H = (V_H, E_H) \in \mathcal{H}$ with $|V_H| = |C|$ and a bijective map $\varphi : V_H \to C$ such that each edge $\{u, v\} \in V_H$ is *visible* in the current coalition structure. Note that in the case of matching, there is no need to define a separate visibility notion for edges, since all coalitions are single edges.

To define visibility for an edge, we construct a visibility graph for a state by adding L and a clique of *temporary links* for every existing coalition. Formally, the *visibility graph* for coalition structure

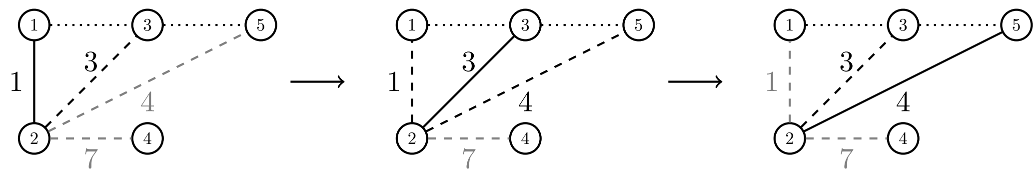

Figure 2: A sequence of local improvement steps. Vertices are agents, thick edges indicate current matching edges, dotted lines symbolize links. The dashed edges show additional pairs in E, where black edges are accessible and gray edges are not accessible in the current state. Benefits are indicated by numbers alongside the edges. From left to right: Initial matching $\{\{1,2\}\}$, although link $\{1,3\}$ is not a matching edge, it serves to make the unique local blocking pair $\{2,3\}$ accessible; resolution of local blocking pair $\{2,3\}$ removes $\{1,2\}$, reduces the distance between 2 and 5 to 2, and makes $\{2,5\}$ accessible; resolution of local blocking pair $\{2,5\}$, a locally stable matching is reached. Note that the attractive pair $\{2,4\}$ is not accessible throughout.

$\mathcal{S}$ is $G(\mathcal{S}) = (V, L \cup \{\{w,w'\} \mid w,w' \in S \text{ and } S \in \mathcal{S}\})$. The direct extension of local stability for matchings would be to define an edge as visible when the nodes are within a hop-distance of 2 in the visibility graph – formally $\{u,v\} \in V_H \Rightarrow dist(\varphi(u),\varphi(v),G(\mathcal{S})) \leq 2$. However, this condition is arguably quite a stretch for larger coalitions, since here each connection for organizing the new coalition might only be a 2-hop path in terms of current visibility structure. As such, it might be the case that for each edge of H, the coalition C must rely on paths in $G(\mathcal{S})$ involving outside agents. We briefly consider this model in Section 4.2 and prove an exponential lower bound for very simple games of this kind.

For this reason, our main interest is in a more direct approach. In a *local coalition formation game*, an edge is termed *visible* if it exists in $G(\mathcal{S})$. Consequently, a coalition C is *accessible* in state $\mathcal{S}$ if there is some $H = (V_H, E_H) \in \mathcal{H}$ and some bijective map $\varphi : V_H \to C$ such that $\{u,v\} \in V_H \Rightarrow \{\varphi(u),\varphi(v)\} \in G(\mathcal{S})$. Again by abuse of notation we will denote this situation by $H(C) \in L \cup \mathcal{S}$. A coalition C is a *local blocking coalition* if it is a blocking coalition and accessible, and a *locally stable state* is a state $\mathcal{S}$ without a locally blocking coalition.

Observe that if $\mathcal{H}$ does not solely consist of cliques, this definition implies an exploration aspect of the resulting dynamics, similar to locally stable matching. However, with this direct approach to visibility the exploration aspect relies on the presence of larger coalitions. When we consider the special case of matching, the scenario becomes socially stable matching (and not locally stable matching).

2.2.3 Considerate Stability

In *considerate stable matching* we model the influence of friendship relations or contract partnerships on matching dynamics. The relations are again captured via an undirected network $N = (V, L)$ with link set L. In this case, however, the links express friendships and positive externalities. Every edge $e \in E$ is available for formation throughout. However, agents deviate and form a new match only if none of their friends (neighbors) in the network N suffers from this.

A blocking pair $e = \{u,v\} \in E$ is *not accessible* if at least one agent in e has a neighbor in N, for whom benefit decreases when $\{u,v\}$ is resolved. Such pairs are not available for resolution, even if they constitute blocking pairs. Formally, a pair $\{u,v\} \in E$ is *not accessible* in state M if there is an agent v' such that $\{u,v'\} \in M$, and (a) $\{u,v'\} \in L$ or (b) $\{v,v'\} \in L$. Otherwise, the pair is called accessible in M. A pair $\{u,v\}$ is a *considerate blocking pair* if it is a blocking pair and accessible. A *considerate improvement step* is the resolution of such a considerate blocking pair, and a state M without considerate blocking pair is a *considerate stable matching* (see Fig. 3 for an example).

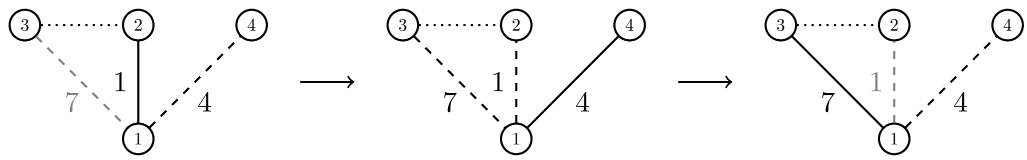

Figure 3: A sequence of considerate improvement steps. Vertices are agents, thick edges indicate current matching edges, dotted lines symbolize links. The dashed edges show additional pairs in E, where black edges are accessible and gray edges are not accessible in the current state. Benefits are indicated by numbers alongside the edges. From left to right: Initial matching $\{\{1,2\}\}$, where the pair $\{1,3\}$ is a blocking pair but not a considerate blocking pair because of the link $\{2,3\}$; resolution of considerate blocking pair $\{1,4\}$ removes $\{1,2\}$ and makes $\{1,3\}$ accessible; resolution of considerate blocking pair $\{1,3\}$, a considerate stable matching is reached.

For larger coalitions we generalize this definition in the following way. In *considerate coalition formation* there is a limit l for each agent on the number of neighbors it is willing to hurt when being part of a resolved blocking coalition. For considerate matching this number is set to $l = 0$. For matching games at most two drop their partners if a blocking pair is resolved. Thus, a limit of $l = 1$ would only have influence on constellations where one agent is friends with both his own matching partner and the current partner of the desired new matching partner. In contrast, for larger coalitions the number of agents influenced by an improvement step can grow quadratically. If we consider coalitions of size Δ, the formation of one new coalition can result in up to $\Delta(\Delta - 1)$ agents being dropped from the coalition structure. In these settings, it is reasonable to consider $l > 0$, that is, agents might be willing to hurt some small number of their friends by deviating to a new coalition.

Formally, for a coalition C in state $\mathcal{S}$, consider for $v \in C$ the penalty $pen(v, C) = |\{v' : v' \in C' \in \mathcal{S}, v' \notin C, C \cap C' \neq \emptyset, \{v, v'\} \in L\}|$, i.e., the number of neighbors in N that are part of some coalition in $\mathcal{S}$ and become single when C is added to and all overlapping coalitions are removed from the state. Then C is accessible in $\mathcal{S}$ if

$$\max_{v \in C} pen(v, C) \leq l.$$

A coalition C is a *considerate blocking coalition* if it is a blocking coalition and accessible, and a state $\mathcal{S}$ without considerate blocking coalitions is a *considerate stable state*.

2.2.4 Friendship Stability

In considerate stable matching we formulated a binary condition on externalities among friends. *Friendship stable matching* represents an aggregative setting where every agent $u \in V$ expresses for every agent $v \in V$ a numerical value $c_{u,v}$ to which extent it cares about the benefit of agent $v \in V$. We normalize the values by assuming $c_{u,u} = 1$. Agent u now strives to optimize the *perceived benefit*

$$b_p(u, \mathcal{S}) = \sum_{v \in V} c_{u,v} b(v, \mathcal{S}) = b(u, \mathcal{S}) + \sum_{v \in V \setminus \{u\}} c_{u,v} b(v, \mathcal{S})$$

in state $\mathcal{S}$. In contrast to all other scenarios treated above, this definition inherently relies on cardinal benefit values for coalitions and cannot be applied directly to ordinal preferences.

We directly introduce the general approach for coalitions or arbitrary size. In a *friendship coalition formation game*, a state $\mathcal{S}$ has a *perceived blocking coalition* $C \in \mathcal{C}$ if for every $v \in C$ we have

$$b_p(v, \mathcal{S}) < b_p(v, (\mathcal{S} \setminus \{C' \mid C \cap C' \neq \emptyset\}) \cup \{C\}),$$

that is, the perceived benefit needs to increase for every agent involved in C. A perceived blocking coalition is a set of agents that have an incentive to deviate and form C, since the trade-off (expressed

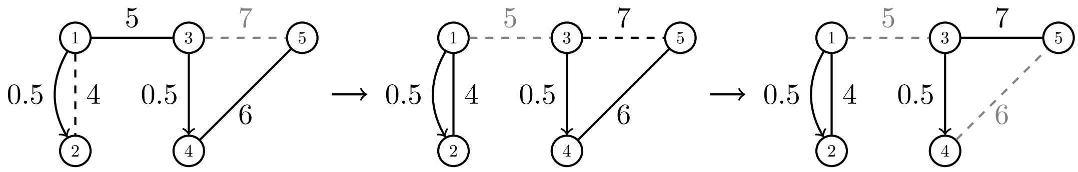

Figure 4: A sequence of perceived improvement steps. Vertices are agents, thick edges indicate current matching edges. The dashed edges show additional pairs in E, where black edges are perceived blocking pairs. Benefits are indicated by numbers alongside the edges. The values of $c_{u,v}$ are illustrated by arrows, where we omit an arrow for $c = 0$. The value of an arrow pointing from u to v gives the value of $c_{u,v}$. Here $c_{1,2} = c_{3,4} = 0.5$ while all other values are 0. From left to right: Initial matching $\{\{1,3\},\{4,5\}\}$, where the pair $\{3,5\}$ is not a perceived blocking pair since agent 3 gains less than it would lose by 5 leaving 4; resolution of considerate blocking pair $\{1,2\}$ of benefit 4 removes $\{1,3\}$ of benefit 5, represents an improvement in terms of perceived benefits for both 1 and 2, and results in $\{3,5\}$ becoming a perceived blocking pair; resolution of perceived blocking pair $\{3,5\}$, a friendship stable matching is reached.

in perceived benefit) between the improvement of their individual benefit and the deterioration of individual benefit of their friends is strictly positive for each of them. A *perceived improvement step* is the resolution of such a perceived blocking coalition. A state $\mathcal{S}$ without perceived blocking coalition is a *perceived* or *friendship stable state* (for an example see Fig. 4). Friendship stable matching was proposed and studied by Anshelevich et al [4].

Based on the values $c_{u,v}$, it is possible to capture a variety of phenomena – from altruism when $c_{u,v} \geq 1$ for all v to pure egoism for $c_{u,v} = 0$ for all v, or spiteful behavior with $c_{u,v} < 0$. Moreover, in general we could imagine asymmetric relations among agents when using $c_{u,v} \neq c_{v,u}$. In this paper, however, we restrict our analysis to symmetric and non-negative values, i.e., $c_{u,v} \geq 0$ and $c_{u,v} = c_{v,u}$ for all $u, v \in V$. If values are negative or asymmetric, we show in Appendix A.4 that there are instances in which friendship stable matchings do not exist.

3 Coalition Formation Games with Constraints

In this section we introduce a general model of *coalition formation games with constraints*, where the structural constraints are captured by two simple sets of rules: one determines which coalitions can be *generated* as a deviation in the current state, the other determines which coalitions are *dominated* and therefore not possible as deviation in the current state. For a given state, a given coalition is a blocking coalition if and only if it can be generated and is not dominated.

Our main results show that when generation and domination rules satisfy a small number of *consistency* properties, then for every initial state there is a path to stability of polynomial length. Moreover, for every initial state, if a state is reachable by a sequence of improvement steps, then there is also one such sequence with a *polynomial* number of steps. The set of axioms to define consistency is a minimal set – removing any one of the axioms allows for instances with initial states such that there is no path to stability (of polynomial length).

In Appendix A we show that the coalition formation games defined above give rise to consistent generation and domination rules. This includes all matching variants, as well as, social, considerate and friendship coalition formation games, which allow coalitions of larger size.

The case is different for local coalition formation games, which in general do not fall within the framework of consistent generation and domination rules. They will be the subject of Section 4, where we characterize the existence of polynomial paths to stability based on the set of formation graphs.

3.1 Coalition Formation Games with Consistent Constraints

In *coalition formation games with constraints* we supplement a hedonic coalition formation game with sets $T, D \subseteq \{(\mathcal{R}, C) \mid \mathcal{R} \subset \mathcal{C}, C \in \mathcal{C}\}$, where we term T the transition or *generation rules* and D the *domination rules*. Here $\mathcal{R}$ is a set of coalitions that serves as prerequisite (termed the *precondition*) and C is a coalition that (due to presence of $\mathcal{R}$) is generated or dominated, respectively.

More formally, given a state $\mathcal{S}$, if there is a rule $(\mathcal{R}, C) \in T$ such that $\mathcal{R} \subseteq \mathcal{S}$ and $C \notin \mathcal{S}$, then C is called a *candidate coalition*. For convenience, we exclude generation rules of the form $(\emptyset, C)$ from T and capture these rules in a set $\mathcal{C}_g \subseteq \mathcal{C}$ of *self-generating coalitions*. As no precondition has to be fulfilled to form it, a coalition $C \in \mathcal{C}_g$ is a candidate coalition for every state $\mathcal{S}$ with $C \notin \mathcal{S}$.

Similarly, given a state $\mathcal{S}$, if there is a rule $(\mathcal{R}, C) \in D$ such that $\mathcal{R} \subseteq \mathcal{S}$, then C is *dominated*. To express the underlying (correlated) preferences of the agents, we assume that D always includes at least the set $D_b = \{(\{C_1\}, C_2) \mid b(C_1) \geq b(C_2), C_1 \cap C_2 \neq \emptyset, C_1 \neq C_2\}$ of all *benefit domination rules*. Benefit domination rules simply capture the usual conditions of stability and myopic improvement, where coalition C_2 is dominated by another coalition C_1 if they overlap and the benefit of C_1 is at least that of C_2. Thus, existence of C_1 prohibits emergence of C_2.

A *feasible state* does not contain any dominated coalitions. If $\mathcal{S}$ is such that there is a rule $(\mathcal{R}, C) \in D$ with $\mathcal{R} \subset \mathcal{S}$ and $C \in \mathcal{S}$, then the state $\mathcal{S}$ is infeasible. We assume that infeasible states are excluded from the game. Note that benefit domination rules never make a state infeasible, because states cannot contain overlapping coalitions.

For a feasible state $\mathcal{S}$, a blocking coalition C is an undominated candidate coalition. D_b ensures that C can only become a blocking coalition if every agent $v \in C$ prefers resolution of C in terms of its benefit. In an improvement step, a blocking coalition is inserted and an auxiliary state $\mathcal{S}' = \mathcal{S} \cup \{C\}$ is created. Then, every dominated coalition in $\mathcal{S}'$ is removed. Note that D_b ensures that we remove at least every coalition that overlaps with C and has strictly smaller benefit. In particular, the new set of coalitions constitutes a feasible state.

We now define *consistent* generation and domination rules.

Definition (Consistent Generation Rules) *The generation rules of a coalition formation game with constraints are called* consistent *if* $T \subseteq \{(\{C'\}, C) \mid C, C' \in \mathcal{C}, C \cap C' \neq \emptyset\}$. *Every generation rule involves only a single coalition in the precondition, and it shares at least one agent with the candidate coalition.*

Definition (Consistent Domination Rules) *The domination rules of a coalition formation game with constraints are called* consistent *if* $D \subseteq \{(\mathcal{R}, C) \mid \mathcal{R} \subset \mathcal{C}, C \in \mathcal{C}, C \notin \mathcal{R}, \exists\, C' \in \mathcal{R} : C \cap C' \neq \emptyset\}$. *At least one coalition in* $\mathcal{R}$ *intersects with the dominated coalition.*

Observe, in particular, that benefit domination rules D_b are consistent. Note further that the definition of consistent generation rules is only meaningful since we exclude generation rules for self-generating coalitions. We note some direct consequences of consistency in the following observations.

Observation 1: For generation rules the definition requires that there is only a single coalition in the precondition and that this coalition overlaps with the candidate coalition. By the benefit domination rules, such a coalition is undominated only if the precondition coalition is of strictly smaller benefit than the candidate coalition. In the subsequent improvement step, we replace the precondition coalition with the candidate coalition. Therefore, the only meaningful generation rules are those where the precondition coalition is of strictly smaller benefit than the candidate coalition.

Observation 2: For domination rules we allow an arbitrary number of coalitions in the precondition, but at least one of them has to overlap with the dominated coalition. In consequence, when adding a blocking coalition C into $\mathcal{S}'$, several coalitions might become dominated. However, since $\mathcal{S}$ was a feasible state, there exist no overlaps between coalitions in $\mathcal{S}$. Hence, all dominated coalitions in $\mathcal{S}'$ must intersect C. Moreover, since C is a blocking coalition, it is undominated in $\mathcal{S}'$. Hence, all dominated coalitions in $\mathcal{S}'$ must intersect C and be of smaller benefit. Thus, we may restrict our attention to D_b to identify the dominated coalitions in $\mathcal{S}'$.

These observations show that for the implementation of the improvement step, we can concentrate on generation and domination rules that execute the standard benefit improvement of agents in overlapping coalitions. While we only need D_b for this, the rules in $D \setminus D_b$ can still make a significant difference in the definition of blocking coalition.

For example, in considerate stable matching, the intuition is that all matching edges can form, but agents only deviate if this improves their benefit and does not hurt their friends. Hence, we do not need generation rules (i.e., $T = \emptyset$) and have $\mathcal{C}^g = E$. A domination rule $(e', e) \in D \setminus D_b$ expresses that a matching edge $e = \{u, v\}$ is dominated if $e' = \{v, w\} \in M$, since in such a deviation u would "steal" the partner v from his friend w. Note that in this case, the agent that gets stolen yields an overlap $v = e' \cap e$, which makes the domination rules consistent.

As another example, in locally stable matching, the main idea is that generation of matching edges depends on the current state. Here the domination rules are simply the benefit domination rules D_b. A generation rule $(e', e) \in T$ expresses that a matching edge $e = \{u, v\}$ can be generated when $e' = \{v, w\} \in E$ and there is some $\{w, u\} \in L$. Such a deviation would imply agent v switches from partner u to u's friend w, which yields an overlap $v = e' \cap e$. Since $\{u, w\} \in L$ is a static condition, only existence of the single coalition $\{v, w\} \in M$ is needed to generate $\{u, v\}$. Hence, these generation rules are consistent. Note that in a variant where w can have two matching partners, $e' = \{v, w\} \in M$ *and* $e'' = \{u, w\} \in M$ together can lead to $\{u, v\}$ becoming a local blocking pair. As such, the generation rule $(\{e', e''\}, e)$ is not consistent, and there exist exponential lower bounds on the path to stability for this variant [31]

More broadly, in the above examples, socially and locally stable matching rely only on benefit domination $D = D_b$. For socially stable matching we simply have $\mathcal{C}^g = E$ and $T = \emptyset$. For locally stable matching we use consistent generation rules as specified above. In considerate and friendship stable matching, we do not make use of generation rules (i.e., $T = \emptyset$). Every every coalition can always be generated $\mathcal{C}^g = E$. Here the domination rules are consistent and express the externality constraints, thereby making certain deviations unavailable. For considerate stable matching, we saw above that the domination rules are of the form (e', e) with a single coalition in the precondition. In contrast, friendship stable matching has an aggregative nature of externality. As such, there are cases in which agents u and v have no incentive to deviate to $\{u, v\} \notin M$ only if two edges $e' = \{v, w\}$ and $e'' = \{u, z\}$ are *both present* in M. Hence, we obtain consistent domination rules $(\{e', e''\}, e)$ with more than one coalition in the precondition.

Note that the framework of consistent rules is much more powerful. For example, it allows to study instances, in which we can combine the effects outlined above into one instance of, say, "considerate locally" stable matching. There is a single agent set and two sets of links, where one set is used to express considerate stable matching and the other one locally stable matching. The rule sets T and D for this game are unions of the corresponding rules for the instances of considerate and locally stable matching. Obviously, T and D for the combined instance are again consistent.

3.1.1 Short Paths to Stability

Our existence proof for a short path to stability generalizes a previous construction for locally stable matchings [31]. To explain the construction of a short sequence, we will find it convenient to think of the dynamics in the form of moving set of tokens over a graph. More formally, we equivalently represent a coalition formation game with consistent constraints as a directed *coalition movement* *hypergraph* $G_{mov} = (V_{mov} \cup V_{mov,g}, T_{mov} \cup D_{mov})$. The vertices are the possible coalitions: $V_{mov,g} = \{v_C \mid C \in \mathcal{C}_g\}$ contains a vertex for every self-generating coalition, and $V_{mov} = \{v_C \mid C \in \mathcal{C} \setminus \mathcal{C}_g\}$ a vertex for every other possible coalition.

We represent a coalition structure $\mathcal{S}$ as a set of tokens being placed on the vertices. We put a single token on every vertex v_C with $C \in \mathcal{S}$. Resolution steps are simulated by the creation of new tokens or the (re-)movement of existing ones. To correctly express the consequences of an improvement step in the game, we use two edges sets corresponding to generation and domination rules. The directed *transition edges*

$$T_{mov} = \{(v_{C'}, v_C) \mid (\{C'\}, C) \in T, b(C') < b(C)\} \ .$$

capture generation rules. By Observation 1, note that T_{mov} can be restricted to transition edges resulting from generation rules $(\{C'\}, C)$ with $b(C') < b(C)$. The transition edges induce a directed acyclic graph (DAG).

In addition, the directed *domination hyperedges*

$$D_{mov} = \{(\{v_{C'} \mid C' \in \mathcal{R}\}, v_C) \mid (\mathcal{R}, C) \in D\} \ .$$

capture the domination rules. They express on which vertices we cannot simultaneously put tokens. In particular,

$$D_{mov,b} = \{(\{C'\}, C) \mid C \neq C' \in \mathcal{C}, C \cap C \neq \emptyset, b(C') \geq b(C)\} \subseteq D_{mov}$$

is the set of all domination edges given by benefit domination rules. We call a vertex v in G_{mov} *undominated* if for every hyperedge $(U, v) \in D_{mov}$ at least one vertex in U has no token. For the initial state $\mathcal{S}_0$ we put a token on every vertex in $V_{mov,0} = \{v_C \mid C \in \mathcal{S}_0\}$. Since $\mathcal{S}_0$ is a feasible state, all vertices with tokens are undominated.

Now consider an improvement step that resolves a blocking coalition C. We express the improvement step by the following *token adjustment*. Since C is added to the state, we need to put a token on v_C. Note that a blocking coalition is not part of the current state and undominated, i.e., v_C must have no token and be undominated in G_{mov}. If $C \in \mathcal{C}_g$, then we can simply create the coalition and thus create a new token at $v_C \in V_{mov,g}$. Otherwise, C is generated as a result of some generation rule $(C', C) \in T$, where $C' \in \mathcal{S}$ has a token. Due to Observation 1 the precondition coalition C' will be removed. As such, we *move the token* from $v_{C'}$ along the corresponding transition edge to v_C. Finally, we delete all dominated coalitions from the state. Note that any dominated coalition C' results from a domination rule $(\mathcal{R}, C')$. Since all the coalitions in $\mathcal{R}$ exist, all vertices in $U = \{v_S \mid S \in \mathcal{R}\}$ have tokens. Hence, $v_{C'}$ is dominated due to the corresponding edge $(U, v_{C'}) \in D_{mov}$. We delete every token at dominated vertices. As a result, if we execute this token adjustment, the location of tokens correspond exactly to the existing coalitions after the resolution of blocking coalition C.

Observation 3: The token adjustment in the coalition movement hypergraph correctly captures the creation and deletion of coalitions in every improvement step of the coalition formation game with constraints.

To prove the existence of a short sequence of improvement steps to a stable state, we show that only a polynomial number of token adjustments is necessary to reach a stable state. To compute the sequence consider Algorithm 1.

Phase 1 is a greedy procedure. Phases 2 and 3 generate a new token at a vertex of $v \in V_{mov,g}$ and move it to a vertex u such that by reaching u, an existing token gets removed. Vertex u is *undominated reachable* from v if there is a directed path $(v = v_1, e_1, v_2, \ldots, e_{\ell-1}, v_\ell = u)$ in T_{mov} such that v_{i+1} is undominated if v_i has a token (along with all other vertices currently having tokens), for every $i = 1, \ldots, \ell - 1$. Note that for any vertex $v \in V_{mov,g}$ the set of vertices that are undominated reachable from v can be computed efficiently by BFS.

Algorithm 1: Short Path to Stability

Input: Hypergraph G_{mov} based on a coalition formation game with consistent rules, initial tokens on $V_{mov,0}$

Output: Polynomial sequence of token adjustments ending in a stable state

Phase 1: In a round of phase 1, check whether there is a transition edge from a vertex with a token to an undominated one. If this is the case, we move the token along the transition edge, remove tokens at dominated vertices, and start the next round of Phase 1. Phase 1 ends when there is no such transition edge. Then proceed to Phase 2

Phase 2: In a round of phase 2, consider the set $V_{cand,g} \subseteq V_{mov,g}$ of undominated vertices without tokens. For each vertex $v \in V_{cand,g}$, check via BFS if there is a vertex u that is undominated reachable from v, and when a token is created at v that leads to removal of an existing token. If for all $v \in V_{cand,g}$ no such vertex u exists, proceed to Phase 3. Otherwise, create a token at v and move it along the undominated path to u. Then restart Phase 1.

Phase 3: In a round of phase 3, again consider the set $V_{cand,g}$. For each $v \in V_{cand,g}$, compute the set $V_{reach}(v)$ of vertices that are undominated reachable from v. Pick a vertex $u_C \in V_{reach} = \bigcup_{v \in V_{cand,g}} V_{reach}(v)$ with maximum benefit $b(C)$. Create a token at the corresponding $v \in V_{cand,g}$ and move it along the undominated path to u_C. Then, start a new round of Phase 3. Phase 3 ends when V_{reach} is empty.

Lemma 1 *If the tokens are located at vertices that represent a feasible initial state, Algorithm 1 constructs $O(nm^2)$ token manipulations such that tokens are placed at vertices that represent a stable state. The algorithm runs in polynomial time.*

Proof. We consider each phase separately. In each round of Phase 1, we move an existing token to a vertex of higher value by using a transition edge. Hence, the total number of tokens does not increase. There are at most n tokens, T_{mov} represents a DAG, and each token can move to at most m vertices. Hence, Phase 1 finishes after at most $n \cdot (m-1)$ rounds, where each round corresponds to a single improvement step.

Phase 2 can be seen as an extension of Phase 1. In a round of Phase 2, we create a new token and move it to a vertex u where it leads to removal of an existing token at some vertex, say, u'. Due to Observation 2, u must correspond to a coalition with higher benefit than u'. Thus, at the end of the round, the number of tokens is not increased, and one token has moved from u' to u with higher coalition benefit. Note that a round of Phase 2 corresponds to at most m improvement steps, which are required to move the token to u. Consequently, the overall number of rounds in Phase 1 and 2 is at most $n \cdot (m-1)$, where each round corresponds to at most m improvement steps.

When Phase 2 terminates, consider the set $V_{2,end}$ of vertices that have tokens. The tokens at $V_{2,end}$ cannot simply be moved to vertices with higher benefit by transition edges. Moreover, by introducing new tokens at $V_{mov,g}$ and moving them via undominated reachable vertices, we will never reach a vertex such that some $v \in V_{2,end}$ becomes dominated (otherwise Phase 2 would not have ended). We say that the tokens in $V_{2,end}$ correspond to *stable coalitions*: No sequence of token adjustments in the graph (improvement step in the game) can lead to any of these tokens (coalitions) being removed. While the coalitions corresponding to $V_{2,end}$ are stable, they do not represent a stable state. There might be additional tokens that can be created. In particular, in Phases 1 and 2 overall the number of tokens is non-increasing.

Now in Phase 3, we create and move additional tokens in G_{mov}. In a round of Phase 3, we generate a token at some vertex in $V_{mov,g}$ and move it such that it reaches the undominated reachable vertex of highest benefit. For that reason, and since the set of undominated reachable vertices only shrinks in the presence of additional tokens, the created token cannot be (re)moved in subsequent rounds. Hence, each round of Phase 3 expands the set of tokens that correspond to stable coalitions by one. Consequently, in Phase 3 there are at most n rounds, where each round corresponds to at most m improvement steps. The phase finishes with a set of tokens corresponding to a stable state.

For efficient computation of the sequence, the relevant tasks are constructing the graph G_{mov}, checking edges in T_{mov} for possible improvement of tokens, and computing the reachable subgraphs of $V_{mov,g}$. All these tasks can be executed in time polynomial in n, m, $|T|$ and $|D|$ using standard algorithmic techniques. □

Combining Observation 3 with Lemma 1 we obtain the desired result for consistent generation and domination rules:

Theorem 1 *In every correlated coalition formation game with constraints and consistent generation and domination rules, for every initial structure $\mathcal{S}$ there is a sequence of $O(nm^2)$ improvement steps which results in a stable coalition structure. The sequence can be computed in polynomial time.*

3.1.2 Short Paths to Every Reachable Coalition Structure

The previous theorem shows the existence and, in particular, an algorithm for efficient computation of a short path to some stable state. In this section, our first result shows that consistent rules always imply short sequences of improvement steps. More formally, we show that assuming a coalition structure is reachable from a given initial state, then it can be reached in a polynomial number of improvement steps. However, our second result shows that *deciding* reachability of a *given* stable matching from a given initial state is NP-hard. In fact, the decision problem is also in NP (and, hence, NP-complete) – if the matching is reachable, our first result implies there is a polynomial path to stability, which is a polynomial-time checkable proof. This hardness result holds in all examples of two-sided matching games with constraints discussed above. Moreover, it extends to standard domains like two-sided matching with correlated preferences and ties, or two-sided matching with strict preferences.

To prove short sequences to all reachable states, we observe in the subsequent Theorem 2 that if there is a longer sequence of improvement steps, it contains steps which are not relevant for reaching the final state and can be omitted.

Note that Theorem 2 is somewhat more general and implies parts of Theorem 1. However, the latter presents a polynomial-time algorithm, and thus we decided to present it first. More precisely, in games with consistent domination rules, every sequence of improvement steps is finite due to a lexicographic improvement property (see, in particular, Proposition 3 below). Hence, an initial (possibly exponentially long) sequence to some stable state can always be constructed by arbitrarily executing improvement steps as long as possible. Then Theorem 2 can be applied to yield a short path to stability and implies the existence conditions of Theorem 1. However, this is not an efficient algorithm since the initial sequence might be too long. Instead, Theorem 1 shows how to derive a short path to stability in time polynomial in the size of the description of the game.

Theorem 2 *In a correlated coalition formation game with constraints and consistent generation and domination rules, for every coalition structure $\mathcal{S}$ that is reachable from some initial state $\mathcal{S}_0$ through some sequence of improvement steps, there is also a sequence of polynomially many improvement steps from $\mathcal{S}_0$ to $\mathcal{S}$.*

Proof. Consider a sequence I of improvement steps that transforms $\mathcal{S}_0$ into $\mathcal{S}$. The proof uses the interpretation of tokens in the coalition formation graph introduced in the last section. At any point in time, the coalition structure is represented via tokens being placed on every vertex corresponding to a coalition in the structure.

Consider a single improvement step. In the first part, either a new token τ is introduced (when generating a coalition in $\mathcal{C}^g$) or an existing token τ is moved (when applying a generation rule) from the vertex of the precondition coalition to the vertex of the target coalition. In the second part, the application of domination rules removes tokens from the graph (corresponding to dominated coalitions that are removed from the coalition structure). We say token τ *collects* all tokens that are removed in the second part. Note that all tokens collected by τ are located at vertices with strictly smaller benefit.

Now consider the sequence I of improvement steps. We classify the tokens that appear during the sequence. Each token that is initially placed for some coalition in $\mathcal{S}_0$ is called *initial.* Each token that represents a coalition of $\mathcal{S}$ in the end is *final.* Note that there can be tokens that are both initial and final. For a token that is neither initial nor final, we say it is *justified* if during the time of its existence it collects at least one other token. Finally, all remaining tokens are *superfluous.*

Consider a superfluous token. It corresponds to a subsequence of (not necessarily consecutive) improvement steps, which create some coalition, apply a number of improvement steps, and then delete the resulting coalition due to creation of a dominating one (which was generated from a different precondition coalition). None of these steps is necessary to reach $\mathcal{S}$ from $\mathcal{S}_0$ – the coalitions that are created during these steps do not alter or remove any other coalitions, and they do not lead to a coalition that survives in the end. If we just omit all improvement steps that correspond to creation and movement of a superfluous token, we obtain a shorter feasible improvement sequence I' that starts in $\mathcal{S}_0$ and ends in $\mathcal{S}$.

Suppose we remove all superfluous tokens. If a justified token τ in I was collecting only superfluous tokens, it becomes superfluous. Hence, we keep iterating the removal of superfluous tokens until only initial, final, and justified tokens remain. In the resulting *compact sequence*, a token that is initial but not final must be collected by a justified token. When it is collected, the justified token must be located at a vertex with higher benefit. If the collecting token is not final, it must be collected by another justified or final token, again at a vertex with higher benefit, and so on, until the collecting token is final. In addition, there might be final tokens that do not collect any initial or justified tokens.

There are $|\mathcal{S}|$ final tokens, and each one takes at most m improvement steps to reach the final location in the graph. Initial or justified tokens take at most m steps each until they are collected. If a token is collected, the collecting token is located at a vertex with higher benefit. Hence, if we track the collection events from an initial token to the collecting justified token, to the collecting justified token, etc, to the collecting final token, then there are at most m such collection events for each initial token. Hence, for each initial token, the sequence has at most m^2 improvement steps. In conclusion, this proves that any compact sequence consists of at most $|\mathcal{S}_0| \cdot m^2 + |\mathcal{S}| \cdot m \leq nm^2 + nm$ steps. □

While every reachable state can be reached fast, it is not easy to decide whether a state is reachable at all. This even holds if we only consider reachability of *stable* states. Our next theorem presents a generic reduction which can be adjusted to prove the NP-completeness of this problem for socially, locally, considerate, and friendship matching, even in the two-sided case. It also applies to ordinary two-sided stable matching games that have either correlated preferences with ties, or non-correlated strict preferences. Note that the problem is trivially solvable for stable matching with strict correlated preferences (without ties), as there is a unique stable matching that can always be reached using a simple greedy sequence [1, 2]. Moreover, it trivially implies hardness also for all variants with more than two agents per coalition. It even applies to local coalition formation games studied in the subsequent Section 4, since this model includes social stable matching as a special case (as discussed in Section 2.2.2).

Theorem 3 *It is* NP*-complete to decide if for a given matching game, initial matching* M_0 *and* *stable matching M, there is a sequence of improvement steps leading from M_0 to M. This holds even for two-sided games with strict correlated preferences and*

1. *socially stable matching,*
2. *locally stable matching,*
3. *considerate matching, and*
4. *friendship matching for symmetric c-values in* $[0,1]$.

In addition, it holds for ordinary two-sided stable matching and

5. *correlated preferences with ties,*
6. *strict preferences.*

Proof. The proof is via reduction from 3SAT. We rely on a central construction for all cases. We then adapt the structure of the clause gadgets to the specific settings. Each clause gadget will have the property that one particular agent v_C has to be matched to an agent of the central construction at some (arbitrary) point during the sequence and subsequently must be left single again. Otherwise, the clause gadget cannot be transformed into the state of the desired final matching.

We first outline the universal proof construction including only the one particular agent v_C per clause C. We show that it is NP-hard to decide whether there is a sequence of improvement steps such that each of the clause vertices gets matched and afterwards unmatched at least once. Our proof is then completed by providing for every setting the exact clause gadget and explain why it is necessary to match v_C to some agent outside the clause gadget to reach the final state.

Given a 3SAT formula with k variables $x_1, \ldots, x_k$ and l clauses $C_1, \ldots, C_l$, where clause C_j contains the literals $l_{1,j}, l_{2,j}$ and $l_{3,j}$, for the central construction we have

$$\begin{aligned} U = &\quad \{u_{x_i} \mid i = 1 \ldots k\} \cup \{u_{\overline{x}_i} \mid i = 1 \ldots k\} \cup \{v_{C_j} \mid j = 1 \ldots l\}, \\ W = &\quad \{w_{x_i} \mid i = 1 \ldots k\} \cup \{w_{\overline{x}_i} \mid i = 1 \ldots k\}. \end{aligned}$$

Further $E = E_1 \cup E_2 \cup E_3$ with

$$\begin{aligned} E_1 = &\quad \{\{u_{x_i}, w_{x_i}\}, \{u_{\overline{x}_i}, w_{\overline{x}_i}\} \mid i = 1 \ldots k\}, \\ E_2 = &\quad \{\{u_{x_i}, w_{\overline{x}_i}\}, \{u_{\overline{x}_i}, w_{x_i}\} \mid i = 1 \ldots k\}, \text{ and} \\ E_3 = &\quad \{\{v_{C_j}, w_{l_{i,j}}\} \mid j = 1 \ldots l, i = 1 \ldots 3\}, \end{aligned}$$

and benefits as given in Table 1.

Table 1: Edge benefits in the central construction of Theorem 3

U	W	$b(\{u,w\})$	
v_{C_j}	$w_{l_{i,j}}$	$i \cdot l + j$	$j = 1 \ldots l$, $i = 1 \ldots 3$
u_{x_i}	$w_{\overline{x}_i}$	$4l + i$	$i = 1 \ldots k$
$u_{\overline{x}_i}$	w_{x_i}	$4l + k + i$	$i = 1 \ldots k$
u_{x_i}	w_{x_i}	$4l + 2k + i$	$i = 1 \ldots k$
$u_{\overline{x}_i}$	$w_{\overline{x}_i}$	$4l + 3k + i$	$i = 1 \ldots k$.

For a schematic example, see Figure 5. In the case of locally and socially stable matching, we will have social links between all vertices of U and W to make sure that all edges of E are available for matching at all times. In the case of friendship matching we set all c to 0 to ensure that benefit is also perceived benefit.

We start from $M_0 = E_2$ and need to decide whether we can reach $M = E_1$. Note that this is the only stable state of this graph. For any i, we can either first match w_{x_i} to $u_{\overline{x_i}}$, and then match u_{x_i} to $w_{\overline{x_i}}$, or the other way around. In any case, after the first step, either $w_{\overline{x_i}}$ or w_{x_i} will be available to match to some v_{C_j}. Suppose w_{x_i} first matches to $u_{\overline{x_i}}$, then $w_{\overline{x_i}}$ is single and can become matched to v_{C_j}. Now, however, $w_{\overline{x_i}}$ obtains more benefit when he matches to u_{x_i}. In this way, v_{C_j} becomes unmatched again. Similarly, when $w_{\overline{x_i}}$ first matches to u_{x_i}, then v_{C_j} can become matched to w_{x_i} and unmatched later on when w_{x_i} matches to $u_{\overline{x_i}}$. In this way, we fulfill the condition that the agent v_{C_j} becomes matched and unmatched again.

The hardness emerges from the decision whether we can construct a sequence which involves matching and unmatching all the v_{C_j} and reaches the desired final matching. Note that we have to create some edge $\{v_{C_j}, w_{l_{i,j}}\}$ of E_3 for every clause C_j. In the beginning, all those edges are blocked through E_2. During the sequence, per variable we can switch one edge of E_2 to E_1, freeing the other w-agent. Then, this agent can be used to sequentially match to v_{C_j} for all adjacent clauses in increasing order before creating the second edge of E_1. But the w-agent that switched first remains blocked and thus cannot be used for matching to any of the v_{C_j}. Hence, the choice whether to first create $\{u_{x_i}, w_{x_i}\}$ or $\{u_{\overline{x}_i}, w_{\overline{x}_i}\}$ can be interpreted as the choice whether to set x_i true or false (by creating the "opposite" edge first). This implies the equivalence to solving the 3Sat formula.

Let us now formally prove correctness of the reduction.

Assume that the 3Sat formula is satisfiable. Then pick a satisfying assignment, and for each variable generate the edges of E_1 which symbolize the inverses of the variable assignment. Now, for every variable the w-agent corresponding to the assigned value is unmatched. We sequentially generate the incident edges leading to the clause variables in increasing order starting from the smallest unblocked edge. For every clause, at least one literal is satisfied, and the edges are created in increasing order. Thereby, at the end of this sequence all vertices v_{C_j} were matched at least once. It remains to generate the second edge for every variable gadget. This yields a sequence to M of the desired form.

Assume that we can reach M from M_0 with a sequence, in which we match and unmatch each v_{C_j} at least once. For each clause C_j, pick an agent $w_{l_{i,j}}$ which was matched to v_{C_j}. We claim that for no variable x_i both vertices w_{x_i} and $w_{\overline{x}_i}$ are picked: In the beginning, both vertices are matched through an edge that is more preferred than any edge to a clause agent. Thus, to match one of these vertices to some v_{C_j}, it first has to become single, that is, its matching partner $u_{\overline{x}_i}$ respectively u_{x_i} has to deviate to a better partner. But the only better partner for $u_{\overline{x}_i}$ is $w_{\overline{x}_i}$ and the only better partner for u_{x_i} is w_{x_i}. Furthermore, both edges then represent stable choices, since they are the top choice of both partners. Hence, to make w_{x_i} available, we have to block $w_{\overline{x}_i}$ for the rest of the sequence, and to make $w_{\overline{x}_i}$ available we have to block w_{x_i} for the rest of the sequence. Since at most one w-agent of each variable is chosen, we assign each of these variables a value according to the chosen agent. To each remaining variable, we assign an arbitrary value. This implies that for each clause at least one literal is evaluates to true, i.e., the formula is satisfied.

Finally, we design appropriate clause gadgets for each case:

1. For *socially stable matching* we add an agent y_{C_j} to W and an edge $\{v_{C_j}, y_{C_j}\}$ of benefit j to E for every clause C_j. Further we also add all the edges $\{v_{C_j}, y_{C_j}\}$ to the initial state M_0 but keep M. Note that we did not add any social links for y_{C_j}. Thus, M is stable and can be reached if and only if we rematch every y_{C_j} at least once (and hence delete $\{v_{C_j}, y_{C_j}\}$).

2. For *locally stable matching* we add the same vertices and edges as for socially stable matching. We add y_{C_j} to W and $\{v_{C_j}, y_{C_j}\}$ of benefit j to E for every clause C_j. Then we add all the

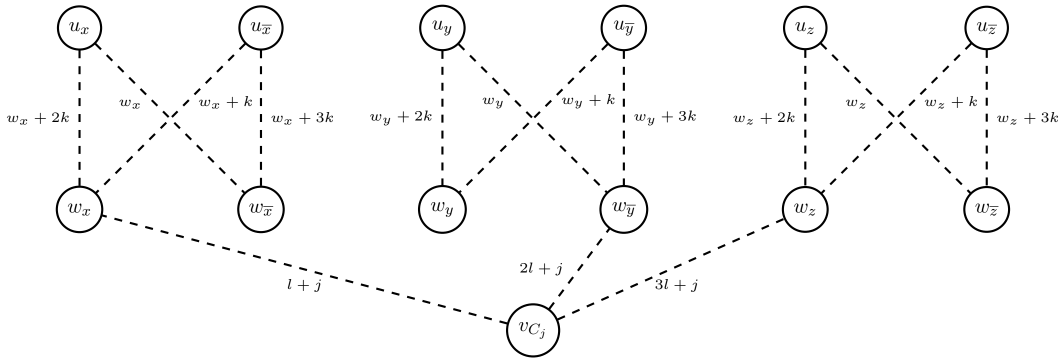

Figure 5: Central gadget with variables x, y, z and clause $C_j = x \vee \overline{y} \vee z$. Dashed edges represent possible matches, edge labels indicate match benefits.

edges $\{v_{C_j}, y_{C_j}\}$ to the initial state M_0. M stays the same. Again, M is stable and can be reached if and only if we rematch every y_{C_j} at least once (and hence delete $\{v_{C_j}, y_{C_j}\}$).

3. For *considerate matching* we add two vertices y_{C_j} and y'_{C_j} to W and edges $\{v_{C_j}, y_{C_j}\}$ of benefit $j - \frac{1}{2}$ and $\{v_{C_j}, y'_{C_j}\}$ of benefit j to E for every clause C_j. Further, we also add all the edges $\{v_{C_j}, y_{C_j}\}$ to the initial state M_0 and all the edges $\{v_{C_j}, y'_{C_j}\}$ to M. Finally we introduce a social link between y_{C_j} and y'_{C_j}. Now v_{C_j} cannot switch from y_{C_j} to y'_{C_j}, as y'_{C_j} is friends with y_{C_j} and would thus reject v_{C_j}. But if v_{C_j} is single, y'_{C_j} does not reject v_{C_j}. Hence, again for every clause C_j, v_{C_j} must be matched to some agent outside the clause gadget and then become unmatched to reach M.

4. For *friendship matching* we add two vertices y_{C_j} and y'_{C_j} to W and edges $\{v_{C_j}, y_{C_j}\}$ of benefit $j - \frac{1}{2}$ and $\{v_{C_j}, y'_{C_j}\}$ of benefit j to E for every clause C_j. The only friendship value $\neq 0$ is $c_{v_{C_j}, y_{C_j}} = \frac{1}{2j-1}$. We add all the edges $\{v_{C_j}, y_{C_j}\}$ to the initial state M_0 and all the edges $\{v_{C_j}, y'_{C_j}\}$ to M. Note that by the choice of $c_{v_{C_j}, y_{C_j}}$, the perceived benefit for v_{C_j} from $\{v_{C_j}, y_{C_j}\}$ now is $(1 + c_{v_{C_j}, y_{C_j}})(j - \frac{1}{2}) = j - \frac{1}{2} + (j - \frac{1}{2})\frac{1}{2j-1} = j = b(\{v_{C_j}, y'_{C_j}\})$, that is, there is a tie in v_{C_j}'s preference list regarding y_{C_j} and y'_{C_j}. Hence, M is stable but v_{C_j} will not switch directly from y_{C_j} to y'_{C_j}. However, once v_{C_j} becomes single, we can match it to y'_{C_j} as desired.

5. For *correlated matching with ties* we add two vertices y_{C_j} and y'_{C_j} to W and edges $\{v_{C_j}, y_{C_j}\}$ and $\{v_{C_j}, y'_{C_j}\}$ – both of benefit j – to E for every clause C_j. Further, we also add all the edges $\{v_{C_j}, y_{C_j}\}$ to the initial state M_0 and all the edges $\{v_{C_j}, y'_{C_j}\}$ to M. Then v_{C_j} does not switch from y_{C_j} to y'_{C_j} as it yields no improvement. However, if v_{C_j} is single, we can match it to y'_{C_j}.

6. For *matching with strict preferences*, we note that all edge values in the central gadget are distinct. Hence, we can derive a strict preference order over all possible matching partners for each agent. Now, for each clause C_j, we add one agent v'_{C_j} to U and two vertices y_{C_j} and y'_{C_j} to W and edges $\{v_{C_j}, y_{C_j}\}$, $\{v_{C_j}, y'_{C_j}\}$, $\{v'_{C_j}, y_{C_j}\}$ and $\{v'_{C_j}, y'_{C_j}\}$ to E. For v_{C_j} we add $y_{C_j} >_{v_{C_j}} y'_{C_j}$ to the bottom of the preference list, that is, all vertices of the central gadget are preferred. For the other preferences we have

$$y'_{C_j} >_{v'_{C_j}} y_{C_j}, v'_{C_j} >_{y_{C_j}} v_{C_j} \text{ and } v_{C_j} >_{y'_{C_j}} v'_{C_j}.$$

To M_0 we add $\{v_{C_j}, y_{C_j}\}$ and $\{v'_{C_j}, y'_{C_j}\}$ and to M we add $\{v_{C_j}, y'_{C_j}\}$ and $\{v'_{C_j}, y_{C_j}\}$. The clause gadget has two stable states: $\{\{v_{C_j}, y_{C_j}\}, \{v'_{C_j}, y'_{C_j}\}\}$ and $\{\{v_{C_j}, y'_{C_j}\}, \{v'_{C_j}, y_{C_j}\}\}$. To switch, again we first have to destabilize the initial state by matching v_{C_j} to some agent of the central gadget and then leave v_{C_j} single. Then y'_{C_j} can switch to its preferred choice v_{C_j} which frees v'_{C_j} for y_{C_j}, resulting in the desired final state.

□

For locally stable matching, the problem of reaching a given stable matching is known to be NP-hard, even from the empty initial matching $M_0 = \emptyset$ [37]. In contrast, our reduction here applies a unified framework to many different variants of the problem. In this level of generality, we need $M_0 \neq \emptyset$, since, e.g., for ordinary two-sided stable matching reaching a given stable matching from $M_0 = \emptyset$ is trivial.

3.1.3 Inconsistent Generation and Domination Rules

In the presence of the positive results for consistent generation and domination rules, it is a natural question if the consistency conditions are also necessary. In this section, we analyze each axiom individually and prove that every relaxation admits instances where stable states cannot be reached within a polynomial number of steps. For relaxed domination rules we even give an example without a stable state where improvement dynamics cycle.

Proposition 1 *Suppose the generation rules are allowed to contain more than one coalition in the precondition and the domination rules are consistent. Then there are instances and initial states such that every sequence to a stable state requires an exponential number of improvement steps.*

Proof. Given an instance of locally stable matching with graph $G = (V, E)$, (social) links L, correlated preferences $b(e)$ based on edge benefits, we define the parameters of the framework as follows. The set of agents is V, the set of possible coalitions is $\mathcal{C} = E$. The coalitions which can always be generated are the ones connected by at most 2 links, i.e., $\mathcal{C}_g = E \cap \{\{u, v\} \mid dist_L(u, v) \leq 2\}$. The benefits remain the same. For the generation rules, we have $T = T_1 \cup T_2$, where

$$T_1 = \{(\{\{u, v\}\}, \{u, v'\}) \mid \{u, v\}, \{u, v'\} \in E, \{v, v'\} \in L\}$$
$$T_2 = \{(\{\{u, v\}, \{v, v'\}\}, \{u, v'\}) \mid \{u, v\}, \{u, v'\}, \{v, v'\} \in E\}$$

Here T_1 captures accessible pairs with 2 hops composed of one matching edge and one link and T_2 captures accessible pairs where both hops are composed of matching edges. The latter generation rules are obviously using two coalitions as precondition. For the domination rules we have $D = D_b$, i.e., only the necessary preference-based domination. □

Proposition 2 *Suppose the generation rules are allowed to have non-overlapping precondition- and target-coalitions and the domination rules are consistent. Then there are instances and initial states such that every sequence to a stable state requires an exponential number of improvement steps.*

Proof. We will build an instance with an initial state that requires a unique exponential improvement sequence to a stable state. The instance consists of k gadgets of 9 agents each and a suitable initial state. The main property in each gadget i is that to create coalition $C_{6,i}$, we need to generate $C_{1,i}$ twice. Further, the gadget will not reach a stable state unless $C_{6,i}$ exists.

For gadget i, we define the set of agents $N_i = \{0_i, \ldots, 8_i\}$, and the set of possible coalitions $\mathcal{C}_i = \{C_{1,i}, \ldots, C_{6,i}\}$ with $C_{1,i} = \{0_i, 1_i, 2_i\}, C_{2,i} = \{1_i, 3_i\}, C_{3,i} = \{3_i, 4_i, 5_i\}, C_{4,i} = \{4_i, 6_i\}, C_{5,i} = \{2_i, 6_i, 7_i\}, C_{6,i} = \{5_i, 7_i, 8_i\}$. The benefits are $b(C_{1,i}) = x_i + 1$, $b(C_{2,i}) = x_i + 2$, $b(C_{3,i}) = x_i + 4$,

$b(C_{4,i}) = x_i + 3$, $b(C_{5,i}) = x_i + 2$ and $b(C_{6,i}) = x_i + 5$ with $x_i = 5(i-1)$. The generation rules are

$$T_i = \{\{\{C_{1,i}\}, C_{2,i}\}, \{\{C_{1,i}\}, C_{5,i}\}, \{\{C_{2,i}\}, C_{3,i}\}, \{\{C_{3,i}\}, C_{1,i}\}, \{\{C_{4,i}\}, C_{1,i}\}, \{\{C_{5,i}\}, C_{6,i}\}\},\quad \text{if } i = 1$$

$$T_i = \{\{\{C_{1,i}\}, C_{2,i}\}, \{\{C_{1,i}\}, C_{5,i}\}, \{\{C_{2,i}\}, C_{3,i}\}, \{\{C_{3,i}\}, C_{4,i-1}\}, \{\{C_{4,i}\}, C_{4,i-1}\}, \{\{C_{5,i}\}, C_{6,i}\}\},\ \text{if } i > 1.$$

For the domination rules we have $D = D_b$, i.e., only the necessary preference-based domination.

We introduce k such gadgets. For every $i = 1, \ldots, k-1$, we merge agent 8_i with 0_{i+1} to become the same agent. Also, we add a generation rule $\{\{C_{1,i+1}\}, C_{6,i}\}$. Consider the coalition movement graph. Figure 6 shows the graph for the first two gadgets to visualize the dynamics inside the gadgets and their interaction.

As initial coalition structure we have $\{C_{4,k}\}$. The proof follows from the property that creation of $C_{6,k}$ (without which the state is not stable), $C_{1,1}$ has to be created at least 2^k times.

Initially, every $C_{4,i}$ for $i > 1$ can only be used to generate $C_{4,i-1}$. Note that in this step $C_{4,i}$ is not deleted. Thus, in the beginning, the unique start of the sequence is to create $C_{4,k-1}, C_{4,k-2}, \ldots, C_{4,1}$, and none of these coalitions is deleted. When $C_{4,1}$ exists, the unique next step is creation of $C_{1,1}$. Next, as $C_{5,1}$ is dominated by $C_{4,1}$, the only option is to generate $C_{2,1}$ and thus remove $C_{1,1}$. Then we can only generate $C_{3,1}$ while deleting $C_{2,1}$ and $C_{4,1}$. With the remaining coalition $C_{3,1}$ we create $C_{1,1}$ a second time, which now can be used to create $C_{5,1}$ and delete $C_{1,1}$ again. Next, we can only create $C_{6,1}$, which causes the deletion of $C_{3,1}$ and $C_{5,1}$.

Hence, we need the existence of $C_{4,1}$ to create $C_{1,1}$, which is needed to create $C_{5,1}$ and $C_{6,1}$. To create $C_{5,1}$ and $C_{6,1}$, however, we also need the absence of $C_{4,1}$. Hence, we first have to create $C_{2,1}$ and $C_{3,1}$ that lead to removal of $C_{4,1}$. Then, via a second creation of $C_{1,1}$ we can create $C_{6,1}$. This idea is now applied subsequently in the next gadgets.

$C_{6,1}$ can now be used to create $C_{1,2}$ which leaves gadget 1 empty, i.e., no coalition $C_{\cdot,1}$ exists. To create $C_{6,2}$, we have to first remove $C_{4,2}$ by creating $C_{2,2}$ and $C_{3,2}$. By a second creation of $C_{1,2}$ (and due to the absence of $C_{4,2}$) we can then create $C_{5,2}$ and $C_{6,2}$. Note, however, that the second creation of $C_{1,2}$ again needs the existence of $C_{6,1}$. However, when we created $C_{1,2}$ for the first time, gadget 1 was empty. Thus, to create $C_{1,2}$ a second time, we have to also create $C_{6,1}$ a second time. Hence, we must create $C_{4,1}$ a second time (before removing $C_{4,2}$) and run through the all of the dynamics for gadget 1 described above. Only then we can finally create $C_{5,2}$ and $C_{6,2}$. Hence, to create $C_{6,2}$ we have to create $C_{1,1}$ a total of four times.

Now $C_{6,2}$ can be used to create $C_{1,3}$. At this point, both gadgets 1 and 2 are empty. To create $C_{6,3}$ we need to remove $C_{4,3}$ and, thus, create $C_{1,3}$ twice. Since, both gadgets 1 and 2 are empty, we first need to recreate $C_{6,2}$ for the second creation of $C_{1,3}$. The second creation of $C_{6,2}$ requires another four creations of $C_{1,1}$ as described above. Hence, to create $C_{6,3}$ we need to create $C_{1,1}$ a total of eight times. This reasoning can be applied to the remaining gadgets and shows the proposition. □

The previous two results show that convergence can take exponential time if domination rules are not consistent. For convergence in finite time, however, we only need consistent domination rules. The following proposition shows that we encounter the standard *lexicographic potential function*: The vector of coalition benefits (ordered non-decreasingly) increases lexicographically in every improvement step. This is property has received interest in correlated matching and coalition formation games without structural constraints [2, 46].

Proposition 3 *If the domination rules are consistent, then there is a lexicographic potential function, and every sequence of improvement steps is finite.*

Proof. Suppose the domination rules are consistent and consider an improvement step. If coalition C gets added, then C can cause one or more coalitions C'_i, $i = 1, 2, \ldots$ to become removed. By Observation 2, however, we have $b(C) > b(C'_i)$ and $C \cap C'_i \neq \emptyset$ for all i.

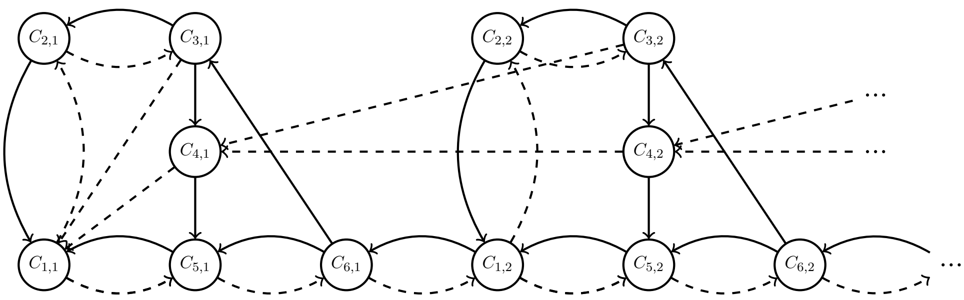

Figure 6: Coalition movement graph of gadgets 1 and 2 in Proposition 2. The thick edges symbolize domination rules, the dashed edges symbolize generation rules

Hence, any novel coalition that gets added to the state causes only (overlapping) coalitions with strictly smaller benefit to be removed. Therefore, we have the lexicographic potential function. Hence, convergence in finite time can always be guaranteed whenever domination rules are consistent. □

In the following proposition, we show that the argument from the previous proof breaks when domination rules are allowed to contain non-overlapping coalitions. Then a coalition with smaller benefit can dominate a non-overlapping coalition with larger benefit. Note that benefit domination rules D_b apply only for overlapping coalitions. In the following instance, however, benefit domination rules are ineffective, since we create three coalitions that are mutually non-overlapping.

Proposition 4 *If the domination rules include target coalitions that do not overlap with any coalition in the precondition, there are instances and starting states such that every sequence of improvement steps cycles.*

Proof. Consider the following example with $N = \{1, \ldots, 6\}$, $\mathcal{C} = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$, $\mathcal{C}_g = \mathcal{C}$, and benefits $b(C) = 1$ for all $C \in \mathcal{C}$. There are no generation rules $T = \emptyset$ (but $\mathcal{C}_g = \mathcal{C}$). For the domination rules, we consider non-overlapping coalitions in precondition and target:

$$D = \{(\{\{1, 2\}\}, \{3, 4\}), (\{\{3, 4\}\}, \{5, 6\}), (\{\{5, 6\}\}, \{1, 2\})\} \ .$$

The initial state is $\mathcal{C}_0 = \{\{1, 2\}\}$.

Since $\{1, 2\}$ exists, $\{3, 4\}$ is dominated and cannot be formed. The other candidate coalition $\{5, 6\}$ is undominated and represents the unique improvement step. As $\{5, 6\}$ dominates $\{1, 2\}$ (but not vice versa), $\{1, 2\}$ gets removed when $\{5, 6\}$ is formed. Now $\{4, 3\}$ is the unique undominated candidate coalition and is formed. Thereby, $\{5, 6\}$ is removed and $\{1, 2\}$ becomes undominated. Next $\{1, 2\}$ is formed, $\{3, 4\}$ is deleted and the cycle is complete. □

We conclude that relaxing the consistency conditions will encompass settings where (fast) convergence cannot be guaranteed. Thus, in this sense the choice of axioms is minimal.

However, let us point out that consistency is not a complete characterization of the set of games that have short improvement sequences for every pair of reachable states (in the sense of if-and-only-if). There also exist non-consistent games that have this property (a trivial example are games with a constant number of agents). As such, there could be orthogonal characterizations of coalition formation games with constraints. We derive such an example below for local coalition formation games.

4 Local Coalition Formation Games

Unlike local matching games, local coalition formation games are not fully described by consistent generation and domination rules. In these games, a coalition might be accessible only if several other coalitions are present, which implies inconsistent generation rules. Nevertheless, polynomial-time convergence still might be possible.

In Section 4.1 we analyze local coalition formation games depending on the choice of the set $\mathcal{H}$ of formation graphs. If $\mathcal{H}$ only consists of cliques, we encounter the problem that a coalition might be accessible only if several coalitions exist. However, since we require a clique structure to form a coalition, improvement sequences do not conduct exploration. More formally, the edge set of the visibility graph $G(\mathcal{S})$ is shrinking monotonically. This allows to show convergence in polynomial time. If $\mathcal{H}$ only consists of stars, we can define consistent generation and domination rules. Thereby, all results of the previous section apply. For all other choices of $\mathcal{H}$, we provide an instance and a starting state, from which there is a unique sequence to a stable state, which takes an exponential number of improvement steps.

Subsequently, in Section 4.2 we provide a lower bound for the model using longer paths in the visibility graph. In this approach (recall the discussion in Section 2.2.2), a coalition is accessible if it can be mapped to a formation graph, such that each edge of the formation graph is represented by a path with hop-distance at most 2 in the visibility graph. In such games, there are instances with maximal coalition size 3 and starting states, from which the unique sequence to a stable state requires an exponential number of improvement steps, even if the set of formation graphs consists only of cliques or stars.

4.1 Characterizing Formation Graphs

In this section, we consider local coalition formation games, where we parametrize the games with the set $\mathcal{H}$ of formation graphs. In this way, we analyze how different choices of formation graphs influence the convergence properties, i.e., the existence of short paths to stability. We start with instances where $\mathcal{H}$ consists solely of cliques:

Theorem 4 *Let $\mathcal{H} \subseteq \{K_i \mid i = 1 \dots n\}$ where K_i is a clique of i vertices. In every local coalition formation game using $\mathcal{H}$, for every initial state $\mathcal{S}$ there is a sequence of n local improvement steps which results in a locally stable coalition structure. If local improvement steps are chosen at random in every step, the dynamics converges in expected time $O(mn)$.*

Proof. We observe that the set of edges in the visibility graph is monotonically shrinking. A coalition C is only accessible if all pairs of agents in C are connected by permanent (through L) or temporary links (currently in the same coalition). Thus, if C is a local blocking coalition and gets added, this does not introduce any new temporary links. On the other hand, some temporary links might get deleted if coalitions overlapping with C are deleted.

Now assume C is a local blocking coalition of maximal benefit among all local blocking coalitions in $\mathcal{S}$. Then, once C is formed, it will not be deleted by any other coalition created through blocking dynamics as no new coalitions of higher value can become accessible. Thus, the natural greedy approach of resolving the local blocking coalition that adds the most valuable coalition to the state results in a stable state after at most n steps.

If, instead of picking the most valuable one, we pick a local blocking coalition at random, with probability at least $\frac{1}{m}$ we pick one of the most valuable local blocking coalitions. Hence, every m steps in expectation we pick a most valuable local blocking coalition. Due to the monotonic shrinking visibility graph, the set of available coalitions is also monotonically decreasing. Hence, whenever we pick a most valuable blocking coalition, this coalition will not get removed subsequently. Thus, after at most m steps in expectation we enlarge the set of stable coalitions by at least one. Overall, after at most mn steps in expectation, we reach a stable state. □

Let us now turn to star formation graphs. Here use the fact that in a star all edges share the same (center) vertex. As a consequence, when embedding the star into the visibility graph, it uses temporary links introduced by at most one coalition of $\mathcal{S}$ – otherwise the coalitions would have to share the center vertex. Thus, for $\mathcal{H} \subseteq \{H_i \mid i = 1 \dots n\}$, where by H_i we denote a star consisting of a center and $i-1$ leaves, all results from the previous section apply.

Theorem 5 *Let $\mathcal{H} \subseteq \{H_i \mid i = 1 \dots n\}$ where H_i is a star with a center and $i-1$ leaves. Every local coalition formation game using $\mathcal{H}$ is equivalent to a coalition formation game with constraints and consistent generation and domination rules.*

Proof. We keep the set of possible coalitions as well as their benefits the same. The main insight of the proof is to express the locality constraints using generation and domination rules. The domination rules is easy to define – the only reason an accessible coalition should not be formed is because one of the involved agents is already part of a better or equally preferred coalition. Thus, we define

$$D = D_b = \{(\{C'\}, C) \mid C, C' \in \mathcal{C}, C \cap C' \neq \emptyset, b(C') \geq b(C)\}.$$

For the generation rules, observe that coalitions cannot overlap and in a star all edges share the center vertex. Hence, no accessible coalition can rely on temporary links of more than one existing coalition when embedding a star formation graph. Hence, for the generation rules we set

$$\begin{aligned} T = &\{(\emptyset, C) \mid C \in \mathcal{C}, H_{|C|} \subseteq (V, L)\} \\ &\cup \{(\{C'\}, C) \mid C, C' \in \mathcal{C}, C \cap C' \neq \emptyset, H_{|C|} \subseteq (V, L \cup \{C'\})\} \ . \end{aligned}$$

Note that these generation and domination rules are consistent. We now verify that using these rules, the resulting coalition formation game correctly expresses the original local coalition formation game.

First, assume that C is a local blocking coalition for $\mathcal{S}$. Then C is accessible, that is, $H_{|C|}$ appears in the visibility graph. As discussed above, temporary links of at most one coalition are used in embedding $H_{|C|}$. Thus, either C is always accessible via L (i.e., $C \in \mathcal{C}_g$), or it relies on temporary links of some $C' \in \mathcal{S}$. Then $(\{C'\}, C) \in T$ which makes C a candidate coalition in $\mathcal{S}$. Further, if C is a local blocking coalition, there is no coalition $C'' \in \mathcal{S}$ such that $C \cap C'' \neq \emptyset$ and $b(C'') \geq b(C)$. Consequently, the candidate coalition C is undominated in $\mathcal{S}$. Hence, if C is a local blocking coalition in the original game, it is also a blocking coalition in $\mathcal{S}$ for the coalition formation game with constraints.

Conversely, let C be a blocking coalition in $\mathcal{S}$ for the coalition formation game with constraints. Then C is undominated, that is, C is a blocking coalition. Further, C is a candidate coalition. Thus, C is accessible in $\mathcal{S}$. In consequence, if C is a blocking coalition for the coalition formation game with constraints, it is a local blocking coalition in the original game.

Now, resolving the local blocking coalition C results in deleting all overlapping coalitions. By definition of blocking coalitions, all existing overlapping coalitions are of smaller value than C. Similarly, resolving the undominated candidate coalition C results in deleting all coalitions dominated by C. By definition, these are exactly the coalitions overlapping with C of less or equal value than C. As C was undominated, all these coalitions have to be of smaller value than C. Thus, the set of deleted coalitions coincides in both games. □

As the main result, by applying Theorem 1 we obtain short paths to stability to locally stable states.

Corollary 1 *Let $\mathcal{H} \subseteq \{H_i \mid i = 1 \dots n\}$. In every local coalition formation game using $\mathcal{H}$, for every initial state $\mathcal{S}$ there is a sequence of $O(nm^2)$ local improvement steps which results in a locally stable coalition structure.*

Finally, we analyze the case where the formation graph is any connected graph. It turns out that for every such graph that is neither a star nor a clique, there is an instance and a starting state with only exponentially sequences to a locally stable state. To prove this property, we first extract a single subgraph structure which is present in every connected graph except for stars and cliques. We will then construct an instance without short paths to stability using only this subgraph structure. To adjust the instance to an arbitrary graph, one can simply add the "missing" vertices and edges using separate auxiliary agents and links for every coalition.

We start by observing that if some graph H is neither a star nor a clique, then there is some path of length 3 in H such that the first and the third vertex are not directly connected by some edge.

Lemma 2 *Let $H = (V_H, E_H)$ be any simple, undirected, connected graph. If for every path $v_1v_2v_3v_4$ with $\{v_1, v_2\}$, $\{v_2, v_3\}$, $\{v_3, v_4\} \in E_H$ there also exists an edge $\{v_1, v_3\} \in E_H$, then H is either a clique or a star.*

Proof. If H does not have any paths of length ≥ 3, then H is a star. Further, if H is connected and has less than 4 vertices, H is either a star or a clique.

Now assume that H holds some path $v_1v_2v_3v_4$ with $\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_4\} \in E_H$. Then, considering this path forward and backward, we know that $\{v_1, v_3\}$ and $\{v_2, v_4\}$ are in E_H as well. Thus, we also have the path $v_1v_2v_4v_3$ and conclude $\{v_1, v_4\} \in E_H$, that is, v_1, v_2, v_3, and v_4 form a clique C_1 in H. Now, as H is connected, if $H \neq C_1$, there exists some edge e connecting C_1 to the rest of H. Let v_5 be the vertex in $e \cap (V_H \setminus \{v_1, v_2, v_3, v_4\})$. W.l.o.g. let v_5 be connected to v_1, that is, $e = \{v_1, v_5\}$. Using the edges in C_1, we have the paths $v_5v_1v_2v_3$, $v_5v_1v_3v_2$, and $v_5v_1v_4v_2$ and thus also edges from v_5 to all other vertices in C_1. Hence, v_1, v_2, v_3, v_4 and v_5 form a clique C_2 in H. Now we can apply the same arguments to the bigger clique repeatedly until each vertex of V is included. Thus, H has to be a clique. □

Hence, there is a path of length 3 with a missing edge between the first and the third vertex in all formation graphs H we want to analyze. We use the substructure to establish the desired lower bound by concentrating on this structure. The rest of H can then be added separately for every potential coalition using auxiliary agents and links. We also ensure that additional edges between the vertices of the path, that is, edges between the first and the fourth vertex, and edges between the second and the fourth vertex, have no influence on the dynamics.

Theorem 6 *Let $H = (V_H, E_H)$ be any simple, undirected, connected graph which is neither a clique nor a star. Then there are instances and initial states of local coalition formation games using $\mathcal{H} = \{H\}$ such that every sequence to a stable state requires an exponential number of improvement steps.*

Proof. By Lemma 2, we know that there are vertices $v_1, v_2, v_3, v_4 \in V_H$ such that $\{v_1, v_2\}$, $\{v_2, v_3\}$, $\{v_3, v_4\} \in E_H$ but $\{v_1, v_3\} \notin E_H$. Let $H_{rest} = (V_{H,rest} = V_H \setminus \{v_1, v_2, v_3, v_4\}, E_{H,rest} = E_H \setminus \{\{v_1, v_2\}$, $\{v_2, v_3\}$, $\{v_3, v_4\}\}$.

We show how to construct a network of links involving only the vertices v_1, v_2, v_3, and v_4 as well as $\{v_1, v_2\}, \{v_2, v_3\}$, and $\{v_3, v_4\}$ for each coalition. For each coalition C, all other vertices $V_{H,rest}$ are unique (denoted by $V_{H,C,rest}$) and are connected by links according to $E_{H,rest}$ (denoted by $E_{H,C,rest}$) with each other as well as with v_1, v_2, v_3 and v_4. The network is composed of a concatenation of identical gadgets, each of which have a *starting link* $\{u_{1,i}, u_{2,i}\}$ and two *final links* $\{u_{3,i}, u_{5,i}\}$ and $\{u_{9,i}, u_{11,i}\}$. These three links are temporary and not in L, that is, they are only available when their incident vertices are currently in the same coalition. The dynamics are designed such that the starting link must be created and deleted once to create the first final link and a second time to create the second final link. The coalition providing the starting link of gadget $i+1$ can only be created when both final links of gadget i are available. Thus, to create the coalition which provides the final link of gadget k, the starting link of gadget 1 must be formed and removed 2^k times, that is, coalition C_0 has to be formed 2^k times.

For gadget G_i we have 7 coalitions $C_{1,i}, \dots, C_{7,i}$. The agent set of G_i consists of

$$\{u_{1,i}, \dots, u_{14,i}\} \cup \bigcup_{j=1}^{7} V_{H,C_{j,i},rest}$$

and the link set of

$$\{\{u_{1,i}, u_{5,i}\}, \{u_{1,i}, u_{9,i}\}, \{u_{2,i}, u_{3,i}\}, \{u_{2,i}, u_{11,i}\}, \{u_{3,i}, u_{4,i}\}, \{u_{3,i}, u_{11,i}\}, \{u_{5,i}, u_{6,i}\}, \{u_{5,i}, u_{7,i}\},$$
$$\{u_{7,i}, u_{8,i}\}, \{u_{9,i}, u_{10,i}\}, \{u_{11,i}, u_{12,i}\}, \{u_{11,i}, u_{13,i}\}, \{u_{13,i}, u_{14,i}\}\} \cup \bigcup_{j=1}^{7} E_{H,C_{j,i},rest} \ .$$

Gadget 1 additionally has agents w_1 and w_2 with links $\{w_1, u_{1,1}\}$, $\{w_1, u_{2,1}\}$, and $\{w_2, u_{2,1}\}$ and a coalition $C_0 = \{w_1, w_2, u_{1,1}, u_{2,1}\} \cup V_{H,C_0,rest}$ of value $\frac{1}{2}$. The transition from gadget i to gadget $i+1$ is realized by associating $u_{5,i}$ with $u_{1,i+1}$ and $u_{9,i}$ with $u_{2,i+1}$.

The possible coalitions in gadget i are the following ones.

$$\begin{aligned}
C_{1,i} &= \{u_{1,i}, u_{2,i}, u_{3,i}, u_{4,i}\} \cup V_{H,C_{1,i},rest} & \text{with benefit} \quad & 4i+1\\
C_{2,i} &= \{u_{1,i}, u_{3,i}, u_{5,i}, u_{6,i}\} \cup V_{H,C_{2,i},rest} & & 4i+2\\
C_{3,i} &= \{u_{3,i}, u_{5,i}, u_{7,i}, u_{8,i}\} \cup V_{H,C_{3,i},rest} & & 4i+3\\
C_{4,i} &= \{u_{1,i}, u_{2,i}, u_{9,i}, u_{10,i}\} \cup V_{H,C_{4,i},rest} & & 4i+1\\
C_{5,i} &= \{u_{2,i}, u_{9,i}, u_{11,i}, u_{12,i}\} \cup V_{H,C_{5,i},rest} & & 4i+2\\
C_{6,i} &= \{u_{9,i}, u_{11,i}, u_{13,i}, u_{14,i}\} \cup V_{H,C_{6,i},rest} & & 4i+3\\
C_{7,i} &= \{u_{3,i}, u_{5,i}, u_{9,i}, u_{11,i}\} \cup V_{H,C_{7,i},rest} & & 4i+4.
\end{aligned}$$

The initial state is $\mathcal{S}_0 = \emptyset$.

We will describe the dynamics of gadget 1 as well as the transition to the next gadget. The other gadgets work similarly. If no agent of gadget 1 is involved in any coalition, the only accessible coalition is C_0. Once C_0 is formed, $C_{1,1}$ and $C_{4,1}$ become accessible because the starting link $\{u_{1,1}, u_{2,1}\}$ is in the visibility graph. As both coalitions are of higher benefit than C_0, they represent local blocking coalitions. W.l.o.g. assume that $C_{1,1}$ is formed first. Now $C_{4,1}$ is not a blocking coalition anymore. The temporary link $\{u_{1,1}, u_{3,1}\}$ is present, which makes $C_{2,1}$ a local blocking coalition. Forming $C_{2,1}$ results in $u_{2,1}$ being single again. Additionally, the first final temporary link $\{u_{3,1}, u_{5,1}\}$ exists, and $C_{3,1}$ becomes a local blocking coalition. Now, forming $C_{3,1}$, agent $u_{1,1}$ becomes single again. Thereby C_0 (which is always accessible) becomes a blocking coalition again. Further, the link $\{u_{3,1}, u_{5,1}\}$ is still present. At this point, the only local blocking coalition again is C_0. This time, after C_0 is formed, $C_{1,1}$ is not a blocking coalition because $u_{3,1}$ is involved in a better coalition. Thus, the unique improvement step is to create $C_{4,1}$. As above, after forming $C_{4,1}$, $C_{5,1}$ becomes accessible. By adding $C_{5,1}$, the second final link $\{u_{9,1}, u_{11,1}\}$ is created. Now we can either directly form $C_{7,1}$, or we can first add $C_{6,1}$ and then $C_{7,1}$ (and thereby remove $C_{6,1}$ again). Note that $C_{6,1}$ is necessary to make C_0 available again for the case we would have chosen $C_{4,1}$ instead of $C_{1,1}$ first. The existence of $C_{7,1}$ creates the temporal link $\{u_{5,1}, u_{9,1}\} = \{u_{1,2}, u_{2,2}\}$ for the fist time. This initiates the same dynamics in the second gadget.

Observe that the dynamics cannot terminate early by resolving local blocking coalitions in a different order. Thus in the end $C_{7,k}$ is necessary for the coalition structure to be stable. Due to the structure, 2^k creations of C_0 are necessary to form $C_{7,k}$.

Note that we heavily rely on the temporary links in the dynamics being temporary and not part of $E_{H,C,rest}$. As H holds a path of 3 edges with no connection between the first and the third vertex, we can actually arrange $E_{H,C,rest}$ such that the "suitable" links are missing when forming a coalition and only being established when adding this same coalition. However, the structure of H might require an edge between the first and the fourth and/or the second and the fourth vertex of the path. These are two vertices which are part of the shared gadget and not the part which is unique for each coalition anyway. Luckily, these edges – namely

$$\begin{array}{lll} \{u_{1,i}, u_{4,i}\}, \{u_{2,i}, u_{4,i}\} & \text{for} & C_{1,i} \\ \{u_{1,i}, u_{6,i}\}, \{u_{3,i}, u_{6,i}\} & & C_{2,i} \\ \{u_{3,i}, u_{7,i}\}, \{u_{3,i}, u_{8,i}\} & & C_{3,i} \\ \{u_{1,i}, u_{10,i}\}, \{u_{2,i}, u_{10,i}\} & & C_{4,i} \\ \{u_{2,i}, u_{12,i}\}, \{u_{9,i}, u_{12,i}\} & & C_{5,i} \\ \{u_{9,i}, u_{13,i}\}, \{u_{9,i}, u_{14,i}\} & & C_{6,i} \\ \{u_{3,i}, u_{11,i}\}, \{u_{5,i}, u_{11,i}\} & & C_{7,i} \end{array}$$

all connect vertices which only share this one coalition. Hence the (permanent) existence of such links does not make any other coalitions accessible. □

As discussed above, our dichotomy results are strongest if $m \in \mathsf{poly}(n)$. Note that this follows directly if the formation graphs are of constant size, since only coalitions of constant size can form. Under this assumption, our results above show there are paths of length $\mathsf{poly}(n)$ for cliques and stars. In contrast, our exponential lower bound already holds when, e.g., the formation graph is a path with 4 nodes.

4.2 Visibility via Longer Paths

In this section, we consider the case, where for an accessible coalition in state $\mathcal{S}$, a formation graph H can be embedded into the visibility graph $G(\mathcal{S})$ such that each edge of H is represented by a path of hop-distance of 2 in $G(\mathcal{S})$. Our results here are negative, even if the formation graphs are restricted to cliques or stars.

Theorem 7 *For local coalition formation games with visibility over paths of distance* 2 *or more, there are instances and initial states such that every sequence to a stable state requires an exponential number of improvement steps. This is true even when (1) all coalitions have size at most 3 and (2)* $\mathcal{H}$ *consists of all cliques, or* $\mathcal{H}$ *consists of all stars.*

Proof. We compose the instances by attaching a number of identical gadgets. In our construction, every coalition has size at most 3. Each gadget has a distinct start coalition and two distinct final coalitions. To create one of the final coalitions, the start coalition has to be formed and then be deleted. It is not possible to form both final coalitions using only one creation of the start coalition. Thus, for both final coalitions to exist, the start coalition must have been formed and deleted twice.

The gadgets are composed such that the start coalition of gadget i can only be formed if both final coalitions of gadget $i-1$ exist. Hence, to form both final coalitions of the k^{th} gadget, the start coalition of gadget 1 has to be created 2^k times.

Note that since all coalitions have size at most 3, we can restrict to formation graphs of at most 3 vertices. Our construction is slightly different depending on whether we consider

$$\mathcal{H}_1 = \{(\{u,v\}, \{\{u,v\}\}), (\{u,v,w\}, \{\{u,v\},\{v,w\}\})\}$$

the set of all stars of size at most 3, or

$$\mathcal{H}_2 = \{(\{u,v\}, \{\{u,v\}\}), (\{u,v,w\}, \{\{u,v\},\{v,w\},\{u,w\}\})\}$$

the set of all cliques of size at most 3.

Gadget i consists of agents $V_i = \{v_{1,i}, \ldots, v_{7,i}\}$. The link set L_i includes the set

$$\{\{v_{1,i}, v_{4,i}\}, \{v_{2,i}, v_{5,i}\}, \{v_{2,i}, v_{7,i}\}, \{v_{3,i}, v_{6,i}\}, \{v_{5,i}, v_{7,i}\}\} \ .$$

If $\mathcal{H} = \mathcal{H}_2$, L_i contains in addition $\{v_{1,i}, v_{6,i}\}$. The potential coalitions and their benefits are

$$\begin{array}{lll} C_{1,i} = \{v_{1,i}, v_{2,i}, v_{3,i}\} & \text{with benefit} & 3i+1 \\ C_{2,i} = \{v_{2,i}, v_{4,i}\} & & 3i+2 \\ C_{3,i} = \{v_{4,i}, v_{5,i}\} & & 3i+3 \\ C_{4,i} = \{v_{2,i}, v_{6,i}\} & & 3i+2 \\ C_{5,i} = \{v_{6,i}, v_{7,i}\} & & 3i+3 \end{array}$$

$C_{1,i}$ is the start coalition, $C_{3,i}$ and $C_{5,i}$ are the final coalitions of gadget i. To connect gadget i with gadget $i+1$, we identify agent $v_{1,i+1}$ with $v_{4,i}$, agent $v_{2,i+1}$ with $v_{7,i}$, and agent $v_{3,i+1}$ with $v_{6,i}$. Further, since $C_{1,1}$ must be always accessible, we add an agent a and links $\{a, v_{1,1}\}$, $\{a, v_{2,1}\}$, and $\{a, v_{3,1}\}$. The initial state is the empty coalition structure.

Let us analyze the dynamics in gadget 1. Initially, gadget 1 is empty, and the only accessible coalition is $C_{1,1}$. Once $C_{1,1}$ is formed, both $C_{2,1}$ and $C_{4,1}$ become accessible. As both coalitions are more valuable than $C_{1,1}$, exactly one of them is formed in the next step. Let us assume that $C_{2,1}$ is formed. The case in which $C_{4,1}$ is formed first is similar. The formation of $C_{2,1}$ removes $C_{1,1}$ but makes $C_{3,1}$ accessible and thus a blocking coalition. When $C_{3,1}$ is formed (and $C_{2,1}$ removed), $C_{1,1}$ becomes a blocking coalition again, since all its agents are free. The formation of $C_{1,1}$ then makes $C_{4,1}$ accessible again. At this point, $C_{4,1}$ is the only blocking coalition, because $C_{2,1}$ is dominated by $C_{3,1}$. With $C_{4,1}$ formed, $C_{5,1}$ becomes a blocking coalition. In the next step $C_{5,1}$ is formed, that is, both final coalitions exist. Now $C_{2,1}$ becomes a blocking coalition and the dynamics proceed in gadget 2.

For all subsequent gadgets $i \geq 2$, we encounter the same dynamics except that their start coalition is only accessible when both final coalitions of gadget $i-1$ exist. Only then $v_{1,i}$ sees $v_{2,i}$ via $C_{3,i-1}$ and link $\{v_{5,i-1}, v_{7,i-1}\} = \{v_{5,i-1}, v_{2,i}\}$ and $v_{2,i}$ sees $v_{3,i}$ via $C_{5,i-1}$. Thus, in case of $\mathcal{H} = \mathcal{H}_1$, the star required to make $C_{1,i}$ accessible exists in the visibility graph. In case of $\mathcal{H} = \mathcal{H}_2$, $v_{1,i}$ and $v_{3,i}$ also see each other (permanently) via $v_{1,i-1}$, which completes the clique in the visibility graph required to make $C_{1,i}$ accessible.

Note that we can alter the gadgets such that all coalitions have size 3 by adding an auxiliary agent to each coalition of size 2 and connecting it with one (for stars) or both (for cliques) vertices via a path of length 2 (using auxiliary vertices). Similar adjustments allow to extend the gadget to coalitions of arbitrary sizes. □

5 Discussion and Conclusion

In this paper, we have studied convergence properties in stable matching and hedonic games with structural constraints, which generalize a variety of recently proposed models for visibility and agent externality. Our results are a tight axiomatic characterization for the existence of a path to stability of polynomial length in games with correlated preferences. The set of consistency axioms are minimal in the sense that if an axiom is dropped or relaxed, a (polynomial) path to stability cannot be guaranteed. Moreover, we have studied a model of local visibility in hedonic games based on the notion of formation graphs, and we provided a tight characterization of the existence of polynomial paths to stability based on the formation graphs.

Our upper bounds on the convergence time are polynomial in the number of agents n and the number of coalitions m. Our lower bound gadgets involve coalitions of constant size and show bounds that are exponential in n. As such, the dichotomy in our results is strongest if $m \in \mathsf{poly}(n)$ as is the case, e.g., in the prominent case of matching.

There are a variety of interesting problems that arise from our work. A natural question is whether additional constraints that arise in matching or coalition formation problems can be formulated within our framework of consistent generation and domination rules. For example, there are interesting additional variants of local coalition formation games, e.g., when each existing coalition forms an organization graph in the network (instead of a clique as in Section 4). Finally, beyond correlated preferences, efficient algorithms for the computation of stability concepts have recently been derived for several other classes of hedonic games [51–54]. It would be interesting to see under which conditions one can also guarantee (polynomial) paths to stability (with and without additional structural constraints).

References

[1] David Abraham, Ariel Levavi, David Manlove, and Gregg O'Malley. The stable roommates problem with globally ranked pairs. *Internet Math.*, 5(4):493–515, 2008.

[2] Heiner Ackermann, Paul Goldberg, Vahab Mirrokni, Heiko Röglin, and Berthold Vöcking. Uncoordinated two-sided matching markets. *SIAM J. Comput.*, 40(1):92–106, 2011.

[3] Kemal Akkaya, Ismail Guneydas, and Ali Bicak. Autonomous actor positioning in wireless sensor and actor networks using stable matching. *Intl. J. Parallel, Emergent and Distrib. Syst.*, 25(6):439–464, 2010.

[4] Elliot Anshelevich, Onkar Bhardwaj, and Martin Hoefer. Stable matching with network externalities. *Algorithmica*, 78(3):1067–1106, 2017.

[5] Esteban Arcaute and Sergei Vassilvitskii. Social networks and stable matchings in the job market. In *Proc. 5th Workshop Internet & Network Economics (WINE)*, pages 220–231, 2009.

[6] Georgios Askalidis, Nicole Immorlica, Augustine Kwanashie, David Manlove, and Emmanouil Pountourakis. Socially stable matchings in the hospitals/residents problem. In *Proc. 13th Workshop Algorithms and Data Structures (WADS)*, pages 85–96, 2013.

[7] Haris Aziz and Florian Brandl. Existence of stability in hedonic coalition formation games. In *Proc. 11th Conf. Auton. Agents and Multi-Agent Systems (AAMAS)*, pages 763–770, 2012.

[8] Haris Aziz, Felix Brandt, and Paul Harrenstein. Fractional hedonic games. In *Proc. 13th Conf. Auton. Agents and Multi-Agent Systems (AAMAS)*, pages 5–12, 2014.

[9] Haris Aziz, Felix Brandt, and Hans Georg Seedig. Computing desirable partitions in additively separable hedonic games. *Artif. Intell.*, 195:316–334, 2013.

[10] Haris Aziz and Rahul Savani. Hedonic games. In Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel Procaccia, editors, *Handbook of Computational Social Choice*, pages 356–376. Cambridge University Press, 2016.

[11] Suryapratim Banerjee, Hideo Konishi, and Tayfun Sönmez. Core in a simple coalition formation game. *Social Choice & Welfare*, 18(1):135–153, 2001.

[12] Davide Bilò, Luciano Gualà, Stefano Leucci, and Guido Proietti. Locality-based network creation games. *ACM Trans. Parallel Comput.*, 3(1):6, 2016.

[13] Péter Biró, Matthijs Bomhoff, Petr A. Golovach, Walter Kern, and Daniël Paulusma. Solutions for the stable roommates problem with payments. *Theoret. Comput. Sci.*, 540:53–61, 2014.

[14] Péter Biró, Katarína Cechlárová, and Tamás Fleiner. The dynamics of stable matchings and half-matchings for the stable marriage and roommates problems. *Int. J. Game Theory*, 36(3–4):333–352, 2008.

[15] Péter Biró and Gethin Norman. Analysis of stochastic matching markets. *Int. J. Game Theory*, 42(4):1021–1040, 2013.

[16] Yosef Blum, Alvin Roth, and Uriel Rothblum. Vacancy chains and equilibration in senior-level labor markets. *J. Econom. Theory*, 76:362–411, 1997.

[17] Yosef Blum and Uriel Rothblum. "Timing is everything" and marital bliss. *J. Econom. Theory*, 103:429–442, 2002.

[18] Anna Bogomolnaia and Matthew Jackson. The stability of hedonic coalition structures. *Games Econom. Behav.*, 38:201–230, 2002.

[19] Philipp Brandes and Roger Wattenhofer. On finding better friends in social networks. In *Proc. 14th Symp. Stabilization, Safety, Security of Distrib. Syst. (SSS)*, pages 266–278, 2012.

[20] Katarina Cechlárova. Stable partition problem. In *Encyclopedia of Algorithms*. 2008.

[21] Po-An Chen, Bart De Keijzer, David Kempe, and Guido Schäfer. Altruism and its impact on the price of anarchy. *ACM Trans. Econom. Comput.*, 2(4):17, 2014.

[22] Christine Cheng. On the stable matchings that can be reached when the agents go marching in one by one. *SIAM J. Disc. Math.*, 30(4):2047–2063, 2016.

[23] Christine Cheng and Eric McDermid. Maximum locally stable matchings. *Algorithms*, 6(3):383–395, 2013.

[24] Effrosyni Diamantoudi, Eiichi Miyagawa, and Licun Xue. Random paths to stability in the roommates problem. *Games Econom. Behav.*, 48(1):18–28, 2004.

[25] Jacques Dreze and Joseph Greenberg. Hedonic coalitions: Optimality and stability. *Econometrica*, 48(4):987–1003, 1980.

[26] Martin Gairing and Rahul Savani. Computing stable outcomes in hedonic games. In *Proc. 3rd Symp. Algorithmic Game Theory (SAGT)*, pages 174–185, 2010.

[27] David Gale and Lloyd Shapley. College admissions and the stability of marriage. *Amer. Math. Monthly*, 69(1):9–15, 1962.

[28] Michel Goemans, Li Li, Vahab Mirrokni, and Marina Thottan. Market sharing games applied to content distribution in ad-hoc networks. *IEEE J. Sel. Areas Commun.*, 24(5):1020–1033, 2006.

[29] Dan Gusfield and Robert Irving. *The Stable Marriage Problem: Structure and Algorithms*. MIT Press, 1989.

[30] Jana Hajduková. Coalition formation games: A survey. *Intl. Game Theory Rev.*, 8(4):613–641, 2006.

[31] Martin Hoefer. Local matching dynamics in social networks. *Inf. Comput.*, 222:20–35, 2013.

[32] Martin Hoefer, Michal Penn, Maria Polukarov, Alexander Skopalik, and Berthold Vöcking. Considerate equilibrium. In *Proc. 22nd Intl. Joint Conf. Artif. Intell. (IJCAI)*, pages 234–239, 2011.

[33] Martin Hoefer and Alexander Skopalik. Social context in potential games. In *Proc. 8th Workshop Internet & Network Economics (WINE)*, pages 365–378, 2012.

[34] Martin Hoefer and Alexander Skopalik. Altruism in atomic congestion games. *ACM Trans. Econom. Comput.*, 1(4):21, 2013.

[35] Martin Hoefer, Daniel Vaz, and Lisa Wagner. Hedonic coalition formation in networks. In *Proc. 29th Conf. Artif. Intell. (AAAI)*, pages 929–935, 2015.

[36] Martin Hoefer and Lisa Wagner. Matching dynamics with constraints. In *Proc. 10th Conf. Web and Internet Economics (WINE)*, pages 161–174, 2014.

[37] Martin Hoefer and Lisa Wagner. Locally stable marriage with strict preferences. *SIAM J. Disc. Math.*, 31(1):283–316, 2017.

[38] Moshe Hoffman, Daniel Moeller, and Ramamohan Paturi. Jealousy graphs: Structure and complexity of decentralized stable matching. In *Proc. 9th Conf. Web and Internet Economics (WINE)*, pages 263–276, 2013.

[39] Ayumi Igarashi and Edith Elkind. Hedonic games with graph-restricted communication. In *Proc. 15th Conf. Auton. Agents and Multi-Agent Systems (AAMAS)*, pages 242–250, 2016.

[40] Robert Irving. An efficient algorithm for the “stable roommates” problem. *J. Algorithms*, 6(4):577–595, 1985.

[41] Alexander Kelso and Vincent Crawford. Job matchings, coalition formation and gross substitute. *Econometrica*, 50(1483–1504), 1982.

[42] Donald Knuth. *Marriages stables et leurs relations avec d'autres problemes combinatoires.* Les Presses de l'Université de Montréal, 1976.

[43] John Ledyard. Public goods: A survey of experimental resesarch. In John Kagel and Alvin Roth, editors, *Handbook of Experimental Economics*, pages 111–194. Princeton University Press, 1997.

[44] David Levine. Modeling altruism and spitefulness in experiments. *Rev. Econom. Dynamics*, 1:593–622, 1998.

[45] David Manlove. *Algorithmics of Matching Under Preferences.* World Scientific, 2013.

[46] Fabien Mathieu. Self-stabilization in preference-based systems. *Peer-to-Peer Netw. Appl.*, 1(2):104–121, 2008.

[47] Fabien Mathieu. Acyclic preference-based systems. In Xuemin Shen, Heather Yu, John Buford, and Mursalin Akon, editors, *Handbook of peer-to-peer networking.* Springer Verlag, 2010.

[48] Roger Myerson. Graphs and cooperation in games. *Math. Oper. Res.*, 2:225–229, 1977.

[49] Nhan-Tam Nguyen, Anja Rey, Lisa Rey, Jörg Rothe, and Lena Schend. Altruistic hedonic games. In *Proc. 15th Conf. Auton. Agents and Multi-Agent Systems (AAMAS)*, pages 251–259, 2016.

[50] Guillermo Owen. Values of graph-restricted games. *SIAM J. Algebraic Discrete Methods*, 7(2):210–220, 1986.

[51] Dominik Peters. Complexity of hedonic games with dichotomous preferences. In *Proc. 30th Conf. Artif. Intell. (AAAI)*, pages 579–585, 2016.

[52] Dominik Peters. Graphical hedonic games of bounded treewidth. In *Proc. 30th Conf. Artif. Intell. (AAAI)*, pages 586–593, 2016.

[53] Dominik Peters. Towards structural tractability in hedonic games. In *Proc. 30th Conf. Artif. Intell. (AAAI)*, pages 4252–4253, 2016.

[54] Dominik Peters and Edith Elkind. Simple causes of complexity in hedonic games. In *Proc. 24th Intl. Joint Conf. Artif. Intell. (IJCAI)*, pages 617–623, 2015.

[55] Alvin Roth, Tayfun Sönmezc, and Utku Ünver. Pairwise kidney exchange. *J. Econom. Theory*, 125(2):151–188, 2005.

[56] Alvin Roth and Marilda Oliveira Sotomayor. *Two-sided Matching: A study in game-theoretic modeling and analysis.* Cambridge University Press, 1990.

[57] Alvin Roth and John Vande Vate. Random paths to stability in two-sided matching. *Econometrica*, 58(6):1475–1480, 1990.

[58] Shao-Chin Sung and Dinko Dimitrov. Computational complexity in additive hedonic games. *Europ. J. Oper. Res.*, 203(3):635–639, 2010.

[59] Gerhard Woeginger. Core stability in hedonic coalition formation. In *Proc. 39th Conf. Current Trends in Theory & Practice of Comput. Sci. (SOFSEM)*, pages 33–50, 2013.

A Games with Consistent Local Constraints

In this section we will show that a variety of matching and coalition formation games can be captured by our framework of coalition formation game with constraints and consistent generation and domination rules. For each setting, we will either define an equivalent coalition formation game with constraints, or show that consistent generation rules are impossible via examples that lack short paths to stability. The arguments are very similar in most cases and solely consist of rather straightforward verification that dynamics are captured correctly. As such, we will only flesh out one formal correctness proof in the last domain we consider – friendship stable matching in Section A.4. This is arguably the most complicated setting, since we must construct a coalition formation game in which D_b expresses domination according to *perceived* benefits. As such, we need to define different match benefits than in the original game. Moreover, to keep the domination rules in accordance with our definition of consistency, all coalition formation games with constraints will allow an agent to participate in at most one coalition at a time.

A.1 Socially Stable States

Given an instance of socially stable matching consisting of a graph $G_M = (V_M, E_M)$ of potential matching edges, social link set $L \subseteq E_M$, and edge benefits b, we construct an equivalent coalition formation game with consistent constraints as follows. We keep the agent set, the coalition set, and the benefits. The set of self-generating coalitions $\mathcal{C}_g$ is set to L, and $T = \emptyset$. Further, the only constraints for generation of an available coalition are based on the agents' desire to improve their benefits, that is, $D = D_b$. These generation and domination rules are obviously consistent.

For social coalition formation games, instead of using L directly we have to check whether we can find suiting organization structures in L. Thus, to construct a game with consistent rules, we keep V, $\mathcal{C}$, and the benefits. We choose $T = \emptyset$ and $D = D_b$ as before. For the self-generating coalitions we have

$$\mathcal{C}_g = \{C \mid C \in \mathcal{C}, \exists H \in \mathcal{H}, \varphi \text{ bijective} : V[H] \to C \text{ such that } \{u, v\} \in E[H] \\ \Rightarrow \{\varphi(u), \varphi(v)\} \in L\} \ .$$

Proposition 5 *Social coalition formation games give rise to consistent generation and domination rules.*

A.2 Locally Stable States

Given an instance of locally stable matching consisting of a graph $G_M = (V_M, E_M)$ of potential matching edges, social link set L, and edge benefits b, we construct an equivalent coalition formation game with consistent constraints as follows. We again keep the agent set, the coalition set, and the benefits. The set of self-generating coalitions and the generation rules become

$$\mathcal{C}_g = \{\{u, v\} \mid \{u, v\} \in E_M, dist(u, v, (V, L)) \leq 2\}$$

and

$$T = \{(\{e'\}, e) \mid e = \{u, v\}, e' = \{u, v'\} \in E, dist(u, v, (V, L \cup \{e'\})) \leq 2\}.$$

The domination rules are $D = D_b$.

Using this embedding, the coalition movement graph becomes the edge movement graph defined in [31] to prove the existence of short paths to stability for locally stable matching.

If we extend the hop-distance used for visibility to be larger than 2, then there can be edges that are accessible only if 2 or more matching edges exist. This condition cannot be captured by a consistent generation rule, since we would need 2 or more edges in the precondition. In fact, it is known [31, Theorem 3] that in this case there are instances and initial states such that every sequence of local improvement steps to a locally stable matching is exponentially long. For coalitions of larger size, the case depends on the formation graphs, see our discussion in Section 4.

Proposition 6 *Locally stable matching gives rise to consistent generation and domination rules if visibility is limited to a hop-distance of* 2 *in* $(V, M \cup L)$*. Otherwise, there are instances that imply non-consistent generation rules.*

A.3 Considerate Stable States

Given an instance of considerate stable matching consisting of a graph $G_M = (V_M, E_M)$ of potential matching edges, social link set L, and edge benefits b, we construct an equivalent coalition formation game with consistent constraints as follows. We again keep the agent set, the coalition set, and the benefits. Since there are no visibility constraints, we set $\mathcal{C}_g = E$, $T = \emptyset$. The main restriction is externality, which we capture using the domination rules $D = D_b \cup D_1$ with

$$D_1 = \{(\{\{u,v\}\}, \{u,v'\}) \mid \{u,v\}, \{u,v'\} \in E, v \neq v', \{u,v\} \in L \text{ or } \{v,v'\} \in L\} \ .$$

Note that the precondition and the target coalitions of every rule in D_1 overlap, which implies they are consistent.

The generalization for larger coalitions is straightforward. Let V be the agent set, $\mathcal{C}$ the set of potential coalitions, L the social links, b the benefits, and l be the limit of friends an agent is willing to hurt. To improve readability, we set $f(v, C) = |\{v' \mid v' \in C, \{v, v'\} \in L\}|$ to be the number of friends v has in coalition C. We define $\mathcal{C}_g = \mathcal{C}$ and $T = \emptyset$. For the domination rules we need to consider all combinations of coalitions which overlap with the target coalition and exceed the limit of friends for some participating agent. Formally, $D = D_b \cup D_1$ with

$$D_1 = \{(\mathcal{S}, C) \mid \mathcal{S} \subset \mathcal{C} \text{ coalition structure}, C \in \mathcal{C} \setminus \mathcal{S}, \exists v \in C : \sum_{C' \in \mathcal{S} : C \cap C' \neq \emptyset} f(v, C' \setminus C) > l\}.$$

Proposition 7 *Considerate coalition formation gives rise to consistent generation and domination rules.*

A.4 Friendship Stable States

When representing friendship coalition formation by a coalition formation game with consistent constraints, we must pay special attention to the benefit structure. Based on the perceived benefits, agents are willing to switch to coalitions of lower (direct, non-perceived) benefit. Thus, the domination rules must capture the perceived value of each coalition. Additionally, we have to take into account all perceived gains and losses caused by the formation of a new coalition.

Our proofs above rely on correlated preferences, that is, all participating agents obtain the same benefit from forming a coalition. In friendship coalition formation, it is not enough to require correlated preferences to obtain this property for perceived benefits. We also need to assume that the function c of friendship values is non-negative and symmetric among coalitions. For matching, this simply implies that (1) agents feel positive about the benefit of others, and (2) for each pair both incident agents feel equally strongly about their relation.

The following two examples show that without conditions (1) and (2), there are instances in which convergence does not occur. More fundamentally, under these conditions no stable state exists.

Example: The game is depicted in Figure 7. There are 3 agents. Each pair of agents generates a direct benefit of 1 for both incident agents. Based on the friendship values, agent 1 values coalition $\{1, 2\}$ more than coalition $\{1, 3\}$, agent 2 values coalition $\{2, 3\}$ more than coalition $\{1, 2\}$, and agent 3 values coalition $\{1, 3\}$ more than coalition $\{2, 3\}$. This implies a preference cycle and no stable state exists. ■

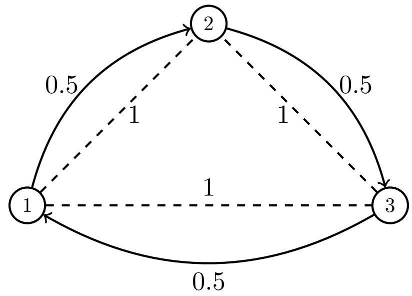

Figure 7: A friendship matching game with correlated preferences and asymmetric non-negative friendship values that does not have a friendship stable matching

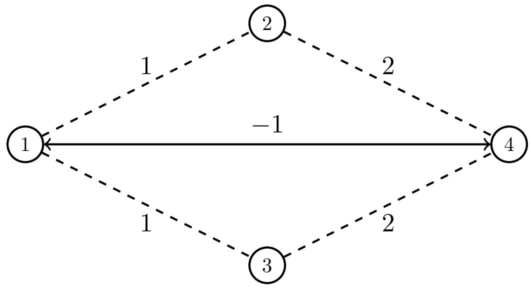

Figure 8: A friendship matching game with correlated preferences and symmetric non-positive friendship values that does not have a friendship stable matching

Example: The game is depicted in Figure 8. There are four agents. Friendship values are symmetric and agent 1 and agent 4 feel negatively towards each other. Both agents can potentially match with agent 2 or agent 3 and are indifferent between these options. Agent 2 and 3 prefer 4 over 1. Now, 4 cannot match both 2 and 3 at the same time. If 1 is single, there is a perceived blocking pair. W.l.o.g. let us assume 2 is not matched. Once $\{1, 2\}$ is formed, agent 4 is negatively effected, as he receives -1 of perceived benefit from agent 1. Thus, even if he is currently matched to 3, it is profitable for 4 to switch to 2, because this way 1 is free again and the negative perceived benefit vanishes. Now 1 is free and the same sequence repeats with the roles of 2 and 3 exchanged. ■

In the remainder, we assume friendship values are nonnegative and symmetric regarding all potential coalitions. Given an instance of friendship stable matching consisting of a graph $G = (V_M, E_M)$, edge benefits b and symmetric friendship values c, we construct an equivalent coalition formation game with consistent constraints as follows. We keep the agent set and the coalition set. Since the constraints are in terms of externalities, we set $\mathcal{C}_g = E_M$ and $T = \emptyset$. To express the perceived benefits, we alter the benefits to

$$b'(\{u, v\}) = b(\{u, v\}) + c_{u,v} b(\{u, v\}) = b(\{u, v\}) + c_{v,u} b(\{u, v\}) \ .$$

For the domination rules, we have $D = D_1 \cup D_2 \cup D_b$, where

$$\begin{aligned} D_1 &= \{(\{\{u, v\}\}, \{u, v'\}) \mid \{u, v\}, \{u, v'\} \in E_M, (c_{v',v} + c_{v',u}) b(\{u, v\}) \geq (1 + c_{v',u}) b(\{u, v'\})\}, \\ D_2 &= \{(\{\{u, v\}, \{u', v'\}\}, \{u, v'\}) \mid \{u, v\}, \{u, v'\}, \{u', v'\} \in E_M \\ &\qquad (1 + c_{v',u'}) b(\{u', v'\}) + (c_{v',v} + c_{v',u}) b(\{u, v\}) \geq (1 + c_{v',u}) b(\{u, v'\})\} \ . \end{aligned}$$

Every rule in D_1 captures a situation, where some agent v' gains so much perceived benefit through a coalition $\{u, v\}$ he is not involved in that it is not profitable for him to break this coalition and form

$\{u, v'\}$. In D_2 we additionally express all situations where two edges would have to be destroyed to form $\{u, v'\}$, but their combination more profitable. Observe that every rule of D_1 also appears in D_2 in combination with any other coalition. However, when a state consists of only one coalition, we could not apply those rules of D_2. Thus, we cannot forgo D_1. The situation that an agent does not want to switch from a coalition to a more worthy one because of perceived benefit is captured neither in D_1 nor in D_2 – this is already expressed by the altered benefits.

For this construction we give a formal correctness proof. All other constructions above can be proven similarly.

Proposition 8 *The described coalition formation game with consistent generation and domination rules is equivalent to friendship matching.*

Proof. We have to show both directions: Every resolution of a perceived blocking pair in the original game is correctly mirrored by a resolution of an equivalent blocking coalition in the associated coalition formation game with constraints $\mathcal{G} = (V, \mathcal{C}, w, \mathcal{C}_g, T, D)$. Conversely, every resolution of a blocking coalition in the coalition formation game is captured by a resolution of the corresponding perceived blocking pair in the original game. We start with the first direction.

Let M be a matching in G. Assume there is a perceived blocking pair $e' = \{u, v'\}$ for M, which we intend to resolve. Note that domination can only occur through edges involving u or v'.

Firstly, if u and v' are both unmatched, then $\{u, v'\}$ is undominated, as no overlapping coalitions are present in the current state. Then, because we can generate all matching edges as candidate coalitions via $\mathcal{C}_g = \mathcal{C}$, $\{u, v'\}$ is a blocking coalition. After adding e', no edge is removed. In the same manner generating $\{u, v'\}$ does not result in the deletion of any existing coalitions. Hence, the set of coalitions resulting from the rules above exactly represents the matching after resolving the perceived blocking pair e'.

Secondly, assume that agent u is matched. He wants to replace the incident edge $e = \{u, v\}$ with e' while v' is unmatched. Then $\{u, v\}$ is part of the current state, and v' is not part of any existing coalition. As e' is a perceived blocking pair, we know that u improves by switching from v to v', that is,

$$\begin{aligned} & b(e) + \alpha_{u,v} b(e) + \sum_{u' \in V} \alpha_{u,u'} b(M \setminus \{e\}, u') \\ < \quad & b(e') + \alpha_{u,v'} b(e') + \sum_{u' \in V} \alpha_{u,u'} b(M \setminus \{e\}, u'), \end{aligned}$$

which cancels out to $b(e) + \alpha_{u,v} b(e) < b(e') + \alpha_{u,v'} b(e')$. Thus, $\{u, v'\}$ is not dominated by $\{u, v\}$ through D_b. Now, $\{u, v\}$ might still dominate $\{u, v'\}$ through D_1. But then $\alpha_{v,v'} b(e) + \alpha_{u,v'} b(e) \geq b(e') + \alpha_{u,v'} b(e')$, that is, the gain v' receives through its friendships with v and u from e is at least as large as the gain it would receive by matching with u (directly and through friendship). This contradicts the assumption that e' is a perceived blocking pair. Other coalitions involving u, v' are not present. Hence, $\{u, v'\}$ is an undominated candidate coalition. After adding $\{u, v'\}$, $\{u, v\}$ is dominated through benefit and hence gets dropped. Again, the set of coalitions resulting from our rules exactly corresponds to the matching after resolving the perceived blocking pair e'.

Thirdly, assume both u and v' are already matched. Agent u wants to drop $e = \{u, v\}$ and v wants to drop $e'' = \{u', v'\}$ to form the new edge e'. Then $\{u, v\}$ and $\{u', v'\}$ are part of the current state. The previous arguments for edges that dominate $\{u, v'\}$ through D_1 or D_b can again be applied. It remains to check whether domination via D_2 is possible. But the domination rules in D_2 imply that the loss caused by giving up e **and** e'' for v' is at least as large as the gain generated from e'. Thus, as e' is a perceived blocking pair, there is no rule in D_2 relating $\{\{u, v\}, \{u', v'\}\}$ to $\{u, v'\}$. Again, no other coalitions involving u, v' are present, and we can generate $\{u, v'\}$. Now $\{u, v\}$ and $\{u', v'\}$ are dominated via D_b and thus dropped. The same happens with M which becomes $M \setminus \{e, e''\} \cup \{e'\}$.

Again, the set of coalitions resulting from our rules exactly correspond to the matching after resolving the perceived blocking pair $\{u, v'\}$. This proves the first direction.

Conversely, let $\mathcal{S}$ be a feasible coalition structure in our coalition formation game with constraints. Observe that feasibility regarding the rules defined above implies that $\mathcal{S}$ must correspond to a feasible matching M. Further, let $\{u, v'\} \notin \mathcal{S}$ be an undominated coalition. Assume for contradiction that $e' = \{u, v'\}$ is not a perceived blocking pair.

Firstly, let v' and u both be unmatched. Then the edge can be formed without removing any edges. As the benefit caused by e' is strictly positive, e' hence is a perceived blocking pair.

Secondly, let v' be unmatched but u be matched. Then there is some edge $e = \{u, v\} \in \mathcal{S}$. Obviously, $\{u, v\}$ dominates $\{u, v'\}$ neither through D_b nor through D_1. Thus, $b(e) + \alpha_{u,v} b(e) < b(e') + \alpha_{u,v'} b(e')$ and $\alpha_{v,v'} b(e) + \alpha_{u,v'} b(e) < b(e') + \alpha_{u,v'} b(e')$. But then e' is a perceived blocking pair.

Thirdly, assume that both v' and u are both matched in M. Then there are edges $e = \{u, v\}$ and $e'' = \{u', v'\}$ in $\mathcal{S}$. As $\{u, v\}$ is a blocking coalition, neither one nor both of those two coalitions combined form the precondition of a domination rule with target coalition $\{u, v\}$. As a consequence, $(1 + c_{v',u'} b(\{u', v'\})) + (c_{v',v} + c_{v',u}) b(\{u, v\}) < (1 + c_{v',u}) b(\{u, v'\})$ and $(1 + c_{u,v} b(\{u, v\})) + (c_{u,v'} + c_{u,u'}) b(\{u', v'\}) < (1 + c_{u,v'}) b(\{u, v'\})$, that is, it is profitable for both u and v' to drop v and u' and form e'. Again, e' is a perceived blocking pair in M. □

When we consider larger coalitions, the symmetry condition for every coalition C implies $c_{u,v} = c_{v,u} = c_{u,w}$ for all $u, v, w \in C$. In consequence, there must be one value c_C that defines the mutual perceived value of all pairs in C. Moreover, $c_C = c_{C'}$ whenever $|C \cap C'| \geq 2$. Arguably, this seems quite a strong condition to assume. Nevertheless, for the sake of completeness, we briefly show that this general domain also implies consistent rules.

As for considerate coalition formation, we now have to consider for every coalition all collections of overlapping coalitions that form a coalition structure. For the same reasons as in the matching case it is not sufficient to consider inclusion maximal sets of coalitions. We keep the set of agents and the set of potential coalitions, and define $\mathcal{C}_g = \mathcal{G}$ and $T = \emptyset$. Further, we set $b'(C) = b(C) + (|C| - 1) c_{u,v} b(C)$ for some $u, v \in C$. Note that this definition is only consistent since we assume equal c-values for all pairs of agents within a coalition. We define a function $p(v, C) = \sum_{v' \in C} c_{v,v'} b(C)$ for all $C \in \mathcal{C}$ to capture the amount of perceived benefit agent v receives from some coalition C. Note that if $v \in C$, we have $p(v, C) = b(C) + (|C| - 1) c_{u,v} b(C) = b'(C)$. Then

$$D = \{(\mathcal{S}, C) \mid \mathcal{S} \subset \mathcal{C} \text{ coalition structure}, C \in \mathcal{C} \setminus \mathcal{S}, \exists v \in C : \sum_{C' \in \mathcal{S} : C \cap C' \neq \emptyset} p(v, C) \geq b'(C)\}.$$

This set encompasses D_b by including all rules where $\mathcal{S} = \{C'\}$ with $C \cap C' \neq \emptyset$ and $b'(C') \geq b'(C)$.

Proposition 9 *Friendship coalition formation gives rise to consistent generation and domination rules for non-negative, symmetric friendship values.*