
Clustering via Hedonic Games: New Concepts and Algorithms

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 We study fundamental connections between coalition formation games and cluster-
2 ing, illustrating the cross-disciplinary relevance of these concepts. We focus on
3 graphical hedonic games where agents’ preferences are compactly represented by
4 a friendship graph and an enmity graph. In the context of clustering, friendship
5 relations naturally align with data point similarities, whereas enmity corresponds to
6 dissimilarities. We consider two stability notions based on single-agent deviations:
7 local popularity and local stability. Exploring these concepts from an algorithmic
8 viewpoint, we design efficient mechanisms for finding locally stable or locally
9 popular partitions. Besides gaining theoretical insight into the computational
10 complexity of these problems, we perform simulations that demonstrate how our
11 algorithms can be successfully applied in clustering and community detection. Our
12 findings highlight the interplay between coalition formation games and data-driven
13 clustering techniques, offering fresh perspectives and applications in both areas.

14 1 Introduction

15 Coalition formation and clustering are two research areas that capture different aspects of grouping a
16 set of entities in a meaningful or optimal way. While coalition formation studies autonomous agents
17 who have preferences over the various ways in which they can be partitioned into groups, clustering
18 aims to unearth hidden structure and similarities within a set of datapoints. Despite the shared theme
19 of group formation, there is no substantial interplay between these two research fields.

20 This study aims to investigate how different solution concepts emerging from coalition formation—in
21 particular, from hedonic games—can be used in data clustering. Traditional clustering metrics
22 measure the quality of the boundaries between the clusters or the average similarities within clusters.
23 We propose a criterion that is based on a form of stability: we aim for a clustering where no deviation
24 by a single agent (or data point) should lead to a more preferred outcome. We aim to investigate how
25 such stability notions can be harnessed in clustering algorithms.

26 Similarities and dissimilarities, the baseline concepts used by all clustering algorithms, can be
27 associated with relations of compatibility and of conflict between two agents. Leveraging this idea,
28 we focus on a model in coalition formation due to Dimitrov et al. [16] where agents’ preferences over
29 different coalitions are determined by their friendships and their enmities, represented as two graphs.
30 More precisely, we adopt the model of hedonic *friends–enemies–neutrals* (FEN) games [32, 36, 26]
31 where agents classify all other agents as a friend, an enemy, or a neutral, and each agent evaluates a
32 given partition based on the number of friends and the number of enemies in their coalition.

33 Unfortunately, finding “optimal” or, in some sense, *stable* partitions often turns out to be a computa-
34 tionally intractable task. Since computational efficiency is often crucial in clustering applications,
35 we focus our attention on *local* stability concepts requiring only that re-assigning a single agent

to a different coalition does not yield a partition deemed better for the majority of agents. Such single-agent deviations were studied in FEN games by Kerkmann and Rothe [27] and, more recently, in friends–enemies (FE) games—a model excluding neutrality—by Brandt et al. [10] who showed that in a model where two partitions are compared by taking the majority vote of the agents, such deviations converge into a *locally stable* partition in a polynomial number of rounds. We extend these results to FEN games and also to a slightly more stringent stability concept, *local popularity*. As opposed to the FE games studied by Brandt et al. [10], the option of neutrality introduces a flexibility that allows focusing on relations between points that are either very similar, or very different. Furthermore, neutrality yields smaller degrees in the friendship and enmity graphs which, as we show, has a huge benefit in terms of the convergence speed. Depending on the application under focus, different emphasis might be given to grouping together similar data points versus separating dissimilar data points. In the context of a FEN game, agents may put a significantly larger weight on the objective of being grouped together with friends than being separated from enemies, or vice versa; a third, more balanced approach is when agents treat these two objectives with the same importance. These options give rise to three preference domains—friend-appreciating, enemy-averse, and balanced—whose study may facilitate the application of our results to different clustering domains.

By identifying “similarity” with “friendship” and “difference” with “enmity,” we demonstrate via simulations how our algorithms for finding locally stable or locally popular partitions can be used in the context of data clustering. The stability property enforced by our algorithm is particularly useful in clustering tasks where agents have meaningful local interactions or preferences. This includes standard benchmarks such as citation networks (e.g., Cora [35]), email communication graphs (e.g., Enron [29]), and social networks, where misclassifying a node can distort community structure. By ensuring that no agent prefers to switch to a different cluster based on local relationships, our method yields clusters that are robust against local deviations. This distinguishes our approach from traditional methods like k -means or DBSCAN, which may prioritize global fit over local consistency. As our simulations attest, this interdisciplinary approach can be surprisingly successful when compared to the most prevalent clustering algorithms. Thus we hope our work can spark a stronger cooperation between these fields to explore the applicability of coalition formation games in clustering.

Related work. Due to the interdisciplinary nature of this paper, we only point out the most important precursors of our work and a few key references for more background on the relevant research areas. Coalition formation is a central topic in cooperative game theory. Starting with the seminal work of Drèze and Greenberg [17], researchers’ focus shifted to so-called hedonic games where agents’ preferences depend only on the coalition they are contained in; see also [5, 8, 13] and the book chapters by Aziz and Savani [4] and Bullinger et al. [11, Section 3.6]. FE games were introduced by Dimitrov et al. [16] as a subclass of hedonic games with a concise description (namely, the agents’ friendship relations). FEN games, the extension of their model that allows for neutrality, have also been studied [14, 26, 27, 32, 36, 39]. Single-agent deviations and their induced dynamics were investigated in FE games by Brandt et al. [10] who, in various settings, showed that such deviations converge to a partition that fulfills certain stability concepts. Among the stability concept they considered is what we call *local stability*; our work extends their results on this notion to FEN games. Our stability concepts stem from the idea of popularity as introduced by Gärdenfors [21]. Popularity has been widely studied in matching and allocation problems [1, 6, 15, 24, 25] and has also been considered in the context of hedonic games by, e.g., Aziz et al. [3], Kerkmann et al. [26], Brandt and Bullinger [9], Kerkmann and Rothe [28], and Bullinger and Gilboa [12].

For a recent survey on clustering—a fundamental problem in unsupervised learning aimed at discovering structure in data by grouping similar elements—see [19]. Classical clustering methods such as k -means, introduced by MacQueen [33], optimize intra-cluster similarity under geometric assumptions, while DBSCAN by Ester et al. [18] identifies dense regions in data and is robust to noise. The most relevant subfield of clustering for our purposes is community detection, which seeks to partition the nodes of a graph into densely connected groups with sparse interconnections. Here, the Louvain algorithm of Blondel et al. [7] and its improved variant Leiden by Traag et al. [40] are state-of-the-art methods that optimize modularity to reveal community structure in networks. Our main metric to compare our algorithms with k -means, DBSCAN, Louvain, and Leiden is the Rand distance [38], which—assuming some true labeling for the data points—computes the fraction of pairs of points that are correctly clustered together or separately.

Our contribution. We study two stability concepts in hedonic FEN games: local popularity and local stability. The former—introduced in this paper—assures that no single agent can deviate such that

more agents improve than get worse, while the latter requires this only for Nash deviations (where the deviating agent improves). Our main result is a simple algorithm that always finds a locally popular partition in polynomial time if friendship and enmity relationships are symmetric in all three settings (friend-appreciative, enemy-averse, and balanced). For asymmetric relationships, we show that a locally stable partition can still be found efficiently in the balanced case, but all other settings yield NP-hardness. See Table 1 for a summary of our results.

We also conduct an experimental study comparing our algorithms to common clustering and community detection algorithms, in particular, to k -means, DBSCAN, Louvain, and Leiden. Our simulations provided sufficient data to show that our algorithms are well suited for clustering and community detection tasks, often outperforming existing solutions.

2 Preliminaries

Let N denote the set of agents. A subset $C \subseteq N$ is called a *coalition*; in particular, N is the *grand coalition*. In this paper, *partition* refers to a family of pairwise disjoint nonempty subsets of N whose union is N . Given a partition π , the coalition in π that contains $i \in N$ is denoted by $\pi(i)$.

A *hedonic game* is a pair (N, \succeq) , where $\succeq = (\succeq_i)_{i \in N}$ is a preference profile that contains for each agent i a weak preference list \succeq_i over the possible coalitions containing i . That is, for any two coalitions C, C' with $i \in C \cap C'$, either $C' \succ_i C$ (meaning that C' is preferred to C by i), or $C \succ_i C'$ (meaning that C is preferred to C' by i), or $C' \sim_i C$ (meaning that i is indifferent between the two coalitions). In a hedonic game, the preference lists of the agents only depend on the coalitions they are in. Hence, \succeq_i also induces a weak ranking of agent i over the possible partitions.

Given an agent $i \in N$, a partition π , and a coalition $C \in \pi \cup \{\emptyset\}$, the deviation where i leaves its original coalition $\pi(i)$ and joins C (or forms a new coalition $\{i\}$ in the case when $C = \emptyset$) is referred to as the *switch* $i \rightarrow C$. We use $\pi_{i \rightarrow C}$ to denote the partition resulting from π by the switch $i \rightarrow C$, that is, $\pi_{i \rightarrow C} = (\pi \setminus \{\pi(i), C\}) \cup \{\pi(i) \setminus \{i\}, C \cup \{i\}\}$. We say that i has an incentive to join C if $C \cup \{i\} \succ_i \pi(i)$; in this case, $i \rightarrow C$ is a *Nash deviation*. More generally, extending this notion to partitions, we say that an agent j votes for (or, j votes against) a switch $i \rightarrow C$ if j prefers $\pi_{i \rightarrow C}$ to π (or, j prefers π to $\pi_{i \rightarrow C}$, respectively). Further, we define

$$\text{vote}_j(\pi, \pi') = \begin{cases} +1 & \text{if } \pi'(j) \succ_j \pi(j), \\ -1 & \text{if } \pi(j) \succ_j \pi'(j), \\ 0 & \text{if } \pi(j) \sim_j \pi'(j) \end{cases}$$

for partitions π and π' over N . Let $\Lambda(\pi, \pi') = \sum_{j \in N} \text{vote}_j(\pi, \pi')$ be the sum of the agents' votes for π' against π . We will further write $\Lambda_{-i}(\pi, \pi') = \sum_{j \in N \setminus \{i\}} \text{vote}_j(\pi, \pi')$ for the sum of votes for π' against π among all agents excluding agent i .

Definition 1. Given a partition π , an agent i , and a coalition $C \in \pi \cup \{\emptyset\}$, we say that the switch $i \rightarrow C$ *locally dominates* π if $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$; and the switch $i \rightarrow C$ *locally blocks* π if it is a Nash deviation that locally dominates π . A partition π is *popular* if for any partition π' , we have that $\Lambda(\pi, \pi') \leq 0$. Moreover, π is *locally popular* if there is no switch that locally dominates π , and π is *locally stable* if there is no switch that locally blocks π .

Note that a single deviation by one agent can also change the situation—and thus the voting behavior on these two partitions—of other agents, giving rise to the above notions of local popularity and local stability. Note further that local popularity implies local stability.

Preferences in FEN games. Dimitrov et al. [16] introduced a preference domain for hedonic games where agents' preferences over the coalitions containing them depend only on the number of friends and enemies they have in their coalition. Thus, in a *FEN game* over agent set N , each agent i classifies all remaining agents into one of three disjoint categories: those that i considers either *friends*, *enemies*, or *neutrals*; we denote these agent sets by $F(i)$, $E(i)$, and $N(i)$, respectively. The friendship and enmity relations can be represented by two directed graphs, the *friendship digraph* $G^F = (N, A^F)$ and the *enmity digraph* $G^E = (N, A^E)$, where an arc $(i, j) \in A^F$ means that i considers j a friend, whereas an arc $(i, j) \in A^E$ means that i considers j an enemy. We denote by Δ the maximum number of arcs incident to any agent in G^F or in G^E . A FEN game is *symmetric* if friendship and enmity relations are both symmetric; it is *asymmetric* otherwise.

Table 1: Summary of our results on the computational complexity of LOCAL-POPULARITY-EXISTENCE and LOCAL-STABILITY-EXISTENCE. We remark that a locally stable or locally popular partition always exists in the settings where the corresponding problem is proven to be in P. By contrast, a locally stable or locally popular partition may not exist in those settings where the corresponding problem is proven to be NP-complete (abbreviated as “NP-c”).

	LOCAL-STABILITY-EXISTENCE		LOCAL-POPULARITY-EXISTENCE	
	symmetric	asymmetric	symmetric	asymmetric
FEN-B game	P (Thm. 2)	P (Thm. 4)	P (Thm. 2)	NP-c (Thm. 9)
FEN-AF game	P (Thm. 2)	NP-c (Thm. 7)	P (Thm. 2)	NP-c (Thm. 7)
FEN-AE game	P (Thm. 2)	NP-c (Thm. 13)	P (Thm. 2)	NP-c (Thm. 13)

141 FEN games are subclasses of so-called *additively separable hedonic games* [5] where each agent i
 142 has a *valuation function* v_i over $N \setminus \{i\}$ that is defined as

$$v_i(j) = \begin{cases} \alpha_f & \text{if } i \text{ considers } j \text{ a friend, i.e., } j \in F(i), \\ \alpha_e & \text{if } i \text{ considers } j \text{ an enemy, i.e., } j \in E(i), \\ 0 & \text{if } i \text{ considers } j \text{ a neutral, i.e., } j \in N(i) \end{cases}$$

143 for some integers $\alpha_f > 0$ and $\alpha_e < 0$, and agent i 's *utility in a given coalition C containing i* is
 144 defined as $u_i(C) = \sum_{j \in C \setminus \{i\}} v_i(j)$; we refer to the pair (α_f, α_e) as the *type* of the FEN game. Then
 145 the preferences of agent i are based on i 's utilities: for two coalitions C and C' both containing i ,
 146 $C \succeq_i C'$ if and only if $u_i(C) \geq u_i(C')$. For a coalition $C \subseteq N$ with $i \in C$, we let $f_C^i = |F(i) \cap C|$
 147 and $e_C^i = |E(i) \cap C|$ denote the number of i 's friends and enemies within C , respectively; then
 148 $u_i(C) = f_C^i \cdot \alpha_f + e_C^i \cdot \alpha_e$. The *utility of agent i in a partition π* is $u_i(\pi) = u_i(\pi(i))$.

149 **Friend-appreciating, enemy-averse, and balanced preferences.** Following Dimitrov et al. [16],
 150 we introduce the following preference domains for FEN games. A *FEN game with appreciation of*
 151 *friends* (or a *FEN-AF game*, for short) is a FEN game over agent set N whose type is $(|N|, -1)$.
 152 That is, $C \succeq_i C'$ holds for two coalitions C, C' with $i \in C \cap C'$ if and only if either $f_C^i > f_{C'}^i$, or
 153 $f_C^i = f_{C'}^i$ and $e_C^i \leq e_{C'}^i$. By contrast, a *FEN game with aversion to enemies* (or a *FEN-AE game*,
 154 for short) is a FEN game over agent set N whose type is $(1, -|N|)$. That is, $C \succeq_i C'$ holds for two
 155 coalitions C, C' with $i \in C \cap C'$ if and only if either $e_C^i < e_{C'}^i$, or $e_C^i = e_{C'}^i$ and $f_C^i \geq f_{C'}^i$. Finally,
 156 a *FEN game with balanced preferences* (or a *FEN-B game*, for short) is a FEN game with type $(1, -1)$.
 157 That is, $C \succeq_i C'$ holds for two coalitions C, C' with $i \in C \cap C'$ if and only if $f_C^i - e_C^i \geq f_{C'}^i - e_{C'}^i$.
 158 We will call $u_i^B(C) = f_C^i - e_C^i$ the *balanced utility of i in C* .

159 **Problem definitions.** Let us define the decision problems related to the existence of locally popular
 160 and locally stable partitions. Given a FEN game I of type (α_f, α_e) defined over a set N of agents with
 161 a friendship digraph G^F and an enmity digraph G^E , the LOCAL-POPULARITY-EXISTENCE problem
 162 asks if there exists a locally popular partition for I ; analogously, the LOCAL-STABILITY-EXISTENCE
 163 problem—given the same input—asks if there exists a locally stable partition for I .

164 3 Local heuristics

165 We present two local heuristics, one to find a locally popular partition, the other to find a locally
 166 stable partition. Both heuristics start from an arbitrary partition and iteratively modify it using only
 167 single-agent deviations; the conditions that determine whether or not to perform a given switch is a
 168 local condition that can be checked efficiently.

- 169 **LocPop:** 1. Start from an arbitrary partition π .
 2. While there exists a coalition $C \in \pi \cup \{\emptyset\}$ and an agent $i \notin C$ such that
 $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$, replace π with $\pi_{i \rightarrow C}$ and continue.
- 170 **LocStab:** 1. Start from an arbitrary partition π .
 2. While there exists a coalition $C \in \pi \cup \{\emptyset\}$ and an agent $i \notin C$ such that
 $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$ and $C \cup \{i\} \succ_i \pi(i)$, replace π with $\pi_{i \rightarrow C}$ and continue.

171 LocPop and LocStab can both be modified such that the number of coalitions within the partition
 172 remains fixed: it suffices to consider switches of the form $i \rightarrow C$ where $C \in \pi$ and $\{i\} \notin \pi$ in Step 2.

173 In the rest of this section, we show that the above heuristics converge in certain FEN games, that is,
 174 they arrive at a locally popular or locally stable partition in a finite number of steps. We remark that
 175 LocPop and LocStab cannot guarantee to achieve popularity; in fact, we show in Appendix A that
 176 deciding whether there exists a popular partition—or even just verifying popularity—is coNP-hard.

177 Our convergence results both for symmetric and asymmetric FEN games rely on a *potential function*
 178 that we define as the total balanced utility of agents: $f(\pi) = \sum_{i \in N} u_i^B(\pi(i)) = \sum_{i \in N} f_{\pi(i)}^i - e_{\pi(i)}^i$.

179 3.1 Symmetric FEN games

180 **Lemma 1.** *For a FEN game I over agent set N , let π be a partition, $i \in N$ an agent, and $C \in \pi \cup \{\emptyset\}$
 181 a coalition with $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$. Then (a) $\sum_{j \in N \setminus \{i\}} u_j^B(\pi_{i \rightarrow C}) - u_j^B(\pi) = \Lambda_{-i}(\pi_i, \pi_{i \rightarrow C})$.
 182 Moreover, if I is a symmetric FEN game, then (b) $u_i^B(C) - u_i^B(\pi(i)) = \Lambda_{-i}(\pi_i, \pi_{i \rightarrow C})$ and
 183 (c) $f(\pi_{i \rightarrow C}) = f(\pi_i) + 2\Lambda_{-i}(\pi_i, \pi_{i \rightarrow C})$.*

184 *Proof.* First, consider the agents in $N \setminus \{i\}$. Each such agent j may gain or lose a friend or an enemy
 185 by the switch $i \rightarrow C$, and thus its balanced utility increases exactly by the value of $\text{vote}_j(\pi, \pi_{i \rightarrow C}) \in$
 186 $\{-1, 0, 1\}$ as a result of the switch $i \rightarrow C$. Summing up the increase in the balanced utility of all
 187 agents in $N \setminus \{i\}$, we therefore obtain $\Lambda_{-i}(\pi, \pi_{i \rightarrow C})$, proving statement (a) of the lemma.

188 Assume now that I is symmetric. Consider the contribution of agent i to $f(\pi_{i \rightarrow C}) - f(\pi)$. The
 189 change in the balanced utility of i as a result of the switch $i \rightarrow C$ is exactly $u_i^B(C) - u_i^B(\pi(i)) =$
 190 $\Lambda_{-i}(\pi, \pi_{i \rightarrow C})$ as friendships are symmetric, so each agent who gains (or loses) a friend or an enemy
 191 due to $i \rightarrow C$ is in turn a friend or an enemy gained (or lost) by i ; this yields statement (b).

192 Statement (c) then follows by summing up (a) and (b). \square

193 **Theorem 2.** *In a symmetric FEN-AF, FEN-AE, or FEN-B game, irrespective of the initial partition,
 194 heuristic LocPop always converges in $\mathcal{O}(n\Delta^2)$ steps (in $\mathcal{O}(n\Delta)$ steps for FEN-B games) to a locally
 195 popular (and thus also locally stable) partition, and heuristic LocStab always converges in $\mathcal{O}(n\Delta)$
 196 steps to a locally stable partition.*

197 *Proof.* Consider a step in heuristic LocPop or LocStab where a partition π is replaced by $\pi_{i \rightarrow C}$ for
 198 some agent $i \in N$ and coalition $C \in \pi \cup \{\emptyset\}$. Then $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$ implies $\Lambda_{-i}(\pi, \pi_{i \rightarrow C}) \geq 0$.
 199 By statement (b) of Lemma 1, this immediately yields that $f(\pi_{i \rightarrow C}) \geq f(\pi)$, meaning that f is
 200 nondecreasing during the execution of the heuristic. As the total balanced utility is within the range
 201 $[-n\Delta, n\Delta]$, there can be at most $2n\Delta$ steps when $f(\pi)$ strictly increases.

202 Let us now consider the case when $f(\pi_{i \rightarrow C}) = f(\pi)$, i.e., the potential function does not increase;
 203 our aim is to bound the number of such steps during the execution of the heuristic. By statement (b)
 204 of Lemma 1, by $f(\pi_{i \rightarrow C}) = f(\pi)$ we must have $\Lambda_{-i}(\pi, \pi_{i \rightarrow C}) = 0$. By statement (a) of Lemma 1,

$$\Lambda_{-i}(\pi, \pi_{i \rightarrow C}) = f_C^i - e_C^i - f_{\pi(i)}^i + e_{\pi(i)}^i = 0. \quad (1)$$

205 We introduce $g^F(\pi) = |\{(i, j) \in A^F : \pi(i) = \pi(j)\}|$ and $g^E(\pi) = -|\{(i, j) \in A^E : \pi(i) = \pi(j)\}|$.
 206 We distinguish between the following three cases depending on the type of the FEN game.

207 **Case 1: Balanced preferences.** By Lemma 1, the switch $i \rightarrow C$ increases the balanced util-
 208 ity of i by exactly $\Lambda_{-i}(\pi, \pi_{i \rightarrow C}) = 0$. Since agent i has balanced preferences, this implies
 209 $\text{vote}_i(\pi, \pi_{i \rightarrow C}) = 0$. However, then $\Lambda(\pi, \pi_{i \rightarrow C}) = 0$ follows, contradicting our assumption that the
 210 condition $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$ holds. This proves $f(\pi_{i \rightarrow C}) > f(\pi)$ for balanced preferences.

211 **Case 2: Friend-appreciating preferences.** Since we assumed friendships to be symmetric,
 212 $g^F(\pi_{i \rightarrow C}) - g^F(\pi) = 2f_C^i - 2f_{\pi(i)}^i$. We claim that $g^F(\pi_{i \rightarrow C}) > g^F(\pi)$. Indeed, as the switch $i \rightarrow C$
 213 was performed by the heuristic based on the condition $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$, it follows that agent i must
 214 prefer $\pi_{i \rightarrow C}$ to π ; hence, either $f_C^i > f_{\pi(i)}^i$ holds as promised, or $f_C^i = f_{\pi(i)}^i$ and $e_C^i < e_{\pi(i)}^i$;
 215 however, (1) excludes the latter.

216 **Case 3: Enemy-averse preferences.** Since we assumed enmities to be symmetric, we have
 217 $g^E(\pi_{i \rightarrow C}) - g^E(\pi) = 2e_C^i - 2e_{\pi(i)}^i$. We claim that $g^E(\pi_{i \rightarrow C}) > g^E(\pi)$. As in Case 2, we

218 have $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$, which implies that agent i must prefer $\pi_{i \rightarrow C}$ to π ; hence, either $e_C^i < e_{\pi(i)}^i$
 219 holds as promised, or $e_C^i = e_{\pi(i)}^i$ and $f_C^i > f_{\pi(i)}^i$; however, (1) again excludes the latter.

220 **Total number of steps.** We have shown that the number of steps when the potential function f
 221 strictly increases is at most $2n\Delta$. In each such step, g^F and g^E might each decrease by at most Δ ,
 222 yielding an upper bound of $2n\Delta^2$ on the total decrease during the algorithm. Since g^F takes values
 223 within $[0, n\Delta]$ and g^E between $[-n\Delta, 0]$, the number of steps when $f(\pi)$ does not increase is at
 224 most $2n\Delta^2 + n\Delta$. This gives a bound of $\mathcal{O}(n\Delta^2)$ on the number of steps taken by the heuristics.

225 It remains to show that heuristic LocStab (and LocPop for FEN-B games) converges in $\mathcal{O}(n\Delta)$ steps.
 226 In FEN-B games, f always strictly increases, as we showed, hence we need $\mathcal{O}(n\Delta)$ steps. Otherwise,
 227 in FEN-AE and FEN-AF games, a switch $i \rightarrow C$ is only performed in LocStab if $f_C^i \geq f_{\pi(i)}^i$ (or if
 228 $e_C^i \leq e_{\pi(i)}^i$); hence g^F (or g^E , respectively) never decreases. Thus the number of steps when f does
 229 not increase is at most $n\Delta$ in FEN-AF and FEN-AE games, proving the claim for LocStab. \square

230 By applying an efficient way to decide if a switch is locally dominating and by using a carefully
 231 maintained data structure for quickly finding such switches, our heuristics can be implemented in
 232 near-linear time if the maximum degree Δ of the friendship and enmity digraphs is constant; see
 233 Appendix B. As efficiency is often paramount in clustering, we believe that this result significantly
 234 widens the possibilities for applying our heuristics in practical clustering instances. Theorem 3 is
 235 proven in Appendix B, while all further results marked with (\star) are proven in Appendix D.

236 **Theorem 3 (\star) .** LocPop can be implemented to run in $\mathcal{O}(n\Delta^3 \log n)$ time (or $\mathcal{O}(n\Delta^2 \log n)$ for
 237 FEN-B games) and LocStab can be implemented to run in $\mathcal{O}(n\Delta^2 \log n)$ time.

238 3.2 Asymmetric FEN games

239 The following example shows that heuristics LocPop and LocStab may not stop in an asymmetric
 240 FEN game with friend-appreciating or enemy-averse preferences.

241 **Example 1.** In an *asymmetric* FEN-AF or FEN-AE game, heuristics LocPop and LocStab can run into
 242 a cycle and never stop. Let there be seven agents: $0, 1, \dots, 6$. Each agent i considers $(i + 1) \bmod 7$
 243 a friend but $(i + 2) \bmod 7$ and $(i + 3) \bmod 7$ enemies; see Figure 1a for an illustration. (Here and
 244 anywhere else, all relations that are not explicitly mentioned are tacitly assumed to be neutral.)

245 In the AF domain, let the initial partition be $\{\{0, 1, 2, 3\}, \{4, 5, 6\}\}$. Then agent 3 has an incentive
 246 to join $\{4, 5, 6\}$ and this switch takes place, as agents 0, 1, and 3 vote for it and only agent 2 votes
 247 against it. By symmetry, this creates a loop: agent 6 switches next, then agent 2, then agent 5, etc.

248 In the AE domain, let the initial partition be $\{\{0, 1\}, \{2, 3\}, \{4, 5, 6\}\}$. Then agent 4 has an incentive
 249 to join $\{2, 3\}$ and this switch takes place, as agents 4 and 3 vote for it and only agent 2 votes against
 250 it. Again, this creates a loop: agent 2 switches next, then agent 0, then agent 5, etc.

251 **Example 2.** In an *asymmetric* FEN-B game, heuristic LocPop can run into a cycle and never stop.
 252 Let the agent set be $\{0, 1, 2\}$, where each agent i considers $(i - 1) \bmod 3$ a friend and $(i + 1) \bmod 3$
 253 an enemy. Let the initial partition consist of two coalitions: $\{0, 1\}$ and $\{2\}$. Then both 0 and 2 vote
 254 for the switch of $1 \rightarrow \{2\}$, which hence takes place. Next, the switch $2 \rightarrow \{0\}$ takes place, followed
 255 by the switch $0 \rightarrow \{1\}$, and the heuristic arrives back at the initial partition.

256 Contrasting Examples 1 and 2, heuristic LocStab always stops if agents' preferences are balanced.

257 **Theorem 4 (\star) .** In a possibly asymmetric FEN-B game, the total balanced utility of all agents
 258 strictly increases in each step of heuristic LocStab; consequently, heuristic LocStab always converges
 259 in $\mathcal{O}(n\Delta)$ steps to a locally stable partition, irrespective of the initial partition.

260 4 Hardness results for local popularity and local stability

261 **Friend-appreciating FEN games.** Let us introduce the gadget which lies at the heart of our hardness
 262 results on FEN-AF games; see Figure 1b for an illustration.

263 **Definition 2.** Let $k \geq 5$ be an integer with $(2k + 1) \bmod 3 \neq 0$. A set $\{0, 1, \dots, 2k\}$ of agents
 264 forms an *AF-gadget* of size $2k + 1$ if each agent i considers $(i + 1) \bmod (2k + 1)$ a friend (so that the
 265 friendship graph is a directed cycle) and considers $(i + 2), (i + 3), \dots, (i - 4) \bmod (2k + 1)$ enemies.

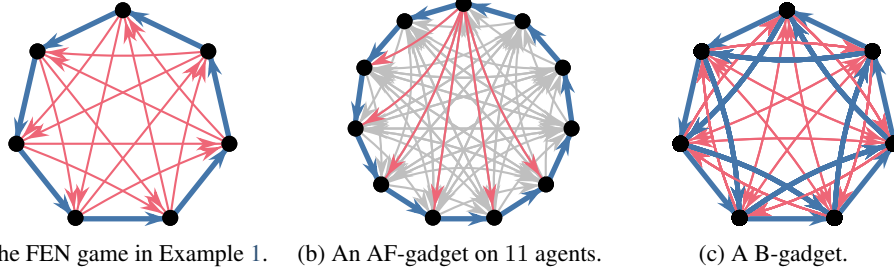


Figure 1: Illustration for Example 1 and Definitions 2 and 3. We use thick, blue arrows for friendship arcs and thin, red ones for enmity arcs (some enmity arcs in (b) are shown in gray for better visibility).

Theorem 5 (*). *The asymmetric FEN-AF game consisting solely of an AF-gadget admits no locally popular partition.*

Building on AF-gadgets, a more involved construction shows that even a locally stable solution may fail to exist. In fact, Theorem 7 shows that it is NP-hard to decide whether a locally popular or a locally stable partition exists; our reductions are from 5-SAT and make ample use of AF-gadgets.

Theorem 6 (*). *There exists an asymmetric FEN-AF game that admits no locally stable partition.*

Theorem 7 (*). *LOCAL-POPULARITY-EXISTENCE and LOCAL-STABILITY-EXISTENCE are NP-complete for FEN-AF games.*

Balanced FEN games. Turning our attention to balanced FEN games, recall that given a FEN-B game I , heuristic LocStab always finds a locally stable partition for I in polynomial time, as stated in Theorem 4. By contrast, we introduce a gadget that is similar to the AF-gadget from Definition 2 but is tailored for showing that FEN-B games may not admit a locally popular partition. Relying on B-gadgets, by a reduction from the NP-complete problem 3-COLORING we can show the NP-hardness of the related decision problem, as stated in Theorem 9.

Definition 3. A set $\{0, 1, \dots, 6\}$ of agents forms a *B-gadget* if each agent i considers $(i + 1) \bmod 7$ and $(i + 2) \bmod 7$ a friend and considers $(i + 3)$, $(i + 4)$, $(i + 5) \bmod 7$ enemies.

Theorem 8 (*). *The asymmetric FEN-B game consisting solely of a B-gadget admits no locally popular partition.*

Theorem 9 (*). *LOCAL-POPULARITY-EXISTENCE is NP-complete for FEN-B games.*

As we have shown in Theorem 4, heuristic LocStab increases the agents' total balanced utility—the potential function f —in each step when running on a FEN-B game. Hence, a partition π maximizing $f(\pi)$ is locally stable but, as Theorem 8 shows, not necessarily locally popular. Interestingly, not even a popular partition is guaranteed to reach the maximum total balanced utility, even in symmetric FEN-B games; see Proposition 10. Moreover, deciding whether there exists a partition that achieves a given total balanced utility is NP-hard, as shown by Theorem 11.

Proposition 10 (*). *There exists a symmetric FEN-B game that admits a partition that is popular but does not maximize the agents' total balanced utility.*

Theorem 11 (*). *Given a symmetric FEN-B game I and an integer t , the problem of deciding whether I admits a partition π whose total balanced utility is at least t is NP-complete.*

Enemy-averse FEN games. Similarly to the gadgets from Definitions 2 and 3, we introduce an AE-gadget for FEN-AE games for proving that such games may not admit a locally stable partition. A reduction from 3-SAT based on AE-gadgets shows the intractability of deciding the existence of locally popular or locally stable partitions. For $k \in \mathbb{N} \setminus \{0\}$, we write $[k]$ as a shorthand for $\{1, 2, \dots, k\}$.

Definition 4. An *AE-gadget* over a player set $K = \{c\} \cup \{b_i : i \in [12]\} \cup \{a_i, d_i, x_i : i \in [3]\}$ has the following friendship and enemy relations: c considers a_i and d_i for $i \in [3]$ friends and agents x_1, x_2, x_3, b_{11} , and b_{12} enemies; b_i considers c , b_{i+1} , and b_{i+2} a friend for each $i \in [8]$; b_9 considers b_{10} a friend, and b_{10} considers b_{11} a friend; b_{11} and b_{12} consider each other and all agents in $\{b_i : i \in [10]\}$ friends and c an enemy; a_i considers x_i, d_i , and c a friend for each $i \in [3]$; d_1, d_2 , and d_3 consider c a friend; and x_i considers a_i a friend for each $i \in [3]$. See Appendix D.3 for an illustration.

307 **Theorem 12** (*). *The asymmetric FEN-AE game consisting solely of an AE-gadget admits no locally*
 308 *stable partition.*

309 **Theorem 13** (*). *LOCAL-POPULARITY-EXISTENCE and LOCAL-STABILITY-EXISTENCE are NP-*
 310 *complete for FEN-AE games.*

311 5 Simulations

312 We complement our theoretical study with a range of experiments.

313 **Setup.** We implemented LocPop and LocStab in Python and run the simulations on a computer with
 314 AMD Ryzen 7735HS CPU and 16GB RAM. All codes are available in the supplementary material.

315 For our algorithms, we chose parameters $f \leq e \in [0, 1]$ to create the friendship and enmity graphs: for
 316 two data points x and y at distance $d(x, y)$ in an instance I ,¹ we added (x, y) to the friendship graph
 317 if $d(x, y) \leq f \cdot \text{diam}(I)$ and to the enmity graph if $d(x, y) > e \cdot \text{diam}(I)$, where $\text{diam}(I)$ denotes
 318 the *diameter* of I , i.e., the maximum distance between any two points in I . We used the parameter
 319 values $(f, e) \in \{(0.2, 0.2), (0.25, 0.35), (0.4, 0.4)\}$. For each parameterization, we considered the
 320 *appreciation-of-friends* (AF), *aversion-to-enemies* (AE), and the *balanced* (B) preference domains.

321 **Community Detection.** For community detection, we used four different datasets.

- 322 1. **Karate club** [22, 30, 42]: a 34-node benchmark dataset for community detection.
- 323 2. **Jazz musicians** [23, 31]: collaboration network of 198 jazz musicians; nodes represent
 324 musicians, edges represent co-membership in a band.
- 325 3. **Cora dataset** [35, 34]: a citation network of 2708 machine learning papers classified into
 326 seven classes; edges denote citation links.
- 327 4. **Random-25**: an instance containing 25 disjoint Erdős-Rényi graphs (10 nodes each, with
 328 $p = 0.2$) as communities; inter-community edges added independently with probability 0.05.

329 We compared our heuristics with two state-of-the-art algorithms, called Louvain and Leiden [2]. We
 330 tested our heuristics with respect to several different parameters. In particular, we considered the
 331 following variations for the initial clustering: (i) putting each agent into a singleton cluster (LocPop-S,
 332 LocStab-S), (ii) dividing agents randomly into k clusters where k is the predicted number of clusters
 333 (LocPop-P, LocStab-P), and (iii) using the output of the Leiden algorithm (LocPop-Ld, LocStab-Ld).

334 For evaluation, we used *Rand index* and *modularity*, both commonly used metrics. The Rand index
 335 assumes underlying true labels for the data points, and measures the fraction of such (x, y) pairs
 336 where x and y are correctly put together or into different clusters. See Figure 2 and Appendix C.1.

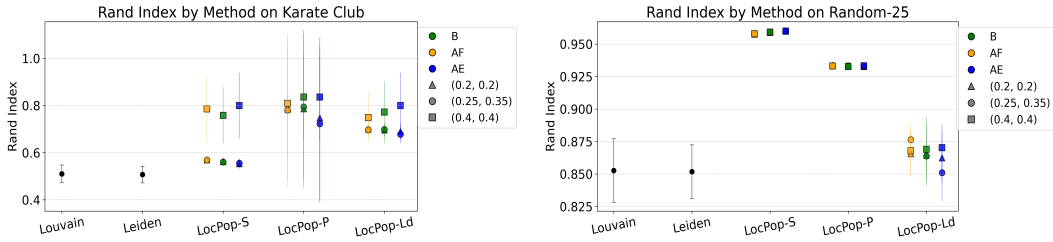


Figure 2: Comparing Rand index of Leiden, Louvain, and LocPop variants on the Karate club dataset (left) and the synthetic dataset Random-25 (right). In all figures henceforth, 2-sigma error bars show the standard deviation of the plotted data.

337 **Clustering.** For our clustering simulations, we used four datasets.

- 338 1. **Iris dataset** [20]: 150 samples from three Iris species, features are sepal and petal sizes.

¹When points are described via their coordinates, we take the Euclidean distance, while in an instance described as a network, the distance of x and y is measured by the number of edges of a shortest xy -path.

- 339 2. **Breast cancer Wisconsin dataset** [41]: 569 samples from diagnostic images (30 dimensional datapoints).
- 340
- 341 3. **Moons dataset** [37]: two half-moons generated by `make_moons` (300 points, 0.05 noise).
- 342 4. **3-Circles dataset**: 300 points in three slightly overlapping circles centered at (0.5, 0.5),
- 343 (0.7, 0.3), and (0.1, 0.7) with radius 0.2 and Gaussian noise ($\text{std} = 0.05$).

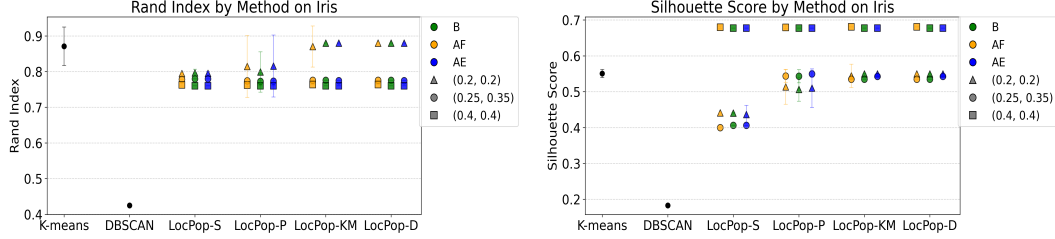


Figure 3: Comparing Rand index (left) and silhouette score (right) of k -means, DBSCAN, and LocPop variants for the Iris dataset.

344 We compared our algorithms to two widely used clustering algorithms, k -means and DBSCAN,
 345 using their initial parameters. We tested our heuristic with respect to several different parameters. In
 346 particular, we considered the following variations for the initial clustering: (i) putting every agent
 347 into a singleton cluster (LocPop-S, LocStab-S), (ii) dividing the agents randomly into k clusters
 348 where k is the predicted number of clusters (LocPop-P, LocStab-P), (iii) using the output of k -means
 349 (LocPop- k M, LocStab- k M), and (iv) using the output of DBSCAN (LocPop-D, LocStab-D).

350 To evaluate the outputs, we considered the Rand index and *silhouette score*, both being commonly
 351 used metrics in clustering. See Figure 3 and Appendix C.2.

352 **Key insights from experimental results.** Our algorithms run within seconds, except for Cora, where
 353 computation takes 3-5 minutes, probably due to high degrees in the friendship and enmity graphs.
 354 LocPop and LocStab performed very similarly, with LocStab depending more on the preference
 355 domain, so we discuss LocPop here. A more detailed evaluation can be found in Appendix C.3.

356 LocPop demonstrated strong and consistent empirical performance, performing very similarly over
 357 the three preference domains (AF, AE, and B). We observed that parameter tuning can significantly
 358 affect outcomes, usually either the setting (0.2, 0.2) or (0.4, 0.4) performed best, and (0.25, 0.35)
 359 performed consistently between them. Interestingly, stronger performance on standard clustering
 360 metrics (silhouette score) often did not align with being closer to the true labels (Rand index).

361 In community detection tasks, LocPop matched or outperformed Louvain and Leiden with respect to
 362 the Rand index, particularly excelling on the Karate club ($\sim 50\%$ better) and Random-25 ($\sim 10\%$
 363 better) datasets. The LocPop-Ld variant also matched them with respect to modularity. In clustering
 364 benchmarks, it mostly surpassed DBSCAN in both silhouette score and Rand index, and also matched
 365 or outperformed k -means (with up to 20% in 3 Circles and Iris in silhouette score). These findings
 366 indicate that LocPop is not only theoretically grounded but also competitive in practice, offering a
 367 flexible and efficient approach for structure discovery while providing additional robustness against
 368 strategic deviations.

369 6 Conclusion

370 We have investigated the computational complexity of finding a popular, a locally popular, or a locally
 371 stable partition in FEN-AF, FEN-AE, and FEN-B games and identified which cases were solvable
 372 efficiently and which were NP-hard. Our algorithms can also be interpreted as dynamics of agents
 373 based on single-agent deviations and majority-based decision making.

374 Our experimental study provided sufficient background to show that social-choice concepts like local
 375 stability and local popularity may lead to efficient clustering and community detection techniques.
 376 Furthermore, by their nature, these concepts may often provide fairer solutions that are also more
 377 stable against potential deviations, which—as argued in the introduction—is particularly relevant in
 378 scenarios where data points correspond to agents capable of making independent actions.

References

- [1] D. J. Abraham, R. W. Irving, T. Kavitha, and K. Mehlhorn. Popular matchings. *SIAM Journal on Computing*, 37(4):1030–1045, 2007.
- [2] D. Albert. Louvain and Leiden implementation. <https://git.esclear.de/esclear/louvain-leiden>, Sept. 2023.
- [3] H. Aziz, F. Brandt, and H. Seedig. Computing desirable partitions in additively separable hedonic games. *Artificial Intelligence*, 195:316–334, 2013.
- [4] H. Aziz and R. Savani. Hedonic games. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 15, pages 356–376. Cambridge University Press, 2016.
- [5] S. Banerjee, H. Konishi, and T. Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18(1):135–153, 2001.
- [6] P. Biró, R. W. Irving, and D. F. Manlove. Popular matchings in the marriage and roommates problems. In *Proceedings of the 7th International Conference on Algorithms and Complexity*, pages 97–108. Springer LNCS #6078, 2010.
- [7] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
- [8] A. Bogomolnaia and M. Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.
- [9] F. Brandt and M. Bullinger. Finding and recognizing popular coalition structures. *Journal of Artificial Intelligence Research*, 74:569–626, 2022.
- [10] F. Brandt, M. Bullinger, and L. Tappe. Single-agent dynamics in additively separable hedonic games. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, pages 4867–4874. AAAI Press, 2022.
- [11] M. Bullinger, E. Elkind, and J. Rothe. Cooperative game theory. In J. Rothe, editor, *Economics and Computation. An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, Classroom Companion: Economics, chapter 3, pages 139–229. Springer, 2nd edition, 2024.
- [12] M. Bullinger and M. Gilboa. Settling the complexity of popularity in additively separable and fractional hedonic games. Technical Report arXiv:2411.05713 [cs.GT], ACM Computing Research Repository (CoRR), Nov. 2024.
- [13] K. Cechlárová and A. Romero-Medina. Stability in coalition formation games. *International Journal of Game Theory*, 29(4):487–494, 2001.
- [14] J. Chen, G. Csáji, S. Roy, and S. Simola. Hedonic games with friends, enemies, and neutrals: Resolving open questions and fine-grained complexity. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems*, pages 251–259. IFAAMAS, May/June 2023.
- [15] A. Cseh. Popular matchings. In U. Endriss, editor, *Trends in Computational Social Choice*, chapter 6, pages 105–122. AI Access Foundation, 2017.
- [16] D. Dimitrov, P. Borm, R. Hendrickx, and S. Sung. Simple priorities and core stability in hedonic games. *Social Choice and Welfare*, 26(2):421–433, 2006.
- [17] J. Drèze and J. Greenberg. Hedonic coalitions: Optimality and stability. *Econometrica*, 48(4):987–1003, 1980.
- [18] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 226–231. AAAI Press, 1996.

- [19] A. E. Ezugwu, A. M. Ikotun, O. O. Oyelade, L. Abualigah, J. O. Agushaka, C. I. Eke, and A. A. Akinyelu. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, 110:104743, 2022.
- [20] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.
- [21] P. Gärdenfors. Match making: assignments based on bilateral preferences. *Behavioural Science*, 20:166–173, 1975.
- [22] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.
- [23] P. M. Gleiser and L. Danon. Community structure in jazz. *Advances in Complex Systems*, 6(4):565–573, 2003.
- [24] S. Gupta, P. Misra, S. Saurabh, and M. Zehavi. Popular matching in roommates setting is NP-hard. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2810–2822, 2019.
- [25] T. Kavitha, J. Mestre, and M. Nasre. Popular mixed matchings. *Theoretical Computer Science*, 412(24):2679–2690, 2011.
- [26] A. Kerkmann, J. Lang, A. Rey, J. Rothe, H. Schadrack, and L. Schend. Hedonic games with ordinal preferences and thresholds. *Journal of Artificial Intelligence Research*, 67:705–756, 2020.
- [27] A. Kerkmann and J. Rothe. Stability in FEN-hedonic games for single-player deviations. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems*, pages 891–899. IFAAMAS, May 2019.
- [28] A. Kerkmann and J. Rothe. The complexity of verifying popularity and strict popularity in altruistic hedonic games. *Journal of Autonomous Agents and Multi-Agent Systems*, 38(2):Article 47, 2024.
- [29] B. Klimt and Y. Yang. The Enron Corpus: A new dataset for email classification research. In *Proceedings of the 15th European Conference on Machine Learning*, volume 66, pages 217–226. Springer *LNAI* #3201, 2004.
- [30] J. Kunegis. KONECT, Jazz musicians dataset. <http://konect.cc/networks/arenas-jazz>, 1977.
- [31] J. Kunegis. KONECT, Zachary karate club dataset. <http://konect.cc/networks/ucidata-zachary>, 2003.
- [32] J. Lang, A. Rey, J. Rothe, H. Schadrack, and L. Schend. Representing and solving hedonic games with ordinal preferences and thresholds. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems*, pages 1229–1237. IFAAMAS, May 2015.
- [33] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, volume 5.1, pages 281–297. University of California press, 1967.
- [34] A. K. McCallum. Cora. <https://dx.doi.org/10.21227/jsg4-wp31>, 2024.
- [35] A. K. McCallum, K. Nigam, J. Rennie, and K. Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3:127–163, 2000.
- [36] K. Ohta, N. Barrot, A. Ismaili, Y. Sakurai, and M. Yokoo. Core stability in hedonic games among friends and enemies: Impact of neutrals. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 359–365. AAAI Press, Feb. 2017.

- 472 [37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel,
473 P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher,
474 M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine*
475 *Learning Research*, 12:2825–2830, 2011.
- 476 [38] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American*
477 *Statistical Association*, 66(336):846–850, 1971.
- 478 [39] J. Rothe, H. Schadrack, and L. Schend. Borda-induced hedonic games with friends, enemies,
479 and neutral players. *Mathematical Social Sciences*, 96:21–36, 2018.
- 480 [40] V. A. Traag, L. Waltman, and N. J. Van Eck. From Louvain to Leiden: Guaranteeing well-
481 connected communities. *Scientific Reports*, 9(1):5233, 2019.
- 482 [41] W. Wolberg, O. Mangasarian, N. Street, and W. Street. Breast Cancer Wisconsin (Diagnostic).
483 UCI Machine Learning Repository, 1993. <https://doi.org/10.24432/C5DW2B>.
- 484 [42] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of*
485 *Anthropological Research*, 33(4):452–473, 1977.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: In the abstract, we claim to study algorithmic questions related to graphical hedonic games—this is done in detail in Sections 3 and 4—and argue that our algorithms are also applicable for general clustering and community detection tasks—which is backed up by our simulations in Section 5.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: The main limitation for our algorithms is the relatively slow running time for very dense graphs, as stated in Theorems 3 and 4. This is discussed in detail in Appendix C.3 along with a possible way to overcome this issue.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: We provide the formal definitions for our framework in Section 2 and prove each of our statements rigorously in Sections 3 and 4. Note that due to space constraints, many of these proofs are deferred to Appendix D.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide all the codes used for our simulations and synthetic datasets in the supplementary material files of the submission. For the publicly available datasets, we provided citations for their availability.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

591 (d) We recognize that reproducibility may be tricky in some cases, in which case
592 authors are welcome to describe the particular way they provide for reproducibility.
593 In the case of closed-source models, it may be that access to the model is limited in
594 some way (e.g., to registered users), but it should be possible for other researchers
595 to have some path to reproducing or verifying the results.

596 5. Open access to data and code

597 Question: Does the paper provide open access to the data and code, with sufficient instruc-
598 tions to faithfully reproduce the main experimental results, as described in supplemental
599 material?

600 Answer: [Yes]

601 Justification: We provided all the codes used for our simulations in our supplementary
602 material files. We only used publicly available datasets, and synthetic datasets—for which
603 the codes are also included—rendering our simulations completely reproducible.

604 Guidelines:

- 605 • The answer NA means that paper does not include experiments requiring code.
- 606 • Please see the NeurIPS code and data submission guidelines ([https://nips.cc/
607 public/guides/CodeSubmissionPolicy](https://nips.cc/public/guides/CodeSubmissionPolicy)) for more details.
- 608 • While we encourage the release of code and data, we understand that this might not be
609 possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not
610 including code, unless this is central to the contribution (e.g., for a new open-source
611 benchmark).
- 612 • The instructions should contain the exact command and environment needed to run to
613 reproduce the results. See the NeurIPS code and data submission guidelines ([https:
614 //nips.cc/public/guides/CodeSubmissionPolicy](https://nips.cc/public/guides/CodeSubmissionPolicy)) for more details.
- 615 • The authors should provide instructions on data access and preparation, including how
616 to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- 617 • The authors should provide scripts to reproduce all experimental results for the new
618 proposed method and baselines. If only a subset of experiments are reproducible, they
619 should state which ones are omitted from the script and why.
- 620 • At submission time, to preserve anonymity, the authors should release anonymized
621 versions (if applicable).
- 622 • Providing as much information as possible in supplemental material (appended to the
623 paper) is recommended, but including URLs to data and code is permitted.

624 6. Experimental setting/details

625 Question: Does the paper specify all the training and test details (e.g., data splits, hyper-
626 parameters, how they were chosen, type of optimizer, etc.) necessary to understand the
627 results?

628 Answer: [Yes]

629 Justification: Our simulations did not include training, nor data splits. Regarding the tests,
630 we provide all the parameters and setting we used in Section 5.

631 Guidelines:

- 632 • The answer NA means that the paper does not include experiments.
- 633 • The experimental setting should be presented in the core of the paper to a level of detail
634 that is necessary to appreciate the results and make sense of them.
- 635 • The full details can be provided either with the code, in appendix, or as supplemental
636 material.

637 7. Experiment statistical significance

638 Question: Does the paper report error bars suitably and correctly defined or other appropriate
639 information about the statistical significance of the experiments?

640 Answer: [Yes]

641 Justification: We provide 2-sigma error bars for all our plots in Section 5. The caption of
642 Figure 2 describes the type of error bars used. We note that in many cases, the bars are too
643 small (e.g. 10^{-4}) to be visible in the figure.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: In Section 5, we provided sufficient details regarding the hardware used for our simulations and the execution times.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: All our research was conducted in line with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: Our algorithms have no immediate societal impacts. However, we briefly discuss in Section 1 that using our algorithms for clustering agents, capable of performing actions and deviations, helps to avoid instability and strategic deviations. This may lead to some positive societal impacts in such cases. We believe that there are no negative societal impacts of our work.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper proposes no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We provide citations for all the (publicly available) datasets that we used in Section 5.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: All details to reproduce the synthetic datasets that we generated are included in Section 5. Furthermore, the codes for these are also available in the supplementary materials of the submission.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: There was no crowdsourcing nor research with human subjects in this project.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

801 Answer: [NA]
 802 Justification: There was no research with human subjects.
 803 Guidelines:
 804 • The answer NA means that the paper does not involve crowdsourcing nor research with
 805 human subjects.
 806 • Depending on the country in which research is conducted, IRB approval (or equivalent)
 807 may be required for any human subjects research. If you obtained IRB approval, you
 808 should clearly state this in the paper.
 809 • We recognize that the procedures for this may vary significantly between institutions
 810 and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the
 811 guidelines for their institution.
 812 • For initial submissions, do not include any information that would break anonymity (if
 813 applicable), such as the institution conducting the review.

814 **16. Declaration of LLM usage**

815 Question: Does the paper describe the usage of LLMs if it is an important, original, or
 816 non-standard component of the core methods in this research? Note that if the LLM is used
 817 only for writing, editing, or formatting purposes and does not impact the core methodology,
 818 scientific rigorousness, or originality of the research, declaration is not required.

819 Answer: [NA]
 820 Justification: LLMs were not used in core methodology.
 821 Guidelines:
 822 • The answer NA means that the core method development in this research does not
 823 involve LLMs as any important, original, or non-standard components.
 824 • Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>)
 825 for what should or should not be described.

Appendix A Existence of a popular partition in FEN games

Recall that Theorems 2 and 3 say that *locally* popular partitions *always* exist and can be efficiently found by a simple heuristic in symmetric FEN-AF and FEN-AE games. In this section, we show that the same does not hold for popularity: there exist symmetric FEN games where for each possible partition π , there is a deviation involving multiple agents that yields a partition π' with $\Lambda(\pi, \pi') > 0$. Specifically, we show that the verification problem for popularity in symmetric FEN games is coNP-complete. Furthermore, we show that deciding whether a popular partition exists in such games is computationally intractable as well; specifically, the existence problem for popularity in symmetric FEN games is coNP-hard. These results on the complexity of popularity verification and existence are the same as those known for other types of hedonic games, such as the altruistic hedonic games studied by Kerkmann and Rothe [28]. Interestingly, Bullinger and Gilboa [12] recently settled the complexity of whether there exist popular partitions in either additively separable or fractional hedonic games by showing that these two problems are complete for NP^{NP} , the second level of the polynomial hierarchy—the first results closing the gap between the computational upper and lower bounds for popularity existence in hedonic games. For our problems, we leave this question open.

We prove the following result for all FEN games, among them FEN-AF, FEN-AE, and FEN-B games.

Theorem 14. *Given a FEN game I of some fixed, arbitrary type with symmetric preferences, verifying whether a partition is popular in I is coNP-complete, while deciding whether a popular partition exists for I is coNP-hard.*

Proof. It is clear that verifying whether a given partition π is popular in a given FEN game I of some fixed, arbitrary type with symmetric preferences is in coNP, because we can nondeterministically guess a partition π' and check in polynomial time whether $\Lambda(\pi, \pi') > 0$. This is the case for some partition π' exactly if π is not popular in I .

To prove coNP-hardness, we reduce from the EXACT 3-COVER problem whose input is a set $U = \{u_1, \dots, u_{3n}\}$ and a set family $\mathcal{F} = \{F_1, \dots, F_m\}$ with each F_i being a size-3 subset of U ; the task is to decide whether there exists a set family $\mathcal{S} \subseteq \mathcal{F}$ of size n such that $U = \bigcup_{F_i \in \mathcal{S}} F_i$.

We construct a FEN game I of the given type as follows; our reduction is correct irrespective of the type of I . For each element $u_i \in U$, we create an agent \bar{u}_i , and for each set $F_j \in \mathcal{F}$, we create an agent \bar{F}_j ; we use the notation $\bar{U} = \{\bar{u} : u \in U\}$ and $\bar{\mathcal{F}} = \{\bar{F}_j : F_j \in \mathcal{F}\}$. We further add two sets A and A' of *anchor* agents with $|A| = |A'| = 5n + m$ and a set B of *guard* agents with $|B| = 2n - 1$. Hence, the set of agents in I is $A \cup A' \cup B \cup \bar{U} \cup \bar{\mathcal{F}}$.

The relationships for I are defined as follows; recall that all relationships in I are symmetric:

- Each anchor agent in A is friends with all other anchor agents in A and all agents in \bar{U} .
- Each anchor agent in A' is friends with all other anchor agents in A' and all guard agents.
- Each guard agent is friends with all agents in $\bar{\mathcal{F}}$.
- An agent $\bar{u}_i \in \bar{U}$ is friends with an agent $\bar{F}_j \in \bar{\mathcal{F}}$ if and only if $u_i \in F_j$.
- Each anchor agent in A and each anchor agent in A' are enemies.

We claim that I admits a popular partition if and only if $\pi^* = \{A \cup \bar{U}, A' \cup B \cup \bar{\mathcal{F}}\}$ is a popular coalition, which in turn happens if and only if our input instance $I_{\text{X3C}} = (U, \mathcal{F})$ of EXACT 3-COVER does *not* have a solution, i.e., there is no exact cover of B in \mathcal{F} .

First, let us show that if $\mathcal{S} \subseteq \mathcal{F}$ is a solution to I_{X3C} , then π^* is not popular. To this end, let $\bar{\mathcal{S}} = \{\bar{F}_j : F_j \in \mathcal{S}\}$. We are going to show that $\Lambda(\pi, \pi^*) > 0$ where the partition

$$\pi = \{A \cup \bar{U} \cup \bar{\mathcal{S}}, A' \cup B \cup \bar{\mathcal{F}} \setminus \bar{\mathcal{S}}\} \quad (2)$$

is obtained from π by a deviation of agents in $\bar{\mathcal{S}}$; this implies that π^* is not popular. Note that each element $u_i \in U$ is contained in some set $F_j \in \mathcal{S}$, which means that each agent $\bar{u}_i \in \bar{U}$ has more friends in $\pi(\bar{u}_i)$ than in $\pi^*(\bar{u}_i)$ (and zero enemies in both) and thus prefers π to π^* . By contrast, each guard agent $b \in B$ has fewer friends in $\pi(b)$ than in $\pi^*(b)$ (and zero enemies

in both) and thus prefers π^* to π . Furthermore, each agent $\overline{F}_j \in \overline{\mathcal{S}}$ gains three new friends and loses all of its guard friends when switching from π^* to π (and has zero enemies in both) and thus prefers π^* to π . Since every remaining agent is indifferent between π and π^* , we obtain that $\Lambda(\pi, \pi^*) = |\overline{U}| - |B| - |\mathcal{S}| = 3n - (2n - 1) - n = 1 > 0$.

Next, we show that if π^* is not popular, then I_{X3C} has a solution. Let π be a partition that is more popular than π^* , i.e., $\Lambda(\pi^*, \pi) > 0$. Observe that for an anchor agent $a \in A \cup A'$, all friends and no enemies of a are in $\pi^*(a)$, so anchor agents cannot prefer π to π^* . Moreover, if some anchor agent prefers π^* to π , then at least $|A| = |A'| = 5n + m$ anchor agents do so; however, at most $|\overline{U}| + |\overline{\mathcal{F}}| + |B| = 3n + m + 2n - 1$ agents may prefer π to π^* , which contradicts $\Lambda(\pi^*, \pi) > 0$. Thus anchor agents must be indifferent between π and π^* , which yields that π contains two coalitions C and C' such that $C \supseteq A \cup \overline{U}$ and $C' \supseteq A' \cup B$.

Define $\overline{\mathcal{S}} = C \cap \overline{\mathcal{F}}$; note that $|\overline{\mathcal{S}}| > 0$, as otherwise, $\pi = \pi^*$. Note that each guard agent $b \in B$ has fewer friends in $\pi(b)$ than in $\pi^*(b)$ (and zero enemies in both) and thus prefers π to π^* , and so do the agents in $\overline{\mathcal{S}}$ as well (they gain three friends each from \overline{U} but lose all guards from among their friends when deviating to π). Let \overline{V} be the set of agents in \overline{U} who prefer π to π^* ; notice that $\overline{V} = \{\overline{u}_i : \exists \overline{F}_j \in \overline{\mathcal{S}} \text{ such that } u_i \in F_j\}$ by construction.

Then the agents preferring π to π^* are those in \overline{V} while those that prefer π^* to π are those in $B \cup \overline{\mathcal{S}}$. Hence, $\Lambda(\pi^*, \pi) = |\overline{V}| - |B| - |\mathcal{S}| \geq 1$, which implies that

$$|\overline{V}| \geq (2n - 1) + |\mathcal{S}| + 1. \quad (3)$$

Recall also that

$$|\overline{V}| \leq 3n \quad \text{and} \quad |\overline{V}| \leq 3|\mathcal{S}|, \quad (4)$$

because each agent in $\overline{\mathcal{S}}$ has three friends in \overline{V} . Taking the linear combination of these two inequalities, we obtain that $|\overline{V}| \leq 2n + |\mathcal{S}|$, and thus the three inequalities in (3) and (4) must all hold with equality. In particular, $|\overline{V}| = 3n$ and $|\mathcal{S}| = n$, and the former implies $\overline{V} = \overline{U}$, that is, $\bigcup_{\overline{F}_j \in \overline{\mathcal{S}}} F_j = U$. Therefore, the set family $\mathcal{S} = \{F_j : \overline{F}_j \in \overline{\mathcal{S}}\}$ is a solution to the EXACT 3-COVER instance I_{X3C} .

It remains to show that if π^* is not popular, then there is no popular partition π at all. For the sake of contradiction, assume that π is popular. We have already proven that $\Lambda(\pi, \pi^*) \geq 0$ implies that π must be of the form (2) for some $\overline{\mathcal{S}} \subseteq \overline{\mathcal{F}}$. By our assumption that $\pi^* \neq \pi$, we also know that $\overline{\mathcal{S}}$ is nonempty. However, then the switch $\overline{F}_j \rightarrow A' \cup B \cup \overline{\mathcal{F}} \setminus \overline{\mathcal{S}}$ locally blocks π for any agent $\overline{F}_j \in \overline{\mathcal{S}}$, because all guard agents as well as \overline{F}_j votes for it, whereas at most three agents in \overline{U} vote against it. This shows that no partition other than π^* may be popular in I . \square

Next, we show that popular partitions may fail to exist even in a symmetric *FE game* where neutral relations are not allowed and thus every agent that is not a friend of i is an enemy of i , i.e., the friendship digraph and the enmity digraph are complements of each other. This suggests that the intractability shown in Theorem 14 does not stem from the possibility of neutrality.

Proposition 15. *There exists a FE game with symmetric and friend-appreciative (or, equivalently, with symmetric and enemy-averse) preferences that does not admit a popular partition.*

Proof. Consider the FE game I with symmetric and friend-appreciative (or equivalently, symmetric and enemy-averse) preferences on agent set $N = \{0, 1, \dots, 6\}$ where the friendship graph is a bidirected cycle of length 7: every agent $i \in N$ considers both $(i - 1) \bmod 7$ and $(i + 1) \bmod 7$ a friend and each other agent in $N \setminus \{i\}$ an enemy. We claim that I does not admit a popular partition.

First, the grand coalition $\pi^N = \{N\}$ is not popular, as $\Lambda(\pi^N, \pi_{i \rightarrow \emptyset}^N) > 0$ for each agent i : even though i and both friends of i vote against the switch $i \rightarrow \emptyset$, all four remaining agents vote for it.

Second, if π is a partition such that there are at least four agents $i \in N$ with $f_{\pi(i)}^i \leq 1$, then $\Lambda(\pi, \pi^N) > 0$, because these four agents prefer the grand coalition to π . It follows that any popular partition must be of the form $\pi^i = \{\{i\}, N \setminus \{i\}\}$ for some $i \in N$. However, in this case, the switch $(i + 1) \bmod 7 \rightarrow \{i\}$ locally blocks π^i , because all agents vote for this switch except for the agent $(i + 2) \bmod 7$; in particular, $\Lambda(\pi^i, \pi_{(i+1) \bmod 7 \rightarrow \{i\}}^i) > 0$, contradicting the popularity of π^i . Hence, no popular partition exists for I . \square

Appendix B Efficient implementation for LocPop and LocStab

In this section, we present an efficient implementation of LocPop and LocStab that proves Theorem 3.

Theorem 3 (*). LocPop can be implemented to run in $\mathcal{O}(n\Delta^3 \log n)$ time (or $\mathcal{O}(n\Delta^2 \log n)$ for FEN-B games) and LocStab can be implemented to run in $\mathcal{O}(n\Delta^2 \log n)$ time.

Proof. Regarding the running time of LocPop and LocStab, it is clear that each step can be performed in $\mathcal{O}(n\Delta k)$ time even in asymmetric FEN games, where k denotes the number of coalitions in the current partition: to check for each agent $i \in N$ and each coalition $C \in \pi \cup \{\emptyset\}$ whether $i \rightarrow C$ locally dominates or locally blocks, it suffices to know the number of agents in C and in $\pi(i)$ who regard i a friend or an enemy and of those whom i considers a friend or an enemy. As the number of such agents is at most 2Δ , each step can be done in $|N|(k+1)\mathcal{O}(\Delta) = \mathcal{O}(nk\Delta)$ time.

In the following, we show how our heuristics can be implemented much more efficiently; we will focus on symmetric FEN games due to their practical relevance. We start with a way to characterize when a switch $i \rightarrow C$ locally dominates a partition π .

Lemma 16. Given a symmetric FEN game I with a partition π , consider the following conditions for some agent i and coalition $C \in \pi \cup \{\emptyset\}$:

$$\begin{aligned} \text{(c0)} \quad & f_C^i - e_C^i \geq f_{\pi(i)}^i - e_{\pi(i)}^i + 1; & \text{(c3)} \quad & 2f_C^i - e_C^i > 2f_{\pi(i)}^i - e_{\pi(i)}^i; \\ \text{(c1)} \quad & f_C^i - e_C^i \geq f_{\pi(i)}^i - e_{\pi(i)}^i + 2; & \text{(c4)} \quad & f_C^i - 2e_C^i > f_{\pi(i)}^i - 2e_{\pi(i)}^i; \\ \text{(c2)} \quad & f_{\pi(i)}^i - e_{\pi(i)}^i + 1 \geq f_C^i - e_C^i \geq f_{\pi(i)}^i - e_{\pi(i)}^i. \end{aligned}$$

If I is a FEN-B / FEN-AF / FEN-AE game, then the switch $i \rightarrow C$ locally dominates π if and only if condition (c0) / condition (c1) \vee ((c2) \wedge (c3)) / condition (c1) \vee ((c2) \wedge (c4)), respectively, hold.

Proof. Statement (b) of Lemma 1 yields

$$\Lambda(\pi, \pi_{i \rightarrow C}) = f_C^i - e_C^i - f_{\pi(i)}^i + e_{\pi(i)}^i + \text{vote}_i(\pi, \pi_{i \rightarrow C}). \quad (5)$$

We first show that the respective conditions imply that $i \rightarrow C$ locally dominates π .

First, if (c1) holds, then $\Lambda(\pi, \pi_{i \rightarrow C}) \geq 2 + \text{vote}_i(\pi, \pi_{i \rightarrow C}) > 0$ due to (5), so the switch $i \rightarrow C$ locally dominates π even if i votes against it, irrespective of the type of I . If I is a FEN-B game, then (c0) already suffices to ensure that $i \rightarrow C$ locally dominates π , because it implies $\text{vote}_i(\pi, \pi_{i \rightarrow C}) \geq 1$ by the definition of balanced preferences, which in turn yields $\Lambda(\pi, \pi_{i \rightarrow C}) \geq 2$ due to (5).

If (c2) holds, then $\Lambda(\pi, \pi_{i \rightarrow C}) \geq \text{vote}_i(\pi, \pi_{i \rightarrow C})$ due to (5) and, hence, it suffices to show that (c3) implies that $i \rightarrow C$ is a Nash deviation under friend-appreciating preferences, whereas (c4) implies that $i \rightarrow C$ is a Nash deviation under enemy-averse preferences:

- Assume first that I is a FEN-AF game where (c2) and (c3) hold, but $i \rightarrow C$ is not a Nash deviation. Then either $f_{\pi(i)}^i = f_C^i$ and $e_{\pi(i)}^i \leq e_C^i$, or $f_{\pi(i)}^i > f_C^i$. The former case cannot happen, as it contradicts (c3). Hence, we must have $f_{\pi(i)}^i \geq f_C^i + 1$; however, adding this inequality to $f_{\pi(i)}^i - e_{\pi(i)}^i + 1 \geq f_C^i - e_C^i$, we get a contradiction to (c3) again.
- Assume now that I is a FEN-AE game where (c2) and (c4) hold, but $i \rightarrow C$ is not a Nash deviation; our reasoning mirrors the previous case. Then either $e_{\pi(i)}^i = e_C^i$ and $f_{\pi(i)}^i \geq f_C^i$, or $e_{\pi(i)}^i < e_C^i$. The former case cannot happen, as it contradicts (c4). Hence, we must have $e_C^i \geq e_{\pi(i)}^i + 1$; however, adding this inequality to $f_{\pi(i)}^i - e_{\pi(i)}^i + 1 \geq f_C^i - e_C^i$, we get a contradiction to (c4) again.

For the other direction, suppose that $i \rightarrow C$ locally dominates π , i.e., $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$. Then $\Lambda_{-i}(\pi, \pi_{i \rightarrow C}) \geq 0$; recall that $\Lambda_{-i}(\pi, \pi_{i \rightarrow C}) = f_C^i - e_C^i - f_{\pi(i)}^i + e_{\pi(i)}^i$ by statement (b) of Lemma 1, so we get

$$f_C^i - e_C^i \geq f_{\pi(i)}^i - e_{\pi(i)}^i. \quad (6)$$

If $\Lambda_{-i}(\pi, \pi_{i \rightarrow C}) \geq 2$, then (c1) is satisfied. Else, $\Lambda_{-i}(\pi, \pi_{i \rightarrow C}) \in \{0, 1\}$, which implies (c2), and we also get $\text{vote}_i(\pi, \pi_{i \rightarrow C}) \geq 0$ due to (5) and the assumption that $i \rightarrow C$ locally dominates π . In

fact, $\text{vote}_i(\pi, \pi_{i \rightarrow C}) = 0$ is not possible, as that would mean $f_{\pi(i)}^i - e_{\pi(i)}^i = f_C^i - e_C^i$ implying $\Lambda(\pi, \pi_{i \rightarrow C}) = 0$ by (5); a contradiction to our assumption that $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$. Therefore, $i \rightarrow C$ must be a Nash deviation; using this, we now show that the required conditions hold.

- For balanced preferences, this means that (5) implies $\Lambda(\pi, \pi_{i \rightarrow C}) \geq 2$ which, due to (5), implies (c0).
- For friend-appreciating preferences, we get $f_C^i > f_{\pi(i)}^i$, or $f_C^i = f_{\pi(i)}^i$ and $e_C^i < e_{\pi(i)}^i$; taking (6) into account, both of these cases yield $2f_C^i - e_C^i > 2f_{\pi(i)}^i - e_{\pi(i)}^i$, i.e., (c3) holds.
- For enemy-averse preferences, we get $e_C^i < e_{\pi(i)}^i$, or $e_C^i = e_{\pi(i)}^i$ and $f_C^i > f_{\pi(i)}^i$; taking (6) into account, both of these cases imply $f_C^i - 2e_C^i > f_{\pi(i)}^i - 2e_{\pi(i)}^i$, i.e., (c4) holds.

This completes the proof of Lemma 16. \square

Implementation. Let us describe the implementation of heuristics LocPop and LocStab in detail.

The key idea is that for each agent $i \in N$, we maintain two vectors x_i^1 and x_i^2 that are used to store the following values corresponding to each coalition $C \in \pi \cup \{\emptyset\}$. The first vector stores the values $x_i^1(C) = f_C^i - e_C^i - f_{\pi(i)}^i + e_{\pi(i)}^i$, while the values stored in the second vector depend on the type of the game: If condition (c2) of Lemma 16 holds, then x_i^2 stores $x_i^2(C) = 2f_C^i - e_C^i - 2f_{\pi(i)}^i + e_{\pi(i)}^i$ if preferences are friend-appreciating, and $x_i^2(C) = f_C^i - 2e_C^i - f_{\pi(i)}^i + 2e_{\pi(i)}^i$ if preferences are enemy-averse, while we set $x_i^2(C) = -\infty$ if (c2) fails. For balanced preferences, the vector x_i^2 is not used.

Notice that maintaining the values $x_i^1(C)$ and $x_i^2(C)$ for each agent i and for all coalitions $C \in \pi \cup \{\emptyset\}$ for the current partition π would suffice to determine whether there exists a switch $i \rightarrow C$ for some coalition $C \in \pi \cup \{\emptyset\}$ that locally dominates or locally blocks π , due to Lemma 16. However, it is not hard to observe that it suffices to focus for each agent $i \in N$ on the set \mathcal{C}^i of coalitions in π in which i has at least one friend or enemy. Then it is sufficient to maintain the values $x_i^1(C)$ and $x_i^2(C)$ for each agent i only for coalitions $C \in \mathcal{C}^i$, together with the values $f_{\pi(i)}^i$ and $e_{\pi(i)}^i$. Indeed, for every coalition $C \in \pi \cup \{\emptyset\} \setminus \mathcal{C}^i$ we know $e_C^i = f_C^i = 0$, and thus the values $x_i^1(C)$ and $x_i^2(C)$ can be calculated easily: first, we have $x_i^1(C) = -f_{\pi(i)}^i + e_{\pi(i)}^i$; second, to determine $x_i^2(C)$, we first check whether condition (c2) holds (using $e_C^i = f_C^i = 0$ again) and set $x_i^2(C) = -\infty$ if (c2) fails and, otherwise, set $x_i^2(C) = -2f_{\pi(i)}^i + e_{\pi(i)}^i$ in FEN-AF games and $x_i^2(C) = -f_{\pi(i)}^i + 2e_{\pi(i)}^i$ in FEN-AE games. Notice that the values $x_i^1(C)$ and $x_i^2(C)$ therefore are the same for each coalition C in $\pi \cup \{\emptyset\} \setminus \mathcal{C}^i$.

Therefore, the vectors x_i^1 and x_i^2 are indexed by the set of all coalitions in \mathcal{C}^i together with one dummy coalition representing all coalitions in $\pi \cup \{\emptyset\} \setminus \mathcal{C}^i$.

To facilitate an efficient search for a locally dominating or locally blocking switch, we keep the entries in the vectors x_i^1 and x_i^2 ordered. Furthermore, we additionally order the set $X^1 = \{x_i^1 : i \in N\}$ of vectors according to their first coordinates, and order the vectors in $X^2 = \{x_i^2 : i \in N\}$ similarly. Observe that the largest entry occurring in any vector of X^1 is therefore the first coordinate of the first vector in X^1 and, similarly, the largest entry occurring in any vector of X^2 is the first coordinate of the first vector in X^2 . We will maintain this property throughout the algorithm; call it property (\star) .

The initialization. To initialize \mathcal{C}^i and the vectors x_i^1 and x_i^2 for the starting partition π_0 , we start by computing the values f_C^i and e_C^i for all agents $i \in N$ and coalitions $C \in \pi_0$ in which i has a friend or an enemy. This takes $\mathcal{O}(n\Delta)$ time in total for all agents using the friendship and enmity digraphs.

Next, we order the $\mathcal{O}(\Delta)$ values in each of the $2n$ vectors in $X^1 \cup X^2$; this takes $\mathcal{O}(n\Delta \log \Delta)$ time. Then we order the set $X^1 = \{x_i^1 : i \in N\}$ of vectors according to their first coordinates, and order the vectors in $X^2 = \{x_i^2 : i \in N\}$ similarly in $\mathcal{O}(n \log n)$ time. The required running time for the initializations is therefore $\mathcal{O}(n\Delta + n\Delta \log \Delta + n \log n) = \mathcal{O}(n\Delta \log n)$ as $1 \leq \Delta \leq n$.

1006 **The computation of a step in the heuristic.** We check if the largest entry in the vectors in X^1 is
1007 at least 2 by looking at the first coordinate of the first vector in $\mathcal{O}(1)$ time. If yes, we have found
1008 a locally dominating switch $i \rightarrow C$; note that if the coalition achieving the largest entry in X^1 is
1009 a dummy coalition representing all coalitions in $\pi \cup \{\emptyset\} \setminus C^i$ for some agent i , then a coalition
1010 $C' \in \pi \cup \{\emptyset\} \setminus C^i$ can be found in $\mathcal{O}(|C^i|) = \mathcal{O}(\Delta)$ time.² To maintain the vectors x_i^1 and x_i^2 as well
1011 as the sets C^i for all agents $i \in N$, we only need to update these vectors and sets for the $\mathcal{O}(\Delta)$ agents
1012 in $\{i\} \cup E(i) \cup F(i)$ and, regarding the vectors, only at the coordinates corresponding to $\pi(i)$ and
1013 to C ; note that this update may result in adding a new coordinate to the vector or deleting one from it.
1014 Then, in order to maintain property (\star) , we have to reorder X^1 and X^2 again—but we do not have
1015 to perform a complete ordering from scratch; we just have to find the correct places of the updated
1016 elements in the already existing ordering. Hence, this can be implemented in $\mathcal{O}(\Delta(\log \Delta + \log n))$
1017 time, as there are $\mathcal{O}(\Delta)$ elements to update and an insertion of an element to an ordered list of length
1018 ℓ can be done in $\mathcal{O}(\log \ell)$ time.

1019 If the largest entry in the vectors of X^1 is 0 or 1, then we check the largest entry in the vectors of X^2
1020 (recall that x_i^2 stores a nonnegative entry for a coalition C only if condition (c2) holds for i and C)
1021 and check if the first coordinate is larger than 0 or not. If not, then π is locally popular by Lemma 16.
1022 Otherwise, we again have found a locally dominating switch; we then perform this switch and update
1023 the sets C^i and the vectors in X^1 and X^2 in $\mathcal{O}(\Delta \log n)$ time as in the previous case.

1024 Checking whether $i \rightarrow C$ is a Nash deviation for LocStab can also be implemented in the same
1025 running time. We just have to append the condition that i improves to each case in Lemma 16, which
1026 is also a condition that can be checked in $\mathcal{O}(1)$ time. In fact, we can redefine the values $x_i^1(C)$ and
1027 $x_i^2(C)$ to be $-\infty$ if i has no incentive to join C . This way, it remains true that the value of $x_i^1(C)$,
1028 and also of $x_i^2(C)$, coincides for all coalitions where i has neither a friend nor an enemy; thus, our
1029 approach still works.

1030 Summarizing all of the above, we get the running time of $\mathcal{O}(n\Delta \log n + K_{\# \text{steps}} \cdot \Delta \log n)$ where
1031 $K_{\# \text{steps}}$ is the number of steps (i.e., number of switches) performed by the algorithm. Combining this
1032 with Theorem 2 proves Theorem 3. \square

²We can take $C' = \emptyset$ unless we aim for a variant of the algorithm that keeps the number of coalitions constant; in this case, we can iterate over the coalitions of the current partition until we find one that is *not* in C^i in $\mathcal{O}(\Delta)$ time.

1033 Appendix C Additional material for the simulations

1034 We remark that all codes used for our simulations are available in the supplementary material files
 1035 of our submission. In Section C.1, we present our simulation result for community detection in the
 1036 form of some figures; we do the same for clustering in Section C.2. Finally, we present a detailed
 1037 evaluation of our results in Section C.3.

1038 C.1 Community detection: omitted figures

1039 This section contains all figures depicting our simulation results for the Karate club (Figure 4), Jazz
 1040 musicians (Figure 5), Cora (Figure 6), and Random-25 (Figure 7) datasets, comparing our heuristics
 1041 LocPop and LocStab with the Louvain and Leiden algorithms in terms of Rand index and modularity.
 1042 As the Jazz dataset has no true labels, we did not compute the Rand index for it.

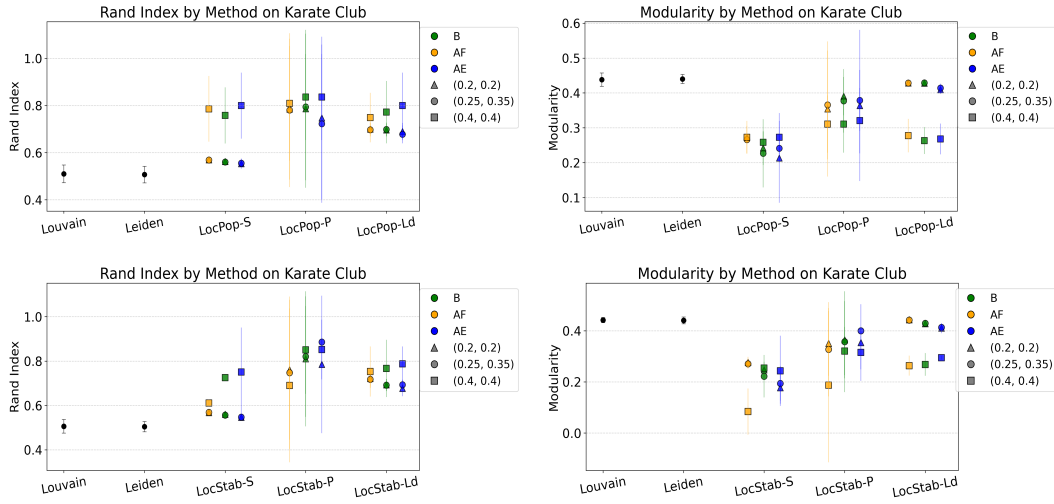


Figure 4: Comparing Rand index (left) and modularity (right) of Louvain, Leiden, and LocPop (top) or LocStab (bottom) variants for the Karate club dataset.

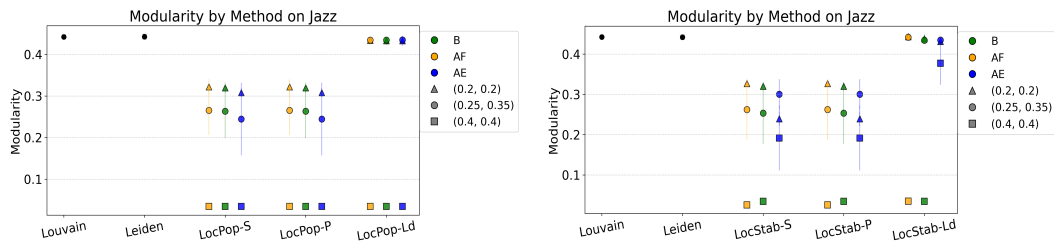


Figure 5: Comparing modularity of Louvain, Leiden, and LocPop (left) or LocStab (right) variants for the Jazz dataset.

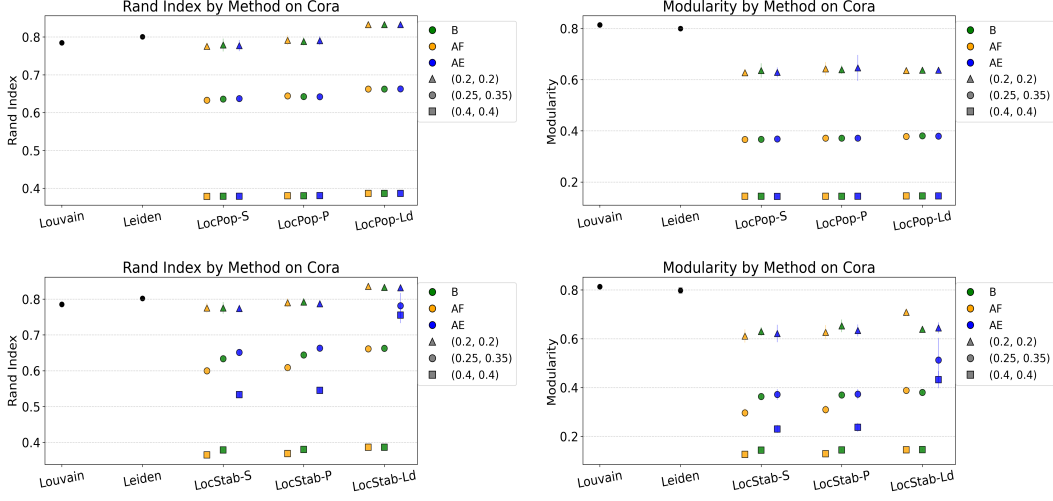


Figure 6: Comparing Rand index (left) and modularity (right) of Louvain, Leiden, and LocPop (top) or LocStab (bottom) variants for the Cora dataset.

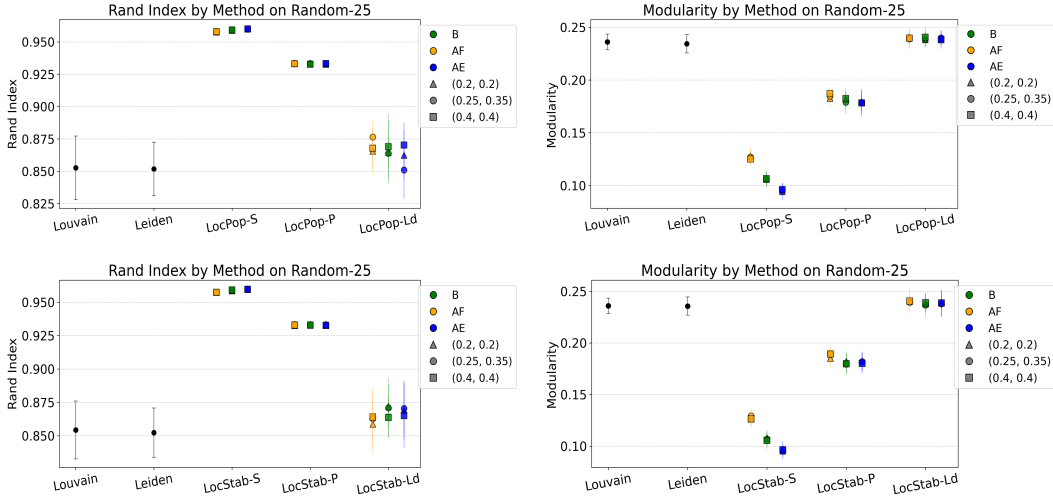


Figure 7: Comparing Rand index (left) and modularity (right) of Louvain, Leiden, and LocPop (top) or LocStab (bottom) variants for the Random-25 dataset.

1043 C.2 Clustering: omitted figures

1044 This section contains all figures depicting our simulation results on the Iris (Figure 8), Breast Cancer
 1045 Wisconsin (abbreviated as *Cancer*, Figure 9), Moons (Figure 10), and 3-Circles (Figure 11) datasets,
 1046 comparing our heuristics LocPop and LocStab with the k -means and DBSCAN algorithms in terms
 1047 of Rand index and silhouette score.

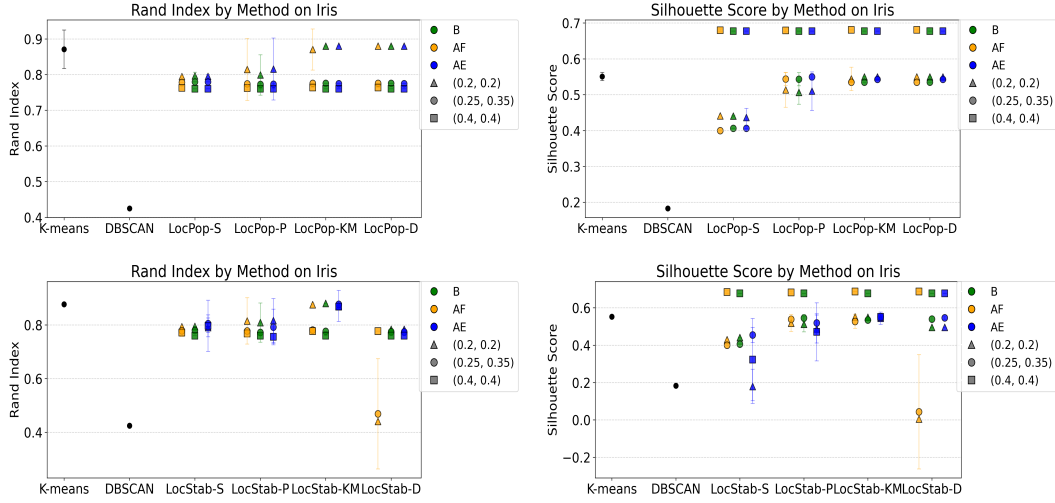


Figure 8: Comparing Rand index (left) and silhouette score (right) of k -means, DBSCAN and LocPop (top) or LocStab (bottom) variants for the Iris dataset.

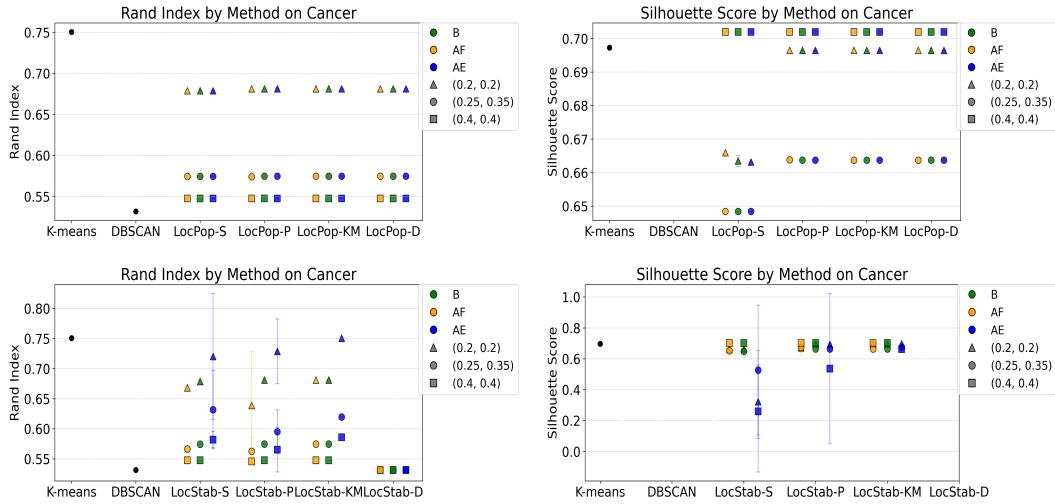


Figure 9: Comparing Rand index (left) and silhouette score (right) of k -means, DBSCAN and LocPop (top) or LocStab (bottom) variants for the Cancer dataset.

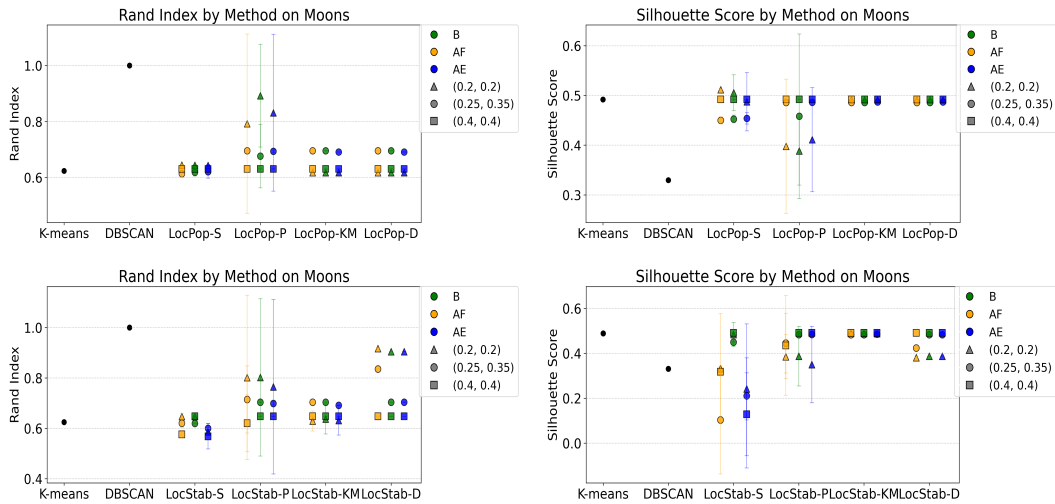


Figure 10: Comparing Rand index (left) and silhouette score (right) of k -means, DBSCAN and LocPop (top) or LocStab (bottom) variants for the Moons dataset.

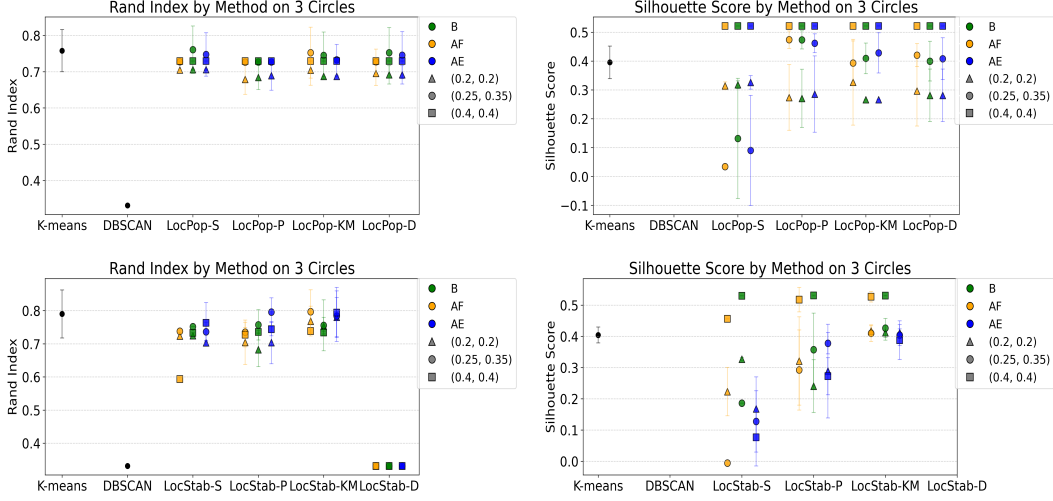


Figure 11: Comparing Rand index (left) and silhouette score (right) of k -means, DBSCAN and LocPop (top) or LocStab (bottom) variants for the 3-Circles dataset.

C.3 Evaluation of results

General takeaways for LocPop and LocStab. LocPop and LocStab produced very similar results, although the scores of LocStab were more dependent on the preference domains. The best threshold parameters were also mostly the same for LocPop and LocStab. In LocPop variants, the three different preferences domains (AF, AE, and B) lead to quite similar results across all instances. This is probably due to the fact that only the agent who switches may vote differently in these cases, and a single vote usually has no substantial influence. In the case of local stability, where a Nash deviation is needed, a larger differentiation was observed. The simulations also showed that—while no preference domain was consistently performing better than the two others—the *balanced* one performed best or close to the best for both LocPop and LocStab. Hence, in light of Theorems 3 and 4, using balanced preferences seems most preferable, as we have a faster runtime guarantee, and it is the only preference domain where the LocStab heuristic converges in polynomial time for asymmetric graphs. Hence, we suggest that future research focuses on the balanced preference domain.

In general, the -P and -S variants both performed well. However, the -S variant often requires much more steps to reach an outcome, which is expected, as it starts with n clusters. Hence, if there is a good estimate (or even an upper bound) on the expected number of clusters, then the -P variant may be the optimal. The LocPop-P and LocStab-P variants we implemented also had a constraint that new coalitions could only be created if the current number of clusters was smaller than their predicted number.

In most cases, the (0.4, 0.4) or the (0.2, 0.2) parameters lead to the best results, but these parameters also often lead to the worst results; while the parameter (0.25, 0.35) performed consistently between them. This suggests that parameter tuning can have a large effect on achieving the best possible outcome.

Since LocPop and LocStab are not deterministic (randomness comes from (i) the initial clustering, (ii) the indexing of the data points, and (iii) the choice of the implemented switches), we did 10 repetitions, each time randomly permuting the data points and with a random initial partition with the predicted number of clusters for LocPop-P and LocStab-P (for the other variants the initial partition is fixed) to estimate the mean and standard deviation of the scores. The algorithm was implemented in a way such that switches involving smaller-indexed agents were preferred, but by randomly permuting the indices, this bias was eliminated. The standard deviations observed were at most 0.01-0.02 on average after doing 10 repetitions, except for the Karate club dataset (~ 0.05). Hence, for that dataset, we increased the number of repetitions to 30 for more robust results. We also note that in many cases, the standard deviations were too small to be visible in the figures (e.g., $< 10^{-3}$).

Comparison with mainstream algorithms. Let us compare the performance of our algorithms with the most widely used mainstream methods, both for community detection and for clustering.

1083 **Community detection.** Using the output of Leiden, variants of LocPop-Ld (and LocStab-Ld)
1084 achieved a similar value in modularity, and slightly outperformed it w.r.t. the Rand index. Hence,
1085 starting from the output of Leiden was very beneficial in terms of achieving better modularity scores.
1086 For the Random-25 dataset, LocPop-S and LocPop-S were able to significantly outperform Louvain
1087 and Leiden (by more than 10%) w.r.t. the Rand index, but LocPop-P and LocPop-S were close to
1088 a 10% improvement too. For the Karate club dataset, LocPop and LocStab variants were able to
1089 outperform Louvain and Leiden by more than 50% in Rand index.

1090 **Clustering.** In clustering, with the parameter $(0.4, 0.4)$, LocPop performed up to 20% better than
1091 k -means w.r.t. the silhouette score in the Iris and 3 Circles datasets (and performed similarly for
1092 the others), while the other parameters led to (sometimes significantly) worse results. DBSCAN
1093 performed the worst in terms of the silhouette score, and it often put all points together, in which
1094 case the silhouette score is undefined. Regarding the Rand index, $(0.4, 0.4)$ turned out to be the
1095 worst parameter in general, which highlights that silhouette score may often not be the best measure
1096 to assess the quality of a clustering with regards to label prediction. This was most apparent in the
1097 Cancer dataset.

1098 With respect to the Rand index, the best performing variants of LocPop and LocStab, which were
1099 the ones with the parameter $(0.2, 0.2)$ (except in the 3 Circles case) were within 10% of k -means,
1100 often matching it, LocPop-P and LocStab-D outperformed it by roughly 50% for the Moons dataset
1101 (LocStab-P being close behind them, but all variants performed better than k -means), where DBSCAN
1102 was able to find the true clusters completely. As opposed to community detection, for LocPop, here
1103 we found no clear benefit of using the output of DBSCAN or k -means to start with, while starting
1104 with a correctly predicted number of random clusters helped significantly for the Moons dataset. With
1105 LocStab-D, starting with these outputs did lead to better results for the Moons dataset.

1106 Summarizing our observations, we conclude that the variants of LocPop and LocStab performed well
1107 and consistently, often leading to the best results in both community detection and clustering.

1108 **Other observations, runtimes and outlook.** For heuristics LocPop and LocStab started with the
1109 outputs of Leiden and k -means, we tracked the Rand index between their initial partition (i.e., the
1110 output of Leiden or k -means, respectively) and the final partition. We observed that even in these
1111 cases, both algorithms often made quite a lot of changes. For LocPop, we observed an average Rand
1112 index—here used to measure the difference between the initial and the final partition—of 0.65 with a
1113 minimum of 0.33 across community detection datasets; for clustering datasets, these values reached
1114 an average of 0.55 with a minimum of 0.34. For LocStab, the obtained numbers were slightly higher,
1115 as expected. For community detection datasets, the average Rand index was 0.67 with a minimum of
1116 0.33 and for clustering datasets, an average of 0.65 with a minimum of 0.43.

1117 We also tracked the running times. Our LocPop algorithms run within a couple of seconds even on
1118 a mid-range hardware, except for Cora, the largest dataset, where computation took 3-5 minutes,
1119 LocPop-S and LocStab-S being the slowest as expected. For these variants, to speed up computation,
1120 we used random clusters of size 6, instead of singleton clusters, still modeling the same behavior.
1121 We emphasize that our implementation is not the most efficient one discussed in Section 3, but a
1122 simpler approach that iterates through the agents and the current clusters in each step. Furthermore,
1123 our approach to constructing the friendship and enmity graphs did not guarantee bounded degrees,
1124 causing a limitation, since then the running time in Theorem 3 are not guaranteed to be linear. For
1125 instances of even larger size, constructions that guarantee bounded degrees—e.g., taking the d nearest
1126 neighbors as friends and the d farthest ones as enemies for each agent—may lead to more efficient
1127 computation. We leave this for future work.

1128 There are many more ways to generate friendship and enmity relations among agents, and it might
1129 be interesting to see whether there are approaches that lead to improved results. Further, another
1130 limitation is that LocPop might have a problem with instances containing clusters of different sizes,
1131 since large clusters can get rid of unwanted agents simply due to their voting power. It will be
1132 interesting to test this in future work. Finally, note that LocStab might be a fix for those instances,
1133 since agents cannot be pushed out of a cluster “against their will” (and maybe that is the reason why
1134 it performs better on the Cancer dataset).

1135 Appendix D Omitted proofs

1136 This section contains all proofs omitted from the main body of our paper, except for the proof of
 1137 Theorem 3 which we proved in Appendix B. Sections D.1, D.2, and D.3 contain the omitted proofs
 1138 of our results on FEN-AF, FEN-B, and FEN-AE games, respectively.

1139 As already mentioned in Example 1, throughout this paper, all relations among agents that are not
 1140 explicitly mentioned are tacitly assumed to be neutral.

1141 D.1 Omitted proofs for our results on FEN-AF games

1142 Before we restate and prove our intractability results concerning FEN-AF games, we state some
 1143 structural observations concerning AF-gadgets in Section D.1.1. We then proceed to provide all
 1144 omitted proofs of our results on FEN-AF games in Sections D.1.2–D.1.4.

1145 D.1.1 A structural lemma on AF-gadgets

1146 Given an AF-gadget G , let us define the *distance* of agents i and j within G as the minimum number
 1147 of arcs on a path from i to j or from j to i within the friendship graph of G . Moreover, whenever
 1148 an AF-gadget is defined over agent set $K = \{0, 1, \dots, 2k\}$, the agent $i + c$ is always interpreted as
 1149 $(i + c) \bmod (2k + 1)$ for each agent $i \in K$ and constant integer c .

1150 **Lemma 17.** *Consider a FEN-AF game that admits a locally stable partition π and contains an*
 1151 *AF-gadget G over agent set K . If some coalition $C \in \pi$ with $C \subseteq K$ contains two agents whose*
 1152 *distance in G is at least 4, then $C = K$.*

1153 *Proof.* Note that any two agents whose distance in G is at least 4 are mutual enemies. Hence, if two
 1154 such agents i and j belong to some coalition $C \subseteq K$, then each of them must have a friend in C :
 1155 indeed, assuming that one of them, say agent i , does not have a friend in C , we get that the switch
 1156 $i \rightarrow \emptyset$ locally blocks π , because i and i 's enemies in C vote for it, while only $i - 1$ may vote against
 1157 it. The distance between the friend of i and the friend of j (i.e., agents $i + 1$ and $j + 1$) within G
 1158 is again at least 4, so by the same reasoning, each of them must have a friend within C . Repeating
 1159 this argument, we get that every agent of C must have a friend within C , which implies $K \subseteq C$. By
 1160 $C \subseteq K$ we get $C = K$, as required. \square

1161 D.1.2 Proof of Theorem 5

1162 **Theorem 5 (\star).** *The asymmetric FEN-AF game consisting solely of an AF-gadget admits no locally*
 1163 *popular partition.*

1164 *Proof.* Consider an AF-gadget G of size $2k + 1$ containing agents $0, 1, \dots, 2k$. Suppose for the sake
 1165 of contradiction that π is a locally popular partition.

1166 If π is the grand coalition, then for an arbitrary agent i , the switch $i \rightarrow \emptyset$ locally dominates π , because
 1167 (recalling that $k \geq 5$) exactly $2k - 4 > 3$ agents vote for it (those who consider i an enemy) and
 1168 only the two agents i and $i - 1$ vote against it (as they both lose a friend). This contradicts the local
 1169 popularity of π .

1170 Thus π is not the grand coalition. By Lemma 17, this implies that for any two agents i and j
 1171 with $\pi(i) = \pi(j)$, the distance of i and j in G is at most 3, so each coalition in π is a subset
 1172 of $\{i, i + 1, i + 2, i + 3\}$ for some agent i .

1173 Suppose now that some agents i and $i + 2$ are both within some coalition $C = \pi(i) = \pi(i + 2)$, but
 1174 $i + 1$ is not in C . Then $i + 1$ has an incentive to join C , that is, $i + 1$ votes for the switch $i + 1 \rightarrow C$,
 1175 and so does i . Since the distance between any two agents of C is at most 3 in G , there can be at most
 1176 one agent in C who considers $i + 1$ an enemy, namely agent $i - 1$. Thus at most one agent in C
 1177 votes against the switch $i + 1 \rightarrow C$, and note also that no agent in $\pi(i + 1)$ votes against it. Hence,
 1178 $i + 1 \rightarrow C$ locally blocks π , a contradiction. This shows that each coalition in π must either contain
 1179 at most four agents who, additionally, appear consecutively on the friendship cycle of G , or it must
 1180 be of the form $\{i, i + 3\}$ for some i .

1181 If there is a coalition of the form $\{i, i + 3\}$, then the switch $i + 3 \rightarrow \pi(i + 4)$ locally blocks π , as
 1182 both i as $i + 3$ vote for it and the only agent who may vote against it is $i + 7$.

1183 If there is a coalition containing four agents $\{i, i + 1, i + 2, i + 3\}$, then again, the switch $i + 3 \rightarrow \pi(i + 4)$
 1184 locally blocks π : the only agents who may vote against it are $i + 2$ and $i + 7$, whereas agents $i, i + 1$,
 1185 and $i + 3$ all vote for it.

1186 Therefore, every coalition must contain at most three agents who appear consecutively on the
 1187 friendship cycle of G . Suppose there is a coalition with exactly three agents. Then, as 3 does not
 1188 divide $2k + 1$, there is one coalition, say $\{i, i + 1, i + 2\}$, which is followed (along the friendship cycle
 1189 of G) by a coalition of size two, namely coalition $\{i + 3, i + 4\}$. Then the switch $i + 2 \rightarrow \pi(i + 3)$
 1190 locally blocks π , as agents i and $i + 2$ vote for it, and only agent $i + 1$ votes against it.

1191 Hence, each coalition contains at most two agents. If there is one containing exactly two agents, then
 1192 there is one coalition, say $\{i, i + 1\}$, which is preceded (on the friendship cycle of G) by a coalition
 1193 of size one, namely $\{i - 1\}$, as $2k + 1$ is odd. Hence, the switch $i - 1 \rightarrow \{i, i + 1\}$ locally blocks π .
 1194 It follows that each coalition in π must be a singleton. But then the switch where any agent i joins
 1195 $\{i + 1\}$ locally blocks π , again a contradiction. It follows that no locally popular partition exists for
 1196 this hedonic game. \square

1197 Note that for the hedonic game defined in the proof of Theorem 5, there does exist a unique locally
 1198 stable partition: the grand coalition containing all agents. In fact, the proof of Theorem 5 even implies
 1199 the following slightly stronger claim.

1200 **Remark 1.** Suppose that a hedonic game with friend-appreciative preferences contains an AF-gadget
 1201 over agent set K , and admits a locally stable partition π . If no coalition of π contains agents both in
 1202 and outside of K , then K itself is a coalition in π .

1203 D.1.3 Proof of Theorem 6

1204 **Theorem 6** (*). *There exists an asymmetric FEN-AF game that admits no locally stable partition.*

1205 *Proof.* As a counterexample, consider a FEN-AF game I that contains an AF-gadget over a set
 1206 $K = \{0, 1, \dots, 2k\}$ of agents together with an additional set $A = \{a_1, \dots, a_{2k-4}\}$ of agents such
 1207 that

- 1208 • a_1, \dots, a_{2k-4} are mutual enemies with $1, 2, \dots, 2k + 1$,
- 1209 • a_2, \dots, a_{2k-4} consider 0 an enemy,
- 1210 • 0 considers a_1 a friend and a_2, \dots, a_{2k-4} enemies,
- 1211 • a_1 and a_2 are mutual friends, and
- 1212 • a_i considers a_{i+1} a friend for $i = 2, 3, \dots, 2k - 5$.

1213 Suppose that π is a locally stable partition for I . First, we claim that if a coalition $C \in \pi$ contains
 1214 both an agent from K and an agent from A , then C contains 0 as well. To see this, consider the
 1215 “largest” agent $j = \max_{i:i \in K \cap C} i$ of K in C . If $0 \notin C$, then the switch $j \rightarrow \emptyset$ locally blocks π ,
 1216 because agent j and all agents in $A \cap C \neq \emptyset$ vote for it, while at most one agent (namely, agent $j - 1$)
 1217 may vote against it.

1218 Next, we claim that if some coalition $C \in \pi$ contains both 0 and a_i for some i , then C also
 1219 contains a_{i+1} if it exists. For $i = 1$, this follows from the fact that, otherwise, a_1 has an incentive
 1220 to join $\pi(a_2)$, as both a_1 and a_2 vote for the switch $a_1 \rightarrow \pi(a_2)$ and only agent 0 votes against it
 1221 (note that only agents in K consider a_1 an enemy, and we have $K \cap \pi(a_2) = \emptyset$ by the previous
 1222 paragraph), contradicting our assumption that π is locally stable. For $i > 1$, we can similarly show
 1223 that if $a_{i+1} \notin C$, then the switch $a_i \rightarrow \pi(a_{i+1})$ locally blocks π : both a_i and 0 vote for it, and only
 1224 agent a_{i-1} may vote against it.

1225 As a consequence, no coalition C in π contains both an agent from K and an agent from A . Indeed,
 1226 assuming otherwise, the coalition C would contain agents 0 and a_{2k-4} by the previous two claims,

1227 in which case the switch $a_{2k-4} \rightarrow \emptyset$ locally blocks π : both a_{2k-4} and 0 vote for it, while at most one
 1228 agent (namely, agent a_{2k-3}) may vote against it.

1229 Due to Remark 1, it follows that the agents of K must form a single partition in π . In particular,
 1230 agent 0 has one friend and $2k - 4$ enemies within $\pi(0)$. Also, agent 0 has at least one friend and at
 1231 most $2k - 5$ enemies in the coalition $\pi(a_1)$, so 0 has an incentive to join $\pi(a_1)$. Finally, the switch
 1232 $0 \rightarrow \pi(a_1)$ locally blocks π , because $2k - 4 + 1$ agents from K vote for it (namely, agent 0 and all
 1233 those regarding 0 an enemy), while only one agent from K (namely, agent $2k$) and $2k - 5$ agents
 1234 from A (all of them except a_1) vote against it. This contradiction shows that π cannot be locally
 1235 stable. \square

1236 D.1.4 Proof of Theorem 7

1237 **Theorem 7** (*). LOCAL-POPULARITY-EXISTENCE and LOCAL-STABILITY-EXISTENCE are NP-
 1238 complete for FEN-AF games.

1239 It is clear that both problems are in NP, as given a FEN-AF game, we can nondeterministically guess
 1240 a partition π and then verify in polynomial time that no possible switch locally dominates or locally
 1241 blocks π (note that the number of switches for π is at most quadratic in the number of agents). We
 1242 complete the proof of Theorem 7 by proving the NP-hardness of LOCAL-STABILITY-EXISTENCE
 1243 and LOCAL-POPULARITY-EXISTENCE separately in Theorems 18 and 24, starting with the former.

1244 **Theorem 18.** LOCAL-STABILITY-EXISTENCE is NP-hard for FEN-AF games.

1245 *Proof.* We provide a reduction from 5-SAT which, given a Boolean formula φ in conjunctive normal
 1246 form with five literals per clause, asks whether φ is satisfiable. Let φ be our input 5-SAT formula
 1247 with variables v_1, \dots, v_n and clauses C_1, \dots, C_m . By adding dummy clauses if needed, we may
 1248 assume that $m \equiv 1 \pmod 3$ and that $m \equiv 1 \pmod 2$.

1249 We create an instance I of LOCAL-STABILITY-EXISTENCE as follows. Let $k = 5 \cdot 11m = 55m$;
 1250 then $k \equiv 1 \pmod 2$ and $k \equiv 1 \pmod 3$. For each clause C_j , we create a *clause gadget* G_j that is an
 1251 AF-gadget of size k . The agents of G_j , ordered along the friendship cycle of G_j , are $x_j^0, \dots, x_j^{55m-1}$.
 1252 Every $11m$ -th agent $x_j^0, x_j^{11m}, \dots, x_j^{44m}$ along this cycle corresponds to one of the literals in $C_j =$
 1253 $(l_{j_1} \vee \dots \vee l_{j_5})$; these are called *literal agents*. We also add $5m$ *auxiliary gadgets* A_1, \dots, A_{5m} , each
 1254 of them an AF-gadget of size $k = 55m$. Each auxiliary gadget A_j has a single distinguished agent
 1255 a_j^0 ; we refer to such agents as *enforcer agents*. We also add an agent T . Intuitively, in a locally stable
 1256 partition, the coalition containing T will also include a literal agent from each clause corresponding
 1257 to a true literal. We proceed by describing the relationships between the agents on different cycles:

- 1258 • Every literal agent considers T a friend and is considered by T an enemy.
- 1259 • Every enforcer agent considers T a friend but is considered by T as a neutral.
- 1260 • T and every non-literal, non-enforcer agent consider each other mutual enemies.
- 1261 • Any two agents y and z from two different clause gadgets are mutual enemies, unless y
 1262 and z are both literal agents that do *not* correspond to a positive and a negative literal of the
 1263 same variable, in which case they are mutually neutral.
- 1264 • Any two agents y and z from two different auxiliary gadgets are mutual enemies, unless
 1265 both are enforcer agents, in which case they are mutually neutral.
- 1266 • Finally, for a vertex x from a clause gadget and a vertex y from an auxiliary gadget,
 1267 – if y is not a literal agent or z is not an enforcer agent, then they are mutual enemies;
 1268 – if y is the literal agent corresponding to the ℓ -th literal of the j -th clause, then y and z
 1269 are mutual enemies if $z = a_{5(j-1)+\ell}^0$ and they are mutually neutral if z is an enforcer
 1270 agent other than $a_{5(j-1)+\ell}^0$.

1271 This completes our construction of the instance I of LOCAL-STABILITY-EXISTENCE from the
 1272 given 5-SAT instance φ . It remains to prove the correctness of the reduction: I is a yes-instance

1273 of LOCAL-STABILITY-EXISTENCE if and only if φ is satisfiable. This will follow from a series of
1274 claims.

1275 **Direction “ \Rightarrow ”:** Let us first assume that there exists a locally stable partition π for I .

1276 **Claim 19.** *Coalition $\pi(T)$ contains at most $20m - 1$ agents. Consequently, $\pi(T)$ cannot contain all*
1277 *agents from a clause or an auxiliary gadget.*

1278 *Proof of claim.* For a contradiction, suppose that more than $20m - 1$ agents are in the same coalition
1279 as T . Then at least $10m$ of them are mutual enemies of T , as only $10m$ agents consider T a friend,
1280 and no agent considers T as neutral. Recall also that T considers no agent as a friend and thus has an
1281 incentive to leave $\pi(T)$. Hence, the switch $T \rightarrow \emptyset$ locally blocks π , a contradiction. \triangleleft

1282 **Claim 20.** *No coalition $C \in \pi$ can contain all agents from some clause or auxiliary gadget.*

1283 *Proof of claim.* For the sake of contradiction, assume that some coalition C contains some clause
1284 or auxiliary gadget K . Claim 19 implies $C \neq \pi(T)$. Let y be a literal or enforcer agent in K .
1285 Since y has one friend and exactly $55m - 5$ enemies within K , y has an incentive to join $\pi(T)$ by
1286 Claim 19. Thus the switch $y \rightarrow \pi(T)$ locally blocks π : at least $(55m - 5) + 1$ agents—all enemies
1287 of y within K and y itself—vote for it, whereas at most $(20m - 1) + 1$ agents—enemies of y in $\pi(T)$
1288 and the unique agent in K who considers y a friend—vote against it, a contradiction. \triangleleft

1289 **Claim 21.** *If a coalition $C \in \pi$ with $T \notin C$ contains agents from two different clause or auxiliary*
1290 *gadgets, then C cannot contain a non-literal, non-enforcer agent.*

1291 *Proof of claim.* Suppose for the sake of contradiction that $C \in \pi$ with $T \notin C$ contains some agent y
1292 from a (clause or auxiliary) gadget K , some agent y' from a different gadget K' , and a non-literal,
1293 non-enforcer agent w . Note that neither K nor K' is fully contained in C due to Claim 20. Therefore,
1294 using also $T \notin C$, we get that C contains an agent $z \in K$ and an agent $z' \in K'$ such that neither z
1295 nor z' considers any agent in C a friend. Moreover, w is a mutual enemy with some $z'' \in \{z, z'\}$.
1296 This implies that the switch $z'' \rightarrow \emptyset$ locally blocks π , because both z'' and w vote for it, while at
1297 most one agent (who considers z'' a friend) may vote against it, a contradiction. \triangleleft

1298 **Claim 22.** *Coalition $\pi(T)$ contains at least one agent from each clause gadget and each auxiliary*
1299 *gadget.*

1300 *Proof of claim.* For a contradiction, suppose that $\pi(T)$ contains no agent from some clause or
1301 auxiliary gadget K . If no agent from K is contained in the same coalition as some agent from a
1302 different gadget, then Remark 1 implies $K \in \pi$, a contradiction to Claim 20. Hence, there is a
1303 coalition C that contains some agent y from K and some agent from a different gadget. By Claim 21,
1304 we get that y is either a literal or an enforcer agent. Suppose that y is a literal agent $y = x_j^i$ for
1305 some clause gadget $K = G_j$; the case when K is an auxiliary gadget and y is an enforcer agent is
1306 analogous.

1307 Note that x_j^{i+1} is the only agent x_j^i considers a friend, while the only agent who considers x_j^i a
1308 friend is G_j^{i-1} . Since both G_j^{i-1} and x_j^{i+1} are non-literal, non-enforcer agents, neither of them is
1309 contained in C due to Claim 21. In particular, x_j^i has an incentive to join $\pi(x_j^{i+1})$. Since the switch
1310 $x_j^i \rightarrow \pi(x_j^{i+1})$ cannot locally block π , we obtain that there must be some agent in $\pi(x_j^{i+1})$ who
1311 considers x_j^i an enemy. Observe that $T \notin \pi(x_j^{i+1})$ by our choice of $K = G_j$, so by Claim 21 we know
1312 that all agents of $\pi(x_j^{i+1})$ belong to G_j . Using Lemma 17 and Claim 20, we get that $\pi(x_j^{i+1})$ cannot
1313 contain two agents whose distance along G_j is more than 3; we refer to this fact as (\spadesuit) . This implies
1314 $i' \in \{i-2, i-1, i+1, i+2, i+3, i+4\}$ for every agent $x_j^{i'} \in \pi(x_j^{i+1})$, so the agent within $\pi(x_j^{i+1})$
1315 who considers x_j^i an enemy must be x_j^{i-2} or x_j^{i+4} . Moreover, $\{x_j^{i-2}, x_j^{i+4}\} \not\subseteq \pi(x_j^{i+1})$ also follows.
1316 We distinguish between two cases and arrive at a contradiction in both, which proves the claim.

1317 **Case A:** $x_j^{i-2} \in \pi(x_j^{i+1})$. Then $\pi(x_j^{i+1}) \subseteq \{x_j^{i-2}, x_j^{i-1}, x_j^{i+1}\}$ due to (\spadesuit) . If $x_j^{i-1} \notin \pi(x_j^{i+1})$, then
1318 x_j^{i+1} has an incentive to join $\pi(x_j^{i+2})$. Since the switch $x_j^{i+1} \rightarrow \pi(x_j^{i+2})$ does not locally block π
1319 but both x_j^{i+1} and x_j^{i-2} vote for it, it follows that at least two agents must consider x_j^{i+1} an enemy
1320 in $\pi(x_j^{i+2})$. However, due to (\spadesuit) , these agents could only be x_j^{i-1} and x_j^{i+5} , but they cannot both be

1321 in $\pi(x_j^{i+2})$, a contradiction. Thus it must be the case that $\pi(x_j^{i+1}) = \{x_j^{i-2}, x_j^{i-1}, x_j^{i+1}\}$. Then x_j^i
 1322 has an incentive to join $\pi(x_j^{i+1})$; recall that $T \notin \pi(x_j^i)$, so x_j^i has no friend in $\pi(x_j^i)$. Observe that
 1323 the switch $x_j^i \rightarrow \pi(x_j^{i+1})$ locally blocks π : both x_j^i and x_j^{i-1} vote for it, but only x_j^{i-2} votes against
 1324 it, a contradiction.

1325 **Case B: $x_j^{i+4} \in \pi(x_j^{i+1})$.** Then $\pi(x_j^{i+1}) \subseteq \{x_j^{i+1}, x_j^{i+2}, x_j^{i+3}, x_j^{i+4}\}$. Then x_j^{i+4} has an incentive to
 1326 join $\pi(x_j^{i+5})$, so both x_j^{i+1} and x_j^{i+4} vote for the switch $x_j^{i+4} \rightarrow \pi(x_j^{i+5})$. Since π is locally stable,
 1327 there must be two agents who vote against this switch. If $x_j^{i+3} \notin \pi(x_j^{i+1})$, then these must be two
 1328 agents in $\pi(x_j^{i+5})$ who consider x_j^{i+4} an enemy. Due to (), these can only be agents x_j^{i+2} and
 1329 x_j^{i+8} , but at most one of them can be in $\pi(x_j^{i+5})$, a contradiction. Hence, $x_j^{i+3} \in \pi(x_j^{i+1})$. Since
 1330 the switch $x_j^{i+2} \rightarrow \pi(x_j^{i+1})$ does not locally block π , it follows that $x_j^{i+2} \in \pi(x_j^{i+1})$ as well, which
 1331 yields $\{x_j^{i+1}, x_j^{i+2}, x_j^{i+3}, x_j^{i+4}\} \in \pi$. Then the switch $x_j^{i+4} \rightarrow \pi(x_j^{i+5})$ locally blocks π , because
 1332 three agents (namely, x_j^{i+1} , x_j^{i+2} , and x_j^{i+4}) vote for it and at most two may vote against it (x_j^{i+3} and
 1333 possibly x_j^{i+8}), a contradiction. \triangleleft

1334 **Claim 23.** *Coalition $\pi(T)$ contains no non-literal, non-enforcer agents. Also, each literal agent*
 1335 *in $\pi(T)$ can have at most one mutual enemy in $\pi(T)$, which is necessarily an enforcer agent.*

1336 *Proof of claim.* For the sake of contradiction, assume that $y \neq T$ is a non-literal, non-enforcer agent
 1337 in $\pi(T)$. Take an agent $z \in \pi(T)$ from a clause G_j not containing y such that the unique friend of z
 1338 within K is not in $\pi(T)$. By Claims 22 and 19, there must be such an agent z . Note that y and z are
 1339 mutual enemies. If z is not a literal agent, then z has no friend in $\pi(T)$, so the switch $z \rightarrow \emptyset$ locally
 1340 blocks π , a contradiction. If $z = x_j^i$ is a literal agent, then z is mutual enemies with every agent from
 1341 at least one auxiliary gadget by construction, so z has at least two mutual enemies in $\pi(T)$.

1342 It now suffices to show that the assumption that z has at least two mutual enemies in $\pi(T)$ leads to a
 1343 contradiction.

1344 Assume first that x_j^i has an incentive to join $\pi(x_j^{i+1})$. Then x_j^i , T , and the two mutual enemies of x_j^i
 1345 in $\pi(T)$ vote for the switch $x_j^i \rightarrow \pi(x_j^{i+1})$ and at most one agent in $\pi(T)$ votes against it. Moreover,
 1346 due to Lemma 17 and Claim 20, each two agents in $\pi(x_j^{i+1})$ are at distance at most 3 from each other
 1347 along G_j , and thus the only agents in $\pi(x_j^{i+1})$ who consider x_j^i an enemy can be x_j^{i-2} and x_j^{i+4} , so
 1348 fewer than four agents vote against the switch, a contradiction.

1349 It follows that x_j^i has no incentive to join $\pi(x_j^{i+1})$, which happens only if x_j^i regards at least
 1350 two agents in $\pi(x_j^{i+1})$ as enemies. This implies that $\pi(x_j^{i+1})$ contains at least two agents from
 1351 $\{x_j^{i+2}, x_j^{i+3}, x_j^{i+4}\}$. As π is locally stable, it follows that $\{x_j^{i+2}, x_j^{i+3}\} \subseteq \pi(x_j^{i+1})$. Suppose
 1352 that $x_j^{i+4} \in \pi(x_j^{i+1})$ as well. Then the switch $x_j^{i+4} \rightarrow \pi(x_j^{i+5})$ locally blocks π , because
 1353 x_j^{i+4} , x_j^{i+2} , and x_j^{i+1} all vote for it, whereas only x_j^{i+3} and x_j^{i+8} may vote against it. Oth-
 1354 erwise, $\pi(x_j^{i+1}) = \{x_j^{i+1}, x_j^{i+2}, x_j^{i+3}\}$. The switch $x_j^{i+3} \rightarrow \pi(x_j^{i+4})$ locally blocks π unless
 1355 $\pi(x_j^{i+4}) = \{x_j^{i+5}, x_j^{i+6}, x_j^{i+7}, x_j^{i+8}\}$. In this case, however, the switch $x_j^{i+8} \rightarrow \pi(x_j^{i+9})$ locally
 1356 blocks π by the argument used in the previous case. \triangleleft

1357 We are now ready to prove that φ is satisfiable.

1358 By Claim 22, at least one literal agent from each clause gadget G_j must be contained in $\pi(T)$. We
 1359 claim that the corresponding literals can be set to true at the same time. For a contradiction, suppose
 1360 that there are two literal agents, y and z , in $\pi(T)$ such that one of them corresponds to literal v_i and
 1361 the other one to its negation \bar{v}_i . By Claim 23, $a_j^0 \in \pi(T)$ for any auxiliary gadget A_j , which means
 1362 that every literal agent is mutual enemies with at least one enforcer agent in $\pi(T)$. Since y and z
 1363 are also mutual enemies by construction, each of these two literal agents have at least two mutual
 1364 enemies, which is a contradiction to Claim 23. Hence, φ is satisfiable.

1365 **Direction “ \Leftarrow ”:** Let us now assume that φ is satisfiable.

1366 Consider a truth assignment that makes φ true. Create a partition π as follows. For each auxiliary
 1367 gadget A_j , group s_j^0 together with T . Also, from each clause gadget, choose a literal agent x_j^i

corresponding to a true literal and group it together with T , too. Finally, for each clause or auxiliary gadget K , create coalitions of size two from the agents of K not put into $\pi(T)$ such that each of these coalitions contains two agents appearing consecutively on K .

We claim that π is locally stable. Agent T has an incentive to join every coalition C in $\pi \cup \{\emptyset\}$ other than $\pi(T)$. However, at least $6m$ agents vote against such a switch $T \rightarrow C$ and at most one votes for it. Note that a literal or enforcer agent in $\pi(T)$ has one friend in $\pi(T)$ and at most one enemy: Indeed, each enforcer agent is mutual enemies with exactly one literal agent—which may or may not be in $\pi(T)$ —and vice versa; moreover, two literal agents are mutual enemies only if they correspond to a literal v_i and its negation \bar{v}_i , respectively. This, however, cannot happen, as only literal agents corresponding to true literals are in $\pi(T)$. Since each literal or enforcer agent has at most one friend in every coalition $C \neq \pi(T)$, and if they do have a friend in C then they also have an enemy in C , so they have no incentive to join C .

For a non-literal, non-enforcer agent y , let $\{y, y'\}$ denote the coalition $\pi(y)$. On the one hand, if y regards y' as a friend, then y has no incentive to join any coalition in $\pi \cup \{\emptyset\}$. If, on the other hand, y' regards y as a friend, then y has an incentive to join the unique coalition containing its sole friend y'' , but the switch $y \rightarrow \pi(y'')$ does not locally block π , because only y votes for it but y' votes against it.

We conclude that π is locally stable, proving the correctness of our reduction. \square

We now turn to the existence problem for local popularity and its computational complexity.

Theorem 24. LOCAL-POPULARITY-EXISTENCE is NP-hard for FEN-AF games.

Proof. We again provide a reduction from 5-SAT by slightly modifying the reduction presented in the proof of Theorem 18.

Let φ be a 5-SAT formula with variables v_1, \dots, v_n and clauses C_1, \dots, C_m . By adding dummy clauses if needed, we may again assume that $m \equiv 1 \pmod 3$ and $m \equiv 1 \pmod 2$.

From φ we create an instance I of LOCAL-POPULARITY-EXISTENCE as follows. Again, let $k = 5 \cdot 11m = 55m$; then $k \equiv 1 \pmod 2$ and $k \equiv 1 \pmod 3$. For each clause C_j , we create a *clause gadget* G_j that is an AF-gadget of size k . The agents of G_j are again referred to as $x_j^0, \dots, x_j^{55m-1}$, ordered along the friendship cycle of G_j , and every $11m$ -th agent $x_j^0, x_j^{11m}, \dots, x_j^{44m}$ along this cycle corresponds to one of the literals in $C_j = (l_{j_1} \vee \dots \vee l_{j_5})$. These are again called *literal agents*. Note that we do not create the auxiliary gadgets and enforcer agents used in the proof of Theorem 18, but we do add an agent T . As before, a locally popular partition will always group T together with an agent corresponding to a true literal from each clause. We proceed by describing the relationships between agents on different gadgets:

- Every literal agent considers T a friend, whereas T considers everyone an enemy and is considered an enemy by every non-literal agent.
- Furthermore, two agents y and z from two different clause gadgets are mutual enemies, unless y and z are both literal agents that do *not* correspond to a positive and a negative literal of the same variable, in which case they are mutually neutral.

We are going to show that the constructed instance I of LOCAL-POPULARITY-EXISTENCE admits a locally popular partition if and only if φ is satisfiable.

Direction “ \Rightarrow ”: Let us first assume that there exists a locally popular partition π for I .

Claim 25. No coalition $C \in \pi$ may contain all agents from some clause gadget.

Proof of claim. For a contradiction, suppose that a coalition in π contains all agents of a clause gadget G_j . Then the switch $x_j^i \rightarrow \emptyset$ for an arbitrary agent x_j^i in G_j locally dominates π , because at least $55m - 5$ agents vote for such a switch and at most one agent votes against it, a contradiction. \triangleleft

Claim 26. If a coalition $C \in \pi$ with $T \notin C$ contains agents from two different clause gadgets, then C cannot contain a non-literal agent.

Proof of claim. Exactly the same arguments as used in the proof of Claim 21 prove the claim. \triangleleft

1415 **Claim 27.** *Coalition $\pi(T)$ contains at least one literal agent from each clause gadget.*

1416 *Proof of claim.* For a contradiction, suppose that there is a clause gadget G_j that contains no agent
 1417 in $\pi(T)$. If no agent of G_j is in the same coalition as some agent from a different gadget, then π
 1418 cannot be locally popular, because we have shown in Theorem 5 that an AF-gadget does not admit a
 1419 locally popular partition.

1420 By Claim 26, it follows that there must be a literal agent x_j^i in G_j such that $C = \pi(x_j^i)$ contains
 1421 only literal agents, with at least one of them not in G_j . By $T \notin C$, we know that x_j^i has an incentive
 1422 to join $\pi(x_j^{i+1})$. Hence, as no agent considers x_j^i a friend in C , there must be an agent in $\pi(x_j^{i+1})$
 1423 who considers x_j^i an enemy, as otherwise, the switch $x_j^i \rightarrow \pi(x_j^{i+1})$ locally blocks π . Note that
 1424 $T \notin \pi(x_j^{i+1})$ by our choice of G_j , and since x_j^{i+1} is a non-literal agent, we know that $\pi(x_j^{i+1}) \subseteq G_j$
 1425 by Claim 26. Therefore, by Lemma 17 and Claim 25, we know that all agents in $\pi(x_j^{i+1})$ are at
 1426 distance at most 3 from x_j^{i+1} along G_j . Hence, the agent in $\pi(x_j^{i+1})$ who considers x_j^i an enemy
 1427 must be x_j^{i-2} or x_j^{i-1} , and they cannot both be contained in $\pi(x_j^{i+1})$. Distinguishing between the
 1428 cases $x_j^{i-2} \in \pi(x_j^{i+1})$ and $x_j^{i-1} \in \pi(x_j^{i+1})$ and using exactly the same arguments as in Cases A
 1429 and B in the proof of Theorem 18 we obtain a contradiction to the local stability and, hence, to the
 1430 local popularity of π . \triangleleft

1431 We are now ready to prove that φ is satisfiable.

1432 By Claim 27, $\pi(T)$ contains at least one literal agent from each clause gadget. We claim that the
 1433 corresponding literals can be made true at the same time. For a contradiction, suppose that there are
 1434 two literal agents, y and z , corresponding to some variable v_i and its negation \bar{v}_i , respectively, that
 1435 are both in $\pi(T)$. Then y and z are mutual enemies. Hence, the switch $y \rightarrow \emptyset$ locally dominates π ,
 1436 as both T and z vote for it and only y votes against it, a contradiction. Hence, this yields a truth
 1437 assignment satisfying φ .

1438 **Direction “ \Leftarrow ”:** Assume now that there is a truth assignment satisfying φ .

1439 Create a partition π as follows. For each clause gadget G_j , choose a literal agent corresponding to
 1440 a literal set to true and group it with T . Then for each clause gadget K , create coalitions of size
 1441 two from the agents of K not put into $\pi(T)$ such that each of these coalitions contains two agents
 1442 appearing consecutively on K .

1443 We claim that π is locally popular. Let $C \in \pi \cup \{\emptyset\}$. Note that the only agent who may vote for a
 1444 switch $T \rightarrow C$ is T , while at least m agents vote against it. For a literal agent $x_j^i \in \pi(T)$, the only
 1445 agents who vote for a switch $x_j^i \rightarrow C$ can be T and possibly agent x_j^{i-1} , while x_j^i itself votes against
 1446 it, and moreover, if $x_j^{i-1} \in C$ votes for the switch, then $x_j^{i-2} \in C$ votes against it. Hence, $x_j^i \rightarrow C$
 1447 does not locally dominate π . For an agent y in a coalition $\{x_j^i, x_j^{i+1}\} \in \pi$ of size two, only one agent
 1448 may vote for the switch $y \rightarrow C$, and this happens only if either $y = x_j^{i+1}$ and y has a friend in C , in
 1449 which case x_j^i votes against the switch, or if $y = x_j^i$ and $C = \{x_j^{i-2}, x_j^{i-1}\}$, in which case both x_j^{i-2}
 1450 and x_j^i vote against the switch. Hence, no switch locally dominates π , so π is locally popular, which
 1451 completes the proof of correctness for our reduction. \square

1452 D.2 Omitted proofs for our results on FEN-B games

1453 We start with restating and proving Theorem 4, stating the polynomiality of LocStab on FEN-B
 1454 games, in Section D.2.1. We move on with some structural observations concerning B-gadgets in
 1455 Section D.2.2. We then proceed to provide all omitted proofs of our results on FEN-B games in
 1456 Sections D.2.3–D.2.6.

1457 D.2.1 Polynomiality of LocStab

1458 **Theorem 4** (\star). *In a possibly asymmetric FEN-B game, the total balanced utility of all agents*
 1459 *strictly increases in each step of heuristic LocStab; consequently, heuristic LocStab always converges*
 1460 *in $\mathcal{O}(n\Delta)$ steps to a locally stable partition, irrespective of the initial partition.*

1461 *Proof.* Let N denote the set of agents. Consider a step of the heuristic where a partition π is replaced
 1462 by $\pi_{i \rightarrow C}$ for some agent i and coalition $C \in \pi \cup \{\emptyset\}$ is such that $C \cup \{i\} \succ_i \pi(i)$ and $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$.
 1463 As $\Lambda(\pi, \pi_{i \rightarrow C}) > 0$ we have $\Lambda_{-i}(\pi, \pi_{i \rightarrow C}) \geq 0$. So, by statement (a) of Lemma 1,

$$\sum_{j \in N \setminus \{i\}} u_j^B(\pi_{i \rightarrow C}) - u_j^B(\pi) = \Lambda_{-i}(\pi, \pi_{i \rightarrow C}) \geq 0.$$

1464 Finally, $u_i^B(C) - u_i^B(\pi(i)) > 0$ because the switch is performed by heuristic LocStab based on
 1465 the condition that it is a Nash deviation, i.e., i has an incentive to join C (recall that preferences
 1466 are balanced). Hence, $f(\pi_{i \rightarrow C}) - f(\pi) = \sum_{j \in N} u_j^B(\pi_{i \rightarrow C}) - u_j^B(\pi) > 0$, which proves the first
 1467 statement of the theorem. Since the total balanced utility is between $-n\Delta$ and $n\Delta$, the claimed
 1468 bound on the number of steps follows as well. \square

1469 D.2.2 A structural lemma on B-gadgets

1470 Let us say that a coalition C in a B-gadget over agent set $K = \{0, 1, \dots, 6\}$ is *consecutive* if
 1471 $C = \{i, i+1, \dots, i+\ell\}$ for some integer $\ell \in [6]$.

1472 **Lemma 28.** *Let π be a partition for a B-gadget, and let $C \in \pi$ be a coalition with either $|C| \geq 4$, or*
 1473 *$|C| = 3$ and C is not consecutive. Then $i \rightarrow \emptyset$ locally dominates π for some agent $i \in C$.*

1474 *Proof.* Let π be a partition for the B-gadget over agent set $K = \{0, 1, \dots, 6\}$ and let $C \in \pi$.

1475 If C is the grand coalition, then $i \rightarrow \emptyset$ locally dominates π , because $i, i+2, i+3$, and $i+4$ all vote
 1476 for it, while only $i-1$ and $i-2$ vote against it.

1477 If $4 \leq |C| \leq 6$, then let $i \in C$ be an agent such that $i-1 \notin C$ and, if C is not consecutive, then
 1478 $\{i, i+1, i+2\} \not\subseteq C$; observe that such an agent i exists. Consider the switch $i \rightarrow \emptyset$. If $i-2 \notin C$,
 1479 then at least two agents in $\{i+2, i+3, i+4\} \cap C$ vote for this switch, while $i+1$ is indifferent
 1480 between π and $\pi_{i \rightarrow \emptyset}$. Hence, even if i votes against it, we get $\Lambda(\pi, \pi_{i \rightarrow \emptyset}) \geq 1$. If $i-2 \in C$ and C
 1481 is consecutive, then $C = K \setminus \{i-1\}$, in which case agents $i, i+2, i+3$, and $i+4$ all vote for the
 1482 switch $i \rightarrow \emptyset$, and only $i-2$ votes against it. If $i-2 \in C$ but C is not consecutive, then due to our
 1483 choice of i , we know that C does not contain both $i+1$ and $i+2$, and thus $f_C^i = 1$, implying also
 1484 $e_C^i \geq 2$ (because $|C| \geq 4$ and $i-1 \notin C$). Hence, i has an incentive to join \emptyset . Moreover, at least one
 1485 agent in $\{i+2, i+3, i+4\} \cap C$ votes for $i \rightarrow \emptyset$, while only $i-2$ votes against it. Hence, we get
 1486 $\Lambda(\pi, \pi_{i \rightarrow \emptyset}) \geq 1$ again, which means that $i \rightarrow \emptyset$ locally dominates π , as required.

1487 It remains to consider the case when $|C| = 3$ and C is not consecutive. If $\{i, j, j+1\} \subseteq C$ for some
 1488 agent j , then $j \notin \{i-2, i-1, i+1\}$. If $j \in \{i+2, i+3\}$, then both j and $j+1$ regard i as an
 1489 enemy, and thus they both vote for the switch $i \rightarrow \emptyset$. If $j = i+4$, then both i and $i+4$ vote for
 1490 the switch $i \rightarrow \emptyset$. Hence, we get $\Lambda(\pi, \pi_{i \rightarrow \emptyset}) \geq 1$ in both cases. Finally, if $\{j, j+1\} \not\subseteq C$ for any
 1491 agent $j \in K$, then C must be of the form $\{i, i+2, i+4\}$ for some $i \in K$. Then both $i+2$ and $i+4$
 1492 vote for $i \rightarrow \emptyset$, yielding again a switch that locally dominates π . \square

1493 D.2.3 Proof of Theorem 8

1494 **Theorem 8 (\star).** *The asymmetric FEN-B game consisting solely of a B-gadget admits no locally*
 1495 *popular partition.*

1496 *Proof.* Assume for the sake of contradiction that π is a locally popular partition for the B-gadget
 1497 over agent set $K = \{0, 1, \dots, 6\}$. By Lemma 28 we know that π contains no coalition of size more
 1498 than 3, and moreover, all coalitions of size 3 must be consecutive. If π contains a nonconsecutive
 1499 coalition C of size 2, then C must be of the form $\{i, i+2\}$, as otherwise, its two agents are mutual
 1500 enemies, and thus either of them switching to \emptyset locally dominates π . We may further assume that
 1501 $i-1 \notin \pi(i+1)$ (as otherwise, we can pick the coalition $\{i-1, i+1\}$ instead of C); hence, either
 1502 $\{i+1\}$ is a singleton in π or $\{i+1, i+3\} \in \pi$. Consider the switch $i+2 \rightarrow \pi(i+1)$: both
 1503 $i+1$ and $i+2$ vote for it, while only i votes against it (note that $i+3$ is indifferent between π and
 1504 $\pi_{i+2 \rightarrow \pi(i+1)}$). Hence, this switch locally dominates π , a contradiction that shows that all coalitions
 1505 in π must be consecutive.

1506 If $C = \{i, i+1\} \in \pi$, then both i and $i+1$ vote for the switch $i+2 \rightarrow C$, and only $i+2$ (but
 1507 neither $i+3$ nor $i+4$) may vote against it, contradicting again the local popularity of π . Thus π can

only contain singletons and (consecutive) coalitions of size 3. If i and $i + 1$ both form singletons in π , then i has an incentive to join $\{i + 1\}$, and $i + 1$ does not vote against it, so $i \rightarrow \{i + 1\}$ locally dominates π . Hence, the only remaining possibility is that π contains a singleton and two coalitions of size 3, say $\{i - 3, i - 2, i - 1\}$ and $\{i, i + 1, i + 2\}$ for some agent i . However, in this case the switch $i \rightarrow \pi(i - 1)$ locally dominates π , because agents $i - 1, i - 2$, and $i + 2$ vote for it, while only i and $i - 3$ vote against it. This contradiction completes the proof. \square

1514 D.2.4 Proof of Theorem 9

Before proving Theorem 9, we start with a simple lemma that gives the bases of our reductions for Theorems 9 and 11.

Lemma 29. *Consider a FEN-B game containing three special agents, s_1, s_2 , and s_3 , a set $D_1 \cup D_2 \cup D_3$ of dummy agents, and a set A of vertex agents with $|D_1| = |D_2| = |D_3| = 3|A|$. For each $i \in [3]$, let $D_i^+ = D_i \cup \{s_i\}$ and let the friendship and enmity relations in I be such that*

- every agent in D_i^+ regards all other agents in D_i^+ a friend and regards every remaining special or dummy vertex a friend, and
- vertex agents are mutually neutral with every dummy agent.

Then every locally stable partition π contains three coalitions, C_1, C_2 , and C_3 , such that $D_i^+ \subseteq C_i$ for each $i \in [3]$.

Proof. First, assume that there exists a coalition $C \in \pi$ such that C contains agents both from D_i^+ and D_j^+ for some indices $i \neq j$; we will call such coalitions *mixed*. Pick i such that $|D_i^+ \cap C| \leq |D_j^+ \cap C|$. If $D_i^+ \cap C$ contains a dummy agent d , then the switch $d \rightarrow \emptyset$ locally blocks π , because all agents in D_j^+ as well as d vote for it, while only the agents in $D_i^+ \setminus \{d\}$ vote against it; a contradiction. Hence, we know that $C \cap D_i^+ = \{s_i\}$. In this case we must have $|C \cap D_j^+| \leq |A|$, as otherwise, $s_i \rightarrow \emptyset$ locally blocks π , because only vertex agents in C may vote against it, while all agents in D_j^+ as well as s_i vote for it. Thus each mixed coalition C contains dummy agents from at most one set D_j among D_1, D_2 , and D_3 (and then must contain a special vertex other than s_j), and moreover, a mixed coalition can contain at most $|A|$ agents from D_j^+ .

Fix some index $j \in [3]$, and assume now that there are coalitions C and C' in π such that both contain agents from D_j , $|C \cap D_j^+| \leq |C' \cap D_j^+|$, and C' is non-mixed. Let us pick an arbitrary agent $d \in C \cap D_j^+$. Then the switch $d \rightarrow C'$ locally blocks π , as all agents in $(C' \cap D_j^+) \cup \{d\}$ vote for it, while the only agents who vote against it are those in $(C' \cap D_j^+) \setminus \{d\}$ because C' is non-mixed. This contradiction shows that there is at most one non-mixed coalition that contains agents from D_j , and moreover, it contains fewer agents from D_j^+ than any mixed coalition containing an agent of D_j .

If π contains at least one mixed coalition C with $C \cap D_j \neq \emptyset$, then the unique (or possibly nonexistent) non-mixed coalition that contains agents from D_j must contain fewer than $|A|$ agents from D_j^+ . However, since there can be at most two mixed coalitions containing agents of D_j (as each of them must contain a special agent other than s_j), we obtain that a total of at most $3|A| - 1$ agents of D_j can be contained in coalitions of π , contradicting $|D_j| = 3|A|$. This proves that there are no mixed coalitions containing agents of D_j and hence all agents of D_j must be contained in a unique non-mixed coalition C_j . Note also that $s_j \in C_j$ follows, as otherwise, $s_j \rightarrow C_j$ locally blocks π . This proves the claim. \square

Theorem 9 (*). LOCAL-POPULARITY-EXISTENCE is NP-complete for FEN-B games.

Proof. It is clear that this problem is in NP, as given a FEN-B game, we can nondeterministically guess a partition π and then verify in polynomial time that no possible switch locally dominates π (note again that the number of switches for π is at most quadratic in the number of agents).

To show NP-hardness, we reduce from the NP-hard 3-COLORING problem: given an undirected graph $G = (V, E)$, the task is to decide whether G admits a coloring $\chi : V \rightarrow [3]$ that is *proper*, i.e.,

1554 there is no edge $\{u, v\} \in E$ with $\chi(u) = \chi(v)$. We will call a proper coloring *tight* if each vertex
1555 has two neighbors whose colors differ from each other.

1556 Given an instance $G = (V, E)$ of 3-COLORING, we start by creating a graph H as follows: for each
1557 vertex $v \in V$, we add new vertices v^1, v^2 , and v^3 to G forming a triangle, and we connect v^1 and v^2
1558 to v by an edge. Observe that a proper coloring of G can be extended into a tight coloring of H and,
1559 in fact, every proper coloring of H is tight. Let $H = (V', E')$ denote the constructed graph and let
1560 $n = |V'|$.

1561 Next, we create a FEN-B game I based on H as follows. We introduce a set $D_1 \cup D_2 \cup D_3$ of *dummy*
1562 *agents* with $|D_1| = |D_2| = |D_3| = 21n$, three *special agents*, s_1, s_2 , and s_3 , and for each vertex
1563 $v \in V'$, a set $A_v = \{v_i : i \in \{0, 1, \dots, 6\}\}$ of seven *vertex agents* who form a B-gadget. We will use
1564 the notation $D_i^+ = D_i \cup \{s_i\}$ for each $i \in [3]$, so the set of agents is $N = \bigcup_{i \in [3]} D_i^+ \cup \bigcup_{v \in V'} A_v$.

1565 The remaining relationships between the agents are as follows:

- 1566 • For each $i \in [3]$, every agent in D_i^+ considers each other agent in D_i^+ a friend and every
1567 remaining special or dummy agent an enemy.
- 1568 • Two vertex agents, $u_i \in A_u$ and $v_j \in A_v$, are mutual enemies if and only if $\{u_i, v_j\} \in E'$;
1569 otherwise, they are mutually neutral.
- 1570 • Each vertex agent is mutual friends with every special agent and mutually neutral with every
1571 dummy agent.

1572 This completes our construction of the instance I of LOCAL-POPULARITY-EXISTENCE from the
1573 given 3-COLORING instance G . We are now going to show that the constructed FEN-B game I
1574 admits a locally popular partition if and only if G admits a proper 3-coloring.

1575 **Direction “ \Rightarrow ”:** Assume first that there exists a locally popular partition π for I .

1576 Let us fix an arbitrary vertex $v \in V'$ and consider the corresponding B-gadget on A_v . Let \mathcal{C}_v denote
1577 the set of all coalitions $C \in \pi$ with $C \cap A_v \neq \emptyset$. Let I_v be the FEN-B game comprising solely of A_v ,
1578 and let $\pi^v = \{C \cap A_v : C \in \mathcal{C}_v\}$ denote the partition for I_v that can be thought of as the projection
1579 of π onto A_v . Let the *level of a coalition* $C \in \mathcal{C}_v$ be

$$\text{lev}(C) = |F(v_i) \cap (C \setminus A_v)| - |E(v_i) \cap (C \setminus A_v)|$$

1580 for an arbitrary $v_i \in C \cap A_v$; note that this notion is well-defined, as all vertex agents in A_v have the
1581 same friends and enemies outside A_v .

1582 Observe now that I satisfies the conditions in Lemma 29: hence, as π is locally stable, we get that
1583 there exist coalitions $C_1, C_2, C_3 \in \pi$ such that $D_i^+ \subseteq C_i$ for each $i \in [3]$. Since vertex agents in A_v
1584 regard only special agents as friends outside A_v , it follows that the level of any coalition in \mathcal{C}_v is at
1585 most 1.

1586 **Claim 30.** *Each coalition $C \in \mathcal{C}_v$ has level at least 0.*

1587 *Proof of claim.* For the sake of contradiction, assume that $C \in \mathcal{C}_v$ has level at most -1 .

1588 If $C \cap A_v$ satisfies the condition of Lemma 28 (i.e., either $|C \cap A_v| \geq 4$, or $|C \cap A_v| = 3$ and $C \cap A_v$
1589 is not consecutive), then we know that there exists an agent v_i in $C \cap A_v$ such that the switch $v_i \rightarrow \emptyset$
1590 Notice that agents in $A_v \setminus \{v_i\}$ do not consider their relationships outside A_v when voting for or
1591 against the switch $v_i \rightarrow \emptyset$, so they vote the same way in I as in I_v . Moreover, we have

$$\text{vote}_{v_i}(\pi, \pi_{v_i \rightarrow \emptyset}) \geq \text{vote}_{v_i}(\pi^v, \pi_{v_i \rightarrow \emptyset}^v)$$

1592 because v_i has more enemies in $C \setminus A_v$ than friends (since C has level at most -1). Finally, both
1593 friendship and enemy relationships between two agents who do not belong to the same B-gadget are
1594 symmetric, so we know that

$$\sum_{a \in C \setminus A_v} \text{vote}_a(\pi, \pi_{v_i \rightarrow \emptyset}) = -\text{lev}(C) \geq 1.$$

1595 Summing up all these insights, we obtain that $\Lambda(\pi, \pi_{v_i \rightarrow \emptyset}) \geq 1 - \text{lev}(C) \geq 2$, contradicting the
1596 local popularity of π .

1597 Assume now that $|C \cap A_v| = 3$ and $C \cap A_v$ is consecutive, that is, $C \cap A_v = \{v_i, v_{i+1}, v_{i+2}\}$ for
1598 some $i \in [6]$ (henceforth, we treat indices within A_v modulo 7). Then $v_i \rightarrow \emptyset$ locally dominates π :
1599 agent v_i votes against it, v_{i+2} votes for it, v_{i+1} does neither, while the total vote of all agents in $C \setminus A_v$
1600 sums up to $-\text{lev}(C) \geq 1$, yielding $\Lambda(\pi, \pi_{v_i \rightarrow \emptyset}) \geq 1$.

1601 If $|C \cap A_v| = 2$ and $C \cap A_v$ is nonconsecutive, then the switch $v_i \rightarrow \emptyset$ for any $v_i \in C$ locally
1602 dominates π , because at least one agent in $C \cap A_v$ votes for it, while the total vote of all agents
1603 in $C \setminus A_v$ sums up to $-\text{lev}(C) \geq 1$, yielding $\Lambda(\pi, \pi_{v_i \rightarrow \emptyset}) \geq 1$.

1604 If $C \cap A_v = \{v_i, v_{i+1}\}$ for some agent $v_i \in C$, then again $v_i \rightarrow \emptyset$ locally dominates π : by
1605 $u_{v_i}^B(\pi) = 0$, both v_i and v_{i+1} are indifferent between π and $\pi_{v_i \rightarrow \emptyset}$, and the total vote of all agents
1606 in $C \setminus A_v$ sums up to $-\text{lev}(C) \geq 1$, yielding $\Lambda(\pi, \pi_{v_i \rightarrow \emptyset}) \geq 1$.

1607 Finally, if $C \cap A_v = \{v_i\}$, then again $v_i \rightarrow \emptyset$ locally dominates π , because v_i votes for it, and the
1608 total vote of all agents in $C \setminus A_v$ sums up to $-\text{lev}(C) \geq 1$, yielding $\Lambda(\pi, \pi_{v_i \rightarrow \emptyset}) = 2$. \triangleleft

1609 We next show that it is not possible for all coalitions in \mathcal{C}_v to have level exactly 0. Assume the
1610 contrary. By Theorem 8, we know that $\pi|_v$ cannot be locally popular in I_v , so there exists an agent v_i
1611 and a coalition $C \in \pi|_v \cup \{\emptyset\}$ such that the switch $v_i \rightarrow C$ locally dominates $\pi|_v$ in I_v . Let $C' \in \pi$
1612 be the coalition satisfying $C = C' \cap A_v$. We will show that the switch $v_i \rightarrow C'$ locally dominates π
1613 in I . First, agents in $A_v \cap (\pi(v_i) \cup C') \setminus \{v_i\}$ vote for or against the switch without regard for the
1614 agents outside A_v . Second, the contribution of all agents outside A_v to the balanced utility of v_i
1615 remains the same in C' as in $\pi(v_i)$ due to $\text{lev}(C') = \text{lev}(\pi(v_i)) = 0$. Third, the total vote of all
1616 agents regarding the switch $v_i \rightarrow C'$ sums up to exactly $-\text{lev}(\pi(v_i)) + \text{lev}(C) = 0$. Summing up all
1617 this, we get $\Lambda(\pi, \pi_{v_i \rightarrow C'}) \geq 1$, as promised.

1618 By Claim 30, this shows that for each $v \in V'$, there exists some $C \in \mathcal{C}_v$ with $\text{lev}(C) = 1$; if there
1619 are multiple such coalitions, then fix one arbitrarily. Notice that C must contain a unique special
1620 agent s_i due to $\text{lev}(C) = 1$; we then define $\chi(v) = i$. We claim that χ is a proper coloring of H :
1621 Assuming that the endpoints of an edge $\{u, v\} \in E'$ (recall that E' is the edge set of H) have the
1622 same color i , we get that $C_i = \pi(s_i)$ contains agents both from A_u and A_v ; however, as they are
1623 mutual enemies and the only friend of vertex agents outside their gadget must be a special agent, we
1624 obtain that agents in $A_v \cap C_i$ have at least as many enemies as friends in $C_i \setminus A_v$, a contradiction to
1625 $\text{lev}(C) = 1$. It follows that H and, hence, its subgraph G admits a proper 3-coloring.

1626 **Direction “ \Leftarrow ”:** Assume now that G admits a proper coloring, so H admits a proper and tight
1627 coloring $\chi : V' \rightarrow [3]$. Create the partition π consisting of three coalitions

$$C_i = D_i^+ \cup \{A_v : v \in V', \chi(v) = i\}$$

1628 for $i = 1, 2, 3$. We claim that π is locally popular. For the sake of contradiction, assume that some
1629 switch $a \rightarrow C$ locally dominates π .

1630 First, if a is a dummy or special agent in C_i for some $i \in [3]$, then all agents in D_i^+ vote against $a \rightarrow C$,
1631 and the only agents that may vote for it are vertex agents (in case $a = s_i$); hence, $\Lambda(\pi, \pi_{a \rightarrow C}) \leq 0$
1632 by $|D_i^+| > 7n$.

1633 Second, if a is a vertex agent in A_v , then three agents in A_v vote for $a \rightarrow C$, while two agents in A_v
1634 vote against it. We also know that $u_a^B(\pi) = 0$, because it has three enemies in A_v and two friends
1635 in A_v plus s_i . Moreover, no vertex agent in A_v votes for the switch $a \rightarrow C$: since χ is a proper
1636 coloring, there are no enemies of a in $C_i \setminus A_v$; the special agent s_i votes against it, and dummies in C_i
1637 are indifferent between π and $\pi_{a \rightarrow C}$. This already shows that $a \rightarrow \emptyset$ does not locally dominate π ,
1638 because $\Lambda(\pi, \pi_{a \rightarrow \emptyset}) = \sum_{a' \in C_i} \text{vote}_{a'}(\pi, \pi_{a \rightarrow \emptyset}) = 0$. To deal with the case when $C = C_j \in \pi$ for
1639 some $j \neq i$, recall that since χ is tight, there is at least one vertex $u \in V'$ among the neighbors of v
1640 with $\chi(u) = j$. This means that all seven agents in $A_u \subseteq C_j$ vote against $a \rightarrow C_j$; while the only
1641 agent in C_j that votes for it is s_j . Therefore, $\Lambda(\pi, \pi_{a \rightarrow C_j}) \leq -6$. This contradiction proves that π is
1642 indeed locally popular. \square

1643 D.2.5 Proof of Proposition 10

1644 **Proposition 10 (\star).** *There exists a symmetric FEN-B game that admits a partition that is popular but*
1645 *does not maximize the agents' total balanced utility.*

1646 *Proof.* Consider the symmetric FEN-B game I on agent set $N = A \cup B$ where $A = \{a_i : i \in [7]\}$
 1647 and $B = \{b_i : i \in [7]\}$. To create the (undirected) friendship graph G^F of I , we add a clique on A ,
 1648 a clique on B , plus the three edges in $\{\{a_i, b_i\} : i \in [3]\}$. The (undirected) enmity graph G^E of I
 1649 contains only the four edges in $\{\{a_7, b_j\} : j = 4, 5, 6, 7\}$.

1650 We claim that the grand coalition $\pi^N = \{N\}$ is popular for I . To see this, assume for the sake
 1651 of contradiction that there is a partition π with $\Lambda(\pi^N, \pi) > 0$. Note first that no agent in $N \setminus$
 1652 $\{a_7, b_4, b_5, b_6, b_7\}$ may prefer π to π^N as they have all their friends with them and no enemies in the
 1653 coalition N . Thus at most five agents may prefer π to π^N . Second, if some agent prefers π to π^N ,
 1654 then π must put a_7 and at least one agent in $\{b_4, b_5, b_6, b_7\}$ in different coalitions.

1655 If there is no coalition in π containing A , then all six agents in $A \setminus \{a_7\}$ prefer π^N to π .

1656 If some agent other than a_7 prefers π to π^N , then B must also be contained in some coalition of π .
 1657 Otherwise, no agent in $\{b_4, b_5, b_6, b_7\}$ prefers π to π^N , as they lose at least one friend and lose at
 1658 most one enemy when switching from π^N to π . Hence, in this case only $\pi = \{A, B\}$ is possible, but
 1659 then all six agents in $\{a_i, b_i : i \in [3]\}$ prefer π^N to π .

1660 Finally, if only a_7 prefers π to π^N , then π must contain a coalition C with $C \subsetneq B$. However, then all
 1661 agents $b_i, i \in [3]$, prefer π^N to π .

1662 This proves that π^N is indeed popular.

1663 The total balanced utility of π^N is $f(\pi^N) = 4 \cdot \binom{7}{2} + 2 \cdot 3 - 2 \cdot 4$, but this is lower than the total balanced
 1664 utility of the partition $\{A, B\}$ by $f(\{A, B\}) = 4 \cdot \binom{7}{2} < f(\pi^N)$. This proves our statement. \square

1665 D.2.6 Proof of Theorem 11

1666 **Theorem 11** (*). *Given a symmetric FEN-B game I and an integer t , the problem of deciding*
 1667 *whether I admits a partition π whose total balanced utility is at least t is NP-complete.*

1668 *Proof.* It is clear that the problem is in NP, as we can compute the total utility of any partition
 1669 efficiently. To prove NP-hardness, we again present a reduction from 3-COLORING based on the
 1670 construction in Lemma 29. Given an input graph $G = (V, E)$ over n vertices, let us create a
 1671 symmetric FEN-B game I as follows. We introduce a set $D_1 \cup D_2 \cup D_3$ of *dummy agents* with
 1672 $|D_1| = |D_2| = |D_3| = 3n$, three *special agents*, s_1, s_2 , and s_3 , and a *vertex agent* v' for each vertex
 1673 $v \in V$. We will again use the notation $D_i^+ = D_i \cup \{s_i\}$ for each $i \in [3]$.

1674 The friendship and enmity relationships between the agents are as follows:

- 1675 • For each $i \in [3]$, every agent in D_i^+ considers each other agents in D_i^+ a friend and every
 1676 remaining special or dummy agent an enemy.
- 1677 • Two vertex agents, u' and v' , are mutual enemies if and only if $\{u, v\} \in E$; otherwise, they
 1678 are mutually neutral.
- 1679 • Each vertex agent is mutual friends with every special agent and mutually neutral with every
 1680 dummy agent.

1681 Additionally, we set our target value as $t = 6\binom{n}{2} + 2n$.

1682 This completes our construction of the instance I from the given 3-COLORING instance G . We are
 1683 going to show that the constructed FEN-B game I admits a partition π with $f(\pi) \geq t$ if and only if
 1684 G admits a proper 3-coloring.

1685 **Direction “ \Rightarrow ”:** Assume first that there exists a popular partition π for I with $f(\pi) \geq t$; without loss
 1686 of generality, we may assume that π maximizes the total balanced utility over all partitions for I . In
 1687 particular, this implies that I is locally stable: otherwise, a switch that locally blocks π would result
 1688 in a strict increase in $f(\pi)$, as shown in Theorem 4; a contradiction. Therefore, Lemma 29 can be
 1689 applied, and we obtain that π contains three coalitions $C_1, C_2, C_3 \in \pi$ such that $D_i^+ \subseteq C_i$ for each
 1690 $i \in [3]$.

1691 Since each vertex agent can belong to at most one of the sets C_1, C_2 , and C_3 , the total balanced
 1692 utility of all dummy and special agents is at most $6\binom{n}{2} + n$. Moreover, the balanced utility of any

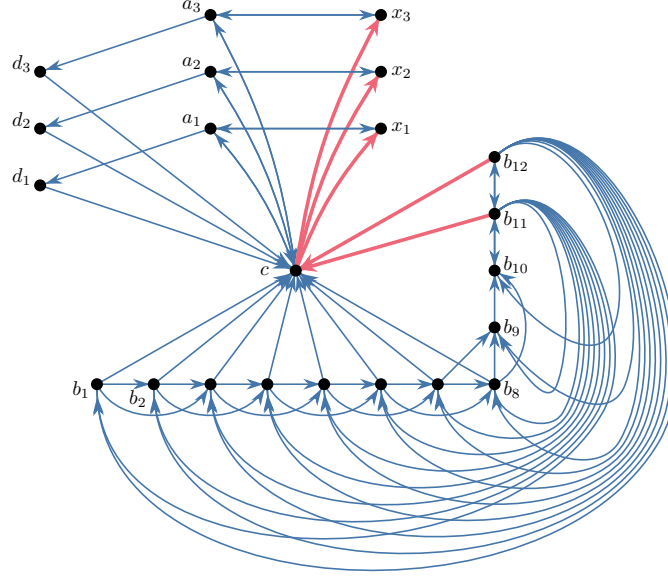


Figure 12: Illustration of the AE-gadget as defined in Definition 4. Here we denote friendship arcs as thin, blue lines and enemy arcs as thick, red lines.

vertex agent v' is at most 1, as the only agents regarded as friends by v' are special agents, and v' can share a coalition with at most one of them. Hence, $f(\pi) \leq 6\binom{n}{2} + 2n = t$, and equality can only be achieved if each vertex agent v is in coalition C_i for some $i \in [3]$ and has no enemies in $\pi(v') = C_i$. However, this implies that each set $\{v \in V : v' \in C_i\}$ is an independent set in G (i.e., no two of its vertices are connected by an edge of G), proving that G admits a proper 3-coloring.

Direction “ \Leftarrow ”: Assume now that G admits a proper 3-coloring $\chi : V \rightarrow [3]$. Define coalitions $C_i = D_i^+ \cup \{v' : v \in V \text{ and } \chi(v) = i\}$ for $i \in [3]$. Then the partition $\pi = \{C_1, C_2, C_3\}$ has total balanced utility exactly $f(\pi) = 6\binom{n}{2} + 2n = t$, proving the correctness of our reduction. \square

D.3 Omitted proofs for our results on FEN-AE games

After stating some structural observations in Section D.3.1, we provide all omitted proofs for FEN-AE games in Sections D.3.2–D.3.3. See Figure 12 for an illustration of an AE-gadget, our building block for all of our nonexistence and hardness results concerning FEN-AE games.

D.3.1 A structural lemma on AE-gadgets

Lemma 31. Consider a FEN-AE game I that contains an AE-gadget as described in Definition 4 over agent set $K = \{c\} \cup \{b_i : i \in [12]\} \cup \{a_i, d_i, x_i : i \in [3]\}$, and let π be a locally stable partition for I . Assume that no agent outside K considers anyone within K a friend, and among agents of K , only x_1, x_2 , and x_3 may have friends outside K . Furthermore, assume that coalitions $\pi(c)$, $\pi(b_{11})$, and $\pi(b_{12})$ only contain agents from K . Then

- (a) $\{a_1, a_2, a_3, d_1, d_2, d_3\} \subseteq \pi(c)$;
- (b) if, for some $i \in [3]$, agent x_i considers no agent in $\pi(x_i) \setminus K$ a friend, then $x_i \in \pi(c)$; and
- (c) $\{b_i : i \in [12]\} \subseteq \pi(b_1) \neq \pi(c)$.

Proof. To show (a), we first show that $d_i \in \pi(c)$ for each $i \in [3]$. By our assumptions, the only agents who consider d_i a friend are a_i and c , whereas d_i has only c as a friend and has no enemies in $\pi(c)$. Therefore, if $d_i \notin \pi(c)$, then d_i has an incentive to join $\pi(c)$, and thus the switch $d_i \rightarrow \pi(c)$ locally blocks π : both d_i and c vote for it and only a_i may vote against it. This proves $d_i \in \pi(c)$.

1718 This implies that among the three agents whom a_i , for some $i \in [3]$, considers friends, two are
 1719 contained in $\pi(c)$, and due to our assumptions, a_i has no enemies in $\pi(c)$. Hence, if $a_i \notin \pi(c)$, then
 1720 a_i has an incentive to join $\pi(c)$. The switch $a_i \rightarrow \pi(c)$ then locally blocks π : both a_i and c vote for
 1721 it, while only x_i (the only agent who considers a_i a friend) may vote against it. This contradiction
 1722 proves $\{a_1, a_2, a_3, d_1, d_2, d_3\} \subseteq \pi(c)$, as desired.

1723 To show (b), assume for the sake of contradiction that $x_i \notin \pi(c)$ and x_i has no friends in $\pi(x_i) \setminus K$.
 1724 Then x_i has no friends in $\pi(x_i)$, because the only agent whom x_i considers a friend is $a_i \in \pi(c)$.
 1725 Since x_i has no enemies in $\pi(c) \subseteq K$ due to our assumptions, x_i has an incentive to join $\pi(c)$. Hence,
 1726 $x_i \rightarrow \pi(c)$ locally blocks π , as both x_i and a_i vote for it, and only c votes against it; a contradiction
 1727 proving (b).

1728 It remains to show (c). First observe that $c \notin \pi(b_{12})$ must hold, as otherwise, the switch $b_{12} \rightarrow \emptyset$
 1729 would locally block π , as both c and b_{12} would vote for it (as they are mutual enemies) and only b_{11}
 1730 may vote against it. Similarly, $c \notin \pi(b_{11})$ also holds, as otherwise, the switch $b_{11} \rightarrow \emptyset$ would locally
 1731 block π , since both c and b_{11} would vote for it and only b_{10} may vote against it (note that $b_{12} \notin \pi(c)$
 1732 as we have just shown).

1733 Next, we claim that $\pi(b_{11}) = \pi(b_{12})$. Suppose the contrary: $\pi(b_{11}) \neq \pi(b_{12})$. Choose j and j' such
 1734 that $\{j, j'\} = \{11, 12\}$ and $\pi(b_j)$ contains at least as many agents from $\{b_i : i \in [10]\}$ as $\pi(b_{j'})$.
 1735 Since b_{11} and b_{12} consider each other a friend and, apart from each other, they have the same set
 1736 of friends, and their only enemy c is not in either of $\pi(b_{11})$ or $\pi(b_{12})$, it follows that the switch
 1737 $b_{j'} \rightarrow \pi(b_j)$ locally blocks π , as both b_{11} and b_{12} vote for it and only b_{10} may vote against it.

1738 It follows that $b_{10} \in \pi(b_{11})$, as otherwise, b_{10} has an incentive to join $\pi(b_{11})$, and the switch
 1739 $b_{10} \rightarrow \pi(b_{11})$ locally blocks π : agents b_{10} , b_{11} , and b_{12} all vote for it and only b_8 and b_9 may vote
 1740 against it. Similarly, we obtain $b_9 \in \pi(b_{11})$. Finally, once we know that $b_{i'} \in \pi(b_{11})$ for all $i' > i$
 1741 for some $i \in [8]$, then $b_i \in \pi(b_{11})$ follows, as otherwise, $b_i \rightarrow \pi(b_{11})$ would locally block π : among
 1742 the three agents whom b_i considers friends, $\pi(b_{11})$ contains two and contains no enemies of b_i by
 1743 $\pi(b_{11}) \subseteq K$, so b_i together with b_{11} and b_{12} vote for $b_i \rightarrow \pi(b_{11})$, while only b_{i-1} and b_{i-2} may
 1744 vote against it. Hence, $\{b_i : i \in [12]\} \subseteq \pi(b_{11}) = \pi(b_{12}) \neq \pi(c)$. \square

1745 D.3.2 Proof of Theorem 12

1746 **Theorem 12 (\star).** *The asymmetric FEN-AE game consisting solely of an AE-gadget admits no locally*
 1747 *stable partition.*

1748 *Proof.* Consider a FEN-AE game I that solely consist of an AE-gadget as described in Definition 4
 1749 over agent set $K = \{c\} \cup \{b_i : i \in [12]\} \cup \{a_i, d_i, x_i : i \in [3]\}$. By Lemma 31, the only partition that
 1750 may be locally stable for I is $\pi = \{\{a_i, d_i, x_i, c : i \in [3]\}, \{b_i : i \in [12]\}\}$. Observe that c has two
 1751 enemies and zero friends in $\pi(b_1)$, whereas c has three enemies in $\pi(c)$, and hence, c prefers $\pi(b_1)$
 1752 to $\pi(c)$. Therefore, the switch $c \rightarrow \pi(b_1)$ locally blocks π : nine agents (namely, c and b_j for $j \in [8]$)
 1753 vote for it, while only eight agents (namely, a_i and d_i for $i \in [3]$ as well as b_{11} and b_{12}) vote against
 1754 it. Thus no partition for I is locally stable. \square

1755 D.3.3 Proof of Theorem 13

1756 **Theorem 13 (\star).** *LOCAL-POPULARITY-EXISTENCE and LOCAL-STABILITY-EXISTENCE are NP-*
 1757 *complete for FEN-AE games.*

1758 *Proof.* It is clear that both problems are in NP, as given a FEN-AE game, we can nondeterministically
 1759 guess a partition π and then verify in polynomial time that no possible switch locally dominates
 1760 or locally blocks π (recall that the number of switches for π is at most quadratic in the number of
 1761 agents).

1762 To show the NP-hardness of these problems, we present a reduction from 3-SAT that works for both
 1763 problems. Let $\varphi = C^1 \wedge \dots \wedge C^m$ be an instance of 3-SAT, i.e., a Boolean formula over variables
 1764 v_1, \dots, v_n with three literals per clause; without loss of generality, we assume that $m \geq 2$.

1765 We create a FEN-AE game I of LOCAL-POPULARITY-EXISTENCE and LOCAL-STABILITY-
 1766 EXISTENCE as follows. For each clause C^j , we create an AE-gadget G^j as described in Definition 4
 1767 over agent set $K^j = \{c^j\} \cup \{b_i^j : i \in [12]\} \cup \{a_i^j, d_i^j, x_i^j : i \in [3]\}$. We will call c^j a *clause agent*

1768 and agents x_1^j , x_2^j , and x_3^j *literal agents*, corresponding to the three literals in C^j . We further add an
1769 agent T .

1770 The friendship and enmity relations between agents from different AE-gadgets are as follows:

- 1771 • x_i^j considers T a friend for each $i \in [3]$ and $j \in [m]$, while all other agents consider T an
1772 enemy.
- 1773 • T considers x_i^j a neutral for each $i \in [3]$ and $j \in [m]$ but considers every other agent an
1774 enemy.
- 1775 • Two agents, u and v , from different AE-gadgets are mutual enemies unless both of them are
1776 literal agents corresponding to literals that are not negations of each other, in which case
1777 they are mutually neutral.

1778 We prove the correctness of the reduction via Claims 32 and 33.

1779 **Claim 32.** *If I admits a locally stable partition, then φ is satisfiable.*

1780 *Proof of claim.* Let π be a locally stable partition for I . For the sake of contradiction, suppose that
1781 an agent $v \in \{c^j, b_{11}^j, b_{12}^j\}$ is contained in the same coalition of π as some agent u from a different
1782 AE-gadget G^h . Let ℓ denote the largest integer such that both u and v have at least ℓ enemies
1783 in $\pi(u) = \pi(v)$. Since u and v are mutual enemies, we know $\ell \geq 1$, and thus both u and v have an
1784 incentive to join \emptyset . However, none of the switches $u \rightarrow \emptyset$ and $v \rightarrow \emptyset$ can locally block π , so there
1785 must exist a set F_u of at least $\ell + 1$ agents in $\pi(u) = \pi(v)$ who consider u a friend and, similarly, a
1786 set F_v of at least $\ell + 1$ agents who consider v a friend in $\pi(v) = \pi(u)$. In particular, u is not a literal
1787 agent $x_i^h \in K^h$ for some $i \in [3]$, as x_i^h is regarded as a friend only by a_i^h . Moreover, as all friendship
1788 relations lie within some AE-gadget, we get that $F_u \subseteq K^h$ and $F_v \subseteq K^j$. However, u must then
1789 be mutual enemies with all $\ell + 1$ agents in $F_v \subseteq \pi(v)$, and v mutual enemies with all $\ell + 1$ agents
1790 in $F_u \subseteq \pi(v)$, which contradicts the definition of ℓ .

1791 Again for the sake of contradiction, suppose now that an agent $v \in \{c^j, b_{11}^j, b_{12}^j\}$ is contained in $\pi(T)$.
1792 Since v and T are mutual enemies but $v \rightarrow \emptyset$ does not locally block π , there must be at least two
1793 agents in $\pi(T)$ who consider v a friend. These two agents cannot be literal agents by the choice
1794 of v , so they are also mutual enemies with T . Hence, at least four agents vote for the switch $T \rightarrow \emptyset$
1795 including T . However, we have already shown that no agent from an AE-gadget other than G^j can
1796 be contained in $\pi(v) = \pi(T)$; therefore, at most three agents in $\pi(T)$ consider T a friend, implying
1797 that $T \rightarrow \emptyset$ locally blocks π , a contradiction.

1798 We conclude that coalitions $\pi(c^j)$, $\pi(b_{11}^j)$, and $\pi(b_{12}^j)$ only contain agents from G^j . This means that
1799 all conditions of Lemma 31 are satisfied. Hence, if no agent $x_i^j \in \{x_1^j, x_2^j, x_3^j\}$ considers at least
1800 one agent in $\pi(x_i^j) \setminus K^j$ a friend, then by Lemma 31 we get $\pi(c^j) = \{x_i^j, a_i^j, d_i^j, c^j : i \in [3]\}$ and
1801 $\pi(b_1^j) = \{b_i^j : i \in [12]\}$. However, then it follows from Theorem 12 that some switch, namely the
1802 switch $c^j \rightarrow \pi(b_1^j)$, locally blocks π , a contradiction.

1803 This means that for each AE-gadget G^j , there exists a literal agent $x_i^j \in K^j$ that considers at least
1804 one agent in $\pi(x_i^j) \setminus K^j$ a friend; by construction, this agent can only be T . Let X_T denote the set
1805 of literal agents in $\pi(T)$; then we have just proven that $X_T \cap K^j \neq \emptyset$ for each $j \in [m]$. Suppose
1806 that two agents in X_T are mutual enemies (which happens if they correspond to literals v_i and \bar{v}_i
1807 for some variable v_i). Then both of them have an incentive to join \emptyset , and thus both of them vote for
1808 the corresponding switches, whereas at most one agent may vote against them (because for each
1809 literal agent x_i^j , there is only one agent, namely a_i^j , who considers x_i^j a friend); thus, either of these
1810 switches locally blocks π , a contradiction.

1811 Therefore, we can create a truth assignment by setting v_i to true if and only if there is a literal agent
1812 corresponding to v_i in X_T . By the above arguments, this is a consistent truth assignment and yields a
1813 true literal in each clause, thus satisfying φ . \triangleleft

1814 **Claim 33.** *If φ is satisfiable, then I admits a locally popular partition.*

1815 *Proof of claim.* Suppose that φ is satisfiable by a truth assignment μ . We create a partition π
 1816 as follows. We let $\pi(T)$ contain—besides T —all literal agents that correspond to true literals.
 1817 Additionally, for each AE-gadget G^j , we create a coalition $\{b_i^j : i \in [12]\}$ and a coalition
 1818 $\{a_i^j, d_i^j, x_i^j, c^j : i \in [3]\} \setminus \pi(T)$.

1819 We claim that π is locally popular. Since μ satisfies φ , each clause agent c^j has at most two enemies
 1820 and six friends in $\pi(c^j)$, has two enemies and zero friends in $\pi(b_1^j)$, and has at least $m + 1 \geq 3$
 1821 enemies in $\pi(T)$. Hence, the only possible coalition in $\pi \cup \{\emptyset\}$ that c^j has an incentive to join is \emptyset
 1822 via the switch $c^j \rightarrow \emptyset$. Regarding the remaining (i.e., all non-clause) agents, observe that none of
 1823 them has enemies in its coalition under π . Using this, we get the following:

- 1824 • Neither T nor any agent d_i^j votes for any possible switch.
- 1825 • An agent b_i^j may only vote for the switch $c^j \rightarrow \pi(b_1^j)$. However, all agents in $\{c^j, a_i^j, d_i^j : i \in [3]\} \cup \{b_{11}^j, b_{12}^j\}$ vote against such a switch, while only the agents in $\{b_i^j : i \in [8]\}$ vote for it, so it does not locally dominate π .
 1826
 1827
- 1828 • As we have argued, a clause agent c^j votes only for the switch $c^j \rightarrow \emptyset$; however, such a
 1829 switch does not locally dominate π , as only c^j votes for it, while all six agents in $\{a_i^j, d_i^j : i \in [3]\}$ vote against it.
 1830
- 1831 • An agent a_i^j votes only for switches of the form $x_i^j \rightarrow \pi(a_i^j)$, where $x_i^j \in \pi(T)$. However,
 1832 such a switch does not locally dominate π , as only a_i^j votes for it, while c^j votes against it.
- 1833 • Finally, an agent x_i^j may only vote for a switch of the form $T \rightarrow \pi(x_i^j)$ or $a_i^j \rightarrow \pi(x_i^j)$.
 1834 A switch $T \rightarrow \pi(x_i^j)$ for some $x_i^j \notin \pi(T)$ is voted for only by the at most two literal
 1835 agents in $\pi(x_i^j)$, while it is voted against by T and all remaining seven agents in $\pi(x_i^j)$. A
 1836 switch $a_i^j \rightarrow \pi(x_i^j)$ for some $x_i^j \in \pi(T)$ is voted for only by the at most three literal agents
 1837 in $\pi(T) \cap K^j$, while T , c^j , and at least $m - 1 \geq 1$ literal agents from $\pi(T) \setminus K^j$ vote
 1838 against it. Thus none of these switches locally dominates π .

1839 We obtain that no switch for which at least one agent votes locally dominates π , that is, π is locally
 1840 popular. \triangleleft

1841 As a locally popular partition is locally stable by definition, NP-hardness of both LOCAL-STABILITY-
 1842 EXISTENCE and LOCAL-POPULARITY-EXISTENCE follows. \square