Robustness in Discrete Preference Games

Vincenzo Auletta
University of Salerno, Italy
auletta@unisa.it

Ioannis Caragiannis
University of Patras & CTI "Diophantus", Greece
caragian@ceid.upatras.gr

Diodato Ferraioli
University of Salerno, Italy
dferraioli@unisa.it

Clemente Galdi
University of Napoli "Federico II", Italy
clemente.galdi@unina.it

Giuseppe Persiano
University of Salerno, Italy
giuper@gmail.com

ABSTRACT

In a *discrete preference game*, each agent is equipped with an internal *belief* and declares her *preference* from a discrete set of alternatives. The payoff of an agent depends on whether the declared preference agrees with the belief of the agent and on the *coordination* with the preferences declared by the neighbors of the agent in the underlying social network. These games have been used to model the formation of opinions and the adoption of innovations in social networks.

Recently, researchers have obtained bounds on the Price of Anarchy and on the Price of Stability of discrete preference games and they have studied to which extent the winning preference reached via best-response dynamics disagrees with the majority of beliefs. In this work, we investigate the robustness of these results to variants of the model.

Our starting point is the observation that bounds on the Price of Anarchy and Stability can be very dependent on the way the quality of an equilibrium is measured. On the other side, results about the disagreement between majority at equilibria and majority among beliefs continue to hold even if we consider different classes of dynamics, such as *no-worse-response* dynamics, best response with multiple players updating at the same time, or with weighted neighbors.

Keywords

Social Network; Opinion Formation; Price of Anarchy; Best-Response Dynamics

1. INTRODUCTION

Opinion formation is a central topic in social science. The literature aims to model how the *opinion* of an individual arises and how it is influenced by the individual's *belief* and her interaction with the environment. A prominent example of these models has been introduced by Friedkin and Johnsen [14] as a refinement to a previous model by DeGroot [11]: it assumes that individuals are on a *social network* representing their social relations and the opinion of each individual is the outcome of a process of repeated averaging between her belief and the opinions of the neighbors.

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

Recently, this model has been the subject of thorough research in Computer Science, mainly focusing on the computational and optimization aspects of the model. In particular, the work of Bindel et al. [7] interprets the repeated averaging process as a *best-response* play in a naturally defined game that leads to a unique equilibrium and they study the *Price of Anarchy* of this equilibrium.

The model considered in [14] and in [7] is continuous, with beliefs and opinions being real numbers in $[0, 1]$. Recently, several works departed from this assumption and considered *discrete* beliefs and opinions. The game theoretic approach of [7] turns out to fit the discrete setting even better than the continuous one, since it is natural to assume that agents are strategic, i.e., they pick the most beneficial preference among the available ones. Games modeling this setting have been named *discrete preference games*, and they have been primarily considered by Ferraioli et al. [12] and by Chierichetti et al. [9].

A well-studied class of discrete preference games restricts beliefs and opinions to only two discrete values, 0 and 1. In this setting, the cost of agent i when the binary strategies of the n agents are given by the vector $\mathbf{x} = (x_1, ...x_n)$ is

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

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, whose relative weight is given by a parameter $\alpha \in [0, 1]$, that depend on the distance of the agent's strategy from her internal belief and from the strategies of her neighbors, respectively.

Chierichetti et al. [9] give bounds on the Price of Anarchy (PoA) and the Price of Stability (PoS) of these games. Specifically, they evaluate the performance of an equilibrium with respect to its ability to minimize the social cost, that is the sum of the costs of agents. Then, they prove that the Price of Anarchy is unbounded (this has also been observed independently by [12]) and they give tight bounds on the Price of Stability. They also show that for $0 \leq \alpha \leq 1/2$ and $\alpha = 2/3$ the states of minimum social cost are always equilibria (that is, the Price of Stability is 1).

A different approach for evaluating the quality of an equilibrium in a discrete preference game has been proposed by Auletta et al. [2]. They consider the extent at which an equilibrium differs from the *truthful* profile, i.e. the one in which the opinion of each agent coincides with her own belief. Their main result shows that for almost every social network it can be the case that *minority becomes majority.* Specifically they show that for $\alpha < 1/2$ and for every social network topology, except the empty graph, the clique and a few "almost-cliques", there is a truthful profile from which a sequence of *best-response deviations* can lead to an equilibrium such that the preference kept by the minority of agents in the truthful profile is the majority of the opinions expressed at the equilibrium. Recently, the same authors extended these results to every $\alpha > 0$, even if each agent i has her own α_i: they show that the only social networks for which minority does not become majority consist of vertices with either too many or too few neighbors [4].

Both approaches for evaluating the performance of equilibria make very precise assumptions on the model: e.g., the Price of Anarchy results consider pure Nash equilibria, whereas the results about convergence to non-truthful majorities are proved to hold only if players update sequentially to their best-response.

Given these assumptions, it is natural to wonder at which extent these results are *robust*: is it possible that a small variation in the model brings to completely different results? Note that this quest for robustness has been advocated very often in game theory. A first example dates back to the Nobel-winning work by Harsanyi and Selten [15], that investigate on Nash equilibria that can be robust to small deviations of players. A more recent example is given by the definition of *robust Price of Anarchy* [18], that looks for Price of Anarchy bounds, that hold not only at Nash equilibria, but also for more general classes of equilibria.

Our Contribution. This work stems out from the observation that results on Price of Anarchy and Price of Stability can vary significantly depending on whether one consider *social cost* or *social welfare.* In [9], agents are cost minimizers and the cost of an agent is given by (1). One could as well consider the equivalent definition of the game in which agents are utility maximizers with the utility of agent i in profile $\mathbf{x}$ defined as

$$u_i(\mathbf{x}) = \alpha\,(1 - |x_i - b_i|) + (1-\alpha) \sum_{j \in N(i)} (1 - |x_i - x_j|). \quad (2)$$

Essentially, the utility of agent i is defined by counting agreements (instead of disagreements) of the opinion of the agent with her internal belief and with the opinions of her neighbors. Clearly, considering utility maximizers instead of cost minimizers does not affect the equilibria of the game. However, the social welfare of a state (i.e., the sum of the agents' utilities) becomes now a natural quality measure for a PoA/PoS analysis. Unfortunately, this type of analysis is not robust to this change and it leads to different conclusions. Specifically, we prove that the PoA for the social welfare is at most 2 (whereas it was unbounded for the social cost). We also prove that PoS for the social welfare is at most 3/2 for values of α close to 1 (whereas for social cost a tight bound of 2 was given). Thus, a slight change in the way we measure equilibria that has, obviously, no effect on the set of equilibria of the game, gives dramatically different results. These results are presented in Section 3.

One can then wonder if also the results about minority becoming majority may be similarly affected by small variations in the model. Note that when we move our focus from social cost/welfare optimization to truthfulness, the quality of a solution turns out to be unaffected from the specific definition of payoffs or costs. However, results in [2] (and in [4]) seem to heavily depend on the strategic play of agents being modeled by sequential best-response deviations on unweighted social networks. May these results be affected by considering a class of deviations that is slightly different from best-response play, or multiple players updating at the same time step, or weighted social networks?

To address this issue, we consider in this work several variants of the model of [2]. First, in place of best-response deviations, we consider *non-worse response deviations* in which an agent can change her strategy not only if the new strategy is strictly better than the current one, but even if the agent evaluates the two strategies as being equivalent. The choice of non-worse response deviations, apart from being the closest class of deviations that need to be evaluated for testing the robustness of the results in [2], is motivated also by the fact that a sequence of these deviations could be triggered in the context of a carefully designed campaign that uses knowledge of the structure of the social network. Interestingly, even though small differences emerge, it turns out that the findings in [2] essentially continue to hold even if we take non-worse responses into account. This will be proved in Section 4.

Next, we consider multiple players updating at the same time step. Specifically, we consider the *independent best response dynamics*, in which at each time step a random subset of players is selected for update and only these selected players are allowed to update their opinion to the best-response, and *concurrent best response dynamics*, according to which at each time step all players concurrently adopt their best-response opinion. The first dynamics is a quite realistic model for the evolution of updates. Indeed, it is usually hard to synchronize the different components of a real-world system so that only one of them updates at each time step, and the remaining ones do not move until they are notified about this change. Note that it is similarly unrealistic to assume that players are able to synchronize themselves so that they all update their opinions at the same time. Still, the concurrent best response dynamics turns out to be an important *stress test* for robustness.

Thus, in order to check at which extent these different dynamics undermine the results of Auletta et al. [2], we designed and run extensive experiments. The outcome of these experiments turns out to be quite surprising: indeed, one can observe that whenever minority becomes majority with sequential updates, it still does with independent and concurrent updates (with very few exceptions in the case of concurrent updates; see below for details). We will discuss these experiments and their results in Section 5.

Finally, we considered *weighted* social networks. In this way, different neighbors can have a different influence on a given player. Weighted graphs are thus able to model more closely real world relationships, where people are affected mainly by close friends and parents than by fleeting friends. We run massive tests even on weighted social networks, from which it emerges that the findings of [2] are robust also against weighted players.

We remark that we consider extensions of the original model along multiple directions: different update rules (best vs. no-worse responses), different selection rules (sequential, concurrent and independent), weighted vs. unweighted networks, and homogeneous vs. heterogeneous agents. It is, in our opinion, very relevant that the property "minority becomes majority" is very robust along all these directions, and even when multiple directions are taken at same type. A partial summary of our results is showed in Table 1.

Related Work. Discrete preference games have attracted a lot of attention recently. This class of games was introduced and analyzed by Ferraioli et al. [12] and, independently, by Chierichetti et al. [9]. The former work mainly focuses on the convergence rate of decentralized dynamics, namely best-response and logit dynamics, to equilibria. The latter mainly studies the Price of Anarchy and Stability, even in the case where players can choose among more than two available opinions. Auletta et al. [2, 4] showed that minority can become majority as an effect of best-response deviations.

Two interesting extensions on discrete preference games have been proposed by Bhawalkar et al. [6] and by Auletta et al. [5]. Bhawalkar et al. [6] consider beliefs and opinions in $[0, 1]$, but, in addition to the opinion expressed, the social relations of an agent are also part of her strategy and are selected so that connections to agents with similar opinions are more preferable. Auletta et al. [5] consider binary opinions, but they model more complex social relations (e.g., to allies and competitors or among more than two agents).

Alternative models are the *HK model* by Hegselmann and Krause [16] and the *DW model* by Weisbuch et al. [23] according to which the social influence of a player is restricted to people whose opinion is close to his own. Recently, Fotakis et al. [13] analyzed the properties of these models on social networks.

Discrete preference games are also related to the large literature on diffusion of information on social networks (we refer the interested reader to [9]), on iterative voting on social networks [10, 21], and on the emergence of social norms [22, 24].

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 $(\mathbf{x}_{-i}, s) = (x_1, \ldots, x_{i-1}, s, x_{i+1}, \ldots, x_n)$ for every $i \in [n]$, every profile $\mathbf{x} \in S$ 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 every $y \in \{0, 1\}$, we denote as $\overline{y}$ the negation of y, i.e., $\overline{y} = 1 - y$.

At a profile $\mathbf{x}$, the utility (or payoff) of agent i is denoted by $u_i(\mathbf{x})$ and it is defined as in (2). A profile is a *pure Nash equilibrium* (or, simply, an equilibrium) if $u_i(\mathbf{x}) \geq u_i(\overline{x}_i, \mathbf{x}_{-i})$ for every agent i. As observed in previous works, discrete preference games always have equilibria. Alternatively, we can define the agents of discrete preference games to be cost-minimizers, following the definition of [9] and [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 [17], we can define the *price of anarchy* (PoA) of the game as the maximum value of the ratio $\frac{SC(\mathbf{x})}{SC(\mathbf{x}^\star)}$ over all equilibria $\mathbf{x}$ of the game, where $\mathbf{x}^\star$ denotes the profile that minimizes the social cost. Furthermore, we can define the more optimistic *price of stability* (PoS), introduced by [1], as the minimum value of the ratio $\frac{SC(\mathbf{x})}{SC(\mathbf{x}^\star)}$ over all equilibria $\mathbf{x}$. Both notions have been proposed in order to assess the impact of selfish behavior on efficiency.

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. In particular, the social welfare of a profile $\mathbf{x}$ is simply the total utility of the agents, i.e., $SW(\mathbf{x}) = \sum_{i \in [n]} u_i(\mathbf{x})$. Then, the PoA/PoS of a game in terms of the social welfare is the maximum/minimum value of the ratio $\frac{SW(\mathbf{x}^\star)}{SW(\mathbf{x})}$ over all equilibria $\mathbf{x}$ of the game, where $\mathbf{x}^\star$ now is the profile that maximizes the social welfare.

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 term *best-response moves* for the former and *non-worse* response moves for the latter. Moreover, we consider three different rules for selecting which players are allowed to update their opinion at each time step: (1) the *sequential selection rule*, according to which a single player is allowed to update at each time step, (2) the *concurrent selection rule*, that instead dictates that all players concurrently update their opinion at each time step, and (3) the *independent selection rule*, according to which the set of players that update their opinion is selected uniformly at random at each time step.

3. PRICE OF ANARCHY AND STABILITY

As it is shown in [12] and [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 for the price of anarchy in terms of the social welfare.

PROPOSITION 1. *The price of anarchy of utility maximizing discrete preference games is at most* 2.

PROOF. Let $\mathbf{x}$ be an equilibrium. Observe that $u_i(\mathbf{x}) \geq \frac{1}{2}(\alpha + (1 - \alpha) \cdot |N(i)|)$ for every i, otherwise i has an incentive to play strategy $\overline{x}_i$. The bound follows, since in the optimal profile $\mathbf{x}^\star$, for every i, $u_i(\mathbf{x}^\star) \leq \alpha + (1 - \alpha) \cdot |N(i)|$. $\square$

It is not hard to see that the above bound is tight. Consider indeed an instance with $\alpha = \frac{k}{k+1}$, with $k > 0$, the social network G being a composition of k matchings, and assume that each agent i has belief $b_i = 0$. Clearly, the truthful profile is the one that maximizes the social welfare. Consider, instead, the profile $\mathbf{x} = (1, \ldots, 1)$. Observe that this is an equilibrium. Indeed, each agent i has exactly k neighbors, and thus the utility is $u_i(\mathbf{x}) = \left(1 - \frac{k}{k+1}\right) k = \frac{k}{k+1}$. Hence, i has not any incentive to adopt opinion 0, since $u_i(0, \mathbf{x}_{-i}) = \alpha = \frac{k}{k+1}$. It would be also possible to prove that these are the only instances achieving this bound.

Similarly, [9] claims that the price of stability for discrete preference games is bounded by 2 and that there exist instances achieving this bound for α 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 the following result.

PROPOSITION 2. *For α close to 1, the price of stability of utility-maximizing discrete preference games is at most 3/2.*

PROOF. Let us set $C = \alpha n + 2(1-\alpha)m$, where m is the number of edges in the social network. Observe that $SW(\mathbf{x}) = C - SC(\mathbf{x})$. Let $\mathbf{x}$ be the equilibrium that maximizes the social welfare and let $\mathbf{x}^\star$ be the optimal profile. Moreover, let $c = SC(\mathbf{x})/SC(\mathbf{x}^\star)$. Finally, consider the truthful profile $\mathbf{b}$. If $\alpha \geq \frac{m}{m+n/6}$, then $SC(\mathbf{x}^\star) \leq SC(\mathbf{b}) \leq 2(1-\alpha)m \leq \frac{1}{4}(\alpha n + 2(1-\alpha)m) \leq \frac{C}{4}$. Hence, it follows that

$$PoS = \frac{SW(\mathbf{x}^\star)}{SW(\mathbf{x})} = \frac{C - SC(\mathbf{x}^\star)}{C - SC(\mathbf{x})} = 1 + \frac{(c-1)\cdot SC(\mathbf{x}^\star)}{C - c\cdot SC(\mathbf{x}^\star)} \leq 1 + \frac{c-1}{4-c} = \frac{3}{4-c} \leq \frac{3}{2},$$

where we used that the PoS in the cost-minimizing version is at most 2. $\square$

Chierichetti et al. [9] also proved that the PoS is 1 if $\alpha \leq 1/2$ or $\alpha = 2/3$. This result carries over to the utility-maximizing definition of the game.

It must be noted that, while robustness of PoA results fails if one redefines discrete preference games as utility-maximizing, it still holds if one instead considers different definitions of social cost and social welfare. In particular, even if we define social cost as the maximum/minimum cost among agents, then it is easy to see that discrete preference games still have unbounded Price of Anarchy: consider, indeed, an instance in which $\alpha \leq \frac{k}{k+1}$, $k > 0$, and every agent i has belief $b_i = 0$ and at least k neighbors. In the optimal profile $\mathbf{b}$, each agent has 0 cost, but there are profiles, e.g., $\mathbf{x} = (1, \ldots, 1)$, that are in equilibrium, and in which each agent has a cost larger than 0. Similarly, if we define social welfare as the maximum/minimum utility among agents, then arguments given above still hold, and thus the Price of Anarchy in the utility-maximizing version is bounded by 2 and this bound is tight.

4. MINORITY BECOMES MAJORITY

In this section we seek social network topologies, for which, starting from a belief assignment where 1 is supported by a minority, we can sequentially convince agents to switch their strategy (without decreasing their payoff), so that an equilibrium with at least half of the players having preference 1 is reached. To this aim, we say that an assignment of beliefs $\mathbf{b}$ to the vertices of a graph G is MWBM (*minority weakly becomes majority*) for G if: (i) the number of vertices with belief 1 in $\mathbf{b}$ is a minority; i.e., $|\{x \in V : b_x = 1\}| < n/2$; (ii) there is a *subverting* sequence of non-worse response moves that starts from $\mathbf{b}$ and converges to an equilibrium $\mathbf{b}'$ in which the number of vertices with preference 1 is a (weak) majority; i.e., $|\{x \in V : b'_x = 1\}| \geq n/2$. Note that this definition is almost equivalent to the definition of MBM belief assignment given in [2], except that we now consider non-worse responses in place of best responses. Auletta et al. [2] prove that if $\alpha < 1/2$, then MBM belief assignments exist for almost all graphs G. Here, we show that this result holds even if we take into account no-worse responses. Specifically, we prove that, if $\alpha \leq 1/2$, MWBM belief assignments exist for all graphs G, except for the following *forbidden* graphs (for which it is immediate to see that no MWBM belief assignment exists): (i) G is the complete graph K_n or consists of n isolated vertices; (ii) n is even and G consists of an isolated vertex and a clique K_{n-1}; (iii) n is even and all vertices of G have degree at least $n-2$. Note that for this last subset of graphs the result of [2] does not hold. We prove the following theorem.

THEOREM 3. *An MWBM belief assignment exists for every non-forbidden graph G. Moreover, this assignment can be computed in polynomial time.*

We prove the theorem for graphs $G = (V, E)$ with an even number $n = |V|$ of vertices. We then show how to extend it to odd n. Let us first 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 $u \in \overline{S}$ and $v \in S$, $W(S, \overline{S}) \leq W(S \cup \{u\} \setminus \{v\}, \overline{S} \cup \{v\} \setminus \{u\})$. The following lemma from [2] will be very useful.

LEMMA 4. *([2, Lemma 1]) For every locally minimal bisection $(S, \overline{S})$ and for every $u \in S$ and $v \in \overline{S}$, $W(u, S) - W(u, \overline{S}) + W(v, \overline{S}) - W(v, S) + 2W(u, v) \geq 0$.*

The following lemma can be found in [3].

LEMMA 5. *([3, Lemma 8]) Let $(S, \overline{S})$ be a bisection in a n-node graph with less than $n-1$ vertices with degree at least $n-2$. If there is a node $z \in S$ with degree $n-2$ and $W(z, S) = W(z, \overline{S}) - 2$ and $W(x, S) = W(x, \overline{S}) = n/2 - 1$ for every $x \in \overline{S}$, then one can compute in polynomial time a bisection $(S', \overline{S'})$ such that $W(S', \overline{S'}) < W(S, \overline{S})$.*

The proof of Theorem 3 distinguishes between two cases. The first case is the one in which there are $n-1$ vertices with degree at least $n-2$.

LEMMA 6. *For every graph in which there are $n-1$ vertices with degree at least $n-2$, there is a MWBM belief assignment.*

The proof of this lemma resembles the one given for [3, Proposition 6], and hence it is omitted.

As for the remaining case, we say that a graph G is of type **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 **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 $W(w, \overline{S}) \leq W(w, S) + 1$ and w is adjacent to two non-adjacent vertices $u, v \in S$. We prove that graphs of types **T1** and **T2** admit an MWBM belief assignment. Later we will prove that every non-forbidden graph is of type **T1** or **T2**.

PROPOSITION 7. *Let G be a graph of type* **T1** *with an even number of vertices. Then, G has an* MWBM *belief assignment.*

PROOF. Let $(S, \overline{S})$ be a bisection that is a witness that G is of type **T1** and let $u \in S$ be a vertex such that $W(u, S) \geq W(u, \overline{S}) + 1$. Consider the belief assignment $\mathbf{b}$ 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 $\mathbf{b}$ is $n/2 - 1$. Consider now vertex u. If u sets its preference equal to its belief then its payoff is $W(u, \overline{S}) + 1$. If u sets its preference to 1, then its payoff is $W(u, S) \geq W(u, \overline{S}) + 1$. The profile $\mathbf{b}' = (\mathbf{1}_S, \mathbf{0}_{\overline{S}})$ after u has switched to 1 has a weak majority of 1.

We complete the proof by verifying that in $\mathbf{b}'$ no vertex of S has an incentive to switch to 0, and thus there is an equilibrium reachable from $\mathbf{b}'$ in which the 1's are a weak majority. This 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 **T1** we have that $W(x, S) + 1 \geq W(x, \overline{S})$. $\Box$

PROPOSITION 8. *Let G be a graph of type* **T2** *with an even number of vertices. Then, G has an* MWBM *belief assignment.*

PROOF. Let $(S, \overline{S})$ be a bisection that is a witness that G is of type **T2** and $u, v \in S$ and $w \in \overline{S}$ be as in the definition of type **T2** above. Consider the belief assignment $\mathbf{b}$ 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 $\mathbf{b}$ 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 $\mathbf{b}$ 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 others payoff. The preference profile $\mathbf{b}'$ after u and v have switched is $\mathbf{b}' = (\mathbf{1}_{S \cup \{w\}}, \mathbf{0}_{\overline{S} \setminus \{w\}})$ and it has a majority of 1.

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

The two propositions above show that it is possible to construct an MWBM belief assignment for a graph of type **T1** and **T2** if a witness for the type is given. Next we prove that, for every non-forbidden graph, such a witness exists.

For a bisection $(S, \overline{S})$ we denote by $\mathcal{B}(S, \overline{S})$ the set containing $(S, \overline{S})$, $(\overline{S}, S)$ and, for every bisection $(S', \overline{S'})$ obtained 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 it minimizes the width among all the bisections obtained from $(S, \overline{S})$ by swapping two pairs of vertices. Finally, $(S, \overline{S})$ is *strong* if it does not satisfy the conditions of Lemma 5.

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

The running time of the algorithm is polynomial: the desired bisection can be computed in polynomial time by local search [20, 19] and the set $\mathcal{B}(S^\star, \overline{S^\star})$ contains at most $2n^2$ bisections, that can be checked in time linear in n.

Finally, we prove that for every non-forbidden graph G there is a bisection in $\mathcal{B}(S^\star, \overline{S^\star})$ 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 9. *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 is forbidden.*

For bisection $(S, \overline{S})$ of a graph G, let $\mathcal{B}'(S, \overline{S}) \subseteq \mathcal{B}(S, \overline{S})$ be the subset of bisections $(S', \overline{S'}) \in \mathcal{B}(S, \overline{S})$ such that at least one isolated vertex is in $\overline{S'}$. We actually prove a stronger statement: for every locally minimal bisection $(S, \overline{S})$ of a graph G if there is no witness that G is of type **T1** in $\mathcal{B}'(S, \overline{S})$, 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 $u \in S$ for which $W(u, S) \leq W(u, \overline{S}) - 1$ and consider bisection $(S', \overline{S'}) \in \mathcal{B}'(S, \overline{S})$, 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 $x \in S$.

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 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 $v \in \overline{S}$, $W(v, \overline{S}) \leq n/2 - 2$ which leaves $c = 2$ as the only possibility.

Let A denote the subset of $\overline{S}$ consisting of all the vertices x with $W(x, \overline{S}) = W(x, S) - 2$; all vertices $x \in \overline{S} \setminus A$ have $W(x, \overline{S}) \geq W(x, S)$. 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 $(S', \overline{S'}) \in \mathcal{B}'(S, \overline{S})$, with $S' = S \cup \{v\} \setminus \{u\}$. Since $v \in A$ is connected to every vertex $x \in S$, then for every vertex $x \in S \setminus \{u, w\}$, the number of neighbor in S' will be at least the same as 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 $A = \overline{S} \setminus \{i\}$).

If instead A is empty, we claim that $W(x, S) = W(x, \overline{S})$ for every vertex $x \in \overline{S}$. Indeed, assume by contradiction that $W(v, \overline{S}) \geq W(v, S) + 1$ for some $v \in \overline{S}$ and let u be any vertex of S. Then, the bisection $(S', \overline{S'}) \in \mathcal{B}'(S, \overline{S})$, 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 $(S', \overline{S'}) \in \mathcal{B}'(S, \overline{S})$, 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 3 for all graphs with an even number of vertices and one isolated vertex. We will now complete the proof of Theorem 3 for all graphs with an even number of vertices by proving the following lemma.

LEMMA 10. *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 G is forbidden.*

PROOF. We first show that if there is a strong 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 there is a non-empty subset $A \subseteq S$, such that $W(u, S) \leq W(u, \overline{S}) - 1$ for every $u \in A$. Notice that, by Lemma 4, it must be the case that $W(u, S) \geq W(u, \overline{S}) - 2$ for $u \in A$.

We first observe that every $u \in A$ must be connected to every vertex of $\overline{S}$ (and, hence, their degree will be at least $n - 2$). Indeed, if this was not the case, let $v \in \overline{S}$ be a vertex that is not adjacent to u. Lemma 4 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 we observe that the vertices of $\overline{S}$ form a clique. Indeed, if $\overline{S}$ contains two non-adjacent vertices v and w, then, by applying Lemma 4 again, $(\overline{S}, S)$ is a witness that G is of type **T2**, that is a contradiction. Hence, $W(x, \overline{S}) = n/2 - 1$ for every $x \in \overline{S}$. Moreover, since the bisection $(\overline{S}, S)$ is not a witness that G is of type **T1**, then it must be the case that $W(x, S) \geq W(x, \overline{S})$ for every $x \in \overline{S}$. Hence, the degree of these vertices is at least $n - 2$. Moreover, $W(S, \overline{S}) \geq \frac{n}{2}\left(\frac{n}{2} - 1\right)$.

The only vertices for which we do not have yet a lower bound on the degree are the ones in $S \setminus A$. Note that $W(x, S) \geq W(x, \overline{S})$ for every such vertex x. Note that if $W(x, \overline{S}) \geq n/2 - 1$, that the degree of x is at least $n - 2$. In order to bound the degree of the remaining vertices, we first note that it is not possible that there exists $u \in A$ with $W(u, S) = W(u, \overline{S}) - 2$. Indeed, if this is the case, the condition of Lemma 5 holds (note that we are assuming that the number of vertices with degree at least $n-2$ is not $n - 1$, since this case was solved by Lemma 6). But this is a contradiction, since $(S, \overline{S})$ is strong.

Hence, for each $x \in S$, we have that $W(x, S) \geq W(x, \overline{S}) - 1$. On the other side, it must be the case that for each such vertex, $W(x, S) \leq W(x, \overline{S})$, otherwise $(S, \overline{S})$ is of type **T1**. Hence, we can partition S in three subset: S_1 contains all vertices x of degree $n - 1$ with $n/2 - 1 = W(x, S) = W(x, \overline{S}) - 1$; S_2 contains all vertices x of degree $n - 2$ with $n/2-1 = W(x, S) = W(x, \overline{S})$; finally, S_3 contains all vertices x of degree less than $n-2$ with $W(x, S) = W(x, \overline{S}) < n/2-1$. Note that either S_3 is empty or $|S_3| \geq 2$, since all vertices in $S_1 \cup S_2$ are connected with each vertex in S. Moreover, for every pair $u, v \in S_3$ we have that $W(u, \overline{S}) + W(v, \overline{S}) > n/2$ and thus they have at least a common neighbor in $\overline{S}$. Indeed, if $W(u, \overline{S}) + W(v, \overline{S}) \leq n/2$, then since $W(S, \overline{S}) \geq \frac{n}{2}\left(\frac{n}{2} - 1\right)$, it must be the case that all the remaining $\frac{n}{2} - 2$ vertices are connected to every vertex in $\overline{S}$. Then they all belong to S_1 and they are connected to each vertex in S. In particular, they are all connected to u and v. Thus, $W(u, S) \geq n/2 - 2$ and $W(u, S) \geq n/2 - 2$. Since $u, v \in S_3$ and thus $W(u, S) = W(u, \overline{S})$ and $W(v, S) = W(v, \overline{S})$, we have that $W(u, \overline{S}) + W(v, \overline{S}) \geq n-4$, that is a contradiction for every $n > 8$. Thus, if S_3 is not empty, then there are two vertices in S_3 that are not adjacent and they have a common neighbor in $\overline{S}$. Then, $(S, \overline{S})$ is of type **T2**, that is a contradiction.

So, we have that $W(x, S) \geq W(x, \overline{S})$ for every $x \in S$. If, in addition, we had $W(u, S) \geq W(u, \overline{S}) + 1$ 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 every vertex $x \in S$. The same argument can be used for all vertices of $\overline{S}$.

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

Finally, we claim that the vertices of $S^\star$ (and, symmetrically, the vertices of $\overline{S^\star}$) form a clique. Assume for sake of contradiction that some vertex $w \in S^\star$ exists that is not adjacent to u. Then, w is not 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 $u' \in N(w) \cap \overline{S^\star}$ (such a vertex exists since w is not isolated and $W(w, S^\star) = W(w, \overline{S^\star})$) and observe that it is adjacent to the non-adjacent vertices u and w in $S^\star$. This triplet shows that $(S^\star, \overline{S^\star})$ is then a witness that G is of type **T2**, that is a contradiction.

In conclusion, since $W(x, S^\star) = W(x, \overline{S^\star}) = n/2 - 1$ 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 9, it follows that G' is of type **T1** and we can find a witness $(S, \overline{S})$ with $i \in \overline{S}$. Then, it is immediate to see that the MWBM belief assignment $\mathbf{b}$ described in Proposition 7, when restricted to the vertices of G', gives an MWBM belief assignment for G.

5. EXPERIMENTAL RESULTS

In this section we will discuss the settings and the results of experiments that we run in order to evaluate the robustness of the results in [2]. Specifically, in [2], it is shown that for all graphs, except few *forbidden* graphs (i.e., the graph with no edges and some "almost-cliques"), it is possible to find an initial assignment of beliefs to vertices and a sequence of best responses that leads to an equilibrium in which minority becomes majority. Actually, their result is stronger: they prove that there is an initial assignment of beliefs to vertices and a *single* vertex u such that, once u plays his best-response, *every* sequence of best responses leads to an equilibrium in which minority becomes majority. Not only, they also show that this initial assignment can be easily found by looking at the neighborhood of a locally minimal bisection.

In our first experiments, we test whether these results still hold when multiple players are allowed to update their strategies at each time step. Specifically, we test if when $\alpha < 1/2$ it holds that the best-response dynamics with concurrent or independent selection rule, starting from an initial assignment of beliefs selected according to some bisection in the neighborhood of a locally minimal bisection of a non-forbidden graph, still reaches an equilibrium in which minority becomes majority.

To this aim, we generated random $G(n,p)$ graphs, with odd n going from 7 to 99 (we focused only on odd n for which minority and majority are always well-defined) and $p = 0.25, 0.5, 0.75$ in order to produce graphs with low, medium, or high density. Hence, in total we considered 141 different pairs (n,p). For each of these pairs, we generated $100n$ different graphs, resulting in a total of 747300 graphs. For each of these graphs, we checked if it is forbidden. Since, as discussed above the set of forbidden graph is very small, it turns out that very few forbidden graphs have been found. Specifically, 740864 graphs out of 747300 (i.e, 99.14%) turn out to be non-forbidden. For each of these graphs we computed a locally minimal bisection B via local search. For graphs with an odd number of vertices we consider bisections $\mathcal{B} = (S, \overline{S})$ in which one side, S, has one element more than the other, $\overline{S}$. Similarly we define the neighborhood of $\mathcal{B}$ as the set of all bisections $\mathcal{B}'$ that can be obtained from $\mathcal{B}$ either by moving a vertex of S in $\overline{S}$ or by swapping one vertex of S with a vertex of $\overline{S}$. Then we considered the bisection $\mathcal{B}$ and every bisection in its neighborhood. We followed the approach of Auletta et al. [2] for building an initial assignment of beliefs from these bisections: that is, we considered a minority of $\frac{n-1}{2}$ vertices consisting of all vertices in S except a single vertex u and a majority of $\frac{n+1}{2}$ vertices consisting of u and of vertices in $\overline{S}$ (for every bisection, we considered all possible choices for u, even though the result in [2] specifies how to choose this vertex). Finally, for each of these initial opinion profiles we run both the concurrent best-response dynamics and the independent best-response dynamics until they reach an equilibrium or until 100 steps have been executed. As we will see, this maximum allowed number of steps is sufficiently high to test the reachability of an equilibrium as most of the failures will occur with small graphs. We say that a graph *fails* for the given dynamics if for every starting opinion profile the dynamics do not converge to an equilibrium in which the minority became the majority. Our experimental results are summarized in Table 1.

	Seq. BR	Seq. NWR	Ind. BR	Con. BR
Unweighted	100%	100%	100%	99.93%
Weighted	100%	100%	100%	99.63%

Table 1: Percentage of graphs in which minority becomes majority

The first finding that comes out from these tests is very surprising: for every non-forbidden graph considered in the experiment, if players update their opinion according to the independent best-response dynamics, then one can find an initial assignment of beliefs from which the dynamics leads to an equilibrium in which minority becomes majority. Roughly speaking, this means that even though the proof of Auletta et al. [2] heavily relies on the selection rule being sequential, their results appear to still hold even if players choose independently and without any synchronization when to update their opinion.

Although we believe that the independent selection rule is far more realistic than the concurrent one, the latter plays an important role in confirming the robustness of the results. Indeed, it turns out that for the 99.93% of the non-forbidden graphs that have been considered, minority becomes majority even if players concurrently update their opinion to the best response at each time step.

The above success percentage is surprisingly large. Still, we believe that it is interesting to investigate which graphs fail for the concurrent best-response dynamics. In Figure 1 we show how the percentage of failure changes as n or p change. It turns out that the large majority of failures occurs when n is small and the graph is dense (i.e., p is large). This suggests that the graphs for which concurrent best-response dynamics cannot lead to an equilibrium with minority becoming majority must be very dense. Indeed, these are the graphs that have a higher probability to occur when n is small and p is large. Thus, although the set of graphs forbidden for the concurrent best-response dynamics is larger compared to the case of sequential updates, it still consists of either almost empty graphs or almost-cliques.

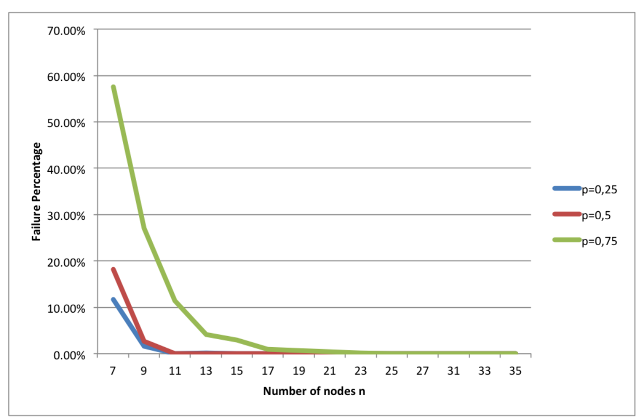

Figure 1: Failure percentage for concurrent best-response dynamics

We also investigated how graphs fail for the concurrent best-response dynamics. This happens either because the dynamics converges to an equilibrium where minority does not become majority, or because it does not converge at all (within 100 steps). It turns out that the latter kind of failures is far more frequent than the former. As reported in Table 2, only 65 among the 569 failures that we observed in our experiments are due to convergence to bad equilibria.

	Failures	Bad Equilibria	No Convergence
Unweighted	569	65	504
Weighted	332	0	332

Table 2: Failures caused by convergence to bad equilibrium or non-convergence

Apart from the effect of different selection rules, we were also interested in understanding if the results in [2] can be extended to edge-weighted graphs. In this case, the cost of agent i is

$$c_i(\mathbf{x}) = \alpha \cdot |x_i - b_i| + (1-\alpha) \cdot \sum_{j \in N(i)} w_{ij} \, |x_i - x_j|,$$

where w_{ij} is the weight on the edge (i, j). To test how the different dynamics perform in this setting, we run experiments similar to the ones described above. Specifically, we considered $G(n,p)$ graphs for 66 different pairs (n,p), namely odd $n = 7, \ldots, 49$ and $p = 0.25, 0.5, 0.75$. For each pair we generated 50 different graphs, for a total of 92400 graphs, among which 90789 turned out to be non-forbidden. We assigned random weights to every edge; the weights were normalized so that the larger edge weight is 1. For these weighted graphs, we computed the initial assignment of beliefs exactly as above, except that the minimality of bisection is evaluated not with respect to the number of edges in the cut, but with respect to the total weight of these edges. For each starting opinion profile we run four different dynamics: best-response dynamics with sequential, independent, and concurrent update rule, and sequential non-worse-response dynamics.

The results of these experiments suggest that the findings of [2] are very robust even against the extension of the model to weighted graphs. Indeed, it turns out that for every non-forbidden graph, there is a belief assignment taken in the neighborhood of a locally minimal bisection, from which minority becomes majority whenever players update their opinion according to sequential best-response, sequential no-worse-response, or independent best-response dynamics. That is, for every graph considered in our experiment and for each of these three dynamics, we have always been able to find a minority that becomes majority at the equilibrium. For concurrent best-response dynamics, the behavior is similar to what we showed for the case of unweighted graphs. Specifically, we found that the dynamics allows minority to become majority in 99.63% of cases, and failure mainly occurs when n is small and p is large, i.e., on graphs that are in some way extreme. Moreover, all these failures are due to the fact that the dynamics do not converge at all.

We conclude this section, by observing that robustness is not limited to the case in which $\alpha < 1/2$ but, instead, the results continue to hold even if each player i is allowed to have her own different value of $\alpha_i \in [0, 1]$. For this setting, Auletta et al. [4] recently proved that minority can become majority for every graph except when every vertex has either too many neighbors or too few neighbors, where the exact thresholds on the number of neighbors depend on the value of α_i. This new result extends the previous one in [2] even with respect to the ability of computing the initial assignment of beliefs from which minority becomes majority: instead of looking at locally minimal bisections and their neighborhood, one needs to consider 3-minimal bisections, that is bisections whose width cannot be decreased by moving at most 3 vertices from each side to the other, and their 3-neighborhood, i.e., the bisections achieved by these moves.

We then tested even the robustness of this extension to changes in the update rule, no-worse-response in place of best-response, and to changes in the selection rule, independent and concurrent updates in place of sequential ones. As above, we generated a large number of $G(n,p)$ graphs, and we assigned to each vertex i a random $\alpha_i \in [0, 1]$. For those among these graphs that turned out to be non-forbidden, we computed a 3-minimal bisection B. For B and every bisection in its 3-neighborhood, we computed the resulting initial assignment of beliefs. From each of these starting opinion profiles, we run the sequential no-worse response dynamics, the independent best-response dynamics, and the concurrent best-response dynamics. A graph is said to fail for a dynamics if it does not converge to an equilibrium in which minority becomes majority for every starting profile.

The results of these tests turn out to be even more surprising than the previous ones: whenever the degree of each vertex is not too low and not too high (with respect to α_i), then in the 3-neighborhood of a 3-minimal bisection there always exists an initial belief assignment from which minority becomes majority for *every* dynamics that we considered, including the concurrent best response dynamics. Thus, as long as we restrict ourselves to graphs that are far away from being forbidden, then the results of [4] appear to be robust against both the change in the update rule and the change in the selection rule.

6. OPEN PROBLEMS

Even though our analysis shows that PoA and PoS depend on whether agents are utility maximizers or cost minimizers, we believe that it is still interesting to further analyze the PoS in terms of the social welfare and explore its dependence on parameter α, as done in [9] for the social cost quality measure, and on the topology of the social network. It would be also interesting to consider, as in [9], extensions of the problem to more than two strategies, and to evaluate PoA and PoS for the utility-maximizing version. Yet another interesting direction, related to the robustness question discussed in this paper, would be to analyze ε-approximated equilibria [8], i.e., profiles in which no player can improve her utility by more than ε.

As for the problem of minority becoming majority, the obvious question is to prove at which extent minority can become majority for the more general processes and game definitions defined in this work. Finally, one can be interested in understanding how probable the minority becomes majority phenomenon is. How is this frequency related to the topological properties of the network?

Acknowledgments. This work was partially supported by a Caratheodory research grant E.114 from the University of Patras and by the "GNCS-INdAM". Part of this work was done while Giuseppe Persiano was visiting Google, New York.

REFERENCES

[1] E. Anshelevich, A. Dasgupta, J. M. Kleinberg, É. Tardos, T. Wexler, and T. Roughgarden. The price of stability for network design with fair cost allocation. *SIAM Journal on Computing*, 38(4):1602–1623, 2008.

[2] V. Auletta, I. Caragiannis, D. Ferraioli, C. Galdi, and G. Persiano. Minority becomes majority in social networks. In *WINE*, pages 74–88, 2015.

[3] V. Auletta, I. Caragiannis, D. Ferraioli, C. Galdi, and G. Persiano. Minority becomes majority in social networks. *CoRR*, abs/1402.4050v4, 2015. Full version of [2].

[4] V. Auletta, I. Caragiannis, D. Ferraioli, C. Galdi, and G. Persiano. Discrete preference games in heterogeneous social networks: Subverted majorities and the swing player. *CoRR*, abs/1603.02971, 2016.

[5] V. Auletta, I. Caragiannis, D. Ferraioli, C. Galdi, and G. Persiano. Generalized discrete preference games. In *IJCAI*, pages 53–59, 2016.

[6] K. Bhawalkar, S. Gollapudi, and K. Munagala. Coevolutionary opinion formation games. In *STOC*, pages 41–50. ACM, 2013.

[7] D. Bindel, J. M. Kleinberg, and S. Oren. How bad is forming your own opinion? In *FOCS*, pages 55–66, 2011.

[8] S. Chien and A. Sinclair. Convergence to approximate nash equilibria in congestion games. In *SODA*, pages 169–178. SIAM, 2007.

[9] F. Chierichetti, J. M. Kleinberg, and S. Oren. On discrete preferences and coordination. In *EC*, pages 233–250, 2013.

[10] S. Chopra, E. Pacuit, and R. Parikh. Knowledge-theoretic properties of strategic voting. In *European Workshop on Logics in Artificial Intelligence*, pages 18–30. Springer, 2004.

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

[12] D. Ferraioli, P. W. Goldberg, and C. Ventre. Decentralized dynamics for finite opinion games. volume 648, pages 96–115. Elsevier, 2016.

[13] D. Fotakis, D. Palyvos-Giannas, and S. Skoulakis. Opinion dynamics with local interactions. In *IJCAI 2016*, pages 279–285, 2016.

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

[15] J. C. Harsanyi and R. Selten. *A General Theory of Equilibrium Selection in Games*. MIT Press, 1988.

[16] R. Hegselmann, U. Krause, et al. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

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

[18] T. Roughgarden. Intrinsic robustness of the price of anarchy. *Journal of the ACM*, 62(5):32, 2015.

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

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

[21] S. Sina, N. Hazon, A. Hassidim, and S. Kraus. Adapting the social network to affect elections. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 705–713. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

[22] S. Swarup, A. Apolloni, and Z. Fagyal. A model of norm emergence and innovation in language change. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 693–700. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[23] G. Weisbuch, G. Deffuant, F. Amblard, and J.-P. Nadal. Meet, discuss, and segregate! *Complexity*, 7(3):55–63, 2002.

[24] C. Yu, M. Zhang, F. Ren, and X. Luo. Emergence of social norms through collective learning in networked agent societies. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 475–482. International Foundation for Autonomous Agents and Multiagent Systems, 2013.