Decentralized Dynamics for Finite Opinion Games

Item type	Article
Authors	Ferraioli, D.; Goldberg, P. W. (Paul); Ventre, C. (Carmine)
Citation	Ferraioli, D., Goldberg, P. W., Ventre, C. (2016) 'Decentralized Dynamics for Finite Opinion Games' Theoretical Computer Science; Available online: 24 Aug 2016
Eprint Version	Post-print
DOI	10.1016/j.tcs.2016.08.011
Publisher	Elsevier
Journal	Theoretical Computer Science
Additional Link	http://www.journals.elsevier.com/theoretical-computer-science/
Rights	Following 12 month embargo author can archive post-print (ie final draft post-refereeing). For full details see http://www.sherpa.ac.uk/romeo/issn/0304-3975/ [Accessed: 18/08/2016]
Downloaded	8-Nov-2018 15:41:14
Link to item	http://hdl.handle.net/10149/618521

Decentralized Dynamics for Finite Opinion Games[*]

Diodato Ferraioli[†]   Paul W. Goldberg[‡]   Carmine Ventre[§]

Abstract

Game theory studies situations in which strategic players can modify the state of a given system, in the absence of a central authority. Solution concepts, such as Nash equilibrium, have been defined in order to predict the outcome of such situations. In multi-player settings, it has been pointed out that to be realistic, a solution concept should be obtainable via processes that are decentralized and reasonably simple. Accordingly we look at the computation of solution concepts by means of decentralized dynamics. These are algorithms in which players move in turns to decrease their own cost and the hope is that the system reaches an "equilibrium" quickly.

We study these dynamics for the class of opinion games, recently introduced by Bindel et al. [10]. These are games, important in economics and sociology, that model the formation of an opinion in a social network. We study best-response dynamics and show upper and lower bounds on the convergence to Nash equilibria. We also study a noisy version of best-response dynamics, called logit dynamics, and prove a host of results about its convergence rate as the noise in the system varies. To get these results, we use a variety of techniques developed to bound the mixing time of Markov chains, including coupling, spectral characterizations and bottleneck ratio.

Keywords: Algorithmic Game Theory, Convergence Rate to Equilibria, Logit Dynamics.

[†]DIEM, Università degli Studi di Salerno, Italy. E-mail: `dferraioli@unisa.it`.
[‡]Department of Computer Science, University of Oxford, UK. E-mail: `P.W.Goldberg@cs.ox.ac.uk`.
[§]School of Computing, Teesside University, UK. E-mail: `C.Ventre@tees.ac.uk`.

1 Introduction

Social networks are widespread in physical and digital worlds. The following scenario therefore becomes of interest. Consider a group of individuals, connected in a social network, who are members of a committee, and suppose that each individual has her own opinion on the matter at hand. How can this group of people reach *consensus* in a decentralized way, that is without a centralized authority dictating the actions of each individual? This is a central question in economic theory, especially for processes in which people repeatedly average their own opinions. This line of work, see e.g. [1, 16, 27, 28], is based on a model defined by DeGroot [15]. In this model, each person i holds an *opinion* equal to a real number x_i, which might for example represent a position on a political spectrum. There is an undirected weighted graph $G = (V, E, w)$ representing a social network, and node i is influenced by the opinions of her neighbors in G (the influence of neighbor j is stronger the higher w_{ij} is). At each time step, node i updates her opinion to be a weighted average of her current opinion with the current opinions of her neighbors. A variation of this model of interest to our study is due to Friedkin and Johnsen [25]. In [25] it is additionally assumed that each node i maintains a persistent *internal belief* b_i, which remains constant even as node i updates her overall opinion x_i through averaging. (See Section 2 for the formal framework.)

However, as recently observed by Bindel et al. [10], consensus is hard to reach, the case of political opinions being a prominent example. The authors of [10] justify the absence of consensus by interpreting repeated averaging as a decentralized dynamics for selfish players. Consensus is not reached as players will not compromise further when this increases their *cost*, defined to measure the distance between a player's opinion and (i) her own belief; (ii) her neighbors' opinions. Therefore, these dynamics will converge to an equilibrium in which players might disagree; Bindel et al. study the cost of disagreement by bounding the price of anarchy in this setting.

In this paper, we continue the study of [10] and ask the question of how quickly equilibria are reached by decentralized dynamics in opinion games. We focus on *finite opinion games*, in which players have only a finite number of opinions available: that is, the opinions do not correspond to real numbers as in [15], but they can be chosen only from a set of finitely many alternatives. This is motivated by the fact that in many cases although players have personal beliefs which may assume a continuum of values, they only have a limited number of strategies available. For example, in political elections, people have only a limited number of parties they can vote for and usually vote for the party which is *closer* to their own opinions. Here we simplify further and consider the case where only two opinions are available, possibly corresponding to an individual's voting intention in a referendum that aims to make a binary decision. This setting already has various novel and interesting technical challenges as outlined below. The restriction to two opinions avoids the issue of different costs associated with different pairs of rival opinions; when more than two opinions are considered, different metrics can be used and for each of them different results can be obtained (see, e.g., [14]).

1.1 Our contribution

We firstly note that finite opinion games are potential game [35, 32] thus implying that these games admit pure Nash equilibria. The set of pure Nash equilibria is then characterized (cf. Lemma 2.4). We also notice the interesting fact that while the games in [10] have a price of anarchy of 9/8, our games have unbounded price of anarchy: indeed, with only finite alternatives, it is clearly more difficult (and sometimes impossible) to find an opinion that is able to balance the distance from neighbors and from the internal belief. We additionally prove that the socially optimal profile is always a Nash equilibrium when the weights of the social network are at least 1, thus implying that the Price of Stability is 1 in this case, and in general give a tight bound of 2 for the Price of Stability when edges have fractional weights.

We then study decentralized dynamics for finite opinion games, where players autonomously update their opinions, without the intervention of a central authority. We first consider the best-response dynamics, where at each time step a single player chooses the opinion that minimizes her cost. We prove that this dynamics quickly converges to pure Nash equilibria in the case of unweighted social networks. For general weights, we prove that the convergence rate is polynomial in the number of players but exponential in the representation of the weights. We also prove that for a specific finite opinion game, there exists an exponentially-long sequence of best responses thus implying that convergence may be exponential in general. The upper bounds are proved by "reducing" a finite opinion game to a version of it in which the internal beliefs can only take certain values. The reduced version is equivalent to the original one, as long as best-response dynamics is concerned. Note that the convergence rate for the version of the game considered in [10] is unknown.

In real life, however, there is some noise in the decision process of players. Arguably, people are not fully rational. Alternatively, even if they were, they might not know exactly what strategy represents the best response

to a given strategy profile due to the incapacity to correctly determine their cost functions. To model this, we study *logit dynamics* [11] for finite opinion games. Logit dynamics features a *rationality level* $\beta \geq 0$ (equivalently, a noise level $1/\beta$) and each player is assumed to play a strategy with a probability which is proportional to the corresponding cost to the player and β. So the higher β is, the less noise there is and the more the dynamics is similar to best-response dynamics. We remark that the parameter β is assumed to describe the noise level in the system, and thus every single player is assumed to have the same rationality level. This is aligned with similar models in Physics, where β represents the temperature in the environment that influences the behavior of particles in that environment.

Logit dynamics for potential games defines a Markov chain that has a nice structure. As in [5, 3] we exploit this structure to prove bounds on the convergence rate of logit dynamics to the so-called *logit equilibrium*. The logit equilibrium corresponds to the stationary distribution of the Markov chain. Intuitively, a logit equilibrium is a probability distribution over strategy profiles of the game; the distribution is concentrated around pure Nash equilibrium profiles[1]. It is observed in [5] how this notion enjoys a number of desiderata of solution concepts.

We prove a host of results on the convergence rate of logit dynamics that give a pretty much complete picture as β varies. We give an upper bound in terms of the cutwidth of the graph modeling the social network. The bound is exponential in β and the cutwidth of the graph, thus yielding an exponential guarantee for some topology of the social network. We complement this result by proving a polynomial upper bound when β takes a small value. We complete the preceding upper bound in terms of the cutwidth with lower bounds. Firstly, we prove that in order to get an (essentially) matching lower bound it is necessary to evaluate the size of a certain subset of strategy profiles. When β is big enough relative to this subset then we can prove that the upper bound is tight for any social network (specifically, we roughly need β bigger than $n \log n$ divided by the cutwidth of the graph). For smaller values of β, we are unable to prove a lower bound which holds for every graph. However, we prove that the lower bound holds in this case at both ends of the spectrum of possible social networks. In detail, we look at two cases of graphs encoding social networks: cliques, which model monolithic, highly interconnected societies, and complete bipartite graphs, which model more sparse "antitransitive" societies. For these graphs, we firstly evaluate the cutwidth and then relate the latter to the size of the aforementioned set of states. This allows us to prove a lower bound exponential in β and the cutwidth of the graph for (almost) any value of β. As far as we know, no previous result was known about the cutwidth of a complete bipartite graph; this might be of independent interest. The result on cliques is instead obtained by generalizing arguments in [31].

To prove the convergence rate of logit dynamics to logit equilibrium we adopt a variety of techniques developed to bound the mixing time of Markov chains. To prove the upper bounds we use some spectral properties of the transition matrix of the Markov chain defined by the logit dynamics, and coupling of Markov chains. To prove the lower bounds, we instead rely on the concept of bottleneck ratio and the relation between the latter and mixing time. (The interested reader might refer to [31] for a discussion of these concepts. Below, we give a quick overview of these techniques and state some useful facts.)

1.2 Related works

Best response dynamics and logit dynamics are prominent examples of decentralized dynamics. We refer the interested reader to [18] for an ampler treatment of decentralized dynamics.

A number of papers study the efficient computation of (approximate) pure Nash equilibria for 2-strategy games, such as *party affiliation games* [21, 6] and *cut games* [8]. The class of games we study here contrasts with those in that for the games considered here, Nash equilibria can be found in polynomial time (Observation 2.2), so that our interest is in the extent to which equilibria can be found easily with simple decentralized dynamic processes. Similarly to these works, we focus on a class of 2-strategy games and study efficient computation of pure Nash equilibria; additionally we also study the convergence rate to logit equilibria.

Recently, much attention has been devoted to games played on social networks. In particular, Dyer and Mohanaraj [17] introduced a class of graphical games, called *pairwise-interaction games*, in which players are placed on vertices of a graph and there is a unique game being played on the edges of this graph. They prove, among other results, quick convergence of best-response dynamics for these games. For special cases of pairwise-interaction games where the edge game is a coordination games, hence called *graphical coordination games*, there are several results about the behavior of the logit dynamics [19, 34, 33]. However, none of these works evaluates the time the logit dynamics takes in order to reach the stationary distribution: this line of research is conducted in [5, 3]. The finite opinion games studied here closely resemble these graphical coordination games. However, we

---

[1]Thus, the solution concept of logit dynamics is different from the one associated with best-response dynamics.

highlight that finite opinion games are not pairwise-interaction games. Indeed, in order to encode the belief of a player we need a different game on every edge (see Section 2 for details).

Several relevant works about opinion formation games appeared recently [14, 26, 9, 2]. In particular, Chierichetti et al. [14] also consider finite opinion games, but they try to extend the model to more than two available opinions. Ghaderi and Srikant [26] consider instead the original model of [25] with an unlimited number of available opinions, but, as in this work, they investigate the behavior of decentralized dynamics (specifically, they consider concurrent best-response dynamics).

2 The game

Let $G = (V, E)$ be a connected undirected graph with $|V| = n$ and for each edge $e = (i, j) \in E$ let $w_{ij} > 0$ be its weight. We set $w_{ij} = 0$ if (i, j) is not an edge of G. Every vertex of the graph represents a player. Each player i has an *internal belief* $b_i \in [0, 1]$.

In the original model of [25] considered by [10], it is assumed that each player can choose an opinion $x_i \in [0, 1]$ and the cost of player i in a strategy profile $\mathbf{x} \in \{0, 1\}^n$ is

$$c_i(\mathbf{x}) = (x_i - b_i)^2 + \sum_{(i,j) \in E} w_{ij}(x_i - x_j)^2. \tag{1}$$

In this work we instead assume that for each player there are only two available strategies or *opinions*, namely 0 and 1. However, the cost of player i in a strategy profile $\mathbf{x} \in \{0, 1\}^n$ is still computed according to (1). We call such a game an n-player *finite opinion game* on a graph G. Let $D_i(\mathbf{x}) = \sum_{j\,:\,x_i \neq x_j} w_{ij}$ be the sum of the weights of edges going from i to players with the opposite opinion to i. Then

$$c_i(\mathbf{x}) = (x_i - b_i)^2 + D_i(\mathbf{x}).$$

We note that finite opinion games can be alternatively defined as follows: For every edge $e = (i, j)$, we consider the following two-player two-strategy game:

<div align="center">Player j</div>

		0	1
Player i	0	$\frac{b_i^2}{\Delta_i}, \frac{b_j^2}{\Delta_j}$	$\frac{b_i^2}{\Delta_i} + w_{ij}, \frac{(1-b_j)^2}{\Delta_j} + w_{ij}$
	1	$\frac{(1-b_i)^2}{\Delta_i} + w_{ij}, \frac{b_j^2}{\Delta_j} + w_{ij}$	$\frac{(1-b_i)^2}{\Delta_i}, \frac{(1-b_j)^2}{\Delta_j}$

where Δ_i represents the degree of i and the values in each cell are the cost of i and j, respectively. It is not too hard to see that the cost c_i of player i is the total cost resulting from each subgame in which she is involved.

We remark that, since costs depend on the identity of the endpoints of the edge, we have a different game played on every edge. Hence, finite opinion games are not pairwise-interaction games.

A (pure) *Nash equilibrium* of a finite opinion game is a profile $\mathbf{x}$ such that for every player i and every $y_i \in \{0, 1\}$, it holds that $c_i(\mathbf{x}) \leq c_i(\mathbf{x}_{-i}, y_i)$, where $\mathbf{x}_{-i} = (x_1, \ldots x_{i-1}, x_{i+1}, \ldots, x_n)$.

2.1 Potential Function and Social Cost

A finite opinion game has an *exact potential function* Φ if for every profile $\mathbf{x}$, every player i and every strategy $y_i \in \{0, 1\}$, it holds that $c_i(\mathbf{x}) - c_i(\mathbf{x}_{-i}, y_i) = \Phi(\mathbf{x}) - \Phi(\mathbf{x}_{-i}, y_i)$. Note that every local minimum of a potential function corresponds to a Nash equilibrium of the game.

Let $D(\mathbf{x}) = \sum_{i,j\,:\,x_i \neq x_j} w_{ij}$ be the sum of the weights of *discording edges* in the strategy profile $\mathbf{x}$, that is the weight of all edges in G whose endpoints have different opinions. Thus, $D(\mathbf{x}) = \frac{1}{2} \sum_i D_i(\mathbf{x})$.

Lemma 2.1. *The function*

$$\Phi(\mathbf{x}) = \sum_i (x_i - b_i)^2 + D(\mathbf{x}) \tag{2}$$

is an exact potential function for the finite opinion game described above.

Proof. Given a strategy profile $\mathbf{x}$, player i experiences a non-negative cost. We show that in a finite opinion game if players minimize their cost, then the function Φ defined in (2) decreases. The difference in the cost to player i when she switches from strategy x_i to strategy y_i is

$$c_i(\mathbf{x}) - c_i(\mathbf{x}_{-i}, y_i) = (x_i - b_i)^2 - (y_i - b_i)^2 + D_i(\mathbf{x}) - D_i(\mathbf{x}_{-i}, y_i).$$

The difference in the potential function between the two corresponding profiles is

$$\Phi(\mathbf{x}) - \Phi(\mathbf{x}_{-i}, y_i) = \sum_j (x_j - b_j)^2 + D(\mathbf{x}) - \sum_{j \neq i}(x_j - b_j)^2 - (y_i - b_i)^2 - D(\mathbf{x}_{-i}, y_i)$$
$$= (x_i - b_i)^2 - (y_i - b_i)^2 + D(\mathbf{x}) - D(\mathbf{x}_{-i}, y_i).$$

Discording edges not incident on i are not affected by the deviation of player i. That is, if we let $K_i(\mathbf{x}) = \sum_{j,k \neq i;\ x_j \neq x_k} w_{jk}$ be the sum of the weights of these edges, then $K_i(\mathbf{x}) = K_i(\mathbf{x}_{-i}, y_i)$. The claim then follows since $D(\mathbf{x}) = K_i(\mathbf{x}) + D_i(\mathbf{x})$ and $D(\mathbf{x}_{-i}, y_i) = K_i(\mathbf{x}_{-i}, y_i) + D_i(\mathbf{x}_{-i}, y_i)$. $\qquad\square$

We remark that a similar function has been used in [10] for analyzing the original opinion games with unlimited opinions.

Notice that the potential function Φ looks similar to (but is not the same as) the social cost

$$\mathsf{SC}(\mathbf{x}) = \sum_{i=1}^n c_i(\mathbf{x}) = \sum_i (x_i - b_i)^2 + 2D(\mathbf{x}). \tag{3}$$

2.2 Computational Complexity

In this section we make a couple of observations about the time that it is necessary for a centralized algorithm, i.e., one that is run by a central authority, to compute Nash equilibria and profiles minimizing the social cost.

Observation 2.2. *Nash equilibria of the finite opinion games can be computed by a centralized algorithm in polynomial time.*

Specifically, consider the following simple greedy algorithm. Start with the pure profile $\mathbf{x}$ where everyone plays 0. As long as the current profile is not an equilibrium take an arbitrary player that prefers to take opinion 1 and let her to play her best response.

Notice that if at some time t during the algorithm player i switches her opinion from 0 to 1, then she will prefer 1 in every $t' > t$. This is because from time t onward, the cost of player i decreases as long as the number of her neighbors with opinion 1 increases. Since during the algorithm the neighbors of i can only switch their opinion from 0 to 1, then i will not have any incentive to change her choice at every time $t' > t$. Hence, the algorithm in at most n steps will reach a Nash equilibrium.

Notice that this algorithm finds an equilibrium that maximizes the number of players playing 0, and we could similarly find one that maximizes the number of players playing 1. It does not necessarily find a socially optimal equilibrium, although it follows from Theorem 2.7 and Observation 2.3 below that when edge weights are at least 1, the lowest-cost equilibrium is computable in polynomial time.

As for the computation of the socially optimal profile, we have the following result.

Observation 2.3 ([20, Theorem 2]). *The profile minimizing the social cost can be computed by a centralized algorithm in polynomial time.*

(A proof of this result can also be found in [23].) Specifically, Escoffier et al. [20] showed that it corresponds to the (s, t)-cut of minimum weight in a suitably built graph. This argument can be extended to prove the efficient computation (by a centralized algorithm) of the equilibrium that minimizes the potential function (but not necessarily the social cost).

2.3 Nash equilibria, Price of Anarchy and Price of Stability

We next give a characterization of Nash equilibria. Let B_i be the integer closer to the internal belief of the player i: that is, $B_i = 0$ if $b_i \leq 1/2$, $B_i = 1$ if $b_i > 1/2$. Moreover, let $W_i = \sum_j w_{ij}$ be the total weight of edges

incident on i and $W_i^s(\mathbf{x}) = \sum_{j\,:\,x_j = s} w_{ij}$ be the total weight of edges going from i to players playing strategy s in the profile $\mathbf{x}$.

The following lemma shows that, for every player, it is preferable to select the opinion closer to her own belief if and only if more than (almost) half of her (weighted) neighborhood has selected this opinion.

Lemma 2.4. *In a Nash equilibrium profile* $\mathbf{x}$*, it holds that for each player* i

$$x_i = \begin{cases} B_i, & \text{if } W_i^{B_i}(\mathbf{x}) \geq \frac{W_i}{2} - \delta; \\ 1 - B_i, & \text{if } W_i^{B_i}(\mathbf{x}) \leq \frac{W_i}{2} - \delta; \end{cases}$$

where $\delta = \frac{1}{2} - |B_i - b_i|$.

Proof. Let us start by observing that $|B_i - b_i| = B_i + b_i - 2b_i B_i$ and then

$$(1 - B_i - b_i)^2 = (B_i - b_i)^2 - 2|B_i - b_i| + 1. \tag{4}$$

Now, we first prove that a profile for which the above conditions hold is a Nash equilibrium, and then we prove that every other profile is not in equilibrium.

Let $\mathbf{x}$ be a profile for which the above conditions hold for every player and i be one such player. We consider first the case that $W_i^{B_i}(\mathbf{x}) \geq W_i/2 - \delta$: then we have $x_i = B_i$. There is no incentive for i to play $1 - B_i$ since

$$\begin{aligned} c_i(\mathbf{x}_{-i}, 1 - B_i) &= (1 - B_i - b_i)^2 + W_i^{B_i}(\mathbf{x}) \\ &\leq (B_i - b_i)^2 + \left(\frac{W_i}{2} + \delta\right) \\ &\leq (B_i - b_i)^2 + \left(W_i - W_i^{B_i}(\mathbf{x})\right) = c_i(\mathbf{x}), \end{aligned}$$

where we used (4) for the first inequality. Similarly, we can prove that if $W_i^{B_i}(\mathbf{x}) \leq \frac{W_i}{2} - \delta$, and thus $x_i = 1 - B_i$, then $c_i(\mathbf{x}_{-i}, B_i) \geq c_i(\mathbf{x})$. Hence, no player has incentive to switch her opinion in $\mathbf{x}$ ad thus $\mathbf{x}$ is a Nash equilibrium.

Now consider a profile $\mathbf{y}$ for which the conditions above do not hold for some player i. It must be the case that $W_i^{B_i}(\mathbf{y}) \neq W_i/2 - \delta$. If $W_i^{B_i}(\mathbf{y}) > W_i/2 - \delta$, this means that $y_i = 1 - B_i$; similarly, if $W_i^{B_i}(\mathbf{y}) < \frac{W_i}{2} - \delta$, we have that $y_i = B_i$. However, it is immediate to check that in the former case $c_i(\mathbf{y}) > c_i(\mathbf{y}_{-i}, B_i)$ and in the latter case $c_i(\mathbf{y}) > c_i(\mathbf{y}_{-i}, 1 - B_i)$. $\square$

Informally, Lemma 2.4 identifies the point at which a player's neighbors dictate her strategy and overcome her internal belief.

Price of Anarchy and Stability. Next we evaluate the performance of Nash equilibria with respect to minimization of the social cost. To this aim, we analyse the Price of Anarchy and the Price of Stability of Nash equilibria in finite opinion games. These concepts measure the performance of Nash equilibria by comparing the social cost of an equilibrium with the optimal social cost. Specifically, the *Price of Anarchy* [30] is defined as the ratio between the maximum social cost among the ones achieved in a Nash equilibrium and the optimal social cost. The *Price of Stability* is instead the ratio between the minimum social cost achieved by a Nash equilibrium and the optimal social cost.

Observation 2.5. *The price of anarchy of finite opinion games is unbounded.*

To see this, consider the finite opinion game on a clique where each player has internal belief 0 and each edge has weight 1: the profile where each player has opinion 0 has social cost 0. By Lemma 2.4, the profile where each player has opinion 1 is a Nash equilibrium and its social cost is $n > 0$. This is in sharp contrast with the bound $9/8$ proved in [10], and this difference is motivated by the fact that with limited available opinions, it is more difficult (and, sometimes, as in the above example, impossible) to find an opinion that compromises between the opinion of the neighbors and the internal belief.

We complete this section by proving bounds on the Price of Stability.

Theorem 2.6. *The price of stability of finite opinion games is* 2.

Proof. We begin by proving the upper bound of 2 on the price of stability for any finite opinion game $\mathcal{G}$. Let $\mathbf{x}^\star$ be the profile of $\mathcal{G}$ minimizing the potential function and OPT be the profile minimizing the social cost. Therefore, we have

$$\mathsf{SC}(\mathbf{x}^\star) = \Phi(\mathbf{x}^\star) + D(\mathbf{x}^\star) \leq 2 \cdot \Phi(\mathbf{x}^\star) \leq 2 \cdot \Phi(OPT) \leq 2 \cdot \mathsf{SC}(OPT).$$

where the first equality and the last inequality follow from the definitions of SC and Φ given in (3) and (2), the first inequality uses that $\Phi(\mathbf{x}) = \sum_i (x_i - b_i)^2 + D(\mathbf{x}) \geq D(\mathbf{x})$, and the second inequality follows from the definition of $\mathbf{x}^\star$ and OPT. The price of stability then follows since $\mathbf{x}^\star$ is a Nash equilibrium.

For the lower bound, consider a finite opinion game $\mathcal{G}$ defined on a star-shaped social network with $n+1$ nodes, with $n > 4$, where each edge is weighted $1/n$. Let each external node, but one, have belief 1. The center and the remaining external node have instead belief 0.

We now argue that the social cost is minimized by the profile in which all nodes apart from the center play their own belief. Indeed, its cost is $1 + 2/n = (n+2)/n$ (we have one discording edge weighted $1/n$ and additionally the center has a cost of 1 since she is playing the strategy opposite to her own belief). All profiles in which at least two nodes play the strategy opposite to their belief cost at least $2 > 1 + 2/n$. All profiles in which there is only one player, different from the center, playing opposite to her belief will have $k > 1$ discording edges and then a cost of $1 + 2k/n > (n+2)/n$. Finally, the profile in which each player plays her belief has $n-1$ discording edges for a social cost of $2(n-1)/n = (2n-2)/n > (n+2)/n$.

Now we note that the latter profile, that we will call $\mathbf{x}$, is the unique Nash equilibrium of the game. Indeed, for all nodes but the center, it is a dominant strategy to play their belief (by doing so, they will have a cost of at most $1/n$ while by switching their cost would be at least 1). But then by Lemma 2.4, the center has a strict incentive to play her belief as well since, letting i be the center, $W_i^0(\mathbf{x}) = 1/n > 0 = W_i/2 - \delta$.

Therefore, the price of stability of this game is $2(n-1)/(n+2)$, which approaches 2 as n increases. $\qquad\square$

We show that above bound can be improved when edge weights are at least 1. Indeed, it turns out that, in this special case, the profile that minimizes the social cost is always a Nash equilibrium. Intuitively, this result follows from the fact that high edge weights make the contribution of the term $D(\mathbf{x})$ prevalent with respect to that of $\sum_i (x_i - b_i)^2$ in both the potential function and the social cost. A formal proof is given in the next theorem.

Theorem 2.7. *For a finite opinion game with weights at least* 1*, the price of stability is* 1*.*

Proof. Let $\mathcal{G}$ be an n-player finite opinion game and let $\mathbf{x}$ be the profile that minimizes the social cost of $\mathcal{G}$. Assume for a contradiction that $\mathbf{x}$ is not a Nash equilibrium. This means there is a player i for which the condition of Lemma 2.4 does not hold, i.e., either $x_i = 1 - B_i$ and $W_i^{B_i}(\mathbf{x}) > W_i/2 - \delta$ or $x_i = B_i$ and $W_i^{B_i}(\mathbf{x}) < W_i/2 - \delta$. Let us consider the first case (the second one can be handled similarly): we will show that the profile $(\mathbf{x}_{-i}, B_i)$ achieves a social cost lower than $\mathbf{x}$ and thus a contradiction. We evaluate the difference between $c_j(\mathbf{x}_{-i}, B_i)$ and $c_j(\mathbf{x})$ for each player j. If $j = i$, then

$$c_i(\mathbf{x}_{-i}, B_i) - c_i(\mathbf{x}) = W_i - W_i^{B_i}(\mathbf{x}) + (B_i - b_i)^2 - W_i^{B_i}(\mathbf{x}) - (1 - B_i - b_i)^2 = W_i - 2W_i^{B_i}(\mathbf{x}) + 2|B_i - b_i| - 1,$$

where we used (4). Consider now a neighbor j of i such that $x_j = B_i$. Then,

$$c_j(\mathbf{x}_{-i}, B_i) - c_j(\mathbf{x}) = W_j - W_j^{B_i}(\mathbf{x}) - w_{ij} + (B_i - b_j)^2 - W_j + W_j^{B_i}(\mathbf{x}) - (B_i - b_j)^2 = -w_{ij}.$$

For a neighbor j of i such that $x_j = 1 - B_i$, we obtain

$$c_j(\mathbf{x}_{-i}, B_i) - c_j(\mathbf{x}) = W_j^{B_i}(\mathbf{x}) + w_{ij} + (1 - B_i - b_j)^2 - W_j^{B_i}(\mathbf{x}) - (1 - B_i - b_j)^2 = w_{ij}.$$

Finally, note that players that are not in the neighborhood of i have the same cost in both profiles. Thus, the difference between social costs is:

$$\mathsf{SC}(\mathbf{x}_{-i}, B_i) - \mathsf{SC}(\mathbf{x}) = W_i - 2W_i^{B_i}(\mathbf{x}) + 2|B_i - b_i| - 1 - \sum_{\substack{j\,:\,(i,j)\in E \\ x_j = B_i}} w_{ij} + \sum_{\substack{j\,:\,(i,j)\in E \\ x_j = 1 - B_i}} w_{ij}$$

$$= 2\left(W_i - 2W_i^{B_i}(\mathbf{x}) + |B_i - b_i|\right) - 1.$$

Observe that by definition of B_i, $|B_i - b_i| \leq 1/2$. We now distinguish two cases.

6

If $|B_i - b_i| < 1/2$ then $W_i^{B_i}(\mathbf{x}) > W_i/2 - \delta$ implies $W_i^{B_i}(\mathbf{x}) \geq W_i/2$ whenever weights are at least one. Therefore, $\mathsf{SC}(\mathbf{x}_{-i}, B_i) - \mathsf{SC}(\mathbf{x}) = 2\left(W_i - 2W_i^{B_i}(\mathbf{x}) + |B_i - b_i|\right) - 1 \leq 2|B_i - b_i| - 1 < 0$ and this concludes the proof in this case.

If $|B_i - b_i| = 1/2$ then $W_i^{B_i}(\mathbf{x}) > W_i/2 - \delta$ is equivalent to $W_i^{B_i}(\mathbf{x}) > W_i/2$ and similarly to the case above we can conclude $\mathsf{SC}(\mathbf{x}_{-i}, B_i) - \mathsf{SC}(\mathbf{x}) < 0$. $\qquad\square$

3   Best-response dynamics

In this section, we analyze the behavior of the best-response dynamics for finite opinion games. Best-response dynamics are a prominent decentralized dynamics in which at each time step a player among the ones that can decrease their cost changes her opinion.

We will evaluate the time that this decentralized dynamics takes for converging to a Nash equilibrium. We will show that this time can be exponential in the representation of the weights. We note that this result is in stark contrast with Observation 2.2 that shows that centralized algorithms compute an equilibrium in polynomial time, regardless of the weights' representation.

In order to bound the convergence time of best-response dynamics, we adopt a quite standard approach (see, e.g., [21] and [13]): namely, we bound the minimum decrease of the potential function in each step. Unfortunately, a naive application of this technique does not work: indeed, it may be the case that the decrease in the cost of a player is very small (for example, when a player has the same number of neighbors with opinion 0 and 1 and her belief b_i is close to $\frac{1}{2}$). To address this issue we will not apply the technique on the original game but on an "equivalent" game. Specifically, we use the following definition, previously used in [17].

Definition 3.1. *Two games $\mathcal{G}, \mathcal{G}'$ are* pure best-response equivalent[2] *if they have the same sets of players and pure-strategies, and for any player and pure-strategy profile, that player's best response is the same in $\mathcal{G}$ as in $\mathcal{G}'$.*

We prove bounds on the time the best-response dynamics for a game $\mathcal{G}$ takes to converge by analyzing the dynamics on a game $\mathcal{G}'$ that is pure best-response equivalent to $\mathcal{G}$ but such that beliefs are "nicely" distributed. To describe these beliefs we need the following definitions.

Definition 3.2. *We say that a belief $b \in [0, 1]$ is* threshold *for player i in finite opinion game $\mathcal{G}$ if player i with belief b in $\mathcal{G}$ is indifferent between playing strategy 0 and strategy 1 for some strategies of the players other than i.*

Given a finite opinion game $\mathcal{G}$ and a player i in $\mathcal{G}$, we then define a finite[3] set $\mathcal{B}_i$ of numbers in $[0, 1]$ as follows: Let $\mathcal{B}'_i$ contain 0 and 1 together with every threshold belief of player i; Let $\mathcal{B}_i$ contains every element of $\mathcal{B}'_i$, and in addition, for every pair of consecutive elements b'_1, b'_2 of $\mathcal{B}'_i$, let $\mathcal{B}_i$ contain at least one element in the interval (b'_1, b'_2) (for example, $\frac{1}{2}(b'_1 + b'_2)$).

For example, consider a player i with three neighbors each one connected with i with an edge of weight $1/3$. For this player there are two threshold beliefs: $1/3$ and $2/3$. Indeed, if $b = 1/3$, then i becomes indifferent between opinion 0 and opinion 1, whenever there is exactly one neighbor with opinion 0. Similarly, if $b = 2/3$, then i becomes indifferent between opinion 0 and opinion 1, whenever there is exactly one neighbor with opinion 1. It is not difficult to see that for every other value of b, player i either strictly prefers opinion 0 over opinion 1 or vice versa. Thus, in this case we will have $\mathcal{B}'_i = \{0, 1/3, 2/3, 1\}$ and $\mathcal{B}_i$ can be $\{0, 1/6, 1/3, 1/2, 2/3, 5/6, 1\}$.

Then we have the following lemma.

Lemma 3.3. *Any finite opinion game $\mathcal{G}$ is pure best-response equivalent to a finite opinion game $\mathcal{G}'$ in which the beliefs of every player i in $\mathcal{G}$ have been replaced by an element of $\mathcal{B}_i$.*

Proof. Fix a finite opinion game $\mathcal{G}$ and player i. Let $\mathcal{G}'$ be a finite opinion game defined on the same social network as $\mathcal{G}$. For each player i in $\mathcal{G}$, if her belief b_i is threshold, then it belongs to $\mathcal{B}_i$, and we keep it the same in $\mathcal{G}'$. If b_i is not a threshold belief, then in $\mathcal{G}'$ it is replaced by one of the elements of $\mathcal{B}_i$ that lies in the subinterval bounded by 2 consecutive elements of $\mathcal{B}'_i$ containing b_i.

We claim that $\mathcal{G}'$ constructed this way, is pure best-response equivalent to $\mathcal{G}$. Consider a player i, and note that for a pair of beliefs b_i and b'_i to result in different best responses, there must be a strategy profile for the remaining players for which the best response under belief b_i is opposite to the one under belief b'_i. Note however that a player's beliefs have been changed in a way that ensures that there is no such pair of pure-strategy profiles. $\qquad\square$

---

[2]In [17] they just use "equivalent".

[3]Note that for every set of strategies played by the neighbors of i, there is a unique threshold whose value is defined by the edge weights (and, clearly, those strategies). Since there are 2^{Δ_i} different strategy profiles for i's neighbors, then it is always possible to define a finite $\mathcal{B}_i$.

3.1 A special case: unitary weights

We start by considering the special case in which $w_{ij} = 1$ for each edge (i, j). This helps to develop the ideas that we use to prove a bound for more general weights.

For a player i, consider $\mathcal{B}'_i$ as defined in Definition 3.2. In this special case, it is easy to see that $\mathcal{B}'_i = \{0, 1\}$ if the neighborhood of i has odd size and $\mathcal{B}'_i = \{0, \frac{1}{2}, 1\}$, otherwise. Thus, by Lemma 3.3, in both cases, $\mathcal{G}$ is pure best-response equivalent to a finite opinion game $\mathcal{G}'$ where each player i has belief in $\mathcal{B}_i = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. The following theorem shows that the best-response dynamics quickly converges to a Nash equilibrium in $\mathcal{G}'$ and hence in $\mathcal{G}$.

Theorem 3.4. *The best-response dynamics for an n-player finite opinion game converges to a Nash equilibrium after a number of steps that is polynomial in n.*

Proof. Let $\mathcal{G}$ be an n-player finite opinion game and let $\mathcal{G}'$ be a pure best-response equivalent game having beliefs in the set $\mathcal{B}_i = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. We show that best-response dynamics converges quickly on $\mathcal{G}'$.

We begin by observing that for every profile $\mathbf{x}$, we have $0 \leq \Phi(\mathbf{x}) \leq W + n$, where $W = \sum_{(i,j) \in E} w_{ij} \leq n^2$. Thus, the theorem follows by showing that at each time step the cost of a player decreases by at least a constant value. Fix $\mathbf{x}_{-i}$, the opinions of players other than i, and let x_i be the strategy currently played by player i and s be her best response. By definition of best response, we have $c_i(\mathbf{x}) > c_i(s, \mathbf{x}_{-i})$. We will show that

$$
\begin{aligned}
\frac{1}{2} \leq \Phi(\mathbf{x}) - \Phi(s, \mathbf{x}_{-i}) &= c_i(x_i, \mathbf{x}_{-i}) - c_i(s, \mathbf{x}_{-i}) = (x_i - b_i)^2 - (s - b_i)^2 + D_i(\mathbf{x}) - D_i(s, \mathbf{x}_{-i}) \\
&= 2|x_i - b_i| - 1 + \Delta D,
\end{aligned}
\tag{5}
$$

where the last equality follows by (4) (with x_i in place of B_i) and by using ΔD as a shorthand for $D_i(\mathbf{x}) - D_i(s, \mathbf{x}_{-i})$. We distinguish three cases based on the value of ΔD.

If $\underline{\Delta D > 1}$, then it will be the case that $\Delta D \geq 2$. Since the difference between the squares is bounded from below by -1, then (5) follows.

If $\underline{-1 < \Delta D \leq 1}$, then it will be the case that $\Delta D \in \{0, 1\}$. If $x_i = 0$, then $c_i(x_i, \mathbf{x}_{-i}) > c_i(s, \mathbf{x}_{-i})$ implies $b_i > \frac{1-\Delta D}{2}$ and, from $b_i \in \mathcal{B}_i$, $b_i \geq \frac{1-\Delta D+1/2}{2}$; for these values of b_i and x_i (5) follows. Similarly, if $x_i = 1$, then $c_i(x_i, \mathbf{x}_{-i}) > c_i(s, \mathbf{x}_{-i})$ implies $b_i < \frac{1+\Delta D}{2}$ and, from $b_i \in \mathcal{B}_i$, $b_i \leq \frac{1+\Delta D-1/2}{2}$; for these values of b_i and x_i (5) follows.

If $\underline{\Delta D \leq -1}$, then we will reach a contradiction. Indeed, since the difference between the squares is bounded from above by 1, we have $c_i(x_i, \mathbf{x}_{-i}) \leq c_i(s, \mathbf{x}_{-i})$. $\qquad \square$

3.2 General weights

We now show how to extend the bound for the convergence of the best-response dynamics to general weights. Our bound will depend on the precision k needed to represent the weights, i.e., the maximum number k of digits after the decimal point in the decimal representation of the weights.

Given a finite opinion game $\mathcal{G}$, we consider a game $\mathcal{G}'$ that is exactly the same as $\mathcal{G}$ except that in $\mathcal{G}'$ the cost of player i in the profile $\mathbf{x}$ is

$$
c'_i(\mathbf{x}) = 10^k \cdot c_i(\mathbf{x}).
$$

We say $\mathcal{G}'$ is the *integer version* of $\mathcal{G}$. Obviously, $\mathcal{G}$ is pure best-response equivalent to $\mathcal{G}'$ and $\mathcal{G}'$ is a potential game with potential function $\Phi' = 10^k \cdot \Phi$. Note that $\mathcal{G}'$ can be equivalently described as follows: each player has two strategies, 0 and 1, and a personal belief b_i as in $\mathcal{G}$; for each edge e of the social graph of $\mathcal{G}$, we set $w'_e = 10^k \cdot w_e$ and $D'_i(\mathbf{x}) = \sum_{j:\, x_i \neq x_j} w'_{ij}$. Then

$$
c'_i(\mathbf{x}) = 10^k(x_i - b_i)^2 + D'_i(\mathbf{x}).
$$

Note that w'_e is an integer for each edge e. Then we have the following lemma.

Lemma 3.5. *Consider a finite opinion game $\mathcal{G}$ and for each player i consider the set $\mathcal{B}'_i$ as defined in Definition 3.2. Then for each $b \in \mathcal{B}'_i$, we have that $10^k \cdot 2b$ is an integer.*

Proof. If $b \in \{0, 1\}$, then the lemma trivially follows. As for b being a threshold belief, we distinguish two cases: if $b \leq 1/2$, then, by Lemma 2.4, there is a profile $\mathbf{x}$ such that

$$b_i = \frac{1}{2} - \frac{1}{2} \sum_j w_{ij} + \sum_{j \,:\, x_j = 0} w_{ij}.$$

Hence,

$$10^k \cdot 2b_i = 10^k - \sum_j w'_{ij} + 2 \sum_{j \,:\, x_j = 0} w'_{ij}.$$

Since each term in the right-hand side of the last equation is an integer then so is also $10^k \cdot 2b_i$. The case $b > 1/2$ can be handled similarly. $\qquad\square$

Now we are ready for proving a bound on the convergence time of the best-response dynamics.

Theorem 3.6. *The best-response dynamics for an n-player finite opinion game $\mathcal{G}$ whose edge weights have bounded precision k converges to a Nash equilibrium in $\mathcal{O}(10^k \cdot n^2 \cdot w_{\max})$, where $w_{\max}$ is the largest edge weight in $\mathcal{G}$.*

Proof. Fix the finite opinion game $\mathcal{G}$ and for each player i consider the set $\mathcal{B}'_i$ as defined in Definition 3.2. Consider, moreover, $\mathcal{B}_i$ containing every element of $\mathcal{B}'_i$, and in addition, the element $\frac{1}{2}(b'_1 + b'_2)$ for every pair of consecutive elements b'_1, b'_2 of $\mathcal{B}'_i$.

Let $\mathcal{G}'$ be a finite opinion game such that each player i has belief in $\mathcal{B}_i$. From Lemma 3.3, $\mathcal{G}$ and $\mathcal{G}'$ are pure best-response equivalent games. Moreover consider $\mathcal{G}''$ the integer version of $\mathcal{G}'$, which is pure best-response equivalent to $\mathcal{G}'$ and then to $\mathcal{G}$. Below all the notation defined so far uses a double prime when it refers to $\mathcal{G}''$. We show that best-response dynamics converges in $\mathcal{O}(10^k \cdot n^2 \cdot w_{\max})$ steps on $\mathcal{G}''$.

We begin by observing that for every profile $\mathbf{x}$, we have $0 \leq \Phi''(\mathbf{x}) \leq 10^k \left(\sum_{(i,j) \in E} w_{ij} + n \right) = \mathcal{O}(10^k \cdot n^2 \cdot w_{\max})$. Thus, the theorem follows by showing that at each time step the cost of a player decreases by at least a constant value. Fix $\mathbf{x}_{-i}$, the opinions of players other than i, and let x_i be the strategy currently played by player i and s be her best response. From the definition of best response, it follows that $c''_i(\mathbf{x}) > c''_i(s, \mathbf{x}_{-i})$. We will show that

$$
\begin{aligned}
1/2 \leq \Phi''(\mathbf{x}) - \Phi''(s, \mathbf{x}_{-i}) &= c''_i(x_i, \mathbf{x}_{-i}) - c''_i(s, \mathbf{x}_{-i}) \\
&= 10^k \left((x_i - b_i)^2 - (s - b_i)^2 \right) + D''_i(\mathbf{x}) - D''_i(s, \mathbf{x}_{-i}) = 10^k \left(2|x_i - b_i| - 1 \right) + \Delta D,
\end{aligned}
\tag{6}
$$

where the last equality follows by (4) (with x_i in place of B_i) and by using ΔD as a shorthand for $D''_i(\mathbf{x}) - D''_i(s, \mathbf{x}_{-i})$. We distinguish three cases based on the value of ΔD.

If $\underline{\Delta D > 10^k}$, then, since all edge weights are integers, it will be the case that $\Delta D \geq 10^k + 1$. Since the difference between the squares is bounded from below by -1, then (6) follows.

If $\underline{-10^k < \Delta D \leq 10^k}$, then, since all edge weights are integers, we have $\Delta D \in \{-10^k + 1, \dots, 10^k\}$. If $x_i = 0$, then $c''_i(x_i, \mathbf{x}_{-i}) > c''_i(s, \mathbf{x}_{-i})$ implies $b_i > \frac{10^k - \Delta D}{2 \cdot 10^k}$. Since $10^k \cdot 2b$ is an integer for each threshold belief b, the smallest threshold belief greater than $\frac{10^k - \Delta D}{2 \cdot 10^k}$ should be at least $\frac{10^k - \Delta D + 1}{2 \cdot 10^k}$. Hence, the first element of $\mathcal{B}_i$ greater than $\frac{10^k - \Delta D}{2 \cdot 10^k}$ will be at least

$$\frac{1}{2} \left(\frac{10^k - \Delta D}{2 \cdot 10^k} + \frac{10^k - \Delta D + 1}{2 \cdot 10^k} \right) = \frac{10^k - \Delta D + 1/2}{2 \cdot 10^k}.$$

Thus, if $x_i = 0$, then $b_i \geq \frac{10^k - \Delta D + 1/2}{2 \cdot 10^k}$ and (6) follows. Similarly, one can prove that if $\mathbf{x}_i = 1$, then $b_i \leq \frac{10^k + \Delta D - 1/2}{2 \cdot 10^k}$ and (6) follows.

If $\underline{\Delta D \leq -10^k}$, then we will reach a contradiction. Indeed, since the difference between the squares is bounded from above by 1, we have $c_i(x_i, \mathbf{x}_{-i}) \leq c_i(s, \mathbf{x}_{-i})$. $\qquad\square$

The bound on the convergence time of the best response dynamics given in previous theorem can be very large if the edge weights are very large or they need high precision. However, in the next section we will show that, in these cases, such a large bound cannot be avoided. On the positive side, we remark that our bounds hold regardless

of the order in which player are selected for playing the best-responses. In particular, they hold even if this order is chosen adversarially.

Interestingly, the machinery used for proving Theorem 3.6 can be useful for proving that best-response dynamics can converge in time that is polynomial on the number of players and in the maximum weight to ε-*approximate equilibria*, i.e., profiles in which no player can improve her utility by more than ε by changing their strategy. Specifically, we have the following theorem.

Theorem 3.7. *The best-response dynamics for an n-player finite opinion game $\mathcal{G}$ converges to a ε-approximate equilibrium in $\mathcal{O}(\log \frac{1}{\varepsilon} \cdot n^2 \cdot w_{\max})$, where $w_{\max}$ is the largest edge weight in $\mathcal{G}$.*

Indeed, one can define an approximate integer version $\mathcal{G}'$ of the finite opinion game $\mathcal{G}$, by setting the cost of player i to $c_i'(\mathbf{x}) = \left\lceil 10^{\left\lceil \log_{10} \frac{1}{\varepsilon} \right\rceil} \cdot c_i(\mathbf{x}) \right\rceil$. Observe that if in the original game $\mathcal{G}$ there is a player that has a best response that diminishes her cost by more than ε, then this strategy remains a best-response for the player in the game $\mathcal{G}'$. Hence, every equilibrium of $\mathcal{G}'$ must be an ε-approximate equilibrium of $\mathcal{G}$. Then, by repeating the arguments of Theorem 3.6 one can prove that the best response takes $\mathcal{O}(\log_{10} \frac{1}{\varepsilon} \cdot n^2 \cdot w_{\max})$ steps to converge to a Nash equilibrium of $\mathcal{G}'$.

3.3 Exponentially many best-response steps for general weights

The following result builds a game with an exponentially large gap between the largest and the smallest edge weight for which there exist exponentially long sequences of best responses, where the choice of the player switching her strategy at each step is made by an adversary. Thus it remains an open question whether exponentially-many steps may be required if, for example, players were allowed to best-respond in round robin manner, or if the best-responding player was chosen randomly at each step. However this does establish that the potential function on its own is insufficient to bound from above the number of steps with any polynomial. The construction uses graphs with bounded degree and path-width, with all players having belief $\frac{1}{2}$.

Theorem 3.8. *The best-response dynamics for finite opinion games may take exponentially many steps.*

Proof. In the following construction, all players have a belief of $\frac{1}{2}$.

We start by giving some preliminary definitions. A *6-gadget G* is a set of 6 players $\{A, B, C, D, E, F\}$ with edges (A, B), (B, C) and (C, D) having weights ε, 2ε, 3ε respectively and edges (D, E), (B, F) and (D, F), all weighting 4ε, for some $\varepsilon > 0$ (see Figure 1).
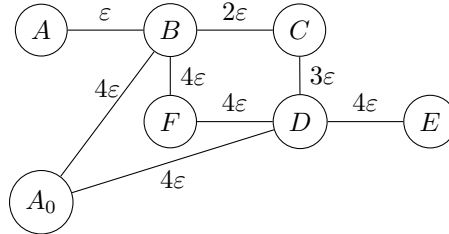

Figure 1: A 6-gadget G with its switch

Consider a 6-gadget and a new player A_0 with edges (A_0, B) and (A_0, D) of weight 4ε. We say that A_0 is a *switch* for the 6-gadget G, since it allows G to switch between the opinion vectors $(0, 0, 0, 0, 0, 1)$ and $(0, 1, 0, 1, 0, 1)$. Specifically, if initially the players (A, B, C, D, E, F) have opinions $(0, 0, 0, 0, 0, 1)$ and A_0 is set to 1, then we can have the following best-response sequence in the 6-gadget, that will be named *switch-on cycle*:

$$(0, 0, 0, 0, 0, 1) \to (0, 1, 0, 0, 0, 1) \to (0, 1, 0, 1, 0, 1).$$

Indeed, in the first step B reduces her cost from $\frac{1}{4} + 8\varepsilon$ to $\frac{1}{4} + 3\varepsilon$, whereas in the second step D reduces her cost from $\frac{1}{4} + 8\varepsilon$ to $\frac{1}{4} + 7\varepsilon$. If instead the players (A, B, C, D, E, F) have opinions $(0, 1, 0, 1, 0, 1)$ and A_0 is set to 0, then we can have the following best-response sequence in the 6-gadget, that will be named *switch-off cycle*:

$$(0, 1, 0, 1, 0, 1) \to (1, 1, 0, 1, 0, 1) \to (1, 0, 0, 1, 0, 1) \to (0, 0, 0, 1, 0, 1) \to (0, 0, 1, 1, 0, 1) \to (0, 1, 1, 1, 0, 1)$$
$$\to (1, 1, 1, 1, 0, 1) \to (1, 1, 1, 0, 0, 1) \to (1, 1, 0, 0, 0, 1) \to (1, 0, 0, 0, 0, 1) \to (0, 0, 0, 0, 0, 1).$$

10

To see that in the switch-off cycle all steps consists of best-responses, observe that: in the first step, A reduces her utility from $\frac{1}{4} + \varepsilon$ to $\frac{1}{4}$; in the second step, B reduces her utility from $\frac{1}{4} + 6\varepsilon$ to $\frac{1}{4} + 5\varepsilon$; in the third step, A reduces her utility from $\frac{1}{4} + \varepsilon$ to $\frac{1}{4}$; in the fourth step, C reduces her utility from $\frac{1}{4} + 3\varepsilon$ to $\frac{1}{4} + 2\varepsilon$; in the fifth step, B reduces her utility from $\frac{1}{4} + 6\varepsilon$ to $\frac{1}{4} + 5\varepsilon$; in the sixth step, A reduces her utility from $\frac{1}{4} + \varepsilon$ to $\frac{1}{4}$; in the seventh step, D reduces her utility from $\frac{1}{4} + 8\varepsilon$ to $\frac{1}{4} + 7\varepsilon$; in the eighth step, C reduces her utility from $\frac{1}{4} + 3\varepsilon$ to $\frac{1}{4} + 2\varepsilon$; in the ninth step, B reduces her utility from $\frac{1}{4} + 6\varepsilon$ to $\frac{1}{4} + 5\varepsilon$; in the last step, A reduces her utility from $\frac{1}{4} + \varepsilon$ to $\frac{1}{4}$.

Notice also that A's opinion does not change in the switch-on cycle, whereas it follows the sequence $0 \to 1 \to 0 \to 1 \to 0$ during the switch-off cycle.

We now define the game. Consider n 6-gadgets G_i with players $\{A_i, B_i, C_i, D_i, E_i, F_i\}$ and edge weights parametrized by ε_i, where $1 \leq i \leq n$. For each $i = 2, \ldots, n$, we connect G_i with G_{i-1} by having A_{i-1} acting as a switch for G_i. We finally add the switch player A_0 for G_1. Thus, the total number of players is $6n + 1$.

In order that the behavior of A_i in gadget G_i, for $i = 1, \ldots, n-1$ is not influenced by the new edges, (A_i, B_{i+1}) and (A_i, D_{i+1}) of weight $4\varepsilon_{i+1}$, we need that the weights of these edges are small enough with respect to the weight of the unique edge incident on A_i in the gadget G_i, namely (A_i, B_i) of weight ε_i. Specifically, it is sufficient to set $\varepsilon_i > 8\varepsilon_{i+1}$. Hence, the largest edge-weight is $4\varepsilon_1$ and the smallest edge-weight is ε_n and their ratio is greater than $4 \cdot 8^{n-1} = 2^{3n-1}$.

Consider now the following starting profile: players B_1 and D_1 have opinion 1, all players F_i have opinion 1, and all other players start with opinion 0. Note that G_1 is in the starting configuration of a switch-off cycle.

We finally specify an exponentially-long sequence of best-response: we start the switch-off cycle of G_1; as long as A_i, for $i = 1, \ldots, n-1$, switches her opinion from 0 to 1, we execute the switch-on cycle of G_{i+1}; as long as A_i, for $i = 1, \ldots, n-1$, switches her opinion from 1 to 0, we execute the switch-off cycle of G_{i+1}. Note that the last two cases occur two times during the switch-off cycle of G_i. Thus, G_2 goes through 2 switch-on cycles and 2 switch-off cycles, G_3 goes through 4 switch-on cycles and 4 switch-off cycles and, hence, G_n goes through 2^{n-1} switch-on cycles and 2^{n-1} switch-off cycles. $\qquad\square$

4 Logit Dynamics for Finite Opinion Games

Let $\mathcal{G}$ be a finite opinion game and let $S = \{0,1\}^n$ denote the set of all strategy profiles. For two vectors $\mathbf{x}, \mathbf{y} \in S$, we denote with $H(\mathbf{x}, \mathbf{y}) = |\{i \colon x_i \neq y_i\}|$ the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$. The *Hamming graph* of the game $\mathcal{G}$ is defined as $\mathcal{H} = (S, \mathsf{E})$, where two profiles $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) \in S$ are adjacent in $\mathcal{H}$ if and only if $H(\mathbf{x}, \mathbf{y}) = 1$.

The *logit dynamics* is a decentralized dynamics that models the behavior of players that can sometimes take wrong decisions. In the logit dynamics for a finite opinion game $\mathcal{G}$, at each time step a player i is selected uniformly at random. The selected player is assumed to update her strategy according to the *Boltzmann distribution* with parameter β over the set $S_i = \{0, 1\}$ of her strategies. That is, a strategy $s_i \in S_i$ will be selected with probability

$$\sigma_i(s_i \mid \mathbf{x}_{-i}) = \frac{1}{Z_i(\mathbf{x}_{-i})} e^{-\beta c_i(\mathbf{x}_{-i}, s_i)}, \tag{7}$$

where $\mathbf{x}_{-i} \in \{0,1\}^{n-1}$ is the profile of strategies played at the current time step by players other than i, $Z_i(\mathbf{x}_{-i}) = \sum_{z_i \in S_i} e^{-\beta c_i(\mathbf{x}_{-i}, z_i)}$ is the normalizing factor, and $\beta \geq 0$. As mentioned above, from (7), it is easy to see that for $\beta = 0$ player i selects her strategy uniformly at random, for $\beta > 0$ the probability is biased toward strategies promising higher payoffs, and for β that goes to ∞ player i chooses her best response strategy (if more than one best response is available, she chooses one of them uniformly at random).

The above dynamics defines a *logit dynamics Markov chain* $\{X_t\}_{t \in \mathbb{N}}$ on the set of strategy profiles, where the probability $P(\mathbf{x}, \mathbf{y})$ of a transition from profile $\mathbf{x} = (x_1, \ldots, x_n)$ to profile $\mathbf{y} = (y_1, \ldots, y_n)$ is zero if $H(\mathbf{x}, \mathbf{y}) \geq 2$ and it is $\frac{1}{n}\sigma_i(y_i \mid \mathbf{x}_{-i})$ if the two profiles differ exactly at player i. Hence, we have that $P(\mathbf{x}, \mathbf{x}) = \frac{1}{n}\sum_i \sigma_i(x_i \mid \mathbf{x}_{-i})$.

The logit dynamics Markov chain is *ergodic*, i.e. there is T such that for every pair of profiles $\mathbf{x}, \mathbf{y} \in S$ there is a positive probability that the logit dynamics starting from $\mathbf{x}$ reaches $\mathbf{y}$ in T steps. In particular, one can check that this property holds for $T = n$. Hence, from every initial profile $\mathbf{x}$ the distribution $P^t(\mathbf{x}, \cdot)$ of chain X_t starting at $\mathbf{x}$ will eventually converge to a *stationary distribution* π as t tends to infinity.[4] As in [5], we call the stationary

---

[4] The notation $P^t(\mathbf{x}, \cdot)$, standard in Markov chains literature [31], denotes the probability distribution over states of S after the chain has taken t steps starting from $\mathbf{x}$.

distribution π of the logit dynamics Markov chain on a game $\mathcal{G}$, the *logit equilibrium* of $\mathcal{G}$. It is known [11] that the logit equilibrium of a finite opinion games is the well known *Gibbs measure*, i.e.

$$\pi(\mathbf{x}) = \frac{1}{Z} e^{-\beta \Phi(\mathbf{x})} \tag{8}$$

where $Z = \sum_{\mathbf{y} \in S} e^{-\beta \Phi(\mathbf{y})}$ is the normalizing constant.

It is easy to check that the logit dynamics Markov chain for a finite opinion game $\mathcal{G}$ is *reversible* with respect to the logit equilibrium π. That is, if, for all $\mathbf{x}, \mathbf{y} \in S$, it holds that

$$\pi(\mathbf{x})P(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y})P(\mathbf{y}, \mathbf{x}).$$

In this work we bound the rate of convergence of the logit dynamics for a finite opinion game $\mathcal{G}$ to its logit equilibrium, that is the first time step t_{mix}, known as *mixing time*, in which the distance between the distribution of the logit dynamics Markov chain after t steps and the logit equilibrium is below a given parameter ε.

For more details on logit dynamics and logit equilibria and for a discussion of some of their properties we refer the interested reader to [5].

4.1   Known mixing time results

Auletta et al. [3] give bounds on the mixing time of the logit dynamics for potential games. Specifically, they give two bounds that hold for every β: $\text{poly}(n) \cdot e^{\beta \Delta \Phi}$ [3, Theorem 3.3], where $\Delta \Phi = \max_{\mathbf{x}, \mathbf{y}} (\Phi(\mathbf{x}) - \Phi(\mathbf{y}))$, and $\exp(n) \cdot e^{\beta \zeta}$ [3, Theorem 3.5], where ζ is the maximum variation in the potential function that is necessary to observe if one wants to connect any pair of profiles (see [3, Section 3.4] for a formal definition). Note that these two bounds are not comparable: indeed, ζ can be extremely smaller than $\Delta \Phi$ and, hence, one can have that the first bound is better than the second for small β, whereas the latter becomes more stringent when β is large.

Since a finite opinion game is a potential game all these bounds hold also in our setting. However, we will show that for this class of games better bounds (actually tight ones) can be given. Specifically, we prove that for every β it holds that the mixing time of the logit dynamics for potential games is bounded from above by $\text{poly}(n) \cdot e^{\beta \Theta \text{CW}(G)}$, where $\text{CW}(G)$ denotes the cutwidth of the network G underlying the game, and it is defined as follows: Consider the bijective function $\sigma \colon V \to \{1, \dots, |V|\}$ representing an ordering of vertices of G; let $\mathcal{L}$ be the set of all orderings of vertices of G and set $V_i^\sigma = \{v \in V \colon \sigma(v) < i\}$; for any partition (L, R) of V let $W(L, R)$ be the sum of the weights of edges that have an endpoint in L and the other one in R; then, the (weighted) *cutwidth* of G is

$$\text{CW}(G) = \min_{\sigma \in \mathcal{L}} \max_{1 < i \leq |V|} W(V_i^\sigma, V \setminus V_i^\sigma).$$

It is not hard to see that $\text{CW}(G)$ can be at least a linear factor smaller than $\Delta \Phi$. Consider, for example, a finite opinion game played on a cycle with an even number of vertices and with all players having belief 0. Clearly, the potential of the profile $\mathbf{0}$ in which all players adopt opinion 0 is exactly 0. On the other hand, the profile $\mathbf{x}$ in which players on the cycle alternatively select opinion 0 and 1 has potential value $\Phi(\mathbf{x}) = \frac{3n}{2}$. Hence, in this case $\Delta \Phi \geq \frac{3n}{2}$. Nevertheless, it is well known that the cutwidth of the cycle is 2. This implies that our bound can improve on the first of the bounds described above.

As for the second bound in [3], it turns out that $\zeta = \Theta(\text{CW}(G))$. Hence, even the second bound described above, that is exponential in the number of players, is always worse than the one discussed in this work.

Moreover, our bound also has another advantage with respect to the bounds given in [3] for general potential games. While theirs refer to abstract properties of the potential functions, our bounds describe the topological properties of the networks that influence the rate of convergence of the logit dynamics.

In [3], it is also proved that for every potential game the logit dynamics takes $\mathcal{O}(n \log n)$ steps to converge to the logit equilibrium, whenever $\beta \leq \frac{c}{n \delta \Phi}$ [3, Theorem 3.4], where $c < 1$ and $\delta \Phi = \max_{\mathbf{x}, \mathbf{y} \colon H(\mathbf{x}, \mathbf{y}) = 1} (\Phi(\mathbf{x}) - \Phi(\mathbf{y}))$. Observe that in finite opinion games it holds that $\delta \Phi \geq w_{\max}$ and thus a direct application of the result of [3] gives a $\mathcal{O}(n \log n)$ bound to the mixing time of the logit dynamics for finite opinion games only for $\beta < \frac{1}{n w_{\max}}$. In this work, we will prove that the $\mathcal{O}(n \log n)$ bound can proved to hold for every $\beta \leq \frac{1}{\Delta_{\max} w_{\max}}$, where $\Delta_{\max}$ be the maximum degree in the graph. Note that this means that for graphs of bounded degree the range of values of β for which the logit dynamics converges to the equilibrium very quickly is substantially larger than what was previously known.

In [3] several bounds are given about the convergence rate of logit dynamics to logit equilibria in graphical coordination games. We note that these results do not extend to finite opinion game, since, as stated above, they do

not belong to the class of graphical coordination games. Still, as we will show later, some of the tools developed for the latter class of games will be useful also for proving our results.

4.2 Techniques

To derive our bounds, we employ several different techniques: *Markov chain coupling* and *spectral techniques* for the upper bound and *bottleneck ratio* for the lower bound. They are well-established techniques for bounding the mixing time; we next summarize them.

4.2.1 Markov Chain Coupling

A *coupling* of two probability distributions μ and ν on S is a pair of random variables (X, Y) defined on $S \times S$ such that the marginal distribution of X is μ and the marginal distribution of Y is ν. A *coupling of a Markov chain* $\mathcal{M}$ on S with transition matrix P is a process $(X_t, Y_t)_{t=0}^{\infty}$ with the property that X_t and Y_t are both Markov chains with transition matrix P and state space S. When the two coupled chains $(X_t, Y_t)_{t=0}^{\infty}$ start at $(X_0, Y_0) = (\mathbf{x}, \mathbf{y})$, we write $\mathbf{P_{x,y}}(\cdot)$ and $\mathbf{E_{x,y}}[\cdot]$ for the probability and the expectation on the space $S \times S$. We denote by τ_{couple} the first time the two chains meet; that is,
$$\tau_{\text{couple}} = \min\{t\colon X_t = Y_t\}.$$
We consider only couplings of Markov chains with the property that $X_s = Y_s$ for $s \geq \tau_{\text{couple}}$.

Recall that $\mathcal{H} = (S, \mathsf{E})$ is the Hamming graph; for $\mathbf{x}, \mathbf{y} \in S$, we denote by $\rho(\mathbf{x}, \mathbf{y})$ the length of the shortest path in $\mathcal{H}$ between $\mathbf{x}$ and $\mathbf{y}$. Then we have the following theorem.

Theorem 4.1 (Path Coupling [12])**.** *Suppose that for every edge* $(\mathbf{x}, \mathbf{y}) \in \mathsf{E}$ *a coupling* (X_t, Y_t) *of* $\mathcal{M}$ *with* $X_0 = \mathbf{x}$ *and* $Y_0 = \mathbf{y}$ *exists such that* $\mathbf{E_{x,y}}[\rho(X_1, Y_1)] \leq e^{-\alpha}$ *for some* $\alpha > 0$. *Then*
$$t_{\text{mix}} = \mathcal{O}\left(\frac{\log(\operatorname{diam}(\mathcal{H}))}{\alpha}\right)$$
where $\operatorname{diam}(\mathcal{H})$ *is the diameter of* $\mathcal{H}$.

4.2.2 Spectral Techniques

Let P and S be the transition matrix and the state space of the logit dynamics Markov chain for a finite opinion game. We denote the *eigenvalues* of P as
$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|S|}.$$
The mixing time is related to these eigenvalues by the following theorem.

Theorem 4.2.
$$t_{\text{mix}} \leq \log\left(\frac{4}{\pi_{\min}}\right) \frac{1}{1 - \max_{i=2,\ldots,|S|} |\lambda_i|},$$
where $\pi_{\min} = \min_{\mathbf{x} \in S} \pi(\mathbf{x})$.

In order to use the theorem above we need to bound these eigenvalues. To this aim, we will use the following lemma.

Lemma 4.3. *Let* $\mathcal{G}$ *be an* n-*player finite opinion game and let* π *be the stationary distribution of the logit dynamics for* $\mathcal{G}$. *For every pair of profiles* $\mathbf{x}, \mathbf{y}$ *we assign a path* $\Gamma_{\mathbf{x},\mathbf{y}}$ *on the Hamming graph* $\mathcal{H}$. *Then*
$$\frac{1}{1 - \max_{i=2,\ldots,|S|} |\lambda_i|} \leq 2n \max_{\substack{\mathbf{z},\mathbf{w}: \\ H(\mathbf{z},\mathbf{w})=1 \\ \pi(\mathbf{z}) \leq \pi(\mathbf{w})}} \frac{1}{\pi(\mathbf{z})} \sum_{\substack{\mathbf{x},\mathbf{y}: \\ (\mathbf{z},\mathbf{w}) \in \Gamma_{\mathbf{x},\mathbf{y}}}} \pi(\mathbf{x})\pi(\mathbf{y})|\Gamma_{\mathbf{x},\mathbf{y}}|.$$

Proof. Let S and P be respectively the state space and the transition matrix of the Markov chain associated to the logit dynamics for $\mathcal{G}$.

In [3] it is proved that all the eigenvalues of P are non-negative. It follows then that $\frac{1}{1 - \max_{i=2,\ldots,|S|} |\lambda_i|} = \frac{1}{1-\lambda_2}$. Moreover, by reversibility of P, we have that for $\mathbf{z}, \mathbf{w} \in S$ such that $H(\mathbf{z}, \mathbf{w}) = 1$ and $\pi(\mathbf{z}) \leq \pi(\mathbf{w})$ it holds that
$$\pi(\mathbf{z})P(\mathbf{z}, \mathbf{w}) = \pi(\mathbf{w})P(\mathbf{w}, \mathbf{z}) \geq \frac{\pi(\mathbf{z})}{2n}.$$
Thus, the claim follows from [3, Theorem 2.6] (see also [29]). $\qquad\square$

4.2.3 Bottleneck Ratio

Finally, an important concept to establish our lower bounds is represented by the *bottleneck ratio*. Consider an ergodic Markov chain with finite state space S, transition matrix P, and stationary distribution π. The bottleneck ratio of $L \subseteq S$, L non-empty, is

$$B(L) = \frac{\pi(L)P(L, S \setminus L)}{\pi(L)},$$

where $\pi(L) = \sum_{\mathbf{x} \in L} \pi(\mathbf{x})$ and $P(L, S \setminus L) = \sum_{\substack{\mathbf{x} \in L \\ \mathbf{y} \in S \setminus L}} P(\mathbf{x}, \mathbf{y})$. We use the following theorem to derive lower bounds to the mixing time (see, for example, Theorem 7.3 in [31]).

Theorem 4.4 (Bottleneck ratio). *The mixing time of the logit dynamics for a finite opinion game with profile space S is*

$$t_{\mathrm{mix}} \geq \max_{L \subseteq S \,:\, \pi(L) \leq 1/2} \frac{1}{4B(L)}.$$

4.3 Upper bounds

4.3.1 For every β

Theorem 4.5. *Let $\mathcal{G}$ be an n-player finite opinion game on a graph $G = (V, E)$. The mixing time of the logit dynamics for $\mathcal{G}$ is*

$$t_{\mathrm{mix}} \leq (1 + \beta) \cdot \mathsf{poly}\,(n, w_{\mathrm{max}}) \cdot e^{\beta\Theta(\mathsf{CW}(G))}.$$

Proof. Let S, P and π be respectively the state space, the transition matrix and the stationary distribution of the logit dynamics Markov chain for $\mathcal{G}$.

In this proof we use the spectral techniques described above for bounding the mixing time. Specifically, we use Lemma 4.3 for bounding the eigenvalues of P and then Theorem 4.2 for extend this bound to the mixing time.

In order to use Lemma 4.3, it is necessary to describe for every pair of profiles $\mathbf{x}, \mathbf{y} \in S$ a path $\Gamma_{\mathbf{x}, \mathbf{y}}$ between these two profiles. In this work we adopt exactly the same choice for paths as done by Berger et al. [7, Proposition 1.1] for a very specific graphical coordination game and generalized by Auletta et al. [3, Theorem 5.1] to every graphical coordination game. For sake of completeness, we recall here this choice of paths: Consider the ordering of vertices of G that obtains the cutwidth; Fix $\mathbf{x}, \mathbf{y} \in S$ and let $v_1, v_2, \ldots, v_d$ denote the indexes (according to this ordering) of the vertices at which the profiles $\mathbf{x}$ and $\mathbf{y}$ differ; We consider the path $\Gamma_{\mathbf{x}, \mathbf{y}} = (\mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^d)$ on $\mathcal{H}$, where

$$\mathbf{x}^i = \left(y_1, \ldots, y_{v_{i+1}-1}, x_{v_{i+1}}, \ldots, x_n\right).$$

(Above, we assume $v_{d+1} = n + 1$). Notice that $\mathbf{x}^0 = \mathbf{x}$, $\mathbf{x}^d = \mathbf{y}$ and $|\Gamma_{\mathbf{x}, \mathbf{y}}| \leq n$.

The crux of the proof will be then to prove that this choice for $\Gamma_{\mathbf{x}, \mathbf{y}}$ is sufficient to give the desired bound. (We recall that finite opinion games are not graphical coordination games and thus the result of [3] does not apply to this setting.) To this aim, we need two ingredients. First, you can check that the potential function Φ described in (2) can be conveniently expressed as follows: $\Phi(\mathbf{x}) = \sum_{e \in E} \Phi_e(\mathbf{x})$, where, for an edge $e = (i, j)$,

$$\Phi_e(\mathbf{x}) = \begin{cases} \alpha_e := \frac{b_i^2}{\Delta_i} + \frac{b_j^2}{\Delta_j}, & \text{if } x_i = x_j = 0; \\ \beta_e := \frac{b_i^2}{\Delta_i} + \frac{(1-b_j)^2}{\Delta_j} + w_{ij}, & \text{if } x_i = 0 \text{ and } x_j = 1; \\ \gamma_e := \frac{(1-b_i)^2}{\Delta_i} + \frac{b_j^2}{\Delta_j} + w_{ij}, & \text{if } x_i = 1 \text{ and } x_j = 0; \\ \delta_e := \frac{(1-b_i)^2}{\Delta_i} + \frac{(1-b_j)^2}{\Delta_j}, & \text{if } x_i = x_j = 1. \end{cases} \tag{9}$$

The second ingredient, already adopted in [7] and in [3], consists of a function Λ_ξ for every edge $\xi = (\mathbf{x}^i, \mathbf{x}^{i+1})$ of $\mathcal{H}$ that assigns to every pair of profiles $\mathbf{x}, \mathbf{y}$ such that $\xi \in \Gamma_{\mathbf{x}, \mathbf{y}}$, the following new profile

$$\Lambda_\xi(\mathbf{x}, \mathbf{y}) = \begin{cases} \left(x_1, \ldots, x_{v_{i+1}-1}, y_{v_{i+1}}, y_{v_{i+1}+1}, \ldots, y_n\right) & \text{if } \pi(\mathbf{x}^i) \leq \pi(\mathbf{x}^{i+1}); \\ \left(x_1, \ldots, x_{v_{i+1}-1}, x_{v_{i+1}}, y_{v_{i+1}+1}, \ldots, y_n\right) & \text{otherwise.} \end{cases}$$

It is easy to see that Λ_ξ is an injective function (see, for example, [3, Lemma 5.2]).

The proof now proceeds by adapting some of the arguments of [7] and [3] to finite opinion games. Let $E^\star = \{(j, k) \in E \colon j < v_{i+1} \text{ and } k \geq v_{i+1}\}$: observe that $\sum_{(j,k) \in E^\star} w_{jk} \leq \mathsf{CW}(G)$. For any edge $e = (j, k) \in E^\star$, for every $\mathbf{x}, \mathbf{y} \in S$ and for every $\xi = (\mathbf{x}^i, \mathbf{x}^{i+1}) \in \Gamma_{\mathbf{x}, \mathbf{y}}$, we distinguish two cases:

If $x_j = y_j$ or $x_k = y_k$, for all available values of x_j, y_j, x_k and y_k we show

$$\Phi_e(\mathbf{x}) + \Phi_e(\mathbf{y}) - \Phi_e(\arg\min\{\pi(\mathbf{x}^i), \pi(\mathbf{x}^{i+1})\}) - \Phi_e(\Lambda_\xi(\mathbf{x}, \mathbf{y})) = 0.$$

Firstly, assume that $x_j = y_j$ and $\arg\min\{\pi(\mathbf{x}^i), \pi(\mathbf{x}^{i+1})\} = \mathbf{x}^i$ which in turns implies that $\Lambda_\xi(\mathbf{x}, \mathbf{y}) = (x_1, \ldots, x_{v_{i+1}-1}, y_{v_{i+1}}, y_{v_{i+1}+1}, \ldots, y_n)$. We have:

$$
\begin{aligned}
\Phi_e(\mathbf{x}) + \Phi_e(\mathbf{y}) &- \Phi_e(\arg\min\{\pi(\mathbf{x}^i), \pi(\mathbf{x}^{i+1})\}) - \Phi_e(\Lambda_\xi(\mathbf{x}, \mathbf{y})) \\
&= \Phi_e(x_j, x_k) + \Phi_e(y_j, y_k) - \Phi_e(y_j, x_k) - \Phi_e(x_j, y_k) \\
&= \Phi_e(x_j, x_k) + \Phi_e(x_j, y_k) - \Phi_e(x_j, x_k) - \Phi_e(x_j, y_k) = 0.
\end{aligned}
$$

It is not hard to check that the same is true for all the other possible cases arising.

If $x_j \neq y_j$ and $x_k \neq y_k$, similarly to the above, it is not hard to see that for all available values of x_j, y_j, x_k and y_k

$$\Phi_e(\mathbf{x}) + \Phi_e(\mathbf{y}) - \Phi_e(\arg\min\{\pi(\mathbf{x}^i), \pi(\mathbf{x}^{i+1})\}) - \Phi_e(\Lambda_\xi(\mathbf{x}, \mathbf{y})) = \pm(\alpha_e + \delta_e - \beta_e - \gamma_e) = \pm 2w_e,$$

where $\alpha_e, \beta_e, \gamma_e$ and δ_e are defined in (9). Moreover for $e = (j, k) \in E \setminus E^\star$ it holds:

$$\Phi_e(\mathbf{x}) + \Phi_e(\mathbf{y}) - \Phi_e(\arg\min\{\pi(\mathbf{x}^i), \pi(\mathbf{x}^{i+1})\}) - \Phi_e(\Lambda_\xi(\mathbf{x}, \mathbf{y})) = 0$$

since, by construction, one of $\arg\min\{\pi(\mathbf{x}^i), \pi(\mathbf{x}^{i+1})\}$ and $\Lambda_\xi(\mathbf{x}, \mathbf{y})$ has j-th and k-th entry of $\mathbf{x}$ and the other has j-th and k-th entry of $\mathbf{y}$. Thus, we have that for every $\mathbf{x}, \mathbf{y} \in S$ and for every $\xi = (\mathbf{x}^i, \mathbf{x}^{i+1}) \in \Gamma_{\mathbf{x}, \mathbf{y}}$,

$$
\begin{aligned}
\Phi(\mathbf{x}) + \Phi(\mathbf{y}) &- \Phi(\arg\min\{\pi(\mathbf{x}^i), \pi(\mathbf{x}^{i+1})\}) - \Phi(\Lambda_\xi(\mathbf{x}, \mathbf{y})) \\
&= \sum_{e \in E} \left(\Phi_e(\mathbf{x}) + \Phi_e(\mathbf{y}) - \Phi_e(\arg\min\{\pi(\mathbf{x}^i), \pi(\mathbf{x}^{i+1})\}) - \Phi_e(\Lambda_\xi(\mathbf{x}, \mathbf{y})) \right) \\
&= \sum_{e \in E^\star} \left(\Phi_e(\mathbf{x}) + \Phi_e(\mathbf{y}) - \Phi_e(\arg\min\{\pi(\mathbf{x}^i), \pi(\mathbf{x}^{i+1})\}) - \Phi_e(\Lambda_\xi(\mathbf{x}, \mathbf{y})) \right) \\
&\geq -2\mathsf{CW}(G).
\end{aligned}
$$
$$(10)$$

Now let $\xi^\star = (\mathbf{z}, \mathbf{w})$ with $\pi(\mathbf{z}) \leq \pi(\mathbf{w})$ be the edge of $\mathcal{H}$ for which $\sum_{\substack{\mathbf{x},\mathbf{y}: \\ \xi^\star \in \Gamma_{\mathbf{x},\mathbf{y}}}} \frac{\pi(\mathbf{x})\pi(\mathbf{y})}{\pi(\mathbf{z})} |\Gamma_{\mathbf{x},\mathbf{y}}|$ is maximized. Applying Lemma 4.3, we obtain

$$
\begin{aligned}
\frac{1}{1 - \max_{i=2,\ldots,|S|} |\lambda_i|} &\leq 2n \sum_{\substack{\mathbf{x},\mathbf{y}: \\ \xi^\star \in \Gamma_{\mathbf{x},\mathbf{y}}}} \frac{\pi(\mathbf{x})\pi(\mathbf{y})}{\pi(\mathbf{z})} |\Gamma_{\mathbf{x},\mathbf{y}}| \leq 2n^2 \sum_{\substack{\mathbf{x},\mathbf{y}: \\ \xi^\star \in \Gamma_{\mathbf{x},\mathbf{y}}}} \frac{\pi(\mathbf{x})\pi(\mathbf{y})}{\pi(\mathbf{z})\pi(\Lambda_{\xi^\star}(\mathbf{x},\mathbf{y}))} \pi(\Lambda_{\xi^\star}(\mathbf{x},\mathbf{y})) \\
&\leq 2n^2 e^{2\beta \mathsf{CW}(G)} \sum_{\substack{\mathbf{x},\mathbf{y}: \\ \xi^\star \in \Gamma_{\mathbf{x},\mathbf{y}}}} \pi(\Lambda_{\xi^\star}(\mathbf{x},\mathbf{y})) \leq 2n^2 e^{2\beta \mathsf{CW}(G)} \sum_{\mathbf{x}} \pi(\mathbf{x}) \leq 2n^2 e^{2\beta \mathsf{CW}(G)},
\end{aligned}
$$

where the third inequality follows from (8) and (10), and the penultimate from the fact that Λ_ξ is injective.

The theorem follows from Theorem 4.2 and by observing that, since $\Phi(\mathbf{x}) \geq 0$ for any strategy profile $\mathbf{x}$, the definition of π implies

$$
\begin{aligned}
\log\left((\pi_{\min}/4)^{-1}\right) &= \log\left(4 \sum_{\mathbf{x}} e^{-\beta(\Phi(\mathbf{x}) - \Phi_{\max})}\right) \\
&\leq \log\left(2^{n+2} \cdot e^{\beta \Phi_{\max}}\right) \leq \log\left(e^{n+2+\beta \Phi_{\max}}\right) = n + 2 + \beta \Phi_{\max},
\end{aligned}
$$

where $\Phi_{\max} = \max_{\mathbf{x}} \Phi(\mathbf{x}) \leq n + W$. $\qquad\square$

4.3.2 For small β

The following theorem shows that for small values of β the mixing time is polynomial. We remark that there are network topologies for which this result improves on Theorem 4.5, in the sense that sets an higher upper bound on the value of β that is necessary to have polynomial mixing time.

Theorem 4.6. *Let $\mathcal{G}$ be finite opinion game on a connected graph G, with $n > 2$ players. If $\beta \leq 1/(w_{\max}\Delta_{\max})$, then the mixing time of the logit dynamics for $\mathcal{G}$ is $\mathcal{O}(n \log n)$.*

Proof. The proof uses a standard coupling approach (Theorem 4.1). Similar arguments have been used repeatedly for bounding the mixing time of several and different Markov chains (see, e.g., [3, Theorem 3.4]).

However, we prefer to give here the full proof, since the structure of finite opinion games allows to simplify many of previous arguments and makes it possible to adapt the result to properties of the network in place of abstract properties of the potential function.

Consider two profiles $\mathbf{x}$ and $\mathbf{y}$ that differ only in the strategy played by player j. W.l.o.g., we assume $x_j = 1$ and $y_j = 0$. We consider the following coupling for two chains X and Y starting respectively from $X_0 = \mathbf{x}$ and $Y_0 = \mathbf{y}$: Pick a player i uniformly at random; Update the strategies x_i and y_i of player i in the two chains, by setting

$$(x_i, y_i) = \begin{cases} (0,0), & \text{with probability } \min\{\sigma_i(0 \mid \mathbf{x}), \sigma_i(0 \mid \mathbf{y})\}; \\ (1,1), & \text{with probability } \min\{\sigma_i(1 \mid \mathbf{x}), \sigma_i(1 \mid \mathbf{y})\}; \\ (0,1), & \text{with probability } \sigma_i(0 \mid \mathbf{x}) - \min\{\sigma_i(0 \mid \mathbf{x}), \sigma_i(0 \mid \mathbf{y})\}; \\ (1,0), & \text{with probability } \sigma_i(1 \mid \mathbf{x}) - \min\{\sigma_i(1 \mid \mathbf{x}), \sigma_i(1 \mid \mathbf{y})\}. \end{cases}$$

Observe that for every player i, at most one of the updates $(x_i, y_i) = (0,1)$ and $(x_i, y_i) = (1,0)$ has positive probability. Moreover, if $\sigma_i(0 \mid \mathbf{x}) = \sigma_i(0 \mid \mathbf{y})$ and player i is chosen, then, after the update, we have $x_i = y_i$.

Let N_i be the set of neighbors of i in the finite opinion game. Notice that for any player i, $\sigma_i(0 \mid \mathbf{x})$ only depends on x_k, for any $k \in N_i$, and $\sigma_i(0 \mid \mathbf{y})$ only on y_k, for any $k \in N_i$. Therefore, since $\mathbf{x}$ and $\mathbf{y}$ only differ at position j, $\sigma_i(0 \mid \mathbf{x}) = \sigma_i(0 \mid \mathbf{y})$ for $i \notin N_j$.

We start by observing that if position j is chosen for update (this happens with probability $1/n$) then, by the observation above, both chains perform the same update. Since $\mathbf{x}$ and $\mathbf{y}$ differ only for player j, we have that the two chains are coupled (and thus at distance 0). Similarly, if player $i \neq j$ with $i \notin N_j$ is selected for update (which happens with probability $(n - \Delta_j - 1)/n$) we have that both chains perform the same update and thus remain at distance 1. Finally, let us consider the case in which $i \in N_j$ is selected for update. In this case, since $x_j = 1$ and $y_j = 0$, we have that $\sigma_i(0 \mid \mathbf{x}) \leq \sigma_i(0 \mid \mathbf{y})$. Therefore, with probability $\sigma_i(0 \mid \mathbf{x})$ both chains update position i to 0 and thus remain at distance 1; with probability $\sigma_i(1 \mid \mathbf{y}) = 1 - \sigma_i(0 \mid \mathbf{y})$ both chains update position i to 1 and thus remain at distance 1; and with probability $\sigma_i(0 \mid \mathbf{y}) - \sigma_i(0 \mid \mathbf{x})$ chain X updates position i to 1 and chain Y updates position i to 0 and thus the two chains go to distance 2. By summing up, we have that the expected distance $E[\rho(X_1, Y_1)]$ after one step of coupling of the two chains is

$$E[\rho(X_1, Y_1)] = \frac{n - \Delta_j - 1}{n} + \frac{1}{n} \sum_{i \in N_j} [\sigma_i(0 \mid \mathbf{x}) + 1 - \sigma_i(0 \mid \mathbf{y}) + 2 \cdot (\sigma_i(0 \mid \mathbf{y}) - \sigma_i(0 \mid \mathbf{x}))]$$

$$= \frac{n - \Delta_j - 1}{n} + \frac{1}{n} \cdot \sum_{i \in N_j} (1 + \sigma_i(0 \mid \mathbf{y}) - \sigma_i(0 \mid \mathbf{x}))$$

$$= \frac{n - 1}{n} + \frac{1}{n} \cdot \sum_{i \in N_j} (\sigma_i(0 \mid \mathbf{y}) - \sigma_i(0 \mid \mathbf{x})).$$

Let us now evaluate the difference $\sigma_i(0 \mid \mathbf{y}) - \sigma_i(0 \mid \mathbf{x})$ for some $i \in N_j$. Recall that $W_i^s(\mathbf{x})$ denotes the sum of the weights of edges connecting i with neighbors that have opinion s in the profile $\mathbf{x}$. Note that $W_i^0(\mathbf{y}) = W_i^0(\mathbf{x}) + w_{ij}$ and $W_i^1(\mathbf{x}) = W_i^1(\mathbf{y}) + w_{ij} = W_i - W_i^0(\mathbf{x})$. For sake of compactness we will denote with ℓ the quantity $e^{\beta(2b_i - 1 + 2W_i^1(\mathbf{x}) - W_i)}$. By (7) we have

$$\sigma_i(0 \mid \mathbf{x}) = \frac{e^{-\beta(b_i^2 + W_i^1(\mathbf{x}))}}{e^{-\beta(b_i^2 + W_i^1(\mathbf{x}))} + e^{-\beta((1-b_i)^2 + W_i - W_i^1(\mathbf{x}))}} = \frac{1}{1 + \ell},$$

and

$$\sigma_i(0 \mid \mathbf{y}) = \frac{e^{-\beta(b_i^2 + W_i^1(\mathbf{x}) - w_{ij})}}{e^{-\beta(b_i^2 + W_i^1(\mathbf{x}) - w_{ij})} + e^{-\beta((1-b_i)^2 + W_i - W_i^1(\mathbf{x}) + w_{ij})}} = \frac{1}{1 + \ell e^{-2w_{i,j}\beta}}.$$

The function $\frac{1}{1 + \ell e^{-2w_{ij}\beta}} - \frac{1}{1+\ell}$ is maximized for $\ell = e^{w_{ij}\beta}$. Thus

$$\sigma_i(0 \mid \mathbf{y}) - \sigma_i(0 \mid \mathbf{x}) \leq \frac{1}{1 + e^{-\beta}} - \frac{1}{1 + e^{\beta}} = \frac{2}{1 + e^{-\beta}} - 1.$$

16

By using the well-known approximation $e^{-w_{ij}\beta} \geq 1 - w_{ij}\beta$ and since by hypothesis $w_{ij}\beta \leq 1/\Delta_{\max}$, we have

$$\sigma_i(0 \mid \mathbf{y}) - \sigma_i(0 \mid \mathbf{x}) \leq w_{ij}\beta \cdot \frac{1}{2 - w_{ij}\beta} \leq \frac{1}{\Delta_{\max}} \cdot \frac{\Delta_{\max}}{2\Delta_{\max} - 1}.$$

We can conclude that the expected distance after one step of the chain is

$$E[\rho(X_1, Y_1)] \leq \frac{n-1}{n} + \frac{1}{n} \cdot \frac{\Delta_j}{2\Delta_{\max} - 1} \leq \frac{n-1}{n} + \frac{2}{3n} = 1 - \frac{1}{3n} \leq e^{-\frac{1}{3n}}.$$

where the second inequality relies on the fact that $\Delta_{\max} \geq 2$, since the social graph is connected and $n > 2$. Since $\operatorname{diam}(\mathcal{H}) = n$, by applying Theorem 4.1 with $\alpha = \frac{1}{3n}$, we obtain the theorem. $\qquad\square$

4.4 Lower bound

Recall that $\mathcal{H}$ is the Hamming graph on the set of profiles of a finite opinion games on a graph G. The following observation easily follows from the definition of cutwidth.

Observation 4.7. *For every path on $\mathcal{H}$ between the profile $\mathbf{0} = (0,\ldots,0)$ and the profile $\mathbf{1} = (1,\ldots,1)$ there exists a profile for which the weight of the discording edges is at least* $\mathsf{CW}(G)$.

From now on, let us write CW as a shorthand for $\mathsf{CW}(G)$, when the reference to the graph is clear from the context. For sake of compactness, we set $\mathbf{b}(\mathbf{x}) = \sum_i (x_i - b_i)^2$. We denote as $\mathbf{b}^\star$ the minimum of $\mathbf{b}(\mathbf{x})$ over all profiles with CW discording edges.

Let $R_{\mathbf{0}}$ ($R_{\mathbf{1}}$) be the set of profiles $\mathbf{x}$ for which a path from $\mathbf{0}$ (resp., $\mathbf{1}$) to $\mathbf{x}$ exists in $\mathcal{H}$ such that every profile along the path has potential value less than $\mathbf{b}^\star + \mathsf{CW}$. To establish the lower bound we use the technical result given by Theorem 4.4 which requires to compute the bottleneck ratio of a subset of profiles that is weighted at most a half by the stationary distribution. Accordingly, we set $R = R_{\mathbf{0}}$ if $\pi(R_{\mathbf{0}}) \leq 1/2$ and $R = R_{\mathbf{1}}$ if $\pi(R_{\mathbf{1}}) \leq 1/2$. (If both sets have stationary distribution less than one half, the best lower bound is achieved by setting R to $R_{\mathbf{0}}$ if and only if $\Phi(\mathbf{0}) \leq \Phi(\mathbf{1})$ since, in this case, $\mathbf{b}(\mathbf{0}) \leq \mathbf{b}(\mathbf{1})$.) W.l.o.g., in the remaining of this section we assume $R = R_{\mathbf{0}}$.

4.4.1 For large β

Let ∂R be the set of profiles in R that have at least a neighbor $\mathbf{y}$ in the Hamming graph $\mathcal{H}$ such that $\mathbf{y} \notin R$. Moreover let $\mathcal{E}(\partial R)$ the set of edges $(\mathbf{x}, \mathbf{y})$ in $\mathcal{H}$ such that $\mathbf{x} \in \partial R$ and $\mathbf{y} \notin R$: note that $|\mathcal{E}(\partial R)| \leq n|\partial R|$. The following lemma bounds the bottleneck ratio of R.

Lemma 4.8. *For the set of profiles R defined above, we have* $B(R) \leq n \cdot |\partial R| \cdot e^{-\beta(\mathsf{CW}+\mathbf{b}^\star - \mathbf{b}(\mathbf{0}))}$.

Proof. Since $\mathbf{0} \in R$, it holds $\pi(R) \geq \pi(\mathbf{0}) = \frac{e^{-\beta\mathbf{b}(\mathbf{0})}}{Z}$. Moreover, by (7) we have

$$
\begin{aligned}
\pi(R)P(R,\overline{R}) &= \sum_{\substack{(\mathbf{x},\mathbf{y})\in\mathcal{E}(\partial R):\\ \mathbf{y}=(\mathbf{x}_{-i},y_i)}} \frac{e^{-\beta\Phi(\mathbf{x})}}{Z} \frac{e^{-\beta c_i(\mathbf{y})}}{e^{-\beta c_i(\mathbf{x})} + e^{-\beta c_i(\mathbf{y})}} \\
&= \sum_{\substack{(\mathbf{x},\mathbf{y})\in\mathcal{E}(\partial R):\\ \mathbf{y}=(\mathbf{x}_{-i},y_i)}} \frac{e^{-\beta\Phi(\mathbf{x})}}{Z} \frac{e^{-\beta\Phi(\mathbf{y})}e^{-\beta(c_i(\mathbf{x})-\Phi(\mathbf{x}))}}{e^{-\beta\Phi(\mathbf{x})}e^{-\beta(c_i(\mathbf{x})-\Phi(\mathbf{x}))} + e^{-\beta\Phi(\mathbf{y})}e^{-\beta(c_i(\mathbf{x})-\Phi(\mathbf{x}))}} \\
&= \frac{1}{Z}\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{E}(\partial R)} \frac{e^{-\beta\Phi(\mathbf{x})}e^{-\beta\Phi(\mathbf{y})}}{e^{-\beta\Phi(\mathbf{x})} + e^{-\beta\Phi(\mathbf{y})}} = \frac{1}{Z}\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{E}(\partial R)} \frac{e^{-\beta\Phi(\mathbf{y})}}{1 + e^{\beta(\Phi(\mathbf{x})-\Phi(\mathbf{y}))}} \\
&\leq \frac{1}{Z}\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{E}(\partial R)} e^{-\beta\Phi(\mathbf{y})} \leq |\mathcal{E}(\partial R)| \cdot \frac{e^{-\beta(\mathbf{b}^\star+\mathsf{CW})}}{Z}.
\end{aligned}
$$

The second equality follows from the definition of potential function which implies $\Phi(\mathbf{y}) - \Phi(\mathbf{x}) = c_i(\mathbf{y}) - c_i(\mathbf{x})$ for $\mathbf{x}$ and $\mathbf{y}$ as above; last inequality holds because if by contradiction $\Phi(\mathbf{y}) < \mathbf{b}^\star + \mathsf{CW}$ then, by definition of R, it would be $\mathbf{y} \in R$, a contradiction. $\qquad\square$

From Lemma 4.8 and Theorem 4.4 we obtain a lower bound to the mixing time of the logit dynamics for finite opinion games that holds for every value of β, every social network G and every vector $(b_1, \ldots, b_n)$ of internal beliefs. However, it is not clear how close this bound is to the one given in Theorem 4.5. Nevertheless, by taking $b_i = 1/2$ for each player i and β high enough, we obtain the following theorem.

Theorem 4.9. *Let $\mathcal{G}$ be an n-player finite opinion game on a graph G. Then, there exists a vector of internal beliefs such that for $\beta = \Omega\left(\frac{n \log n}{\mathsf{CW}}\right)$ it holds $t_{\mathrm{mix}} \geq e^{\beta \Theta(\mathsf{CW})}$.*

Proof. If $b_i = 1/2$ for every player i, from Lemma 4.8 and Theorem 4.4, since $|\partial R| \leq 2^n$ then

$$t_{\mathrm{mix}} \geq \frac{e^{\beta \mathsf{CW}}}{n 2^n} = e^{\beta \mathsf{CW} - n \log(2n)} = e^{\beta \Theta(\mathsf{CW})}. \qquad \square$$

4.4.2 For smaller β

Theorem 4.9 gives an (essentially) tight lower bound for high values of β for each network topology. It would be interesting to prove a matching bound also for lower values of the rationality parameter: in this section we prove such a bound for specific classes of graphs: complete bipartite graphs and cliques.

We start by considering the class of complete bipartite graphs $K_{m,m}$.

Theorem 4.10. *Let $\mathcal{G}$ be an n-player finite opinion game on $K_{m,m}$. Then, there exist a vector of internal beliefs and edge weights such that, for every $\beta = \Omega\left(\frac{1}{m}\right)$, we have $t_{\mathrm{mix}} \geq \frac{e^{\beta \Theta(\mathsf{CW})}}{n}$.*

To prove the theorem above, we start by evaluating the cutwidth of $K_{m,m}$: we focus on instances of the game with *identical edge weights*. To simplify the exposition, we assume that $w_e = 1$ for all the edges e of $K_{m,m}$ and characterize the best ordering from which the cutwidth is obtained. We will denote with A and B the two sides of the bipartite graph.

Claim 4.11. *For identical edge weights, the ordering that obtains the cutwidth in $K_{m,m}$ is the one that selects alternatively a vertex from A and a vertex from B. Moreover, the cutwidth of $K_{m,m}$ is $\lceil m^2/2 \rceil$.*

Proof. Let $(T, V \setminus T)$ be a cut of the graph, we denote with t the size of T: we also denote t_A as the number of vertices of A in T and t_B as the number of vertices of B in T. Obviously, $t = t_A + t_B$. Given t_A and t_B, the size of the cut $(T, V \setminus T)$ will be $t_A(m - t_B) + t_B(m - t_A) = mt - 2t_A(t - t_A)$. It is immediate to check that for every fixed t the cut is minimized when $\lceil t/2 \rceil$ vertices of T are taken from A and the remaining ones from B. Therefore, the cutwidth is achieved by an ordering which selects alternatively vertices from the two sides of the graph and is then given by the maximum over t of

$$\left\lceil \frac{t}{2} \right\rceil \left(m - \left\lfloor \frac{t}{2} \right\rfloor \right) + \left\lfloor \frac{t}{2} \right\rfloor \left(m - \left\lceil \frac{t}{2} \right\rceil \right).$$

The above function is equal to $mt - \frac{t^2 - 1}{2}$ for t odd and $m - t^2/2$ for t even. Both these fuctions are maximized for $t = m$. However, this may be impossible to achieve when for example t is odd and m is even. Nevertheless, a simple case analysis on the parity of m and t shows that the maximum is achieved for $t = m - 1, m, m + 1$ when m is even and for $t = m$ for m odd, resulting in a cutwidth of $\lceil m^2/2 \rceil$. $\qquad \square$

The following lemma gives a bound to the size of ∂R for this graph.

Lemma 4.12. *For the finite opinion game on the graph $K_{m,m}$ with $b_i = 1/2$ for every player i and identical edge weights, there exists a constant c_1 such that $|\partial R| \leq e^{c_1 \sqrt{\mathsf{CW}}}$.*

Proof. Since $b_i = 1/2$ for every player i, we have that $\mathbf{b}(\mathbf{x}) = n/4$ for every profile $\mathbf{x}$. Therefore, by definition of R, all profiles in R (and therefore ∂R) have less then CW discording edges. Indeed, for $\mathbf{x} \in R$ we have $\mathbf{b}(\mathbf{x}) + |D(\mathbf{x})| = \Phi(\mathbf{x}) < \mathbf{b}^\star + \mathsf{CW}$. Moreover, if a profile $\mathbf{y}$ has less then $\mathsf{CW} - m$ discording edges, then $\mathbf{y}$ is not in ∂R as a state neighbor of $\mathbf{y}$ has at most $m - 1$ additional discording edges.

Consequently, to bound the size of ∂R, we need to count the number of profiles in R that have potential between $\mathbf{b}^\star + \mathsf{CW} - m$ and $\mathbf{b}^\star + \mathsf{CW} - 1$ (i.e., the number of profiles with at least $\mathsf{CW} - m$ and at most $\mathsf{CW} - 1$ discording edges). To count that, we consider two sets L_0 and L_1: we start by setting $L_0 = V$ and $L_1 = \emptyset$. We take vertices from L_0 and sequentially move them to L_1. We can think of L_0 as the set of vertices with opinion

0 and L_1 as the set of vertices with opinion 1: this way we can model a path from $\mathbf{0}$ to $\mathbf{1}$ in the Hamming graph. The number $M(t)$ of edges between L_0 and L_1 after t moves is the number of discording edges in the social graph when vertices in L_0 have opinion 0 and vertices in L_1 have opinion 1. We have

$$\left\lceil \frac{t}{2} \right\rceil \left(m - \left\lfloor \frac{t}{2} \right\rfloor \right) + \left\lfloor \frac{t}{2} \right\rfloor \left(m - \left\lceil \frac{t}{2} \right\rceil \right) \le M(t) \le mt,$$

where the lower bound follows from the structural proof of minimum cuts contained in Claim 4.11.

Let t_1 be the largest integer such that for all possible ways to choose $t_1 - 1$ vertices in L_0 and move them in L_1, the number of edges between L_0 and L_1 is less than $\mathsf{CW} - m$, i.e.

$$(t_1 - 1)m < \mathsf{CW} - m \Rightarrow t_1 = \left\lfloor \frac{\mathsf{CW}}{m} \right\rfloor = \left\lfloor \frac{m}{2} \right\rfloor = \left\lfloor \frac{n}{4} \right\rfloor.$$

Let t_2 be the smallest integer such that for all possible ways to move $t_2 + 1$ vertices from L_0 to L_1, the number of edges between L_0 and L_1 is at least CW, i.e.

$$\left\lceil \frac{t_2+1}{2} \right\rceil \left(m - \left\lfloor \frac{t_2+1}{2} \right\rfloor \right) + \left\lfloor \frac{t_2+1}{2} \right\rfloor \left(m - \left\lceil \frac{t_2+1}{2} \right\rceil \right) \ge \mathsf{CW},$$

that, as showed in Claim 4.11, means $t_2 = m - 2$ for m even and $t_2 = m - 1$ for m odd. Then, we can conclude $t_2 \le m - 1$.

By the definition of t_1, all profiles with at most $t_1 - 1$ players with opinion 1 are not in ∂R and, by definition of t_2, all profiles with at least $t_2 + 1$ players with opinion 1 are not in R. Thus, we have

$$|\partial R| \le \sum_{i=t_1}^{t_2} \binom{n}{i} \le \sum_{i=t_1}^{t_2} \left(\frac{n \cdot e}{i} \right)^i \le \sum_{i=t_1}^{t_2} (5e)^i = \frac{(5e)^{t_2+1} - (5e)^{t_1}}{5e - 1} \le (5e)^{t_2+1} \le (5e)^m \le e^{3m}, \quad (11)$$

where in the third inequality we used the fact that $i \ge t_1 > n/5$, in the penultimate the fact that $t_2 < m$ and lastly the fact that $5^m \le e^{2m}$ for $m \ge 0$. The lemma follows since $m \le \sqrt{2}\sqrt{\mathsf{CW}}$. □

Proof of Theorem 4.10. If $b_i = 1/2$ for every player i, from Lemmata 4.8 and 4.12, we have

$$B(R) \le n \cdot e^{c_1 \sqrt{\mathsf{CW}}} \cdot e^{-\beta \mathsf{CW}} \le n \cdot e^{-\beta \mathsf{CW}(1-c_2)},$$

where $c_2 = \frac{c_1 \sqrt{\mathsf{CW}}}{\beta \mathsf{CW}} < 1$ since by hypothesis $\beta > \frac{c_1}{\sqrt{\mathsf{CW}}} = \Omega(1/m)$; we also notice that c_2 goes to 0 as β increases. The theorem follows from Theorem 4.4. □

We remark that it is possible to prove a result similar to Theorem 4.10 also for the clique K_n: the proof follows from a simple generalization of Theorem 15.3 in [31] and by observing that the cutwidth of a clique is $\lfloor n^2/4 \rfloor$.

5 Conclusions and open problems

In this work we analyze two decentralized dynamics for finite opinion games with only two available opinions: the best-response dynamics and the logit dynamics. As for the best-response dynamics we show that it takes time polynomial in the number of players to reach a Nash equilibrium, the latter being characterized by the existence of clusters in which players have a common opinion. On the other hand, for the logit dynamics we show polynomial convergence when the level of noise is high enough and that it increases as β grows.

It is important to highlight, as noted above, that the convergence time of the two dynamics are computed with respect to two different equilibrium concepts, namely Nash equilibrium for the best-response dynamics and logit equilibrium for the logit dynamics. This explains why the convergence times of these two dynamics asymptotically diverge even though the logit dynamics becomes similar to the best response dynamics as β goes to infinity.

Theorem 4.5 and 4.9 which prove bounds to the convergence of logit dynamics can also be read in a positive fashion. Indeed, for social networks that have a bounded cutwidth, the convergence rate of the dynamics depends only on the value of β. (We highlight that checking if a graph has bounded cutwidth can be done in polynomial time [36].) In general, we have the following picture: as long as β is less than the maximum of (roughly) $\frac{\log n}{\mathsf{CW}}$ and $\frac{1}{\mathbf{w}_{\max}\Delta_{\max}}$ the convergence time to the logit equilibrium is polynomial. Moreover, Theorem 4.9 shows that for β

lower bounded by (roughly) $\frac{n \log n}{\mathsf{CW}}$ the convergence time to the logit equilibrium is super-polynomial. Then for some network topology, there is a gap in our knowledge which is naturally interesting to close.

In [4] the concept of metastable distributions has been introduced in order to predict the outcome of games for which the logit dynamics takes too much time to reach the stationary distribution for some value of β. Furthermore, in [24] a large class of potential games, that includes finite opinion games on well-defined classes of graphs (e.g., cliques, rings, complete bipartite graphs, etc.), has been shown to admit metastable distributions. It would be interesting to specialize the general result of [24] so to investigate more closely the structure of such distributions for our finite opinion games.

References

[1] D. Acemoglu and A. Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49, 2011.

[2] V. Auletta, I. Caragiannis, D. Ferraioli, C. Galdi, and G. Persiano. Minority becomes majority in social networks. In *Proc. of International Workshop on Internet and Network Economics (WINE)*, pages 74–88. Springer, 2015.

[3] V. Auletta, D. Ferraioli, F. Pasquale, P. Penna, and G. Persiano. Convergence to equilibrium of logit dynamics for strategic games. *Algorithmica*, pages 1–33.

[4] V. Auletta, D. Ferraioli, F. Pasquale, and G. Persiano. Metastability of logit dynamics for coordination games. In *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1006–1024, 2012.

[5] V. Auletta, D. Ferraioli, F. Pasquale, and G. Persiano. Mixing time and stationary expected social welfare of logit dynamics. *Theory of Computing Systems*, pages 1–38, 2013.

[6] M. Balcan, A. Blum, and Y. Mansour. Improved equilibria via public service advertising. In *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 728–737, 2009.

[7] N. Berger, C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields*, 131:311–340, 2005.

[8] A. Bhalgat, T. Chakraborty, and S. Khanna. Approximating pure Nash equilibrium in cut, party affiliation, and satisfiability games. In *Proc. of ACM Conference on Electronic Commerce (EC)*, pages 73–82, 2010.

[9] K. Bhawalkar, S. Gollapudi, and K. Munagala. Coevolutionary opinion formation games. In *Proc. of ACM symposium on Theory of computing (STOC)*, pages 41–50. ACM, 2013.

[10] D. Bindel, J. M. Kleinberg, and S. Oren. How bad is forming your own opinion? In *Proc. of Symposium on Foundations of Computer Science (FOCS)*, pages 57–66, 2011.

[11] L. E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.

[12] R. Bubley and M. E. Dyer. Path Coupling: A technique for proving rapid mixing in Markov chains. In *Proc. of Symposium on Foundations of Computer Science (FOCS)*, pages 223–231, 1997.

[13] S. Chien and A. Sinclair. Convergence to approximate Nash equilibria in congestion games. *Games and Economic Behavior*, 71(2):315 – 327, 2011.

[14] F. Chierichetti, J. Kleinberg, and S. Oren. On discrete preferences and coordination. In *Proc. of ACM conference on Electronic Commerce (EC)*, pages 233–250. ACM, 2013.

[15] M. H. DeGroot. Reaching a consensus. *J. American Statistical Association*, 69, 1974.

[16] P. M. DeMarzo, D. Vayanos, and J. Zweibel. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 2003.

[17] M. Dyer and V. Mohanaraj. Pairwise-interaction games. In *Proc. of Automata, Languages and Programming (ICALP)*, pages 159–170, 2011.

[18] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* Cambridge University Press, 2010.

[19] G. Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071, 1993.

[20] B. Escoffier, D. Ferraioli, L. Gourvès, and S. Moretti. Designing budget-balanced best-response mechanisms for network coordination games. In *Proc. of Internation Symposium on Algorithmic Game Theory (SAGT)*, pages 207–218. 2013.

[21] A. Fabrikant, C. H. Papadimitriou, and K. Talwar. The complexity of pure Nash equilibria. In *Proc. of Symposium on the Theory of Computing (STOC)*, pages 604–612, 2004.

[22] D. Ferraioli, P. W. Goldberg, and C. Ventre. Decentralized dynamics for finite opinion games. In *Proc. of International Symposium on Algorithmic Game Theory (SAGT)*, pages 144–155, 2012.

[23] D. Ferraioli, L. Gourvès, S. Moretti, F. Pascual, and O. Spanjaard. *Combinatorial Optimization with Competing Agents*, pages 675–706. John Wiley & Sons, Inc., 2014.

[24] D. Ferraioli and C. Ventre. Metastability of asymptotically well-behaved potential games - (extended abstract). In *Proc. of Mathematical Foundations of Computer Science (MFCS)*, pages 311–323, 2015.

[25] N. E. Friedkin and E. C. Johnsen. Social influence and opinions. *Math. Sociology*, 15(3-4), 1990.

[26] J. Ghaderi and R. Srikant. Opinion dynamics in social networks: A local interaction game with stubborn agents. In *American Control Conference (ACC), 2013*, pages 1982–1987. IEEE, 2013.

[27] B. Golub and M. O. Jackson. Naive learning in social networks: Convergence, influence and the wisdom of crowds. *American Econ. J.: Microeconomics*, 2, 2010.

[28] M. O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.

[29] M. Jerrum and A. Sinclair. Approximating the Permanent. *SIAM Journal on Computing*, 18(6):1149–1178, 1989.

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

[31] D. Levin, Yuval P., and E. L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.

[32] D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124 – 143, 1996.

[33] A. Montanari and A. Saberi. Convergence to equilibrium in local interaction games. In *Proc. of Symposium on Foundations of Computer Science (FOCS)*, 2009.

[34] H. Peyton Young. *The diffusion of innovations in social networks*, chapter in "The Economy as a Complex Evolving System", vol. III, Lawrence E. Blume and Steven N. Durlauf, eds. Oxford University Press, 2003.

[35] R. W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

[36] D. M. Thilikos, M. J. Serna, and H. L. Bodlaender. Constructive linear time algorithms for small cutwidth and carving-width. In *Proc. of International Symposium on Algorithms and Computation (ISAAC)*, pages 192–203, 2000.