

# Decentralized Dynamics for Finite Opinion Games\*

Diodato Ferraioli<sup>†</sup>

Paul W. Goldberg<sup>‡</sup>

Carminé Ventre<sup>§</sup>

## Abstract

Game theory studies situations in which strategic players can modify the state of a given system, due to the absence of a central authority. Solution concepts, such as Nash equilibrium, are defined 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 improve their own utility 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. [9]. 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.

---

\*Work partially supported by EPSRC grant EP/G069239/1 “Efficient Decentralised Approaches in Algorithmic Game Theory” and by PRIN 2008 research project COGENT (COMputational and GAMe-theoretic aspects of uncoordinated NeTworks), funded by the Italian Ministry of University and Research.

The final version of this paper is available at <https://www.sciencedirect.com/science/article/pii/S0304397516304078/pdf?md5=62f7c16052a1187626b6cea216b91ea8&pid=1-s2.0-S0304397516304078-main.pdf>.

A preliminary version of this paper appeared as [18].

<sup>†</sup>LAMSADE, Université Paris Dauphine, France. E-mail: [diodato.ferraioli@dauphine.fr](mailto:diodato.ferraioli@dauphine.fr).

<sup>‡</sup>Department of Computer Science, University of Liverpool, UK. E-mail: [P.W.Goldberg@liverpool.ac.uk](mailto:P.W.Goldberg@liverpool.ac.uk).

<sup>§</sup>School of Computing, Teesside University, UK. E-mail: [C.Ventre@tees.ac.uk](mailto: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 his own opinion on the matter at hand. How can this group of people reach *consensus*? 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, 13, 20, 21], is based on a model defined by DeGroot [12]. 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 his neighbors in  $G$  (the influence of neighbor  $j$  is stronger the higher  $w_{ij}$  is). In each time step, node  $i$  updates his opinion to be a weighted average of his current opinion with the current opinions of his neighbors. A variation of this model of interest to our study is due to Friedkin and Johnsen [19]. In [19] it is additionally assumed that each node  $i$  maintains a persistent *internal belief*  $b_i$ , which remains constant even as node  $i$  updates his overall opinion  $x_i$  through averaging. (See Section 2 for the formal framework.)

However, as recently observed by Bindel et al. [9], consensus is hard to reach, the case of political opinions being a prominent example. The authors of [9] 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 diminishes their *utility*. 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 [9] and ask the question of how quickly equilibria are reached by decentralized dynamics in opinion games. We focus on the setting in which players have only a finite number of strategies available. 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. Motivated by several electoral systems around the world, we concentrate in this study on the case in which players only have two strategies available. This setting already encodes a number of interesting technical challenges as outlined below.

## 1.1 Our contribution

For the finite version of the opinion games considered in [9], we firstly note that this is a potential game [26, 23] 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 [9] have a price of anarchy of  $9/8$ , our games have unbounded price of anarchy, thus implying that for finite games disagreeing has far more deep consequences on the social cost. We additionally prove that the socially optimal profile is always a Nash equilibrium when the weights of the social network are integers, thus implying that the Price of Stability is 1 in this case, and in general a tight bound of 2 for the Price of Stability when edges have fractional weights.

We then study decentralized dynamics for finite opinion games. We first consider the best-response dynamics, by proving that it 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 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” an 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 [9] is unknown.

In real life, however, there is some noise in the decision process of players. Arguably, people are not fully rational. On the other hand, even if they were, they might not exactly know what strategy represents the best response to a given strategy profile due to the incapacity to correctly determine their utility functions. To model this, we study *logit dynamics* [10] for 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 utility to the player and  $\beta$ . So the higher  $\beta$  is, the less noise there is and the more the dynamics is similar to best-response dynamics. 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<sup>1</sup>. It is observed in [5] how this notion enjoys a number of desiderata one would like solution concepts to have.

We prove a host of results on the convergence rate of logit dynamics that give a pretty much complete picture as  $\beta$  varies. We give an upper bound in terms of the cutwidth of the graph modeling the social network. The bound is exponential in  $\beta$  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  $\beta$  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  $\beta$  is big enough relatively to this subset then we can prove that the upper bound is tight for any social network (specifically, we roughly need  $\beta$  bigger than  $n \log n$  over the cutwidth of the graph). For smaller values of  $\beta$ , 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 details, 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 to prove a lower bound exponential in  $\beta$  and the cutwidth of the graph for (almost) any value of  $\beta$ . 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 [22].

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 [22] for a discussion of these concepts. Below, we give a quick overview of these techniques and state some useful facts.)

## 1.2 Related works

In addition to the papers mentioned above, our paper is related to the works on logit dynamics. This dynamics is introduced by Blume [10] and it is mainly adopted in the analysis of graphical coordination games [15, 25, 24], in which players are placed on vertices of a graph embedding social relations and each player wants to coordinate with neighbors: we highlight that a unique game is played on every edge, whereas, for opinion games, we need different games in order to encode beliefs. Asadpour and Saberi [2] adopt the logit dynamics for analyzing a class of congestion games. 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].

A number of papers study the efficient computation of (approximate) pure Nash equilibria for 2-strategy games, such as *party affiliation games* [17, 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 2.3), 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.

Another related work is [14] by Dyer and Mohanaraj. They study graphical games, called *pairwise-interaction games*, and prove among other results, quick convergence of best-response dynamics for these games. However, our games do not fall in their class. The difference is that, in their case, there is a unique game being played on the edges of the graph; as noted above, we instead need a different game to encode the internal beliefs of the players.

## 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*  $b_i \in [0, 1]$  and only two strategies or *opinions* are available, namely 0 and 1. Motivated by

<sup>1</sup>Thus, the solution concept of logit dynamics is different from the one associated with best-response dynamics.

the model in [9], we define the utility of player  $i$  in a strategy profile  $\mathbf{x} \in \{0, 1\}^n$  as

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

We call such a game an  $n$ -player *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

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

## 2.1 Potential function

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*

$$\Phi(\mathbf{x}) = \sum_i (x_i - b_i)^2 + D(\mathbf{x}) \quad (1)$$

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

*Proof.* Given a strategy profile  $\mathbf{x}$ , player  $i$  experiences a non-positive utility or equivalently, a non-negative cost defined as  $c_i(\mathbf{x}) = -u_i(\mathbf{x})$ . We show that in an opinion game if players minimize their cost, then the function  $\Phi$  defined in (1) decreases. The difference in the cost to player  $i$  when he 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

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

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  $D(\mathbf{x}_{-i}, y_i) = K_i(\mathbf{x}_{-i}, y_i) + D_i(\mathbf{x}_{-i}, y_i)$ .  $\square$

A more convenient way to express the potential function above, useful in one of the proofs below, is the following:  $\Phi(\mathbf{x}) = \sum_{e \in E} \Phi_e(\mathbf{x})$ , where, for an edge  $e = (i, j)$ ,

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

and  $\Delta_i$  represents the degree of  $i$ .

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

$$\text{SC}(\mathbf{x}) = - \sum_{i=1}^n u_i(\mathbf{x}) = \sum_i (x_i - b_i)^2 + 2D(\mathbf{x}).$$

**Observation 2.2** ([16]). *The profile minimizing the social cost can be computed in polynomial time by a centralized algorithm.*

Specifically, Escoffier et al. [16] showed that it corresponds to the  $(s, t)$ -cut of minimum weight in a suitably built graph.

## 2.2 Nash equilibria, Price of Anarchy and Price of Stability

We start with a simple observation about the centralized computation of Nash equilibria.

**Observation 2.3.** *Nash equilibria of the opinion games studied here, can be found in polynomial time.*

Specifically, consider the following simple greedy algorithm. Start with the pure profile  $\mathbf{x}$  where everyone plays 0. Check for players who prefer to play 1; notice that any such player must play 1 in *any* Nash equilibrium; modify  $\mathbf{x}$  so that they play 1. Repeat this until no player prefers to change his own strategy.

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 below and Observation 2.2 that when edge weights are integers, the lowest-cost equilibrium is computable in polynomial time.

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 his own belief if and only if more than (almost) half of his (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

$$(1 - B_i - b_i)^2 = (B_i - b_i)^2 - 2|B_i - b_i| + 1. \quad (3)$$

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

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

where we used (3) 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  $u_i(\mathbf{x}_{-i}, B_i) \leq u_i(\mathbf{x})$ . Hence, no player has incentive to switch his opinion in  $\mathbf{x}$  and 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  $u_i(\mathbf{y}) < u_i(\mathbf{y}_{-i}, B_i)$  and in the latter case  $u_i(\mathbf{y}) < u_i(\mathbf{y}_{-i}, 1 - B_i)$ .  $\square$

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

### Price of Anarchy and Stability.

**Observation 2.5.** *The price of anarchy of the social network games studied here is unbounded.*

To see this, consider the 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 [9].

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

**Theorem 2.6.** *For opinion games, the price of stability is 2.*

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

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

Since  $\mathbf{x}^*$  is a Nash equilibrium then the upper bound to the price of stability follows.

For the lower bound, consider an 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 he is playing the strategy opposite to his own belief). All profiles in which at least two nodes play the strategy opposite to their belief have cost at least  $2 > 1 + 2/n$ . All profiles in which there is only one player, different from the center, playing opposite to his 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 his 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 his 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.  $\square$

We show that above bound can be improved when edge weights are integer. Indeed, it turns out that, in this special case, the profile that minimizes the social cost is always a Nash equilibrium.

**Theorem 2.7.** *For an opinion game with integer weights, the price of stability is 1.*

*Proof.* Let  $\mathcal{G}$  be an  $n$ -player 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 (3). 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:

$$\begin{aligned} SC(\mathbf{x}_{-i}, B_i) - SC(\mathbf{x}) &= W_i - 2W_i^{B_i}(\mathbf{x}) + 2|B_i - b_i| - 1 - \sum_{\substack{j: (i,j) \in E \\ x_j = B_i}} w_{ij} + \sum_{\substack{j: (i,j) \in E \\ x_j = 1 - B_i}} w_{ij} \\ &= 2 \left( W_i - 2W_i^{B_i}(\mathbf{x}) + |B_i - b_i| \right) - 1. \end{aligned}$$

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

If  $|B_i - b_i| < 1/2$  then  $W_i^{B_i}(\mathbf{x}) > W_i/2 - \delta$  implies  $W_i^{B_i}(\mathbf{x}) \geq W_i/2$  for integer weights. Therefore,  $SC(\mathbf{x}_{-i}, B_i) - SC(\mathbf{x}) = 2 \left( W_i - 2W_i^{B_i}(\mathbf{x}) + |B_i - b_i| \right) - 1 \leq 2|B_i - b_i| - 1 < 0$  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  $SC(\mathbf{x}_{-i}, B_i) - SC(\mathbf{x}) < 0$ .  $\square$

### 3 Best-response dynamics

To prove the convergence rate of best-response dynamics, we use the following definition, previously used in [14].

**Definition 3.1.** Two games  $\mathcal{G}, \mathcal{G}'$  are pure best-response equivalent<sup>2</sup> 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. We say that a belief  $b \in [0, 1]$  is *threshold* for player  $i$  in 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  $i$ .

**Lemma 3.2.** Given an opinion game  $\mathcal{G}$  and a player  $i$  in  $\mathcal{G}$ , we define a finite 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  $\mathcal{B}_i$  contain 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,  $\frac{1}{2}(b'_1 + b'_2)$ ).

Then, any opinion game  $\mathcal{G}$  is pure best-response equivalent to an 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 an opinion game  $\mathcal{G}$  and player  $i$ . Let  $\mathcal{G}'$  be an opinion game defined on the same social network as  $\mathcal{G}$ . For each player  $i$  in  $\mathcal{G}$ , if his 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.  $\square$

#### 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 Lemma 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 3.2, in both cases,  $\mathcal{G}$  is pure best-response equivalent to an 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 3.3.** The best-response dynamics for an  $n$ -player 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 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 his best response. By definition of best response, we have  $c_i(\mathbf{x}) > c_i(s, \mathbf{x}_{-i})$ . We will show that

$$\begin{aligned} \frac{1}{2} &\leq \Phi(\mathbf{x}) - \Phi(s, \mathbf{x}_{-i}) = c_i(x_i, \mathbf{x}_{-i}) - c_i(s, \mathbf{x}_{-i}) = (x_i - b_i)^2 - (s - b_i)^2 + D_i(\mathbf{x}) - D_i(s, \mathbf{x}_{-i}) \\ &= 2|x_i - b_i| - 1 + \Delta D, \end{aligned} \quad (4)$$

<sup>2</sup>In [14] they just use ‘‘equivalent’’.

where the last equality follows by (3) (with  $x_i$  in place of  $B_i$ ) and by using  $\Delta D$  as a shorthand for  $D_i(\mathbf{x}) - D_i(s, \mathbf{x}_{-i})$ . We distinguish three cases based on the value of  $\Delta D$ .

If  $\Delta D > 1$ , then it will be the case that  $\Delta D \geq 2$ . Since the difference between the squares is bounded from below by  $-1$ , then (4) follows.

If  $-1 < \Delta D \leq 1$ , 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  $x_i$  (4) 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  $x_i$  (4) follows.

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

### 3.2 Finite-precision weights

We now show how to extend the bound for the convergence of the best-response dynamics to opinion games whose edge weights have bounded precision  $k$ , i.e. they can be written with at most  $k$  digit after the decimal point.

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

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

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

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

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

**Lemma 3.4.** *Consider an opinion game  $\mathcal{G}$  and for each player  $i$  consider the set  $\mathcal{B}'_i$  as defined in Lemma 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.  $\square$

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

**Theorem 3.5.** *The best-response dynamics for an  $n$ -player 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 opinion game  $\mathcal{G}$  and for each player  $i$  consider the set  $\mathcal{B}'_i$  as defined in Lemma 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 an opinion game such that each player  $i$  has belief in  $\mathcal{B}_i$ . From Lemma 3.2,  $\mathcal{G}$  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 his best response. By setting  $c_i''(\mathbf{x}) = -u_i''(\mathbf{x})$ , from the definition of best response, it follows that  $c_i''(\mathbf{x}) > c_i''(s, \mathbf{x}_{-i})$ . We will show that

$$\begin{aligned} 1/2 &\leq \Phi''(\mathbf{x}) - \Phi''(s, \mathbf{x}_{-i}) = c_i''(x_i, \mathbf{x}_{-i}) - c_i''(s, \mathbf{x}_{-i}) \\ &= 10^k ((x_i - b_i)^2 - (s - b_i)^2) + D_i''(\mathbf{x}) - D_i''(s, \mathbf{x}_{-i}) = 10^k (2|x_i - b_i| - 1) + \Delta D, \end{aligned} \quad (5)$$

where the last equality follows by (3) (with  $x_i$  in place of  $B_i$ ) and by using  $\Delta D$  as a shorthand for  $D_i''(\mathbf{x}) - D_i''(s, \mathbf{x}_{-i})$ . We distinguish three cases based on the value of  $\Delta D$ .

If  $\Delta D > 10^k$ , 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 (5) follows.

If  $-10^k < \Delta D \leq 10^k$ , 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 (5) follows. Similarly, one can prove that if  $x_i = 1$ , then  $b_i \leq \frac{10^k + \Delta D - 1/2}{2 \cdot 10^k}$  and (5) follows.

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

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 very small (in which case, we need high precision). However, in the next section we will show that, in these cases, such a large bound cannot be avoided.

### 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 his 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 upper-bound the number of steps with any polynomial. The construction uses graphs with bounded degree and pathwidth, with all players having belief  $\frac{1}{2}$ .

**Theorem 3.6.** *The best-response dynamics for 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  $\varepsilon$ ,  $2\varepsilon$ ,  $3\varepsilon$  respectively and edges  $(D, E)$ ,  $(B, F)$  and  $(D, F)$ , all weighting  $4\varepsilon$ , for some  $\varepsilon > 0$ .

Consider a 6-gadget and a new player  $A_0$  with edges  $(A_0, B)$  and  $(A_0, D)$  of weight  $4\varepsilon$ . 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*:

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

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*:

$$\begin{aligned} (0, 1, 0, 1, 0, 1) &\rightarrow (1, 1, 0, 1, 0, 1) \rightarrow (1, 0, 0, 1, 0, 1) \rightarrow (0, 0, 0, 1, 0, 1) \rightarrow (0, 0, 1, 1, 0, 1) \rightarrow (0, 1, 1, 1, 0, 1) \\ &\rightarrow (1, 1, 1, 1, 0, 1) \rightarrow (1, 1, 1, 0, 0, 1) \rightarrow (1, 1, 0, 0, 0, 1) \rightarrow (1, 0, 0, 0, 0, 1) \rightarrow (0, 0, 0, 0, 0, 1). \end{aligned}$$

Notice that  $A$ 's opinion does not change in the switch-on cycle, whereas it follows the sequence  $0 \rightarrow 1 \rightarrow 0 \rightarrow 1 \rightarrow 0$  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  $\varepsilon_i$ , where  $1 \leq i \leq n$ . For each  $i = 2, \dots, n$ , 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  $i = 1, \dots, n - 1$  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  $\varepsilon_i$ . 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  $\varepsilon_n$  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  $i = 1, \dots, n - 1$ , switches his opinion from 0 to 1, we execute the switch-on cycle of  $G_{i+1}$ ; as long as  $A_i$ , for  $i = 1, \dots, n - 1$ , switches his 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.  $\square$

## 4 Logit Dynamics for Opinion Games

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

The *logit dynamics* for  $\mathcal{G}$  runs as follows: at every time step (i) Select one player  $i \in [n]$  uniformly at random; (ii) Update the strategy of player  $i$  according to the *Boltzmann distribution* with parameter  $\beta$  over the set  $S_i = \{0, 1\}$  of his strategies. That is, a strategy  $s_i \in S_i$  will be selected with probability

$$\sigma_i(s_i | \mathbf{x}_{-i}) = \frac{1}{Z_i(\mathbf{x}_{-i})} e^{\beta u_i(\mathbf{x}_{-i}, s_i)}, \quad (6)$$

where  $\mathbf{x}_{-i} \in \{0, 1\}^{n-1}$  is the profile of strategies played at the current time step by players different from  $i$ ,  $Z_i(\mathbf{x}_{-i}) = \sum_{z_i \in S_i} e^{\beta u_i(\mathbf{x}_{-i}, z_i)}$  is the normalizing factor, and  $\beta \geq 0$ . As mentioned above, from (6), it is easy to see that for  $\beta = 0$  player  $i$  selects his strategy uniformly at random, for  $\beta > 0$  the probability is biased toward strategies promising higher payoffs, and for  $\beta$  that goes to  $\infty$  player  $i$  chooses his best response strategy (if more than one best response is available, he chooses one of them uniformly at random).

The above dynamics defines a *Markov chain*  $\{X_t\}_{t \in \mathbb{N}}$  with the set of strategy profiles as state space, and where the probability  $P(\mathbf{x}, \mathbf{y})$  of a transition from profile  $\mathbf{x} = (x_1, \dots, x_n)$  to profile  $\mathbf{y} = (y_1, \dots, y_n)$  is zero if  $H(\mathbf{x}, \mathbf{y}) \geq 2$  and it is  $\frac{1}{n} \sigma_i(y_i | \mathbf{x}_{-i})$  if the two profiles differ exactly at player  $i$ . More formally, we can define the logit dynamics as follows.

**Definition 4.1** (Logit dynamics [10]). *Let  $\mathcal{G}$  be an opinion game as from the above and let  $\beta \geq 0$ . The logit dynamics for  $\mathcal{G}$  is the Markov chain  $\mathcal{M}_\beta = (\{X_t\}_{t \in \mathbb{N}}, S, P)$  where  $S = \{0, 1\}^n$  and*

$$P(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \cdot \begin{cases} \sigma_i(y_i | \mathbf{x}_{-i}), & \text{if } \mathbf{y}_{-i} = \mathbf{x}_{-i} \text{ and } y_i \neq x_i; \\ \sum_{i=1}^n \sigma_i(y_i | \mathbf{x}_{-i}), & \text{if } \mathbf{y} = \mathbf{x}; \\ 0, & \text{otherwise;} \end{cases} \quad (7)$$

where  $\sigma_i(y_i | \mathbf{x}_{-i})$  is defined in (6).

The Markov chain defined by (7) is ergodic. 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*  $\pi$  as  $t$  tends to infinity.<sup>3</sup> As in [5], we call the stationary distribution  $\pi$  of the Markov chain defined by the logit dynamics on a game  $\mathcal{G}$ , the *logit*

<sup>3</sup>The notation  $P^t(\mathbf{x}, \cdot)$ , standard in Markov chains literature [22], denotes the probability distribution over states of  $S$  after the chain has taken  $t$  steps starting from  $\mathbf{x}$ .

*equilibrium* of  $\mathcal{G}$ . In general, a Markov chain with transition matrix  $P$  and state space  $S$  is said to be *reversible* with respect to the distribution  $\pi$  if, for all  $\mathbf{x}, \mathbf{y} \in S$ , it holds that  $\pi(\mathbf{x})P(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y})P(\mathbf{y}, \mathbf{x})$ . If the chain is reversible with respect to  $\pi$ , then  $\pi$  is its stationary distribution. Therefore when this happens, to simplify our exposition we simply say that the matrix  $P$  is reversible. For the class of potential games the stationary distribution is the well known *Gibbs measure*.

**Theorem 4.2** ([10]). *If  $\mathcal{G} = ([n], S, \mathcal{U})$  is a potential game with potential function  $\Phi$ , then the Markov chain given by (7) is reversible with respect to the Gibbs measure  $\pi(\mathbf{x}) = \frac{1}{Z} e^{-\beta\Phi(\mathbf{x})}$ , where  $Z = \sum_{\mathbf{y} \in S} e^{-\beta\Phi(\mathbf{y})}$  is the normalizing constant.*

**Mixing time of Markov chains.** One of the prominent measures of the rate of convergence of a Markov chain to its stationary distribution is the *mixing time*. For a Markov chain with transition matrix  $P$  and state space  $S$ , let us set

$$d(t) = \max_{\mathbf{x} \in S} \|P^t(\mathbf{x}, \cdot) - \pi\|_{\text{TV}},$$

where the *total variation distance*  $\|\mu - \nu\|_{\text{TV}}$  between two probability distributions  $\mu$  and  $\nu$  on the same state space  $S$  is defined as

$$\|\mu - \nu\|_{\text{TV}} = \max_{A \subset S} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{\mathbf{x} \in S} |\mu(\mathbf{x}) - \nu(\mathbf{x})| = \sum_{\substack{\mathbf{x} \in S: \\ \mu(\mathbf{x}) > \nu(\mathbf{x})}} (\mu(\mathbf{x}) - \nu(\mathbf{x})).$$

For  $0 < \varepsilon < 1/2$ , the mixing time is defined as

$$t_{\text{mix}}(\varepsilon) = \min\{t \in \mathbb{N} : d(t) \leq \varepsilon\}.$$

It is usual to set  $\varepsilon = 1/4$  or  $\varepsilon = 1/2e$ . If not explicitly specified, when we write  $t_{\text{mix}}$  we mean  $t_{\text{mix}}(1/4)$ . Observe that  $t_{\text{mix}}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{\text{mix}}$ .

Next we bound the mixing time of the logit dynamics for an opinion game. Note that, bounds on the mixing time of the logit dynamics for general potential games has been given in [3]. However, these bounds are not tight for the class of opinion game. Moreover, they do not highlight the existing connection between the convergence time of the logit dynamics and the topology of the social network underlying the opinion game.

## 4.1 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.

### 4.1.1 Markov chain coupling

A *coupling* of two probability distributions  $\mu$  and  $\nu$  on  $S$  is a pair of random variables  $(X, Y)$  defined on  $S \times S$  such that the marginal distribution of  $X$  is  $\mu$  and the marginal distribution of  $Y$  is  $\nu$ . 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  $\mathbf{P}_{\mathbf{x}, \mathbf{y}}(\cdot)$  and  $\mathbf{E}_{\mathbf{x}, \mathbf{y}}[\cdot]$  for the probability and the expectation on the space  $S \times S$ . We denote by  $\tau_{\text{couple}}$  the first time the two chains meet; that is,

$$\tau_{\text{couple}} = \min\{t : X_t = Y_t\}.$$

We also consider only couplings of Markov chains with the property that  $X_s = Y_s$  for  $s \geq \tau_{\text{couple}}$ .

Recall that  $\mathcal{H} = (S, E)$  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  $\mathbf{y}$ . The following theorem says that it is sufficient to define a coupling only for pairs of Markov chains starting from states *adjacent* in  $\mathcal{H}$  and an upper bound on the mixing time can be obtained if each of these couplings contracts their distance on average.

**Theorem 4.3** (Path Coupling [11]). *Suppose that for every edge  $(\mathbf{x}, \mathbf{y}) \in \mathbf{E}$  a coupling  $(X_t, Y_t)$  of  $\mathcal{M}$  with  $X_0 = \mathbf{x}$  and  $Y_0 = \mathbf{y}$  exists such that  $\mathbf{E}_{\mathbf{x}, \mathbf{y}}[\rho(X_1, Y_1)] \leq e^{-\alpha}$  for some  $\alpha > 0$ . Then*

$$t_{\text{mix}}(\varepsilon) \leq \frac{\log(\text{diam}(\mathcal{H})) + \log(1/\varepsilon)}{\alpha}$$

where  $\text{diam}(\mathcal{H})$  is the diameter of  $\mathcal{H}$ .

We here describe only the coupling that we use in the proof of Theorem 4.9 below. Our exposition follows the one in [5]. For every pair of strategy profiles  $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n) \in \{0, 1\}^n$  we define a coupling  $(X_1, Y_1)$  of two copies of the Markov chain for which  $X_0 = \mathbf{x}$  and  $Y_0 = \mathbf{y}$ . The coupling proceeds as follows: first, pick a player  $i$  uniformly at random; then, update the strategies  $x_i$  and  $y_i$  of player  $i$  in the two chains, by setting

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

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  $\sigma_i(0 | \mathbf{x}) = \sigma_i(0 | \mathbf{y})$  and player  $i$  is chosen, then, after the update, we have  $x_i = y_i$ .

For the path coupling technique (see Theorem 4.3), the coupling described above is applied only to pairs of starting profiles which are adjacent in  $\mathcal{H}$ .

#### 4.1.2 Relaxation time and spectral techniques

Another important measure related to the convergence of Markov chains is given by the *relaxation time*. Let  $P$  be the transition matrix of a Markov chain with finite state space  $S$  and let us label the eigenvalues of  $P$  in non-increasing order  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{|S|}$ . It is well-known (see, for example, Lemma 12.1 in [22]) that  $\lambda_1 = 1$  and, if  $P$  is irreducible and aperiodic, then  $\lambda_2 < 1$  and  $\lambda_{|S|} > -1$ . We set  $\lambda^*$  as the largest eigenvalue in absolute value other than  $\lambda_1$ , that is,  $\lambda^* = \max_{i=2, \dots, |S|} \{|\lambda_i|\}$ . The *relaxation time*  $t_{\text{rel}}$  of a Markov chain  $\mathcal{M}$  is defined as

$$t_{\text{rel}} = \frac{1}{1 - \lambda^*}.$$

The relaxation time is related to the mixing time by the following theorem (see, for example, Theorems 12.3 and 12.4 in [22]).

**Theorem 4.4** (Relaxation time). *Let  $P$  be the transition matrix of a reversible, irreducible, and aperiodic Markov chain with state space  $S$  and stationary distribution  $\pi$ . Then*

$$(t_{\text{rel}} - 1) \log 2 \leq t_{\text{mix}} \leq \log \left( \frac{4}{\pi_{\min}} \right) t_{\text{rel}},$$

where  $\pi_{\min} = \min_{\mathbf{x} \in S} \pi(\mathbf{x})$ .

Bounds on relaxation time can be obtained by using the following lemma (see Corollary 13.24 in [22]).

**Lemma 4.5.** *Let  $P$  be the transition matrix of an irreducible, aperiodic and reversible Markov chain with state space  $S$  and stationary distribution  $\pi$ . Consider the graph  $\mathcal{H} = (S, \mathbf{E})$ , where  $\mathbf{E} = \{(\mathbf{x}, \mathbf{y}) : P(\mathbf{x}, \mathbf{y}) > 0\}$ , and to every pair of states  $\mathbf{x}, \mathbf{y} \in S$  we assign a path  $\Gamma_{\mathbf{x}, \mathbf{y}}$  from  $\mathbf{x}$  to  $\mathbf{y}$  in  $G$ . We define*

$$\rho = \max_{e=(\mathbf{z}, \mathbf{w}) \in \mathbf{E}} \frac{1}{Q(e)} \sum_{\substack{\mathbf{x}, \mathbf{y} \\ e \in \Gamma_{\mathbf{x}, \mathbf{y}}}} \pi(\mathbf{x})\pi(\mathbf{y})|\Gamma_{\mathbf{x}, \mathbf{y}}|.$$

Then  $\frac{1}{1-\lambda_2} \leq \rho$ .

We here mention a corollary of Lemma 4.5 that will be useful for our results.

**Corollary 4.6.** *Let  $\mathcal{G}$  be an  $n$ -player opinion game with profile space  $S$  and let  $P$  and  $\pi$  be the transition matrix and the stationary distribution of the logit dynamics for  $\mathcal{G}$ , respectively. 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*

$$t_{\text{rel}} \leq 2n \max_{\substack{\mathbf{z}, \mathbf{w}: \\ H(\mathbf{z}, \mathbf{w})=1 \\ \pi(\mathbf{z}) \leq \pi(\mathbf{w})}} \frac{1}{\pi(\mathbf{z})} \sum_{\substack{\mathbf{x}, \mathbf{y}: \\ (\mathbf{z}, \mathbf{w}) \in \Gamma_{\mathbf{x}, \mathbf{y}}}} \pi(\mathbf{x})\pi(\mathbf{y})|\Gamma_{\mathbf{x}, \mathbf{y}}|.$$

*Proof.* In [3] it is proved that all the eigenvalues of the transition matrix of the logit dynamics for potential games are non-negative. It follows then that  $t_{\text{rel}} = \frac{1}{1-\lambda_2}$ . 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:

$$Q(\mathbf{z}, \mathbf{w}) = \pi(\mathbf{z})P(\mathbf{z}, \mathbf{w}) \geq \frac{\pi(\mathbf{z})}{2n}.$$

Thus, the claim follows from Lemma 4.5.  $\square$

### 4.1.3 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  $\pi$ . The probability distribution  $Q(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{x})P(\mathbf{x}, \mathbf{y})$  is of particular interest and is sometimes called the *edge stationary distribution*. Note that if the chain is reversible then  $Q(\mathbf{x}, \mathbf{y}) = Q(\mathbf{y}, \mathbf{x})$ . For any  $L \subseteq S$ , we let  $Q(L, S \setminus L) = \sum_{\mathbf{x} \in L, \mathbf{y} \in S \setminus L} Q(\mathbf{x}, \mathbf{y})$ . The bottleneck ratio of  $L \subseteq S$ ,  $L$  non-empty, is

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

We use the following theorem to derive lower bounds to the mixing time (see, for example, Theorem 7.3 in [22]).

**Theorem 4.7** (Bottleneck ratio). *Let  $\mathcal{M} = \{X_t: t \in \mathbb{N}\}$  be an irreducible and aperiodic Markov chain with finite state space  $S$ , transition matrix  $P$ , and stationary distribution  $\pi$ . Then the mixing time is*

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

## 4.2 Upper bounds

### 4.2.1 For every $\beta$

Consider the bijective function  $\sigma: V \rightarrow \{1, \dots, |V|\}$  representing an ordering of vertices of  $G$ . Let  $\mathcal{L}$  be the set of all orderings of vertices of  $G$  and set  $V_i^\sigma = \{v \in V: \sigma(v) < i\}$ . Moreover, 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

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

**Theorem 4.8.** *Let  $\mathcal{G}$  be an  $n$ -player opinion game on a graph  $G = (V, E)$ . The mixing time of the logit dynamics for  $\mathcal{G}$  is*

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

*Proof.* This proof is a generalization of a similar proof given by Berger et al. [7] and by Auletta et al. [3].

Consider the ordering of vertices of  $G$  that obtains the cutwidth. Fix  $\mathbf{x}, \mathbf{y} \in S$  and let  $v_1, v_2, \dots, v_d$  denote the indices (according to this ordering) of the vertices at which the profiles  $\mathbf{x}$  and  $\mathbf{y}$  differ; we consider the path  $\Gamma_{\mathbf{x}, \mathbf{y}} = (\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^d)$  on  $\mathcal{H}$ , where

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

(Above, we assume  $v_{d+1} = n + 1$ ). Notice that  $\mathbf{x}^0 = \mathbf{x}$ ,  $\mathbf{x}^d = \mathbf{y}$  and  $|\Gamma_{\mathbf{x}, \mathbf{y}}| \leq n$ . For every edge  $\xi = (\mathbf{x}^i, \mathbf{x}^{i+1})$  of  $\mathcal{H}$ , we consider the function  $\Lambda_\xi$  that assigns to every pair of profiles  $\mathbf{x}, \mathbf{y}$  such that  $\xi \in \Gamma_{\mathbf{x}, \mathbf{y}}$ , the following new profile

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

where  $\pi$  denotes the stationary distribution (cf. Theorem 4.2). It is easy to see that  $\Lambda_\xi$  is an injective function: indeed, since  $\xi$  is known, if  $\pi(\mathbf{x}^i) \leq \pi(\mathbf{x}^{i+1})$ , then we can retrieve  $v_{i+1}$ , that is the first vertex where  $\mathbf{x}^i$  and  $\mathbf{x}^{i+1}$  differ and thus, selecting the first  $v_{i+1} - 1$  vertices from  $\Lambda_\xi(\mathbf{x}, \mathbf{y})$  and the remaining ones from  $\mathbf{x}^i$  we are able to reconstruct  $\mathbf{x}$  and, similarly, selecting the first  $v_{i+1} - 1$  vertices from  $\mathbf{x}^i$  and the remaining ones from  $\Lambda_\xi(\mathbf{x}, \mathbf{y})$  we are able to reconstruct  $\mathbf{y}$ . Similarly, if  $\pi(\mathbf{x}^i) > \pi(\mathbf{x}^{i+1})$ , we can retrieve  $v_{i+1}$  and we can reconstruct  $\mathbf{x}$  and  $\mathbf{y}$  from  $\Lambda_\xi(\mathbf{x}, \mathbf{y})$  and  $\mathbf{x}^{i+1}$ .

Let  $E^* = \{(j, k) \in E : j < v_{i+1} \text{ and } k \geq v_{i+1}\}$ : observe that  $\sum_{(j,k) \in E^*} w_{jk} \leq \text{CW}(G)$ . For any edge  $e = (j, k) \in E^*$ , 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

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

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

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

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  $y_k$

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

where  $\alpha_e, \beta_e, \gamma_e$  and  $\delta_e$  are defined in (2). Moreover for  $e = (j, k) \in E \setminus E^*$  it holds:

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

since, by construction, one of  $\perp_{\mathbf{x}^i, \mathbf{x}^{i+1}}$  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  $\xi = (\mathbf{x}^i, \mathbf{x}^{i+1}) \in \Gamma_{\mathbf{x}, \mathbf{y}}$ ,

$$\begin{aligned} \Phi(\mathbf{x}) + \Phi(\mathbf{y}) - \Phi(\perp_{\mathbf{x}^i, \mathbf{x}^{i+1}}) - \Phi(\Lambda_\xi(\mathbf{x}, \mathbf{y})) &= \sum_{e \in E} (\Phi_e(\mathbf{x}) + \Phi_e(\mathbf{y}) - \Phi_e(\perp_{\mathbf{x}^i, \mathbf{x}^{i+1}}) - \Phi_e(\Lambda_\xi(\mathbf{x}, \mathbf{y}))) \\ &= \sum_{e \in E^*} (\Phi_e(\mathbf{x}) + \Phi_e(\mathbf{y}) - \Phi_e(\perp_{\mathbf{x}^i, \mathbf{x}^{i+1}}) - \Phi_e(\Lambda_\xi(\mathbf{x}, \mathbf{y}))) \quad (8) \\ &\geq -2\text{CW}(G). \end{aligned}$$

Now let  $\xi^* = (\mathbf{z}, \mathbf{w})$  with  $\pi(\mathbf{z}) \leq \pi(\mathbf{w})$  be the edge of  $\mathcal{H}$  for which  $\sum_{\xi^* \in \Gamma_{\mathbf{x}, \mathbf{y}}} \frac{\pi(\mathbf{x})\pi(\mathbf{y})}{\pi(\mathbf{z})} |\Gamma_{\mathbf{x}, \mathbf{y}}|$  is maximized. Applying Corollary 4.6, we obtain

$$\begin{aligned} t_{\text{rel}} &\leq 2n \sum_{\substack{\mathbf{x}, \mathbf{y}: \\ \xi^* \in \Gamma_{\mathbf{x}, \mathbf{y}}}} \frac{\pi(\mathbf{x})\pi(\mathbf{y})}{\pi(\mathbf{z})} |\Gamma_{\mathbf{x}, \mathbf{y}}| \leq 2n^2 \sum_{\substack{\mathbf{x}, \mathbf{y}: \\ \xi^* \in \Gamma_{\mathbf{x}, \mathbf{y}}}} \frac{\pi(\mathbf{x})\pi(\mathbf{y})}{\pi(\mathbf{z})\pi(\Lambda_{\xi^*}(\mathbf{x}, \mathbf{y}))} \pi(\Lambda_{\xi^*}(\mathbf{x}, \mathbf{y})) \\ &\leq 2n^2 e^{2\beta\text{CW}(G)} \sum_{\substack{\mathbf{x}, \mathbf{y}: \\ \xi^* \in \Gamma_{\mathbf{x}, \mathbf{y}}}} \pi(\Lambda_{\xi^*}(\mathbf{x}, \mathbf{y})) \leq 2n^2 e^{2\beta\text{CW}(G)} \sum_{\mathbf{x}} \pi(\mathbf{x}) \leq 2n^2 e^{2\beta\text{CW}(G)}, \end{aligned}$$

where the third inequality follows from Theorem 4.2 and (8), and the penultimate from the fact that  $\Lambda_\xi$  is injective.

The theorem follows from Theorem 4.4 and by observing that, since  $\Phi(\mathbf{x}) \geq 0$  for any strategy profile  $\mathbf{x}$ , Theorem 4.2 implies

$$\begin{aligned} \log((\pi_{\min}/4)^{-1}) &= \log\left(4 \sum_{\mathbf{x}} e^{-\beta(\Phi(\mathbf{x}) - \Phi_{\max})}\right) \\ &\leq \log(2^{n+2} \cdot e^{\beta\Phi_{\max}}) \leq \log(e^{n+2+\beta\Phi_{\max}}) = n + 2 + \beta\Phi_{\max}, \end{aligned} \quad (9)$$

where  $\Phi_{\max} = \max_{\mathbf{x}} \Phi(\mathbf{x}) \leq n(1 + w_{\max})$ . □

### 4.2.2 For small $\beta$

The following theorem shows that for small values of  $\beta$  the mixing time is polynomial. We remark that there are network topologies for which this theorem gives a bound higher than that guaranteed by Theorem 4.8 on the values of  $\beta$  for which the mixing time is polynomial.

**Theorem 4.9.** *Let  $\mathcal{G}$  be an  $n$ -player opinion game on a connected graph  $G$ , with  $n > 2$ . Let  $\Delta_{\max}$  be the maximum degree in the graph. 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.* 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 coupling described in Section 4.1.1 for two chains  $X$  and  $Y$  starting respectively from  $X_0 = \mathbf{x}$  and  $Y_0 = \mathbf{y}$ . We next compute the expected distance between  $X_1$  and  $Y_1$  after one step of the coupling.

Let  $N_i$  be the set of neighbors of  $i$  in the opinion game. Notice that for any player  $i$ ,  $\sigma_i(0 | \mathbf{x})$  only depends on  $x_k$ , for any  $k \in N_i$ , and  $\sigma_i(0 | \mathbf{y})$  only on  $y_k$ , for any  $k \in N_i$ . Therefore, since  $\mathbf{x}$  and  $\mathbf{y}$  only differ at position  $j$ ,  $\sigma_i(0 | \mathbf{x}) = \sigma_i(0 | \mathbf{y})$  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  $\sigma_i(0 | \mathbf{x}) \leq \sigma_i(0 | \mathbf{y})$ . Therefore, with probability  $\sigma_i(0 | \mathbf{x})$  both chains update position  $i$  to 0 and thus remain at distance 1; with probability  $\sigma_i(1 | \mathbf{y}) = 1 - \sigma_i(0 | \mathbf{y})$  both chains update position  $i$  to 1 and thus remain at distance 1; and with probability  $\sigma_i(0 | \mathbf{y}) - \sigma_i(0 | \mathbf{x})$  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

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

Let us now evaluate the difference  $\sigma_i(0 | \mathbf{y}) - \sigma_i(0 | \mathbf{x})$  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  $\ell$  the quantity  $e^{\beta(2b_i - 1 + 2W_i^1(\mathbf{x}) - W_i)}$ . By (6) we have

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

and

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

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

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

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

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

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  $\text{diam}(\mathcal{H}) = n$ , by applying Theorem 4.3 with  $\alpha = \frac{1}{3n}$ , we obtain the theorem.  $\square$

### 4.3 Lower bound

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

**Observation 4.10.** *For every path on  $\mathcal{H}$  between the profile  $\mathbf{0} = (0, \dots, 0)$  and the profile  $\mathbf{1} = (1, \dots, 1)$  there exists a profile for which the weight of the discarding edges is at least  $\text{CW}(G)$ .*

From now on, let us write  $\text{CW}$  as a shorthand for  $\text{CW}(G)$ , 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  $\mathbf{b}^*$  the minimum of  $\mathbf{b}(\mathbf{x})$  over all profiles with  $\text{CW}$  discarding edges.

Let  $R_0$  ( $R_1$ ) 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  $\mathbf{b}^* + \text{CW}$ . To establish the lower bound we use the technical result given by Theorem 4.7 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  $R = R_0$  if  $\pi(R_0) \leq 1/2$  and  $R = R_1$  if  $\pi(R_1) \leq 1/2$ . (If both sets have stationary distribution less than one half, the best lower bound is achieved by setting  $R$  to  $R_0$  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  $R = R_0$ .

#### 4.3.1 For large $\beta$

Let  $\partial R$  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 4.11.** *For the set of profiles  $R$  defined above, we have  $B(R) \leq n \cdot |\partial R| \cdot e^{-\beta(\text{CW} + \mathbf{b}^* - \mathbf{b}(\mathbf{0}))}$ .*

*Proof.* Since  $\mathbf{0} \in R$ , it holds  $\pi(R) \geq \pi(\mathbf{0}) = \frac{e^{-\beta \mathbf{b}(\mathbf{0})}}{Z}$ . Moreover, by (6) we have

$$\begin{aligned} Q(R, \bar{R}) &= \sum_{\substack{(\mathbf{x}, \mathbf{y}) \in \mathcal{E}(\partial R): \\ \mathbf{y} = (\mathbf{x} - i, y_i)}} \frac{e^{-\beta \Phi(\mathbf{x})}}{Z} \frac{e^{\beta u_i(\mathbf{y})}}{e^{\beta u_i(\mathbf{x})} + e^{\beta u_i(\mathbf{y})}} \\ &= \sum_{\substack{(\mathbf{x}, \mathbf{y}) \in \mathcal{E}(\partial R): \\ \mathbf{y} = (\mathbf{x} - i, y_i)}} \frac{e^{-\beta \Phi(\mathbf{x})}}{Z} \frac{e^{-\beta \Phi(\mathbf{y})} e^{\beta(u_i(\mathbf{x}) + \Phi(\mathbf{x}))}}{e^{-\beta \Phi(\mathbf{x})} e^{\beta(u_i(\mathbf{x}) + \Phi(\mathbf{x}))} + e^{-\beta \Phi(\mathbf{y})} e^{\beta(u_i(\mathbf{x}) + \Phi(\mathbf{x}))}} \\ &= \frac{1}{Z} \sum_{(\mathbf{x}, \mathbf{y}) \in \mathcal{E}(\partial R)} \frac{e^{-\beta \Phi(\mathbf{x})} e^{-\beta \Phi(\mathbf{y})}}{e^{-\beta \Phi(\mathbf{x})} + e^{-\beta \Phi(\mathbf{y})}} = \frac{1}{Z} \sum_{(\mathbf{x}, \mathbf{y}) \in \mathcal{E}(\partial R)} \frac{e^{-\beta \Phi(\mathbf{y})}}{1 + e^{\beta(\Phi(\mathbf{x}) - \Phi(\mathbf{y}))}} \\ &\leq \frac{1}{Z} \sum_{(\mathbf{x}, \mathbf{y}) \in \mathcal{E}(\partial R)} e^{-\beta \Phi(\mathbf{y})} \leq |\mathcal{E}(\partial R)| \cdot \frac{e^{-\beta(\mathbf{b}^* + \text{CW})}}{Z}. \end{aligned}$$

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

From Lemma 4.11 and Theorem 4.7 we obtain a lower bound to the mixing time of the opinion games that holds for every value of  $\beta$ , every social network  $G$  and every vector  $(b_1, \dots, b_n)$  of internal beliefs. However, it is not clear how close this bound is to the one given in Theorem 4.8. Nevertheless, by taking  $b_i = 1/2$  for each player  $i$  and  $\beta$  high enough, we obtain the following theorem.

**Theorem 4.12.** *Let  $\mathcal{G}$  be an  $n$ -player opinion game on a graph  $G$ . Then, there exists a vector of internal beliefs such that for  $\beta = \Omega\left(\frac{n \log n}{\text{CW}}\right)$  it holds  $t_{\text{mix}} \geq e^{\beta \Theta(\text{CW})}$ .*

*Proof.* If  $b_i = 1/2$  for every player  $i$ , from Lemma 4.11 and Theorem 4.7, since  $|\partial R| \leq 2^n$  then

$$t_{\text{mix}} \geq \frac{e^{\beta \text{CW}}}{n2^n} = e^{\beta \text{CW} - n \log(2n)} = e^{\beta \Theta(\text{CW})}. \quad \square$$



### 4.3.2 For smaller $\beta$

Theorem 4.12 gives an (essentially) tight lower bound for high values of  $\beta$  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 4.13.** *Let  $\mathcal{G}$  be an  $n$ -player 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  $t_{\text{mix}} \geq \frac{e^{\beta \Theta(\text{CW})}}{n}$ .*

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 4.14.** *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 functions 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  $\lceil m^2/2 \rceil$ .  $\square$

The following lemma gives a bound to the size of  $\partial R$  for this graph.

**Lemma 4.15.** *For the 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  $|\partial R| \leq e^{c_1 \sqrt{\text{CW}}}$ .*

*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  $\partial R$ ) have less than  $\text{CW}$  discording edges. Indeed, for  $\mathbf{x} \in R$  we have  $\mathbf{b}(\mathbf{x}) + |D(\mathbf{x})| = \Phi(\mathbf{x}) < \mathbf{b}^* + \text{CW}$ . Moreover, if a profile  $\mathbf{y}$  has less than  $\text{CW} - m$  discording edges, then  $\mathbf{y}$  is not in  $\partial R$  as a state neighbor of  $\mathbf{y}$  has at most  $m - 1$  additional discording edges.

Consequently, to bound the size of  $\partial R$ , we need to count the number of profiles in  $R$  that have potential between  $\mathbf{b}^* + \text{CW} - m$  and  $\mathbf{b}^* + \text{CW} - 1$  (i.e., the number of profiles with at least  $\text{CW} - m$  and at most  $\text{CW} - 1$  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

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

where the lower bound follows from the structural proof of minimum cuts contained in Claim 4.14.

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  $\text{CW} - m$ , i.e.

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

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  $\text{CW}$ , i.e.

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

that, as showed in Claim 4.14, means  $t_2 = m - 2$  for  $m$  even and  $t_2 = m - 1$  for  $m$  odd. Then, we can conclude  $t_2 \leq m - 1$ .

By the definition of  $t_1$ , all profiles with at most  $t_1 - 1$  players with opinion 1 are not in  $\partial R$  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

$$|\partial R| \leq \sum_{i=t_1}^{t_2} \binom{n}{i} \leq \sum_{i=t_1}^{t_2} \left( \frac{n \cdot e}{i} \right)^i \leq \sum_{i=t_1}^{t_2} (5e)^i = \frac{(5e)^{t_2+1} - (5e)^{t_1}}{5e - 1} \leq (5e)^{t_2+1} \leq (5e)^m \leq e^{3m}, \quad (10)$$

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

*Proof of Theorem 4.13.* If  $b_i = 1/2$  for every player  $i$ , from Lemmata 4.11 and 4.15, we have

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

where  $c_2 = \frac{c_1 \sqrt{\text{CW}}}{\beta \text{CW}} < 1$  since by hypothesis  $\beta > \frac{c_1}{\sqrt{\text{CW}}} = \Omega(1/m)$ ; we also notice that  $c_2$  goes to 0 as  $\beta$  increases. The theorem follows from Theorem 4.7.  $\square$

We remark that it is possible to prove a result similar to Theorem 4.13 also for the clique  $K_n$ : the proof follows from a simple generalization of Theorem 15.3 in [22] 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 binary opinion games: 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  $\beta$  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  $\beta$  goes to infinity.

Theorem 4.8 and 4.12 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  $\beta$ . (We highlight that checking if a graph has bounded cutwidth can be done in polynomial time [27].) In general, we have the following picture: as long as  $\beta$  is less than the maximum of (roughly)  $\frac{\log n}{\text{CW}}$  and  $\frac{1}{w_{\max} \Delta_{\max}}$  the convergence time to the logit equilibrium is polynomial. Moreover, Theorem 4.12 shows that for  $\beta$  lower bounded by (roughly)  $\frac{n \log n}{\text{CW}}$  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  $\beta$ . It would be interesting to investigate existence and structure of such distributions for our 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] A. Asadpour and A. Saberi. On the inefficiency ratio of stable equilibria in congestion games. In *Proc. of International Workshop on Internet and Network Economics (WINE)*, pages 545–552, 2009.
- [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. Bhargat, 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] 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.
- [10] L. E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.
- [11] 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.
- [12] M. H. DeGroot. Reaching a consensus. *J. American Statistical Association*, 69, 1974.
- [13] P. M. DeMarzo, D. Vayanos, and J. Zweibel. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 2003.
- [14] M. Dyer and V. Mohanaraj. Pairwise-interaction games. In *Proc. of Automata, Languages and Programming (ICALP)*, pages 159–170, 2011.
- [15] G. Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071, 1993.
- [16] B. Escoffier, D. Ferraioli, L. Gourvès, and S. Moretti. Designing budget-balanced best-response mechanisms for network coordination games. In *Proc. of International Symposium on Algorithmic Game Theory (SAGT)*, pages 207–218, 2013.
- [17] 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.
- [18] 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.
- [19] N. E. Friedkin and E. C. Johnsen. Social influence and opinions. *Math. Sociology*, 15(3-4), 1990.
- [20] B. Golub and M. O. Jackson. Naive learning in social networks: Convergence, influence and the wisdom of crowds. *American Econ. J.: Microeconomics*, 2, 2010.
- [21] M. O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.
- [22] D. Levin, Yuval P., and E. L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.

- [23] D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124 – 143, 1996.
- [24] A. Montanari and A. Saberi. Convergence to equilibrium in local interaction games. In *Proc. of Symposium on Foundations of Computer Science (FOCS)*, 2009.
- [25] 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.
- [26] R. W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.
- [27] 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.