Discrete preference games: social influence through coordination, and beyond*

Vincenzo Auletta† Ioannis Caragiannis‡ Diodato Ferraioli§ Clemente Galdi¶ Giuseppe Persiano∥

Abstract

We study discrete preference games that have been used to model issues such as the formation of opinions or the adoption of innovations in the context of a social network. In these games, the payoff of each agent depends on the agreement of her strategy to her internal belief and on its coordination with the strategies of her neighbors in the social network. Recent papers have obtained bounds on the price of anarchy and stability of these games.

Our starting point is the observation that such bounds can be very sensitive on the way the quality of equilibria is quantified. This motivates us to focus on properties of the games that do not depend on these details. In particular, we consider the question of whether the public opinion can be different than the public “belief”. We show that strategic behavior typically leads to changes of the public opinion compared to the public belief. We completely characterize the social networks (graphs) where such changes can happen and furthermore study the complexity of such transitions. We show that deciding whether a minority belief can become the majority opinion is NP-hard even when the initial number of supporters of this belief is very close to 1/4 of the social network size.

Next, motivated by the limited expressive power of discrete preference games, we define and study the novel class of generalized discrete preference games. These games have additional characteristics and can model social relations to allies or competitors, including complex relations between more than two agents, introduce different levels of strength for each relation, and personalize the dependence of each agent to its neighborhood. We show that these games admit generalized ordinal potential functions. More importantly, we show that every game with two strategies per agent that admits a generalized ordinal potential function is structurally equivalent to a generalized discrete preference game. This implies that the games in the novel class capture the full generality of two-strategy games in which the existence of (pure) equilibria is guaranteed by topological arguments.

1 Introduction

Much of the work in social sciences aims to understand how opinions are formed and expressed in a social context. A classical simple related model has been proposed by Friedkin and Johnsen [13] as a refinement to a previous model by DeGroot [10]. Its main assumption is that each individual has an internal belief and the opinion she eventually expresses is the result of a repeated averaging between her

---

*This work was partially supported by the COST Action IC1205 “Computational Social Choice”, by Italian MIUR under the PRIN 2010-2011 project ARS TechnoMedia – Algorithmics for Social Technological Networks, by the European Social Fund and Greek national funds through the research funding program Thales on “Algorithmic Game Theory”, and by the EU FET project MULTIPLEX 317532.

†Università degli Studi di Salerno
‡University of Patras & CTI “Diophantus”
§Sapienza Università di Roma
§Università di Napoli “Federico II”
∥Università degli Studi di Salerno
belief and the opinions expressed by other individuals with whom she has social relations. The standard model is continuous and assumes that all beliefs and outcomes are real numbers. For example, the recent work of Bindel et al. [4] assumes that beliefs and opinions belong to [0, 1]. They interpret the repeated averaging process as a best-response play in a naturally defined game that leads to an equilibrium.

An obvious refinement of this model is to consider discrete beliefs and opinions. For example, by restricting them to the two discrete values 0 and 1, we can model beliefs or opinions over an inherently yes-or-no issue. Indeed, such models have been considered recently by Ferraioli et al. [12] and Chierichetti et al. [9]. Clearly, the discrete nature of the preferences does not allow for averaging anymore and several nice properties of the opinion formation models mentioned above — such as the uniqueness of the outcome — are lost. In contrast, it now seems natural to assume that each agent is strategic and aims to pick the most beneficial strategy for her, given her internal belief and the strategies of her neighbors. This immediately defines a discrete preference game. With the notation that we will use throughout the paper, the cost of agent \( i \) when the binary strategies of the \( n \) agents are given by the vector \( x = (x_1, \ldots, x_n) \) is

\[
c_i(x) = \alpha \cdot |x_i - b_i| + (1 - \alpha) \cdot \sum_{j \in N(i)} |x_i - x_j|,
\]

where \( b_i \in \{0, 1\} \) denotes the belief of agent \( i \) and \( N(i) \) is her set of neighbors. Note that the cost has two components that depend on the distance of the agent’s strategy from her internal belief and from the strategies of her neighbors, respectively. The parameter \( \alpha \in [0, 1] \) adjusts the relative importance of the two terms. Intuitively, the term \( 1 - \alpha \) indicates the degree of coordination the agents seek.

Chierichetti et al. [9] have mostly focused on bounding the price of stability of discrete preference games. They assess the quality of states by using the total (or social) cost of the agents. Their findings include a tight bounds on the price of stability for discrete preference games as well as conditions (e.g., \( 0 \leq \alpha \leq 1/2 \) or \( \alpha = 2/3 \)) that imply that states of minimum social cost are always equilibria. These results come in contrast to the price of anarchy (in terms of the same social cost definition) that can be unbounded (this has also been observed independently by Ferraioli et al. [12]).

The starting point of this work is the observation that this type of analysis might be very sensitive to the social cost definition. Alternatively, we could define agents as payoff maximizers, using the definition

\[
u_i(x) = \alpha \cdot (1 - |x_i - b_i|) + (1 - \alpha) \cdot \sum_{j \in N(i)} (1 - |x_i - x_j|),
\]

for the payoff \( u_i(x) \) of agent \( i \). Essentially, the payoff is now defined by counting agreements (instead of disagreements) of an agent’s strategy to her internal belief and the strategies of her neighbors and, clearly, this equivalent definition does not affect the game in any way. Then, the social welfare of a state (i.e., the sum of agents’ payoffs) is another natural quality measure for a price of anarchy/stability type of analysis. Unfortunately, this type of analysis can lead to different conclusions. For example, the price of anarchy is at most 2 while the price of stability is at most 3/2 for high values of \( \alpha \) (for which a tight price of stability of 2 were given). These observations are presented in Section 3.

To escape from the sensitivity of the price of anarchy/stability, we look for a quality measure that is independent from the specific definition of payoff or costs. Specifically, we do not focus on social cost/welfare optimization, but on truthfulness, according to which the optimal state is the one where the opinion every agent expresses coincides with her belief. We believe that a comparison of the equilibria of a discrete preference game to this truthful profile provides meaningful information of whether strategic issues can affect the quality of outcomes. We refrain from defining another ratio to quantify this quality; instead, we just ask whether the public opinion at an equilibrium can be different than that in the truthful profile. We study this question focusing specifically on the case \( \alpha = 1/2 \), since it already showcases a conceptual difference with the price of stability results: indeed, even if the price of stability is equal to
in this case and, hence, equilibria can be optimal (in terms of the social cost/welfare), it turns out that they can be “far” from the truthful profile.

1.1 Overview of our contribution

We first consider the question of whether the belief supported by a minority can become the opinion supported by the majority after strategic play. We take the term “strategic play” more broadly than usual and allow for non-worse response deviations (instead of the more standard best-response ones). Our rationale behind this choice is that such a sequence of deviations could be triggered in the context of a carefully designed campaign that uses full knowledge of the structure of the social network. We have managed to characterize the set of social networks (graphs) that allow for the transition of a belief of the minority to the opinion of majority after strategic play. Interestingly, this set is extremely large and includes all graphs besides a small set of forbidden graphs. Proving this fact turned out to be a technically challenging task and is heavily based on properties of local optima of graph bisections. Our characterization result is constructive in the sense that, for the networks that allow for the transition, we are able to identify the corresponding initial set of belief assignments; this can be done by a polynomial-time local-search computation of a bisection with locally-minimal width. A high number of initial supporters of the minority belief seem to be necessary. Indeed, given a \( n \)-vertex social network, deciding whether there exists some assignment of 1-beliefs to less than \( 1/4 \) of the vertices that can become the majority opinion after non-worse-response play is an NP-hard problem. The main source of computational hardness seems to be the requirement to compute the belief assignment. If the initial belief assignment is given, computing the maximum or minimum number of 1- or 0-opinions that can be reached (and, hence, deciding whether majority of an opinion can be achieved) turns out to be an easy problem. These results are presented in Section 4.

Next, motivated by the apparent limitations of the discrete preference games, we significantly generalize them. In contrast to Chierichetti et al. [9] who generalized these games to more than two alternatives, we keep the restriction of the number of alternatives and instead consider richer relations between agents. For example, the facts that each agent treats her neighbors equally, she seek agreement to any of them, and all social relations include only two agents are obvious limitations of discrete preference games. In contrast, we would like to (1) model social relations to allies or competitors, including complex relations between more than two agents, (2) introduce different levels of strength for each relation, and (3) personalize the dependence of each agent to its neighborhood. Motivated by games that are inspired by maximum constraint satisfaction optimization problems (see, e.g., Bhalgat et al. [3]), we define the broader class of generalized discrete preference games that have the desired characteristics. In these games, the strategy of each agent is again a binary opinion. Social relations (or social constraints) are weighted boolean formulas over the strategies of subsets of agents. Then, the payoff of each agent depends on the total weight of the true formulas in which she participates; the particular dependence can be different between agents. We give the detailed definition of generalized discrete preference games in Section 5. In this work we present some structural results about these games. First, we kill many birds with one stone: we show that the sum of the weights of the formulas is a generalized ordinal potential function for such a game. Note that the same potential function may correspond to very different games and essentially describes the general structure of the Nash dynamics graph of each of them. In addition, depending on the details in the definition of the agents’ payoffs, the potential function can be proved to be exact, weighted, or ordinal potential. Second, and probably more importantly, we show that every game with two strategies per agent that admits a generalized ordinal potential is structurally equivalent to a generalized discrete preference game. This implies that generalized discrete preference games capture the full generality of two-strategy games in which the existence of pure equilibria is guaranteed by topological arguments. This result is similar in spirit to the equivalence proved by Monderer and Shapley [16] between potential and congestion games. However, as computer scientists, we find this
universal role of a game inspired from satisfiability particularly interesting!

1.2 Other related work

The paper that is most closely related to ours is the one by Chierichetti et al. [9] which we have discussed extensively above. Independently, Ferraioli et al. [12] have considered the same class of games, mostly focusing on the study of (noisy) best-response dynamics with respect to their convergence to equilibria or stable states. Studies on social networks consider several phenomena related to the spread of social influence such as information cascading, network effects, epidemics, and more. The book of Easley and Kleinberg [11] provides an excellent introduction to the theoretical treatment of such phenomena. From a different perspective, problems of this type have also been considered in the distributed computing literature, motivated by the need to control and restrict the influence of failures in distributed systems; e.g., see the survey by Peleg [17] and the references therein.

On the more technical side, best-response dynamics from truthful profiles have been considered before in the context of iterative voting, e.g., see [15] and [5]. In particular, closer to our current work is the paper of Brânzei et al. [5] who present bounds on the quality of equilibria that can be reached from a truthful profile using best-response play and different voting rules. Of course, the important conceptual difference is that there is no underlying network in their work. Assessing the quality of equilibria that are reached by strategic play starting from particular states has been explicitly considered in the context of facility location and multicast games (e.g., see [8] and [7]).

1.3 Roadmap

We begin with preliminary definitions in Section 2. Our contributions are presented in Sections 3, 4, and 5 as described in detail above. We conclude with open questions in Section 6. Due to lack of space, some proofs have been put in appendix.

2 Preliminaries

Discrete preference games are formally defined as follows. There are \( n \) agents; we use \([n] = \{1, 2, ..., n\}\) to denote their set. Each agent corresponds to a distinct vertex of a graph \( G = (V, E) \) that represents the social network, i.e., the network of social relations between the agents. Agent \( i \) has an (internal) belief \( b_i \in \{0, 1\} \) and her strategy set consists of the two preferences that she can declare, i.e., \( x_i \in \{0, 1\} \). A strategy profile (or, simply, a profile) is a vector of strategies, with one strategy per agent. We use bold symbols for profiles, i.e., \( \mathbf{x} = (x_1, \ldots, x_n) \) and the usual game-theoretic notation \((x_{-i}, s) = (x_1, \ldots, x_{i-1}, s, x_{i+1}, \ldots, x_n)\) for any \( i \in [n] \), any profile \( x \in S \) and \( s \in \{0, 1\} \). In particular, we will call the vector \( \mathbf{b} = (b_1, \ldots, b_n) \) of beliefs the truthful profile. Moreover, for any \( y \in \{0, 1\} \), we denote as \( \neg y \) the negation of \( y \), i.e., \( \neg y = 1 - y \).

At a profile \( \mathbf{x} \), the utility (or payoff) of agent \( i \) is denoted by \( u_i(\mathbf{x}) \) and is defined as in [2]. A profile is a pure Nash equilibrium (or, simply, an equilibrium) if \( u_i(\mathbf{x}) \geq u_i(\mathbf{x}_i, \mathbf{x}_{-i}) \) for every agent \( i \). As observed in previous work, discrete preference games always have equilibrium. Alternatively, we can define the agents of discrete preference games to be cost-minimizers, following the definition of Chierichetti et al. [9] and Ferraioli et al. [12] for the cost \( c_i(\mathbf{x}) \) defined in [1]. Clearly, the two definitions are equivalent in the sense that they define the same incentives for the agents.

Following previous work, we can evaluate the quality of a profile using the social cost defined as the total cost of the agents in the profile, i.e., \( SC(\mathbf{x}) = \sum_{i \in [n]} c_i(\mathbf{x}) \). Now, following the classical line of research that was initiated with the seminal work of Koutsoupias and Papadimitriou [14], we can define the price of anarchy (PoA) of the game as the maximum value of the ratio \( \frac{SC(\mathbf{x})}{SC(\mathbf{x}^*)} \) over all equilibria \( \mathbf{x} \) of the game, where \( \mathbf{x}^* \) denotes the profile that minimizes the social cost. Furthermore, we can define the
more optimistic price of stability (PoS), introduced by Anshelevich et al. [2], as the minimum value of the ratio \( \frac{SC(x)}{SW(x)} \) over all equilibria \( x \). Both notions have been proposed in order to assess the impact of selfish behavior on efficiency.

But, clearly, the above definitions of the price of anarchy and stability in terms of the social cost should not be considered as unique. Since discrete preference games have a natural (equivalent) definition with utility-maximizing agents, we can instead use the social welfare to assess the quality of profiles, without deviating significantly from the rationale behind the previous approaches. In particular, the social welfare of a profile \( x \) is simply the total utility of the agents, i.e., \( SW(x) = \sum_{i \in [n]} u_i(x) \). Then, the PoA/PoS of a game in terms of the social welfare is defined as the maximum/minimum value of the ratio \( \frac{SW(x)}{SW(x^*)} \) over all equilibria \( x \) of the game, where \( x^* \) now denotes the profile that maximizes the social welfare.

We conclude these preliminaries definitions with an overview of games admitting different types of potential functions. A game is called

- an exact potential game if there exists a function \( \Phi \) defined over the states of the game such that the following condition holds: for every two states \( (x_{-i}, s) \) and \( (x_{-i}, s') \) that differ in the strategy of agent \( i \),
  \[
  u_i(x_{-i}, s) - u_i(x_{-i}, s') = \Phi(x_{-i}, s) - \Phi(x_{-i}, s');
  \]
- a weighted potential game if there exists a function \( \Phi \) and weights \( (v_i)_{i \in [n]} \) so that the above condition becomes
  \[
  u_i(x_{-i}, s) - u_i(x_{-i}, s') = v_i \cdot (\Phi(x_{-i}, s) - \Phi(x_{-i}, s')) ;
  \]
- an ordinal potential game if there exists a function \( \Phi \) such that the condition becomes
  \[
  u_i(x_{-i}, s) - u_i(x_{-i}, s') > 0 \iff \Phi(x_{-i}, s) - \Phi(x_{-i}, s') > 0 ;
  \]
- a generalized ordinal potential game if there exists a function \( \Phi \) such that
  \[
  u_i(x_{-i}, s) - u_i(x_{-i}, s') > 0 \Rightarrow \Phi(x_{-i}, s) - \Phi(x_{-i}, s') > 0 .
  \]

Note that each class generalizes the previous one. It is a folklore result (e.g., see [16]) that any generalized ordinal potential game has at least one Nash equilibrium and it can be reached through utility-improving strategy updates of the agents. Specifically, any local maximum of the potential function \( \Phi \) corresponds to an equilibrium of the game. The opposite is true for ordinal potential games but not necessarily true for generalized ordinal potential games. In the following, we often consider strategy updates that strictly improve the utility of the deviating agent as well as ones that do not decrease it. We use the standard term best-response moves for the former and non-worse response moves for the latter.

### 3 Sensitivity of price of anarchy and stability

As shown by Ferraioli et al. [12] and Chierichetti et al. [9], the price of anarchy of discrete preference games in terms of the social cost can be unbounded. However, the claim does not hold when we consider the alternative definition of the price of anarchy in terms of the social welfare. It turns out that the price of anarchy is at most 2 in this case. Indeed, let \( x \) be an equilibrium profile and observe that, for every agent \( i \), \( u_i(x) \geq \frac{1}{2} (\alpha + (1 - \alpha) \cdot |N(i)|) \), otherwise \( i \) can improve her utility by adopting the strategy \( \tau_i \). The bound then follows, since in the optimal profile \( x^* \), for every agent \( i \), \( u_i(x^*) \leq \alpha + (1 - \alpha) \cdot |N(i)| \).

Similarly, Chierichetti et al. [9] claim that the price of stability for discrete preference games is bounded by 2 and that there exist instances achieving this bound for \( \alpha \) close to 1. Again, we can
show that this does not hold when we evaluate the price of stability in terms of the social welfare. Specifically, we prove that when $\alpha$ tends to 1, the price of stability is at most $3/2$. Indeed, let us set $C = \alpha n + 2(1 - \alpha)m$. Observe that

$$SW(x) = C - SC(x), \quad (3)$$

where $m$ is the number of edges of $G$. Let $x$ be the equilibrium that maximizes the social welfare and let $x^*$ be the optimal profile. Moreover let $c = SC(x)/SC(x^*)$. From the previous work, we know that $c \in [1, 2]$. Finally, consider the profile $x'$ that sets the preference of each vertex to its belief. If $\alpha \geq \frac{m}{m+n/6}$, then $SC(x^*) \leq 2(1 - \alpha)m \leq \frac{1}{4} (\alpha n + 2(1 - \alpha)m) \leq \frac{C}{4}$. Hence, it follows from (3) that

$$\text{PoS} = \frac{SW(x^*)}{SW(x)} = 1 + \frac{(c - 1) \cdot SC(x^*)}{C - c \cdot SC(x^*)} \leq 1 + \frac{c - 1}{4 - c} \leq \frac{3}{2}.$$  

Chierichetti et al. [9] also proved that the price of stability is 1 whenever $\alpha \leq 1/2$ or $\alpha = 2/3$. This result carries over to the alternative PoS definition.

4 Can minority become majority?

In this section, we consider issues related to the spread of social influence. Informally, we ask whether it is possible, starting from a belief assignment where 1 is supported by a minority, to sequentially convince agents to switch their strategy (without decreasing their payoff) so that an equilibrium in which at least half of the vertices have 1-preference is reached. More precisely, we say that an assignment of beliefs $b$ to the vertices of a graph $G$ is MBM (minority becomes majority) for $G$ if

1. the number of vertices with belief 1 in $b$ is a minority; i.e., $|\{x \in V : b_x = 1\}| < n/2$ and

2. there is a subverting sequence of non-worse response moves that starts from $b$ and converges to an equilibrium $b'$ in which the number of vertices with preference 1 is a (weak) majority; i.e., $|\{x \in V : b'_x = 1\}| \geq n/2$.

In this section we present the following results:

1. given a graph $G$ and a belief assignment $b$, it is possible to decide in polynomial time if $b$ is MBM for $G$. If this is the case then it is possible to compute in polynomial time a subverting sequence of moves for $b$ and $G$.

2. MBM belief assignments exist for all graphs $G$, except some classes of forbidden graphs (defined below). For any non-forbidden graph, an MBM belief assignment can be computed in polynomial time.

3. Finally we strengthen the above results, by asking whether the majority can be subverted starting from a belief assignment in which the number of vertices with belief 1 is significantly smaller than $n/2$. We prove that deciding whether a graph admits such a MBM belief assignment is NP-hard.

4.1 Deciding MBM belief assignments

We start with the following statement. The proof slightly modifies an algorithm that was used to bound the price of stability in [9].

**Theorem 4.1.** There is a polynomial time algorithm that, given a graph $G = (V, E)$ and a belief assignment $b$, decides whether $b$ is MBM for graph $G$ and, if it is, it outputs a subverting sequence of moves.
Proof. Consider the following two-phase algorithm that receives as input an $n$-vertex graph $G = (V, E)$ and a belief assignment $b$ with less than $n/2$ vertices $u$ with $b_u = 1$:

1. While there exists a vertex $v$ with $b_v = 0$ such that switching does not decrease its utility, set $v$’s preference to 1.
2. While there exists a vertex $v$ with $b'_v = 1$ such that switching strictly increases its utility, set $v$’s preference to 0; let $b'$ denote the profile at the end of this phase.
3. If preference 1 is a weak majority in $b'$, then return 1 and output the sequence of moves of the first and second phase; otherwise, return 0.

Clearly, the running time of the algorithm is polynomial in the size of the input graph, since each vertex updates its preference at most twice. The fact that $b'$ is an equilibrium is proved in [9, Lemma 3.3]. We next show that $b'$ is actually the equilibrium that maximizes the number of 1-preference vertices. We refer to moves that change the preference from $x$ to $x$ as $x$-to-$x$ moves.

Consider a sequence $\sigma$ of non-worse response moves. We will construct another sequence of moves $\sigma'$ from $b$ to $b'$ that contains at most two moves per agent and has the following properties. 0-to-1 moves are non-worse response moves and precede the 1-to-0 moves, which may or may not be non-worse response moves. To construct the sequence $\sigma'$, we repeatedly apply the next procedure to the sequence $\sigma$ until this is no longer possible. Pick an agent $i$ that performs a 1-to-0 move just before the 0-to-1 move of agent $i'$. If $i = i'$ (this is possible when applying the process repeatedly), simply remove both moves from the sequence. Otherwise, swap the moves in the sequence. The crucial observation is that 0-to-1 moves are shifted to the beginning of the sequence in this way. Such a move will still be a non-worse move if it survives the application of the swap process mentioned above (because if it was a non-worse response after a 1-to-0 move, it will also be non-worse response if it is executed before it). In contrast, the repeated application of the swap process shifts 1-to-0 moves to the end of the sequence. We can construct a sequence $\sigma''$ by keeping the 0-to-1 moves as in $\sigma'$ and by completing the sequence with additional best-response 1-to-0 moves until reaching an equilibrium. The crucial observation is that the new sequence cannot contain any other 1-to-0 move besides the ones in $\sigma'$ (otherwise, $b'$ would not be an equilibrium). Hence, if $\sigma$ has the maximum number of 1-preferences, so does $\sigma''$. The theorem follows by observing that $\sigma''$ has the form computed by the two-phase algorithm.

4.2 Characterizing the networks with MBM belief assignments

We say that an $n$-vertex graph $G$ is forbidden if one of the following conditions holds:

- $G$ is $K_n$ or its complement (i.e., $n$ isolated vertices);
- $n$ is even and $G$ consists of an isolated vertex and a clique $K_{n-1}$;
- $n$ is even and all vertices of $G$ have degree at least $n - 2$.

We then prove the following theorem.

Theorem 4.2. There exists a polynomial time algorithm that for any non-forbidden graph $G$ returns an MBM belief assignment $b$.

We prove the theorem for graphs $G$ with an even number $n$ of vertices. We then show how to extend it to odd $n$.

Before proceeding with the proof, let us give a few more definitions and fix notation. For subsets $A, B \subseteq V$, we denote by $W(A, B)$ the number of edges with one endpoint in $A$ and the other in $B$. For a singleton $\{u\}$, we will simply write $W(u, B)$ and $W(A, u)$. Thus $W(u, v) = 1$ if $(u, v)$ is an edge and 0
otherwise. A bisection of a graph $G = (V, E)$ with an even number $n$ of vertices is just a partition $(S, \overline{S})$ of $V$ into two disjoint sets $S$ and $\overline{S}$ each of size $n/2$. Thus, $\overline{S} = V \setminus S$. We call $W(S, \overline{S})$ the width of the bisection. We say that $(S, \overline{S})$ is locally minimal if the width cannot be reduced by swapping two vertices between $S$ and $\overline{S}$. That is, for every $u \in S$ and $v \in S$, $W(S, \overline{S}) \leq W(S \cup \{u\} \setminus \{v\}, \overline{S} \cup \{v\} \setminus \{u\})$.

The following lemma will be very useful.

**Lemma 4.3.** For any locally minimal bisection $(S, \overline{S})$ graph $G$ and for every $u \in S$ and $v \in \overline{S}$:

$$W(u, S) - W(u, \overline{S}) + W(v, \overline{S}) - W(v, S) + 2W(u, v) \geq 0.$$ 

**Proof.** Set $A = S \setminus \{u\}$ and $B = \overline{S} \setminus \{v\}$ and consider bisection $(S', \overline{S'})$ with obtained by swapping $u$ and $v$. That is, $S' = A \cup \{v\}$ and $\overline{S'} = B \cup \{u\}$. Note that

$$W(S', \overline{S'}) = W(A, B) + W(u, A) + W(v, B) + W(u, v)$$

and

$$W(S, \overline{S}) = W(A, B) + W(u, B) + W(v, A) + W(u, v).$$

Therefore, by local minimality of $(S, \overline{S})$,

$$0 \leq W(S', \overline{S'}) - W(S, \overline{S}) = W(u, A) + W(v, B) - W(u, B) - W(v, A)$$

$$= W(u, S) + W(v, \overline{S}) - W(u, \overline{S}) - W(v, S) + 2W(u, v).$$

We say that a graph $G$ is of type $\mathbf{T1}$ if it has a bisection $(S, \overline{S})$ such that, for all $x \in S$, $W(x, S) \geq W(x, \overline{S}) - 1$ and there exists at least one vertex $u \in S$ for which $W(u, S) \geq W(u, \overline{S}) + 1$. Instead, a graph $G$ is of type $\mathbf{T2}$ if it has a bisection $(S, \overline{S})$ such that, for all $x \in S$, $W(x, S) \geq W(x, \overline{S}) - 1$ and there exists at least one vertex $w \in \overline{S}$ for which $W(w, S) \leq W(w, \overline{S}) + 1$ and $w$ is adjacent to two non-adjacent vertices $u, v \in S$. The next two propositions prove that graphs of types $\mathbf{T1}$ and $\mathbf{T2}$ admit an MBM belief assignment.

**Proposition 4.4.** Let $G$ be a graph of type $\mathbf{T1}$ with an even number of vertices. Then, $G$ has an MBM belief assignment.

**Proof.** Let $(S, \overline{S})$ be a bisection that is a witness that $G$ is of type $\mathbf{T1}$ and let $u \in S$ be a vertex such that $W(u, S) \geq W(u, \overline{S}) + 1$. Consider the belief assignment $b$ obtained by setting $b_u = 0$, $b_x = 1$ for every $x \in S \setminus \{u\}$, and $b_x = 0$ for every $x \in \overline{S}$. The number of vertices with belief 1 in $b$ is $n/2 - 1$. Consider now vertex $u$. If $u$ sets its preference equal to its belief then its payoff is $W(u, S)$. If instead $u$ sets its preference to 1, then its payoff is $W(u, S) \geq W(u, \overline{S}) + 1$. The preference profile $b'$ after $u$ has switched to 1 is $b' = (1_S, 0_{\overline{S}})$ and it has a weak majority of 1.

We complete the proof by verifying that in $b'$ no vertex of $S$ has an incentive to switch to 0, and thus there is an equilibrium reachable from $b'$ in which then number of 1’s is a weak majority. This clearly holds for $u$. For $x \in S \setminus \{u\}$, we observe that, since it is playing its belief, its payoff is at least $W(x, S) + 1$; if $x$ switches to 0, its payoff is at most $W(x, \overline{S})$. Since $G$ is $\mathbf{T1}$ we have that $W(x, S) + 1 \geq W(x, \overline{S})$. 

**Proposition 4.5.** Let $G$ be a graph of type $\mathbf{T2}$ with an even number of vertices. Then, $G$ has an MBM belief assignment.

**Proof.** Let $(S, \overline{S})$ be a bisection that is a witness that $G$ is of type $\mathbf{T2}$ and $u, v \in S$ and $w \in \overline{S}$ be as in the definition of type $\mathbf{T2}$ above. Consider the belief assignment $b$ obtained by setting, $b_u = b_v = 0$, $b_w = 1$, $b_x = 1$ for every $x \in S \setminus \{u, v\}$ and $b_x = 0$ for every $x \in \overline{S} \setminus \{w\}$. The number of vertices with belief 1 in $b$ is $n/2 - 1$. Now observe that for vertices $u$ and $v$ switching to 1 does not decrease the payoff. Indeed, the payoff of $u$ in $b$ is $W(u, \overline{S}) - 1$ while the payoff obtained by switching to
1 is \( W(u, S) + 1 \geq W(u, \overline{S}) - 1 \). Similarly, for \( v \). Moreover, \( u \) and \( v \) are not adjacent and thus they do not influence each other’s payoff. The preference profile \( b' \) after \( u \) and \( v \) have switched is \( b' = (1_{S \cup \{w\}}, 0_{\overline{S} \setminus \{w\}}) \) and it has a majority of 1.

We complete the proof by verifying that in \( b' \) no vertex of \( S \cup \{w\} \) has an incentive to switch to 0, and thus there is an equilibrium reachable from \( b' \) in which then number of 1’s is at least a majority. This is clearly true for \( u \) and \( v \). For \( w \), we observe that, since 1 is its belief, its payoff is at least \( W(w, \overline{S}) + 1 \); by switching to 0 the payoff would be at most \( W(w, \overline{S}) \leq W(w, S) + 1 \). For \( x \in S \setminus \{u, v\} \), since its belief is 1, we have that the payoff is at least \( W(x, S) + 1 \). In contrast, by switching to 0, the payoff would be at most \( W(x, \overline{S}) \leq W(x, S) + 1 \).

The two propositions above show that it is possible to construct an MBM belief assignment for a graph of type \( T1 \) and \( T2 \) if a witness for the type is given. Next we describe an algorithm that, for any non-forbidden graph, finds such a witness. This shows that a non-forbidden graph is either of type \( T1 \) or \( T2 \).

For a bisection \((S, \overline{S})\) we denote by \( B(S, \overline{S}) \) the set containing \((S, \overline{S})\), \((\overline{S}, S)\) and for any bisection \((S', \overline{S'})\) obtained from it by swapping two vertices between \( S \) and \( \overline{S} \), it contains both \((S', \overline{S'})\) and \((\overline{S'}, S')\). Moreover, we say that bisection \((S, \overline{S})\) is locally 2-minimal if it is locally minimal and, furthermore, it minimizes the width among all the bisections obtained from \((S, \overline{S})\) by swapping two pairs of vertices; that is,

\[
W(S, \overline{S}) \leq W(S \cup \{u, w\} \setminus \{v, z\}, \overline{S} \cup \{v, z\} \setminus \{u, w\})
\]

for all \( u, w \in \overline{S} \) and \( v, z \in S \).

We are now ready to describe our algorithm. On input a non-forbidden graph \( G \), the algorithm starts by computing a locally 2-minimal bisection \((S^*, \overline{S^*})\) of \( G \). Then, for each bisection \((S, \overline{S}) \in B(S^*, \overline{S^*})\), the algorithm checks if either \((S, \overline{S}) \in B(S^*, \overline{S^*})\) is a witness that \( G \) is of type \( T1 \) or of type \( T2 \).

The running time of the algorithm is polynomial. A locally 2-minimal bisection can be computed in polynomial time on unweighted graphs \([19, 18]\) (see also \([1]\)) and the set \( B(S^*, \overline{S^*}) \) contains at most \( 2n^2 \) bisections each one of which can be checked in time linear in \( n \).

Finally, we prove that for any non-forbidden graph \( G \) there is a bisection in \( B(S^*, \overline{S^*}) \) that is a witness of \( G \) being of type \( T1 \) or of type \( T2 \). We first consider the case in which \( G \) has at least one isolated vertex.

**Lemma 4.6.** Let \( G \) be a graph with an even number of vertices and at least one isolated vertex. If the algorithm above does not find a witness that \( G \) is of type \( T1 \) then \( G \) consists of all isolated vertices or of a clique plus an isolated vertex (and thus \( G \) is forbidden).

For bisection \((S, \overline{S})\) of a graph \( G \), let \( B'(S, \overline{S}) \subseteq B(S, \overline{S}) \) be the subset of bisections \((S', \overline{S'}) \in B(S, \overline{S})\) such that at least one isolated vertex is in \( \overline{S} \). We actually prove a stronger statement: for any locally minimal bisection \((S, \overline{S})\) of a graph \( G \) if there is no witness that \( G \) is of type \( T1 \) in \( B'(S, \overline{S}) \), then \( G \) is forbidden. We stress that it is sufficient to restrict ourselves to witnesses \((S, \overline{S})\) in which \( \overline{S} \) contains at least one isolated vertex.

**Proof.** Fix a locally minimal bisection \((S, \overline{S})\) and let \( i \in \overline{S} \) be an isolated vertex. We start by proving that, for all \( x \in S \), \( W(x, S) = W(x, \overline{S}) \). Suppose by contradiction that there exists vertex \( x \in S \) for which \( W(u, S) \leq W(u, \overline{S}) - 1 \) and consider bisection \((S', \overline{S'}) \in B'(S, \overline{S})\), with \( S' = \overline{S} \cup \{u\} \setminus \{i\} \). Its width is \( W(S', \overline{S'}) = W(S, \overline{S}) + W(u, S) - W(u, \overline{S}) \leq W(S, \overline{S}) - 1 \), contradicting the local minimality of \((S, \overline{S})\). Hence, since \((S, \overline{S})\) is not a witness of \( G \) being of type \( T1 \), it must be the case that \( W(x, S) = W(x, \overline{S}) \) for every vertex \( x \in S \).

Next we prove that if for some vertex \( v \in \overline{S} \) we have \( W(v, \overline{S}) = W(v, S) - c \) for some integer \( c \geq 1 \), then \( c = 2 \) and \( v \) is connected to all vertices in \( S \). Indeed, Lemma 4.3 implies that \( c \in \{1, 2\} \) and,
furthermore, that $v$ is connected to every vertex $x \in S$ (i.e., $W(x, v) = 1$). Therefore, $W(v, S) = n/2$ while, since $v \in S$, $W(v, S) \leq n/2 - 2$ which leaves $c = 2$ as the only possibility.

Let $A$ denote the subset of $S$ consisting of all the vertices $x$ with $W(x, \overline{S}) = W(x, S) - 2$; all vertices $x \in S \setminus A$ have $W(x, S) \geq W(x, S)$. We show that if $A$ is not empty, then the vertices of $S$ form a clique. Assume otherwise and let $u, w \in S$ be two non-adjacent vertices of $S$. Pick a vertex $v \in A$ and consider bisection $(S', \overline{S'}) \in B'(S, \overline{S})$, with $S' = S \cup \{v\} \setminus \{u\}$. Recall that $v \in A$ is connected to every vertex $x \in S$. Thus for every vertex $x \in S \setminus \{u, w\}$, the number of neighbor in $S'$ will be at least the number of neighbors in $S$, i.e., $W(x, S') - W(x, \overline{S'}) \geq W(x, S) - W(x, \overline{S}) = 0$. The vertex $w$ is connected to $v$ but not to $u$ and thus $W(w, S') - W(w, \overline{S'}) \geq W(w, S) - W(w, \overline{S}) + 2 = 2$.

But then, the bisection $(S', \overline{S'})$ would be a witness that $G$ is of type $T1$, that is a contradiction.

So, assuming that the set $A$ is not empty, the vertices of $S$ form a clique. We next show that this implies that $G$ is a clique plus an isolated vertex and thus forbidden. Indeed, since $W(x, S) = W(x, \overline{S})$ for every $x \in S$ and $|\overline{S} \setminus \{i\}| = |S| - 1$, we have that every vertex of $S$ is connected to every vertex of $\overline{S}$, except the isolated one. Such a high width for a locally minimal bisection implies that the graph is a clique plus an isolated vertex (and actually $A = \overline{S}$).

If instead $A$ is empty, we claim that $W(x, S) = W(x, \overline{S})$ for every vertex $x \in S$. Indeed, suppose by contradiction that $W(x, \overline{S}) > W(x, S)$ for some $x \in S$ and let $u$ be any vertex of $S$. Then, the bisection $(S', \overline{S'}) \in B'(S, \overline{S})$, with $S' = \overline{S} \cup \{u\} \setminus \{i\}$ would be a witness that $G$ is of type $T1$, that is a contradiction.

To conclude the proof, we will show that if $A$ is empty, then $G$ must consist of $n$ isolated vertices. Assume otherwise that $G$ has some edge; then, since $(S, \overline{S})$ bisects the neighborhood of each vertex, there must be an edge $(u, v)$ between vertices $u \in S$ and $v \in \overline{S}$. Then, the bisection $(S', \overline{S'}) \in B'(S, \overline{S})$, with $S' = \overline{S} \cup \{u\} \setminus \{i\}$, is a witness that $G$ is of type $T1$, that is a contradiction.

This completes the proof of Theorem 4.2 for all graphs with an even number of vertices and one isolated vertex. We will now complete the proof of Theorem 4.2 for all graphs with an even number of vertices by proving the following lemma.

**Lemma 4.7.** Let $G$ be a graph with an even number of vertices and no isolated vertex. If the algorithm above does not find a witness that $G$ is of type $T1$ or of type $T2$, then all vertices of $G$ have degree at least $n - 2$ (and thus $G$ is forbidden).

**Proof.** We first show that if there is a locally minimal (not necessarily 2-minimal) bisection $(S, \overline{S})$ of $G$ that is not a witness for $G$ being of type $T1$ or of type $T2$, then $W(x, S) = W(x, \overline{S})$ of every $x \in V$. Assume for sake of contradiction that, for some vertex $u \in S$, $W(u, S) < W(u, \overline{S}) - 1$. Then, $u$ must be connected to every vertex of $\overline{S}$. Indeed, if this was not the case, let $v \in \overline{S}$ be a vertex that is not adjacent to $u$. Lemma 4.3 implies that $W(v, \overline{S}) \geq W(v, S) + 1$ and $W(x, \overline{S}) \geq W(x, S) - 1$ for all vertices $x \in \overline{S} \setminus \{v\}$. Hence, $(\overline{S}, S)$ is a witness that $G$ is of type $T1$, that is a contradiction.

Next observe that if $\overline{S}$ contains two non-adjacent vertices $v$ and $w$, then $(\overline{S}, S)$ is a witness that $G$ is of type $T2$, that is a contradiction. So, the vertices of $\overline{S}$ form a clique (together with vertex $w$). Hence, $W(x, \overline{S}) = n/2 - 1$ for every $x \in \overline{S}$. Since the bisection $(\overline{S}, S)$ is not a witness that $G$ is of type $T1$, it must be also that $n/2 - 1 = W(x, \overline{S}) \leq W(x, S)$ for every $x \in S$. This implies that $W(S, \overline{S}) = \frac{n}{2} (\frac{n}{2} - 1)$. Such a high width for a locally minimal bisection implies that the graph is a clique and, hence, it is forbidden.

So, we have that $W(x, S) \geq W(x, \overline{S})$ for every $x \in S$. If, in addition, we had $W(u, S) \geq W(u, \overline{S})$ for some $u \in S$, then $(S, \overline{S})$ would be a witness for $G$ being of type $T1$, that is a contradiction. Hence, $W(x, S) = W(x, \overline{S})$ for any vertex $x \in S$. The same argument can be used to show the same for all vertices of $\overline{S}$.

Let now $(S^*, \overline{S}^*)$ be a locally 2-minimal bisection. We now show that any pair of non-adjacent vertices $u \in S^*$ and $v \in \overline{S}^*$ have the same neighborhood; that is, $N(u) = N(v)$. Note that such
a pair of vertices certainly exists: indeed, $|S^*| = |\overline{S^*}|$ and $W(x, S^*) = W(x, \overline{S^*})$ for every vertex $x$, since $(S^*, \overline{S^*})$ is locally minimal and, by hypothesis, not a witness for $G$ being of type $T1$ or of type $T2$. Now let $S' = S \cup \{v\} \setminus \{u\}$ and consider the bisection $(S', \overline{S'}) \in B(S^*, \overline{S^*})$. Observe that $W(S', \overline{S'}) = W(S^*, \overline{S^*})$. Hence and since $(S^*, \overline{S^*})$ is locally 2-minimal, it follows that $(S', \overline{S'})$ is a locally minimal bisection. Moreover, by hypothesis, this bisection is not a witness of $G$ being of type $T1$ or of type $T2$. Hence, it must be that $W(x, S') = W(x, \overline{S'})$ for every vertex $x$. But this implies that every vertex that is adjacent to $u$ is also adjacent to $v$ and vice versa, i.e., $N(u) = N(v)$.

Next we show that $u$ must be adjacent to every vertex in $S^*$ and, symmetrically, that $v$ must be adjacent to every vertex in $\overline{S^*}$. Assume for sake of contradiction that some vertex $w \in S^*$ exists that is not adjacent to $u$. Then, $w$ is not be adjacent to $v$ either and, by repeating the same argument as above, we conclude that $N(w) = N(v)$ and, consequently, $N(w) = N(u)$. Now, pick a vertex $u' \in N(w) \cap \overline{S^*}$ (such a vertex exists since $w$ is not isolated) and observe that it is adjacent to the non-adjacent vertices $u$ and $w$ in $S^*$. This triplet shows that $(S^*, \overline{S^*})$ is then a witness that $G$ is of type $T2$, that is a contradiction.

Finally, we claim that the vertices of $S^*$ (and, symmetrically, the vertices of $\overline{S^*}$) form a clique. Indeed, assume that two vertices in $N(w) \cap S^*$ are not adjacent. Then, since $v$ is adjacent to each vertex in $S^*$, they belongs to the neighborhood of $v$ and this triplet of vertices shows that $(S^*, \overline{S^*})$ is a witness that $G$ is of type $T2$, that is a contradiction.

In conclusion, since $W(x, S^*) = W(x, \overline{S^*})$ for every vertex $x$, we have that all vertices have degree $n - 2$ and, hence $G$ is forbidden. 

Let us now to consider the case of graphs $G$ with an odd number of vertices. If $G$ is non-forbidden then the graph $G'$ obtained by adding one isolated vertex $i$ to $G$ is non-forbidden, has an even number of vertices and at least one isolated vertex. From Lemma 4.6, it follows that $G'$ is of type $T1$ and we can find a witness $(S, \overline{S})$ with $i \in S$. Then, it is immediate to see that the MBM belief assignment $\mathbf{b}$ described in Proposition 4.4 when restricted to the vertices of $G$, gives an MBM belief assignment for $G$.

Remark 4.1. Theorem 4.2 is tight as it is immediate to see that forbidden graphs do not admit MBM belief assignments.

4.3 Starting from a belief with few supporters

Theorem 4.2 shows that it is always possible to find an MBM belief assignment for a non-forbidden graph. By inspecting the proof, we see that most of the MBM belief assignments constructed actually have a number of $1$-beliefs that is very close to $n/2$. A much stronger result would be the characterization of graphs in which it is possible to subvert the majority by starting from a weaker minority. The next statement shows that deciding whether such a minority exists is a computationally hard problem.

Theorem 4.8. For every constant $0 < \varepsilon \leq \frac{1}{5}$, given a graph $G$ with $n$ vertices, it is NP-hard to decide whether there exists an assignment of beliefs in which the number of vertices with belief $1$ is at most $n \left(\frac{1}{2} - \varepsilon\right)$ and from which a sequence of non-worse response moves can converge to a profile in which the number of vertices with preference $1$ is at least $n/2$.

Proof. We will use a reduction from the NP-hard problem 2P2N-3SAT, the problem of deciding whether a 3SAT formula in which every variable appears as positive in two clauses and as negative in two clauses has a truthful assignment or not (the NP-hardness follows by the results of [20]).

Given an instance of 2P2N-3SAT, i.e., a Boolean formula $\phi$ with $C$ clauses and $V$ variables (with $3C = 4V$; observe that $C$ is a multiple of 4), we will construct a graph $G(\phi)$ with $n$ vertices such that there exists an assignment of $1$-beliefs to a minority of at most $n \left(\frac{1}{2} - \varepsilon\right)$ vertices of $G(\phi)$ such that a non-worse response play leads to a profile with $n/2$ $1$-preferences if and only if $\phi$ has a truthful assignment.

The reduction is as follows:
For every variable $x$ of $\phi$, the reduction constructs a \textit{variable gadget} consisting of 35 vertices and 64 edges (see Figure 1).

The vertices of the gadget for variable $x$ are the two \textit{literal vertices}, $x$ and $\overline{x}$, vertices $v_1(x), \ldots, v_8(x)$, vertices $v_1(\overline{x}), \ldots, v_8(\overline{x})$, vertices $v_0(x)$ and $w_0(x)$, and vertices $w_1(x), \ldots, w_{15}(x)$. The edges are $(x, v_i(x))$ and $(\overline{x}, v_i(\overline{x}))$ for $i = 1, \ldots, 8$, $(v_i(x), v_{i+1}(x))$ and $(v_i(\overline{x}), v_{i+1}(\overline{x}))$ for $i = 1, \ldots, 7$, $(v_0(x), v_8(x))$, $(v_0(x), v_8(\overline{x}))$, $(v_0(x), w_0(x))$ $(w_0(x), v_i(x))$ and $(w_0(x), v_i(\overline{x}))$ for $i = 1, \ldots, 8$ and $(w_0(x), w_i(x))$ for $i = 1, \ldots, 15$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{variable_gadget}
\caption{The variable gadget.}
\end{figure}

For every clause $c$ of $\phi$, the reduction constructs a \textit{clause gadget} consisting of 35 vertices and 127 edges (see Figure 2).

The vertices of the gadget are the \textit{clause vertex} $c$, vertices $u_1(c), u_2(c)$ and $u_3(c)$ and vertices $v_1(c), \ldots, v_{31}(c)$. The 127 edges are $(c, u_i(x))$ for $i = 1, 2, 3, (u_i(c), v_j(c))$ with $i = 1, 2, 3$ and $j = 1, \ldots, 31$, while vertices $v_1(c), \ldots, v_{31}(c)$ are connected in a cycle.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{clause_gadget}
\caption{The clause gadget.}
\end{figure}

- Each clause vertex from the gadget corresponding to clause $c$ is connected to the literal vertices corresponding to the literals that appear in clause $c$ in $\phi$.
- There is a clique of odd size $N \geq \frac{91}{4e}C - \frac{109}{2}C - \frac{1}{4e}$ disconnected from the rest of the graph.
- There are $\frac{191C}{4} + N$ additional isolated vertices in $G(\phi)$.

Overall, the total number of vertices in $G(\phi)$ is $n = 35V + 35C + \frac{191C}{4} + 2N = 109C + 2N$.

A belief assignment is called \textit{proper} if:
- for every variable $x$, it assigns belief 1 to vertex $w_0(x)$ and to exactly one literal vertex of the gadget of $x$;
- for every clause $c$, it assigns belief 1 to vertices $u_1(c), u_2(c)$, and $u_3(c)$;
it assigns belief 1 to exactly $\frac{N+1}{2}$ vertices of the clique;

- it assigns belief 0 to all the remaining vertices.

Hence, in a proper assignment the number of vertices with belief 1 is $2V + 3C + \frac{N+1}{2} = \frac{9C}{2} + \frac{N+1}{2} \leq n\left(\frac{1}{4} - c\right)$.

We have the following claim.

**Claim 4.9.** The Boolean formula $\phi$ is satisfiable if and only if there exists a proper assignment of 1-beliefs to $\frac{9C}{2} + \frac{N+1}{2}$ vertices of $G(\phi)$ so that an equilibrium in which the number of 1-preference vertices is at least $n/2$ can be reached by non-worse response play.

**Proof.** We first prove that starting from a proper belief assignment, there is a sequence of non-worse response moves that leads to a profile in which 26 vertices of every variable gadget have preference 1. To see this consider a proper assignment that assigns belief 1 to $x$ (and to $w_0(x)$) and the following sequence of moves: vertex $v_1(x)$ switches from 0 to 1; then, for $i = 1, \ldots, 7$, vertex $v_{i+1}(x)$ switches to 1 immediately after vertex $v_i(x)$; vertex $v_0(x)$ switch to 1 after vertex $v_8(x)$; finally, $w_1(x), \ldots, w_{15}(x)$ can switch in any order. A similar sequence can be constructed for a proper assignment that assign belief 1 to vertex $\pi$ (and $w_0(x)$) of the gadget for variable $x$. Intuitively, the two assignments simulate the assignment of values 1 and 0 to variable $x$, respectively.

Next we observe that, starting from a proper belief assignment, there exist a sequence of non-worse response moves by which 34 vertices of every clause gadget associated to clause $c$ have preference 1; namely, the vertices $u_1(c), u_2(c), u_3(c)$ and vertices $v_1(c), \ldots, v_{31}(c)$.

Finally, observe that any clique vertex switches to 1. Thus with a proper assignment the non-worse response play will end up with preference 1 in 26 vertices of each variable gadget, with exactly one of the literal vertices with preference 1; preference 1 in the 34 non-clause vertices of every clause gadget, and preference 1 in the $N$ clique vertices. Moreover, observe that the clause vertex $c$ of the gadget of clause $c$ switches to 1 if and only if at least one of the literal vertices corresponding to literals that appear in $c$ have preference 1. Hence, the fact that a clause vertex has preference 1 (respectively, 0) corresponds to the clause being satisfied (respectively, not satisfied) by the Boolean assignment induced by the proper assignment of beliefs. Then, the non-worse response play will lead to an additional number of $C$ clause vertices adopting preference 1 if and only if $\phi$ is satisfiable.

In conclusion, we have that if $\phi$ is satisfiable there is a sequence of non-worse response moves converging to a profile with $26V + 34C + N + C = 109C/2 + N = n/2$ vertices with preference 1. Otherwise, if $\phi$ is not satisfiable, any sequence of non-worse response moves converges to a profile with strictly less than $n/2$ vertices with preference 1. \hfill $\Box$

To complete the proof, it remains to show the following.

**Claim 4.10.** For non-proper assignments that assign belief 1 to at most $\frac{9C}{2} + \frac{N+1}{2}$ vertices, there is no sequence of non-worse response moves converging to a profile in which the number of 1-preference vertices is at least $n/2$.

**Proof.** First observe that if there are less then $\frac{N+1}{2}$ vertices with belief 1 in the clique, then it is not possible to make a 0-belief vertex in the clique switch to a 1-preference. Thus, in this case, any sequence of non-worse response moves will converge to an equilibrium with at most $109C + \frac{N+1}{2} < \frac{n}{2}$ vertices with preference 1.

Let us now to focus on the assignment of the remaining $9C/2$ 1-beliefs. Suppose that there is a non-proper assignment of these beliefs such that a sequence of non-worse response may lead to at least $n/2$ vertices with preference 1.

We first show that this assignment will assign at most two vertices with belief 1 to each variable gadget and at most three vertices with belief 1 to each clause gadget. Indeed, suppose that there is a
variable gadget with more than two vertices with belief 1. Note that, a sequence of non-worse response moves from such an assignment may lead to an equilibrium in which every vertex in the gadget has a 1-preference. As a consequence, it may be the case that two more clause vertices adopt a 1-preference. That is, this assignment increases the number of vertices with preference 1 at the equilibrium by at most 11 with respect to a proper assignment. Similarly, you can check that if a clause gadget has more than three vertices with belief 1, then this increases the number of vertices with preference 1 at the equilibrium with respect to a proper assignment by at most 1. In any way, in order to have more than two vertices with belief 1 in a variable gadget or more than three vertices with belief 1 in a clause gadget, there will be either another variable gadget with at most one vertex with belief 1 or another clause gadget with at most two vertices with belief 1. In both cases, any sequence of non-worse response moves will lead to at most one vertex with preference 1 in that gadget, that is either 25 or 33 vertices less than what happens with a proper assignment.

Suppose now that there is a variable gadget with less than two of these vertices. Thus, any non-worse response leads to an equilibrium in which at most one non-literal vertex in this gadget has preference 1. On the other side, the remaining 1-belief must be assigned either to a clique vertex or to an isolated vertex and, hence, it leads to at most one vertex in the equilibrium more than what happens with a proper assignment. A similar argument works also for proving that that no clause gadget with less than three vertices exists.

Then, it must be the case that in the non-proper assignment for each variable gadget there are two vertices with belief 1 and in each clause gadget there are three vertices with belief 1. However, observe that any non-proper assignment of belief 1 to the vertices of a variable gadget leads to at most 18 1-preferences vertices after non-worse response play (with the upper-bound achieved by assigning belief 1 to vertex $w_0(x)$ and either to vertex $va(x)$ or to the vertex $va(x)$). Note that any of these non-proper assignments cannot lead some literal vertex to having preference 1 at the equilibrium and, thus, it cannot increase the number of clause vertices with preference 1. Similarly, it is easy to see that for any non-proper assignment of belief 1 to a clause gadget every sequence of non-worse response leads to an equilibrium, in which at most 1 vertex has preference 1 (i.e, the clause vertex only if it has belief 1, and each of the literals at which it is connected has preference 1 at the equilibrium).

Hence, any non-proper assignment leads to an equilibrium in which the number of vertices with preference 1 is strictly less then in the case of a proper assignment, completing in this way the proof.

5 Generalized discrete preference games

In this section we introduce and study the generalized discrete preference games in which we allow more complex social relationships among agents and more complex utility (or cost) functions than the ones considered in discrete preference games of Section 2.

We still assume that every agent $i$ has a private belief $b_i \in \{0, 1\}$ and that her strategy set consists of preferences $x_i \in \{0, 1\}$. However, differently from what we discussed in Section 2 now agents are not interested in just agreeing with their neighbors, but they take into account more complex social constraints. Specifically, we define a social constraint as a Boolean formula involving the preferences (but not the beliefs) of a subset of agents. For example, social constraint $C(x_1) = (x_1 \land x_2) \lor (\neg x_1 \land x_3)$ involves agents $i$ and $j$ and is satisfied when $i$ and $j$ have different preferences. We denote by $A(C)$ the set of agents involved in the constraint $C$. In addition to social constraints we also consider belief constraints. The belief constraint $B_i(x)$ for agent $i$ is $B_i(x) = (x_i \land \pi_i) \lor (\neg x_i \land \bar{b}_i)$ and, unlike social constraints, involves solely $i$’s preference $x_i$ and $i$’s belief. Sometimes, we will use the equivalent constraint $B_i(x) = \pi_i(B_i(x) = x_i, \text{resp.})$ if $b_i = 0$ ($b_i = 1$, resp.). A set $C$ of constraints is feasible if for every agent $i$ at most one of the two belief constraints $x_i$ and $\pi_i$ belongs to $C$. 14
In a generalized discrete preference game, a constraint $C$ has weight $W(C) > 0$ and the utility $u_i(x)$ of an agent $i$ in profile $x$ is defined through a monotone non-decreasing function $F_i : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$. More precisely, we have

$$u_i(x) = F_i(w_i(x)),$$

where $w_i(x)$ is the sum of the weights of the constraints that involve agent $i$ and that are satisfied at $x$.

In sums, a generalized discrete preference game is a the tuple $([n], C, W, (F_i)_{i \in [n]})$, where $C$ is a feasible set of constraints, $W$ is the weight function and $F_i$ is the function that determines the utility of $i$. Note that it is possible to define the satisfaction of agents in terms of costs (by considering the sum of the weights of the unsatisfied constraints) without any harm to the results given in this section.

Note that the function $F_i$ gets as input the cumulative weight of all satisfied constraints that involve $i$. That is, an agent cares about all constraints she is involved in and cannot disregard a constraint.

### 5.1 Examples

Let us give some examples highlighting the expressiveness of the generalized discrete preference games described above.

Let us start by showing some notable examples of constraints that are allowed by our definition. A constraint $C$ is an equality-constraint if it is satisfied only when all agents involved in $C$ have the same preference. Formally, $C$ is an equality-constraints on $N \subseteq [n]$ if $C(x) = (\bigwedge_{i \in N} x_i) \lor (\bigwedge_{i \in N} \bar{x}_i)$. This allows to easily model a group of agents that want to agree with each other. A constraint $C$ is an or-constraint if it is satisfied in all profiles except the ones in which the agents involved in $C$ have a specific configuration. Formally, $C$ is an or-constraint on $N$ if $C(x) = \bigvee_{i \in N} \ell_i$, where $\ell_i \in \{x_i, \bar{x}_i\}$. This allows to model the natural requirement according to which the agents wants only to avoid a specific configuration.

Note that we require that the weight of a constraint is the same for all agents involved in the constraint. This symmetry is motivated by the aim to represent the social behavior of the components of a group motivated by similar interests and with similar relationships within the group and outside of it. The number of constraints in which an agent is involved can be thus seen as equivalent to the number of groups at which the agent participates.

We also introduce another class of constraints, that, even if it is less natural than previous ones, it will turn out to be sometimes useful for characterizing the behavior of generalized discrete preference games. Specifically, a constraint $C$ is a switch-constraint if every time an agent that is involved in the constraint switches her preference then the constraint switches between satisfied and unsatisfied. Formally, $C$ is a switch-constraint on set $N$ of players if $C(x) = \bigoplus_{i \in N} \ell_i$, where $\ell_i \in \{x_i, \bar{x}_i\}$.

Also, we observe that very natural utility functions can be modeled through specific choices for functions $F_i$. For example, the utility function according to which an agent tries to maximize the sum of the weights of the satisfied constraints at which she is interested, can be modeled by the identity function; that is, $F_i(r) = r$ for each $r \geq 0$.

In another natural example, the utility function assigns the belief constraint an agent-specific weight that is different from the one assigned to social constraints. For example, consider the following utility function

$$u_i(x) = \alpha B_i(x) + (1 - \alpha) \frac{n_i(x)}{d_i},$$

where $d_i$ is the degree of $i$ (i.e., the number of constraints in which $i$ is involved) and $n_i(x)$ is the number of constraints in which $i$ is involved and that are satisfied at $x$. This utility function then balances among the closeness to agent’s own belief and the fraction of satisfied constraints in which she is involved. It turns out that, even if in our definition the constraints weight are the same for each agent, it is possible to implement such an utility functions by means of a weighted identity function $F_i$; i.e., $F_i(r) = \omega_i r$ for
any $r \geq 0$, where $\omega_i > 0$. Indeed, we can set $W(B_i) = \alpha d_i$ for each $i \in [n]$, $W(C) = 1 - \alpha$ for each social constraint $C \in \mathcal{C}$ and $\omega_i = \frac{1}{d_i}$ for each $i \in [n]$.

Another natural class of utility functions is the one according to which agents are happy if and only if the sum of weights of satisfied constraints is more than the sum of weights of unsatisfied constraints. This can be achieved by using the majority function, i.e., $F_i(r) = 1$ if $r > \frac{1}{2} \sum_{C \in \mathcal{C} : i \in C} W(C)$ and $F_i(r) = 0$ otherwise. Similarly, we can consider threshold functions, according to which the agents are happy if and only if the sum of weights of satisfied constraints is more than some threshold $T$; i.e., $F_i(r) = 1$ if $r > T$ and $F_i(r) = 0$ otherwise.

We observe that the discrete preference games described in Section 2 simply use equality constraints defined on pairs of agents and identity utility functions.

### 5.2 Existence of equilibria

The main result of this section is a proof that generalized discrete preference games are generalized ordinal potential games and thus they always admit a pure Nash equilibrium.

Before proceeding, let us fix notation. We recall that for an agent $i$ and for a profile $x$, $w_i(x)$ is the sum of the weights of all satisfied constraints at $x$ that involve agent $i$. Furthermore, for an agent $i$, we denote by $w_i$ the sum of the weights of all constraints that involve agent $i$ and, for a profile $x$, we denote by $w(x)$ the sum of all constraints satisfied at $x$. Finally, we let $w$ denote the sum of the weights of all constraints.

**Theorem 5.1.** Let $G = ([n], \mathcal{C}, W, (F_i)_{i \in [n]})$ be a generalized discrete preference game. Then, the function $w(x)$ is a generalized ordinal potential function for $G$.

**Proof.** Fix agent $i$ and profile $x$. Assume without loss of generality that $u_i(x_{-i}, 0) - u_i(x_{-i}, 1) > 0$. Therefore we have

$$F_i(w_i(x_{-i}, 0)) - F_i(w_i(x_{-i}, 1)) > 0$$

and, since $F_i$ is monotone non-decreasing,

$$w_i(x_{-i}, 0) - w_i(x_{-i}, 1) > 0.$$

Now observe that if $i$ changes its preference all the constraints not involving $i$ are not affected and,

$$w(x_{-i}, 0) - w_i(x_{-i}, 0) = w(x_{-i}, 1) - w_i(x_{-i}, 1) \quad (4)$$

The theorem then follows. \hfill $\square$

Interestingly, the function $w(x)$ gives a stronger characterization of generalized discrete preference games for special classes of utility functions or of social constraints. Specifically, we have the following proposition.

**Proposition 5.2.** Let $G = ([n], \mathcal{C}, W, (F_i)_{i \in [n]})$ be a generalized discrete preference game. Then, the function $w(x)$ is

- an exact potential for $G$ if the $F_i$ are identity functions;
- a weighted potential for $G$ if the $F_i$ are weighted identity functions;
- an ordinal potential for $G$ if every constraint $C \in \mathcal{C}$ is a switch-constraint and the $F_i$ are majority functions.
Therefore we obtain that $u_\omega$ where the last equality follows from (4). The claim for identity functions follows by taking $\omega_i = 1$.

Let us now consider switch constraints and majority functions. Since we have already proved that $w$ is a generalized ordinal potential function, all it is left to prove is that if $w(x_{-i}, 1) - w(x_{-i}, 0) > 0$ then $u_i(x_{-i}, 1) - u_i(x_{-i}, 0) > 0$.

Observe that for switch constraints we have that, for every agent $i$ and every profile $x$,

$$w_i(x_{-i}, 0) + w_i(x_{-i}, 1) = u_i.$$

In addition, $w(x_{-i}, 1) - w(x_{-i}, 0) > 0$ and Eq. (4) imply

$$w_i(x_{-i}, 1) - w_i(x_{-i}, 0) > 0.$$

Therefore we obtain that

$$w_i(x_{-i}, 1) > \frac{u_i}{2} \quad \text{and} \quad w_i(x_{-i}, 0) < \frac{u_i}{2}.$$

Since $F_i$ is the majority function, it follows that $u_i(x_{-i}, 1) - u_i(x_{-i}, 0) > 0$, as desired.

Proof. Fix agent $i$ and profile $x$.

If the $F_i$ are weighted identity functions (that is, $F_i(w_i(x)) = \omega_i \cdot w_i(x)$) we have

$$u_i(x_{-i}, 0) - u_i(x_{-i}, 1) = \omega_i \cdot w_i(x_{-i}, 0) - \omega_i \cdot w_i(x_{-i}, 1) = \omega_i \cdot (w(x_{-i}, 0) - w(x_{-i}, 1)),$$

where the last equality follows from (4). The claim for identity functions follows by taking $\omega_i = 1$.

We observe that Proposition 5.2 is tight, that is the conditions for the game to have an ordinal potential function are both needed. Indeed, first let us show that if $F_i$ are threshold functions, then we can have a game that has not an ordinal potential function.

Consider two agents, the switch constraints $C_1 = x_1 \oplus x_2$ and $C_2 = x_1 \oplus x_2$ and the belief constraints $B_1 = x_1$ and $B_2 = x_2$ with weights $W(C_1) = 1$, $W(C_2) = 2$, $W(B_1) = 1$ and $W(B_2) = 2$. Finally, take $F_1$ using a threshold $T_1 = 3$ and $F_2$ using a threshold $T_2 = 1$. Now suppose that an ordinal potential $\Phi$ exists for this game. It is easy to see that $u_1(0, 0) = u_1(1, 0)$ and, hence, $\Phi(0, 0) = \Phi(1, 0)$. On the other side, it is easy to see that $u_2(0, 0) < u_2(0, 1)$, $u_1(0, 1) = u_1(1, 1)$ and $u_2(1, 1) = u_2(1, 0)$ and hence $\Phi(0, 0) < \Phi(0, 1) = \Phi(1, 1) = \Phi(1, 0)$, a contradiction.

Finally, we show that the same happens if there is a non-switch constraint. Consider two agents, the switch constraint $C_1 = x_1 \oplus x_2$, the or-constraint $C_2 = x_1 \lor x_2$ and the belief constraints $B_1 = x_1$ and $B_2 = x_2$ with weights $W(C_1) = 1$, $W(C_2) = 4$, $W(B_1) = 4$ and $W(B_2) = 2$. Now suppose that an ordinal potential $\Phi$ exists for this game. It is easy to see that $u_1(0, 0) = u_1(1, 0)$ and, hence, $\Phi(0, 0) = \Phi(1, 0)$. On the other side, it is easy to see that $u_2(0, 0) = u_2(0, 1)$, $u_1(0, 1) < u_1(1, 1)$ and $u_2(1, 1) < u_2(1, 0)$ and hence $\Phi(0, 0) = \Phi(0, 1) < \Phi(1, 1) < \Phi(1, 0)$, a contradiction.

5.3 Better-response equivalence

Definition 5.1. Games $\mathcal{G} = ([n], (S_i)_{i \in [n]}, (u_i)_{i \in [n]})$ and $\mathcal{G}' = ([n], (S'_i)_{i \in [n]}, (u'_i)_{i \in [n]})$ with the same set of profiles $S$ are better-response equivalent if for every pair of profiles $x, y \in S$ that differ in the strategy of one agent, say $i$, we have that $u_i(x) > u_i(y)$ if and only if $u'_i(x) > u'_i(y)$.

In this section we prove the following theorem.

Theorem 5.3. For any two-strategy generalized ordinal potential game $\mathcal{G}$, there is a generalized discrete preference game $\mathcal{G}'$ that is better-response equivalent to $\mathcal{G}$.

Without loss of generality, we let $\{0, 1\}$ be the strategy set of agents of $\mathcal{G}$. The profiles of $\mathcal{G}$ can be identified with the vertices of an $n$-dimensional undirected hypercube. An undirected edge $(x, y)$ of the hypercube connects two profiles that differ for exactly one agent; that is, $x = (x_{-i}, 0)$ and $y = (x_{-i}, 1)$, for some $i \in [n]$. Sometimes we denote edge $(x, y)$ by $(x, i)$ or, equivalently, by $(y, i)$ and call such an edge a dimension-$i$ edge.
**Directing and tagging edges** If \((x, y)\) is a dimension-\(i\) edge and \(u_i(x) < u_i(y)\), then we direct it from \(x\) to \(y\). If instead \(u_i(x) = u_i(y)\), the edge is left undirected. Notice that, since \(G\) is a generalized ordinal potential game, the partially directed hypercube contains no cycle consisting solely of directed edges. Then we tag undirected edges as good as long as we can do so without creating a cycle consisting solely of directed and good edges. The edges that are not tagged in this phase are called bad.

**Claim 5.4.** For each pair of profiles \(x, y\) there is a path on the partially directed hypercube going from \(x\) to \(y\) that does not use bad edges.

**Proof.** Suppose that all paths from \(x\) to \(y\) contain one bad edge. Then we can tag at least one bad edge good without creating a cycle consisting of directed and good edges, a contradiction. \(\square\)

**The constraints** Let us now start to describe the generalized discrete preference game \(G'\) that will turn out to be equivalent to \(G\). It has two types of constraints:

**Vertex constraints:** For every vertex \(x\) of the hypercube, \(G'\) has constraint \(V(x)\) that is satisfied at all profiles except at \(x\). It is easy to see that \(V(x)\) can be expressed as an or-constraint and that every agent is involved in \(V(x)\).

**Edge constraints:** For every edge \((x, i)\), \(G'\) has constraint \(E(x, i)\) that is satisfied at all profiles except at each endpoints, \((x_{-i}, 0)\) and at \((x_{-i}, 1)\). It is easy to see that \(E(x, i)\) can be expressed as an or-constraint that involves all agents except \(i\).

**Constraint weights** Before to define the weights of the constraints, let us fix the notation. We will identify a constraint with its weight. Thus the weight of the constraint associated with edge \((x, i)\) is \(E(x, i)\) and the weight of the constraint associated with vertex \(x\) is \(V(x)\). We also denote by \(E(x)\) the weight of all edge constraints incident to \(x\); that is, \(E(x) = \sum_i E(x, i)\). Similarly, for an agent \(i\) we denote by \(E(i)\) the sum of the weights of all dimension-\(i\) edge constraints; that is, \(E(i) = \sum x E(x, i)\).

Finally, we denote by \(V = \sum x V(x)\) the sum of the weights of all vertex constraints and by \(E = \sum x,i E(x, i)\) the sum of the weights of all edge constraints.

We then assign weight 0 to each edge constraints, except for the ones corresponding to bad edges that have weight \(M > 0\) (to be determined later).

The weights of the vertex constraints are determined as follows. We remove all bad edges from the (partially directed) hypercube and contract good edges by merging endpoints into super-vertices. By Claim 5.4 and by the tagging procedure, the resulting graph is a connected DAG; that is, for every two super-vertices \(X\) and \(Y\), the DAG contains either a path from \(X\) to \(Y\) or a path from \(Y\) to \(X\). Weights of the vertex constraints are determined so that the two following properties hold:

- if \(x\) and \(y\) belong to the same super-vertex \(X\) of the DAG, then
  \[ V(x) + E(x) = V(y) + E(y). \]

  We set \(W(X) = V(x) + E(x)\).

- if \(x\) and \(y\) belong to different super-vertices \(X\) and \(Y\) and there is a path from \(X\) to \(Y\), then
  \[ W(X) = V(x) + E(x) > V(y) + E(y) = W(Y). \]

  Note that an assignment satisfying both these properties always exists. Indeed, we can determine the weights of the vertex constraints according to a topological ordering of the super-vertices of the DAG. For every super-vertex \(X\) of in-degree 0, we determine the vertex \(x^*\) with the maximum \(E(x^*)\) among all of its vertices and set \(V(x^*) = N\) (to be determined later). Then we set \(V(x) = N - E(x) + E(x^*)\)
for all the other vertices \( x \) of \( X \). Observe that \( W(X) = N + E(x^*) \). Suppose now that we have set the weights for all super-vertices that have an out-going edge to supervertex \( Y \) and let \( X \) be the such super-vertex that maximizes \( W(X) \). Then for all vertices \( y \) of \( Y \) we set \( V(y) = W(X) - 1 - E(y) \). Observe that \( W(Y) = W(X) - 1 \).

Note that by taking \( N \) sufficiently large, all weights will be non-negative, as desired.

**Determining the** \( F_i \) **Remember that** \( w_i(x) \) **denotes the sum of the weights of the constraints that involve agent** \( i \) **and that are satisfied at profile** \( x \). **Note that**

\[
w_i(x) = V - V(x) + E - E(i) - E(x) + E(x, i).
\]

We next show that, for each agent \( i \), there is a monotone non-decreasing function \( F_i \) such that the resulting game \( G' \) is better-response equivalent to \( G \). It suffices to show that, if \( (x, y) \) is a dimension-\( i \) directed edge then \( F_i(w_i(x)) < F_i(w_i(y)) \); if instead \( (x, y) \) is a (good or bad) undirected edge then \( F_i(w_i(x)) = F_i(w_i(y)) \).

Consider first a directed edge \( (x, y) \) and let us denote by \( X \) and \( Y \) the super-vertices of the endpoints. By construction we have

\[
W(X) = V(x) + E(x) > V(y) + E(y) = W(Y)
\]

which, together with the observation that \( E(x, i) = E(y, i) \), implies \( w_i(x) < w_i(y) \). Therefore, directed edge \( (x, y) \) imposes the constraint that \( F_i \) is not constant in the interval \([w_i(x), w_i(y)]\).

If \( (x, y) \) is a good edge then \( x \) and \( y \) belong to the same super-vertex and thus

\[
V(x) + E(x) = V(x) + E(y),
\]

which implies that \( w_i(x) = w_i(y) \).

So let us consider a bad edge \( (x, y) \) with \( w_i(x) < w_i(y) \). Then function \( F_i \) must be constant in the interval \([w_i(x), w_i(y)]\) and this is possible if and only if for all \((z, w)\) directed dimension-\( i \) edges the interval \([w_i(z), w_i(w)]\) is not entirely contained in \([w_i(x), w_i(y)]\). Indeed we prove that \( w_i(z) < w_i(x) \).

Let \( X \) and \( Z \) be the super-vertices corresponding to \( x \) and \( z \), respectively. Then,

\[
w_i(z) - w_i(x) = V(X) - V(Z) - E(x, i) + E(z, i) = V(X) - V(Z) - M,
\]

where we used that \( (x, i) \) is a bad edge (and hence \( E(x, i) = M \)) and \((z, i) \) is a directed edge (and hence \( E(z, i) = 0 \)). Thus, by taking \( M \) sufficiently large we have \( w_i(z) < w_i(x) \), as desired.

This completes the proof of Theorem 5.3.

**Remark 5.1.** Observe that there exist no or-constraints of cardinality \( n-1 \) for \( n = 2 \). However, the above proof can be adjusted in order to work with belief constraints in place of or-constraints of cardinality \( n-1 \). Note that, in this case, we need to satisfy the further requirement according to which at most one belief constraint for each agent has a positive weight. However, from Claim 5.4 it follows that at most one bad edge can exist in this case. If no bad edge exists, the proof above does not require any change. If exactly one bad edge exists, let 1 be w.l.o.g. the agent involved in this edge. Then, it can be verified that the proof above holds even if agent 2 assigns weight 0 to both her belief constraints, and agent 1 assigns weights 0 and \( M \).

### 5.4 Impossibility of isomorphism

Unfortunately, we show that the reduction given in previous section cannot be pushed further in order to prove isomorphism between generalized ordinal potential games and generalized discrete preference games, where two games \( \mathcal{G} = ([n], (S_i)_{i \in [n]}, (u_i)_{i \in [n]}) \) and \( \mathcal{G}' = ([n], (S_i)_{i \in [n]}, (u'_i)_{i \in [n]}) \) with the same
set of profiles $S$ are said isomorph if for each profile $x \in S$, $u_i(x) = u'_i(x)$. Consider, indeed, the following two-player game $G$:

$$
\begin{array}{c|cc}
0 & 1 & 0.1 \\
1 & 1.1 & 2.0 \\
\end{array}
$$

where the first value in each cell is the payoff of the row player and the second one is the payoff of the column player. You can check that the following is a generalized ordinal potential function for $G$:

$$
\begin{array}{c|cc}
0 & 1 & 0 \\
1 & 3 & 2 \\
\end{array}
$$

In order to build a generalized discrete preference game that is isomorph to $G$, it is necessary to have $F_1(w_1(01)) < F_1(w_1(00))$ and $F_1(w_1(10)) < F_1(w_1(11))$. From monotony of $F_i$, it turns out we need both $w_i(01) < w_i(00)$ and $w_i(10) < w_i(11)$. Recall that $w(x)$ is the sum of the weights of all constraints satisfied at $x$. Let also $\beta_i(b)$ be the weight of the belief constraint of the player $i$ if it is satisfied by $b$, and 0 otherwise. Note that, since there are only two players, $i$ is involved in each of these constraints, except the belief constraints of the player 2. Hence we have

$$w(01) - \beta_2(1) < w(00) - \beta_2(0), \quad (5)$$
$$w(10) - \beta_2(0) < w(11) - \beta_2(1). \quad (6)$$

As for player 2, we need that both $F_2(w_2(00)) < F_2(w_2(01))$ and $F_2(w_2(11)) < F_2(w_2(10))$. As done above, we can rewrite these requirements as follows:

$$w(00) - \beta_1(0) < w(01) - \beta_1(0), \quad (7)$$
$$w(11) - \beta_1(1) < w(10) - \beta_1(1). \quad (8)$$

Then, in order to satisfy both (5) and (7) we need that $\beta_2(1) > \beta_2(0)$. But, in order to satisfy both (6) and (8) we need that $\beta_2(0) > \beta_2(1)$.

6 Open problems

We believe that our work reveals many interesting open problems. An obvious such problem is to understand the transition from the belief of minority to the opinion of strict majority. Extending our MBM definition, let us call a belief assignment satisfying such a property MBSM (standing for minority becomes strong majority). Clearly, for graphs with an odd number of vertices, our characterization from Section 4 applies to MBSMs as well. Unfortunately, a similar extension is not possible for graphs of even size. For example, consider a six-vertex cycle; according to our definition in Section 4, this is clearly a non-forbidden graph but it can be easily verified that no assignment of 1-beliefs to two vertices admits a sequence of non-worse response moves that leads to an equilibrium in which at least four vertices have 1-preference. Coming up with a characterization for MBSM belief assignments seems to be a challenging open question.

We also believe that our findings in Section 5 open new avenues for understanding the complexity of equilibria in more depth. Even though (approximate) equilibria in exact potential games have been studied rather extensively (e.g., see [6] and the references therein), generalized ordinal potential games are much less understood. Whether generalized discrete preference games could play a role analogous to that of congestion games in this direction certainly deserves investigation.

Finally, even though we have used the sensitivity of the price of anarchy/stability analysis as the starting point of our work, we believe that it is still interesting to further analyze the price of stability in terms of the social welfare and explore its dependence on parameter $\alpha$, as done in [2] for the social cost quality measure, and on the topology of the social network.
References

[1] Emile Aarts and Jan K Lenstra. *Local Search in Combinatorial Optimization.* Princeton University Press, 1997.

[2] Elliot Anshelevich, Anirban Dasgupta, Jon M. Kleinberg, Éva Tardos, Tom Wexler, and Tim Roughgarden. The price of stability for network design with fair cost allocation. *SIAM Journal on Computing,* 38(4):1602–1623, 2008.

[3] Anand Bhalgat, Tanmoy Chakraborty, and Sanjeev Khanna. Approximating pure Nash equilibrium in cut, party affiliation, and satisfiability games. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC),* pages 73–82, 2010.

[4] David Bindel, Jon M. Kleinberg, and Sigal Oren. How bad is forming your own opinion? In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS),* pages 55–66, 2011.

[5] Simina Brânzei, Ioannis Caragiannis, Jamie Morgenstern, and Ariel D. Procaccia. How bad is selfish voting? In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI),* pages 138–144, 2013.

[6] Ioannis Caragiannis, Angelo Fanelli, Nick Gravin, and Alexander Skopalik. Efficient computation of approximate pure Nash equilibria in congestion games. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS),* pages 532–541, 2011.

[7] Moses Charikar, Howard Karloff, Claire Mathieu, Joseph Seffi Naor, and Michael Saks. Online multicast with egalitarian cost sharing. In *Proceedings of the 20th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA),* pages 70–76, 2008.

[8] Chandra Chekuri, Julia Chuzhoy, Liane Lewin-Eytan, Joseph Naor, and Ariel Orda. Noncooperative multicast and facility location games. *IEEE Journal on Selected Areas in Communications,* 25:1193–1206, 2007.

[9] Flavio Chierichetti, Jon M. Kleinberg, and Sigal Oren. On discrete preferences and coordination. In *Proceedings of the 14th ACM Conference on Electronic Commerce (EC),* pages 233–250, 2013.

[10] Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association,* 69(345):118–121, 1974.

[11] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* Cambridge University Press, 2010.

[12] Diodato Ferraioli, Paul W. Goldberg, and Carmine Ventre. Decentralized dynamics for finite opinion games. In *Proceedings of the 5th International Symposium on Algorithmic Game Theory (SAGT),* pages 144–155, 2012.

[13] Noah E. Friedkin and Eugene C. Johnsen. Social influence and opinions. *Journal of Mathematical Sociology,* 15(3–4):193–205, 1990.

[14] Elias Koutsoupias and Christos H. Papadimitriou. Worst-case equilibria. *Computer Science Review,* 3(2):65–69, 2009.

[15] Reshef Meir, Maria Polukarov, Jeffrey S. Rosenschein, and Nicholas R. Jennings. Convergence to equilibria in plurality voting. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI),* pages 823–828, 2010.
[16] Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996.

[17] David Peleg. Local majorities, coalitions and monopolies in graphs: A review. *Theoretical Computer Science*, 282:231–257, 2002.

[18] John E. Savage and Markus G. Wloka. Parallelism in graph-partitioning. *Journal of Parallel and Distributed Computing*, 13(3):257–272, 1991.

[19] Alejandro A. Schäffer and Mihalis Yannakakis. Simple local search problems that are hard to solve. *SIAM Journal on Computing*, 20(1):56–87, February 1991.

[20] Ryo Yoshinaka. Higher-order matching in the linear lambda calculus in the absence of constants is np-complete. In Jurgen Giesl, editor, *Term Rewriting and Applications*, volume 3467 of *Lecture Notes in Computer Science*, pages 235–249. Springer Berlin Heidelberg, 2005.