Social Pressure in Opinion Games

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

Carmine Ventre
University of Essex, UK
c.ventre@essex.ac.uk

Abstract

Motivated by privacy and security concerns in on-line social networks, we study the role of social pressure in opinion games. These are games, important in economics and sociology, that model the formation of opinions in a social network. We enrich the definition of (noisy) best-response dynamics for opinion games by introducing the pressure, increasing with time, to reach an agreement.

We prove that for clique social networks, the dynamics always converges to consensus (no matter the level of noise) if the social pressure is high enough. Moreover, we provide (tight) bounds on the speed of convergence; these bounds are polynomial in the number of players provided that the pressure grows sufficiently fast. We finally look beyond cliques: we characterize the graphs for which consensus is guaranteed, and make some considerations on the computational complexity of checking whether a graph satisfies such a condition.

1 Introduction

Opinion games focus on self-interested individuals, each with an *opinion* and all connected in some social network, in need of reaching a decision in a decentralized way (i.e., without a central authority dictating their actions). For example, they might be sitting on some hiring panel, members having their own favorite candidates and a job offer to be made, or they might be co-authors/reviewers deciding about the submission/notification of a paper.

This subject received large attention in economics and sociology literature. In particular, DeGroot [1974] defined the most prominent model for this setting, where the role of discussions in the decision making process is mathematically captured by individuals repeatedly averaging their own opinion with those of their neighbors. In the variant of Friedkin and Johnsen [1990], each individual additionally maintains a persistent *internal belief*, which remains constant even as they update their opinions through averaging. These models have attracted much attention in recent literature, see e.g., [Bhalgat *et al.*, 2010; Ferraioli *et al.*, 2016; Chierichetti *et al.*, 2013]. This line of work identifies the absence of consensus in many real-life situations and gives a game-theoretic explanation: players will not compromise any further when this increases their *cost*, defined as a measure of the distance between a player's opinion and (i) her own belief; (ii) the opinions of her neighbors on the social network.

We here note, however, that in many cases – such as the examples mentioned above – there is a pressure to reach a consensus. Such a pressure augments as time (e.g., length of the meeting, the approaching deadline) goes on. Under which conditions does this pressure facilitate consensus? How long does it take to reach the consensus, if any?

As noted in [Rajtmajer *et al.*, 2016], these questions bear a certain degree of importance for security and privacy in Online Social Networks (OSNs) and, more generally, for distributed (multi-agent) access control policies. Access control to contents shared on OSNs is a central research topic in security and privacy. Typically, the uploader gets to decide the degree of access (e.g., "friends"; "friends of friends"; etc.) of a shared content (e.g., a picture of a group of people). However, the sensibility to privacy issues of the uploader might differ from that of the others interested in the shared content (e.g., the others in the picture). A distributed protocol could involve all these individuals with the objective to reach a consensus on the accessibility of the content, in such a way to cater for the needs of everyone. The social pressure might be enforced, for example, by only publishing on the OSN upon consensus. The authors of [Rajtmajer *et al.*, 2016] introduce a number of important themes related to this applicative scenario, by defining a game-theoretic model reminiscent of opinion games. However, as we discuss below, they fail to give (complete) answers to our questions of interest.

1.1 Our Contribution

We introduce new dynamical models for opinion games. Specifically, we generalize the definitions of (noisy) best-response dynamics in [Bhalgat *et al.*, 2010; Ferraioli *et al.*, 2016] along two main dimensions. Firstly, as detailed above, we introduce a non-decreasing pressure to coordinate opinions with the neighbors. Secondly, we do not restrict the way the players weigh disagreements. Related literature makes restrictive assumptions on the cost of disagreeing (either with a neighbor or with one's own belief). We here instead use any pair of functions, f and g, that measure the cost of having an opinion different from the belief and a neighbor's opinion, respectively.

In the case in which the underlying social network is a clique (as in many real world setting, e.g. the hiring committee example mentioned above), we prove (rate of) convergence to consensus of these dynamics. We are able to describe very well the behavior of best-response dynamics: once players start to deviate from the initial strategy profile they move onto the opinion played by the majority; this way the initial majority keeps increasing until consensus. Incidentally, this also proves that each player moves only once and, therefore, the rate of convergence – once deviations commence – is polynomial in the number of players. We determine the minimum level of social pressure needed to kick off deviations (as the ratio between the cost of having an opinion different from the belief and the cost of disagreeing with a neighbor) and give an instance showing that our bound on convergence is tight.

In many cases, it might not be realistic to assume that the players are always able to choose their best response. For this reason, we consider players with *bounded rationality*. In order to model the behavior of these players, several *noisy* best response update rules have been introduced. In this work we will focus on one of the most successful, namely *logit update rule* [Blume, 1993], according to which the player selected for update can play every strategy with a probability that is proportional to her advantage in adopting that strategy and to the rationality level $\beta > 0$. We prove that this dynamics always converges to consensus in at most n^3 steps as long as the social pressure is above a given threshold (that depends not only on the ratio between disagreements costs, but also on the rationality level β). This result is achieved by reducing the dynamics to a birth and death chain and evaluating the time the latter takes to reach the consensus.

For both dynamics, our results can be extended in several ways to accommodate more general settings (e.g., better rather than best responses; simultaneous rather than sequential moves; alternative noisy update rules; etc.). A full discussion of these extensions is deferred to the full version of the paper.

We then move onto general social networks and characterize the stationary points (different from consensus) of best-response dynamics in terms of the existence of a certain cut of the graph. For a graph to admit an equilibrium in which players' opinions disagree there must exist a partition of vertices in (at least) two sets such that each vertex has more neighbors in its side of the partition than in the other(s). Can we recognize in polynomial-time whether a graph will guarantee consensus (i.e., will not admit such a cut)? We give a partial answer to this question as we are only able to efficiently recognize (a subset of the) graphs for which consensus might not occur. We leave open the problem of establishing the exact computational complexity of this problem.

1.2 Related Works

Understanding how opinions are formed and expressed in a social context has been object of extensive recent study, in AI and multiagent systems [Pryymak *et al.*, 2012; Tsang and Larson, 2014; Grandi *et al.*, 2015; Schwind *et al.*, 2015], CS at large [Acemoglu and Ozdaglar, 2011; Bindel *et al.*, 2011; Mossel and Tamuz, 2014; Auletta *et al.*, 2015], as well as, so-

ciology, economics, physics, and epidemiology. In particular, the work of Friedkin and Johnsen [1990], which represents our starting point, has been largely studied recently and has then emerged as the principal model in the area. For example, Bindel *et al.* [2011] considered this model and proved that, under mild assumptions, whenever beliefs and opinions belong to $[0, 1]$, the repeated averaging process leads to a *unique* equilibrium, describing the opinion that each agent eventually expresses. Chierichetti *et al.* [2013] considered the case that opinions are discrete and bound the price of stability and price of anarchy of the resulting games. Extensions of this model have been proposed by Bhawalkar *et al.* [2013], by Auletta *et al.* [2016] and by Bilò *et al.* [2016]. A work that is closely related to this paper is [Ferraioli *et al.*, 2016]. Indeed, it focuses on the rate of convergence of opinion dynamics to equilibria under both best-response and logit dynamics. However, their model is simpler than ours (e.g., they consider binary opinion and fixed cost for disagreements) and thus extending their results to our setting is not straightforward. Moreover, all these works do not consider the effects on agents of time-varying social pressure.

Logit dynamics has been introduced by [Blume, 1993] to model agents with bounded rationality. This dynamics is founded on well-established measure theory tools for modeling limited cognitive capacities [McFadden, 1974], and it is largely adopted nowadays for modeling the diffusion of information and opinions on social networks [Peyton Young, 2006; Montanari and Saberi, 2009; Ferraioli *et al.*, 2016; Auletta *et al.*, 2013b; 2013a].

It is worthy to compare our results with those in [Rajtmajer *et al.*, 2016] as our models are similar in spirit. There are a number of notable differences. Firstly, they consider simultaneous moves only, while we are able to deal with the arguably more realistic case of asynchronous moves as well. Secondly, their analytical results only apply to cliques and a unique continuous set from which players choose beliefs and opinions; our results, instead, hold no matter which pair of sets players use for beliefs/opinions and the level of granularity of these sets (incidentally, discrete sets seem to fit better the application of opinion games to OSNs). Thirdly, their convergence result requires an infinitely big level of social pressure while no guarantee on the speed of convergence is given; instead, we do give (tight) guarantees of convergence rate with "reasonable" values of pressure. Finally, they only provide experimental results for players with bounded rationality.

2 The Model

Let $G = (V, E)$ be a connected undirected graph with $|V| = n$. Every vertex of the graph represents a player. Each player i has a *belief* $b_i \in B_i$ (e.g., her preferred privacy setting in the OSN scenario) and can choose a (potentially) different *opinion* $x_i \in S_i$ (e.g., privacy setting in the OSN application). For every pair $x, y \in \bigcup_i (S_i \cup B_i)$, we will denote as $x - y \in [0, 1]$ their distance. Specifically, for every $x, y \in \bigcup_i S_i$, we assume that $x - y = 1$ if $x \neq y$, and $x - y = 0$, otherwise. Note that this assumption fits particularly well our application scenarios. In particular, different metrics could be used for the difference between strategies;

1 Let $x_i^{(0)} = b_i$ for all i

2 Let $k = 0$

3 **while** $\exists \, (j, l) \in E$ s.t. $x_l^{(k)} \neq x_j^{(k)}$ **do**

4   **if** $\exists i$ such that $x_i^{(k)} \neq BR_i(\mathbf{x}_{-i}^{(k)})$ **then**

5   Let i be such an agent

6   Set $x_i^{(k+1)} = BR_i(\mathbf{x}_{-i}^{(k)})$

7   Set $x_j^{(k+1)} = x_j^{(k)}$ for $j \neq i$

8   **else** Set $x_i^{(k+1)} = x_i^{(k)}$ for all i

9   Increment k by 1

Algorithm 1: The dynamics

our "binary" definition serves well our focus on consensus since it is not really relevant how "much" an agent is disagreeing with someone else but only that they disagree. Moreover, this definition allows us to consider even beliefs and opinions that cannot be embedded in a metric space (as instead it is the case in most of the previous literature on the topic).

According to well-established works in sociology and economics [DeGroot, 1974; Friedkin and Johnsen, 1990], the cost of player i in a strategy profile $\mathbf{x} \in S_1 \times \cdots \times S_n$ can be described as

$$c_i(\mathbf{x}) = f_i(x_i - b_i) + \rho \cdot \sum_{(i,j) \in E} g_i(x_i - x_j),$$

where $\rho > 0$ represents the *social pressure* to reach consensus, $f_i \colon [0, 1] \to \mathbb{R}_{\geq 0}$ is a non-decreasing function, and $g_i \colon \{0, 1\} \to \mathbb{R}_{\geq 0}$ is a function such that $g_i(1) > g_i(0)$[1]. Here, f_i and g_i measure the effects to player i of disagreeing with her own beliefs and a neighbor's opinion, respectively. We assume that $f_i(0) = g_i(0) = 0$ for every i. However, our results hold even if strategy/belief distances can be larger than 1 or $f_i(0), g_i(0) \neq 0$ by rescaling the functions f_i, g_i.

To simplify the exposition and present the main ideas, we will henceforth assume that $B_i = B$, $S_i = S$, $f_i = f$ and $g_i = g$ for every i. Moreover, we assume that $B = S$ and that this set is finite; however, we stress that our results can be generalized to different and player-specific sets of infinite cardinality and to player-specific functions.

Let $BR_i(\mathbf{x}_{-i}) = \arg\min_{x \in S} c_i(x, \mathbf{x}_{-i})$ for every $\mathbf{x}_{-i} \in S^{n-1}$. The *best-response consensus dynamics* is defined in Algorithm 1 under the assumption that the social pressure is non-decreasing in k, i.e., $0 < \rho^{(0)} \leq \rho^{(1)} \leq \rho^{(2)} \leq \cdots$, with $\lim_{k \to \infty} \rho^{(k)} \geq \rho^*$. As discussed above, one novelty of our definition is in the role of the ρ's which, ultimately, resides on the value of ρ^*. In fact, ρ^* needs to be big enough to incentivize consensus, for otherwise, we either have a situation wherein no player moves from her belief (when ρ^* is too small) or fall back to a generalized notion of opinion games related to previous studies (when ρ^* is not big enough).

We are also interested in a noisy version of the dynamics of Algorithm 1 wherein players' responses are perturbed by some noise. Specifically, we look at *logit dynamics*, according to which player i plays strategy x_i as a response to $\mathbf{x}_{-i}$

---

[1]If $g_i(1) = g_i(0)$ then there would be no incentive to coordinate.

with a probability proportional to $e^{-\beta c_i(x_i, \mathbf{x}_{-i})}$, with $\beta > 0$ (more details can be found in Sect. 3.1).

3 Clique Social Networks

Theorem 1. *If G is a clique and $\rho^* > \frac{f(1)}{g(1)}$, then Algorithm 1 converges in time at most $n - 1 + k^*$, where*

$$k^* = \min \left\{ k \mid \rho^{(k)} > \frac{f(1)}{g(1)} \right\}.$$

Moreover, there is an instance in which the dynamics takes exactly $n - 1 + k^*$ steps to converge to consensus.

Proof. Given $\mathbf{x}^{(k)}$, the strategy profile at round k of the dynamics, we let $n_s^{(k)}$ denote the number of players playing strategy $s \in S$ in $\mathbf{x}^{(k)}$.

Claim 2. *If $\rho^{(k)} > \frac{f(1)}{g(1)}$ and $\mathbf{x}^{(k)}$ is not a consensus then $\mathbf{x}^{(k+1)} \neq \mathbf{x}^{(k)}$.*

Proof. Since $\mathbf{x}^{(k)}$ is not a consensus, $n_s^{(k)} > 0$ for at least two strategies. Let x_+ be the strategy played by more players in $\mathbf{x}^{(k)}$ and x_- be the strategy played by least (non-zero) number of players in $\mathbf{x}^{(k)}$. For player i playing x_- we have $c_i(\mathbf{x}^{(k)}) - c_i(x_+, \mathbf{x}_{-i}^{(k)}) > -f(1) + \frac{f(1)}{g(1)} g(1) = 0$, where the inequality follows from $n_{x_+}^{(k)} \geq n_{x_-}^{(k)}$ and the lower bound to $\rho^{(k)}$. The test in Line 4 of the dynamics will be then true (as at least i will satisfy it) and thus $\mathbf{x}^{(k+1)} \neq \mathbf{x}^{(k)}$. $\qquad \square$

Let now $k_{\min}$ be the first round such that $\rho^{(k_{\min})} > \frac{f(1)}{g(1)}$. Clearly, if $\mathbf{x}^{(k_{\min})}$ is a consensus, we are done as the dynamics has converged in less that k^* steps. Thus we assume that $\mathbf{x}^{(k_{\min})}$ is not a consensus and, by Claim 2, conclude that $\mathbf{x}^{(k_{\min}+1)} \neq \mathbf{x}^{(k_{\min})}$. Then, we set $s_{\max} = \arg\max_{s \in S} |\{i \in V \colon x_i^{(k_{\min}+1)} = s\}|$, i.e., $s_{\max}$ denotes the strategy[2] played in $\mathbf{x}^{(k_{\min}+1)}$ by the majority of players. Moreover, let $D(\mathbf{x}) = |\{i \in V \mid x_i \neq s_{\max}\}|$, i.e., $D(\mathbf{x})$ is the number of players with an opinion different from $s_{\max}$ in $\mathbf{x}$. Note that, in general, $D(\mathbf{x}) \leq n-1$, while for a consensus profile $\mathbf{x}$, $D(\mathbf{x}) = 0$.

Claim 3. *For every $k \geq k_{\min}$, if $\mathbf{x}^{(k+1)} \neq \mathbf{x}^{(k)}$ then $D(\mathbf{x}^{(k+1)}) = D(\mathbf{x}^{(k)}) - 1$.*

Proof. For every $k \geq k_{\min}$, if $\mathbf{x}^{(k+1)} \neq \mathbf{x}^{(k)}$, let i be the player that deviates from $x_i = x_i^{(k)}$ to $y_i = x_i^{(k+1)} = BR_i(\mathbf{x}_{-i}^{(k)})$. Since i deviates in round $k + 1$, we have that

$$f(x_i - b_i) + \rho^{(k)} g(1) \left(\sum_{s \neq x_i, y_i} n_s^{(k)} + n_{y_i}^{(k)} \right) =$$

$$c_i(x_i, \mathbf{x}_{-i}^{(k)}) > c_i(y_i, \mathbf{x}_{-i}^{(k)}) =$$

$$f(y_i - b_i) + \rho^{(k)} g(1) \left(\sum_{s \neq x_i, y_i} n_s^{(k)} + (n_{x_i}^{(k)} - 1) \right).$$

---

[2]In the proof of Claim 3, we prove that this strategy is unique.

Then

$$f(x_i - b_i) - f(y_i - b_i) + \rho^{(k)}g(1)(n_{y_i}^{(k)} - n_{x_i}^{(k)} + 1) \quad (1)$$

is positive.

We now prove a general result on the behavior of the dynamics, that implies the claim. Namely, we show that if $k \geq k_{\min}$, then when players deviate, they move onto $s_{\max}$ and will never switch to a different strategy afterwards. To this aim, let K' be the set of rounds $k' \geq k_{\min}$ in which a player changes her strategy. We prove that there is a unique strategy adopted by the majority of players in $\mathbf{x}^{(k_{\min}+1)}$ and, by induction on $k' \in K'$, that $n_{s_{\max}}^{k'}$ is increasing in k' and $n_s^{k'}$ is decreasing in k' for every $s \neq s_{\max}$.

The base case is $k' = k_{\min}$. Let j denote the player switching from $x_j = x_j^{(k_{\min})}$ to $y_j = x_j^{(k_{\min}+1)}$ in round $k_{\min}$. We want to prove that $y_j = s_{\max}$, i.e., y_j is the unique strategy played by more players in $\mathbf{x}^{(k_{\min}+1)}$. Recall that y_j is the best response, and thus $y_j = \arg\min_{s \neq x_j \in S} c_i(s, \mathbf{x}_{-j}^{(k_{\min})})$. Since, for every $s \in S$ it holds that $s - b_j \leq 1$ and $\rho^{(k_{\min})}g(1) > f(1)$, then we have $n_{y_j}^{(k_{\min})} = \max_{s \neq x_j \in S} n_s^{(k_{\min})}$. Moreover, since the difference between the f's in (1) is at most $f(1)$, $\rho^{(k_{\min})}g(1) > f(1)$, and $n_s^{(k_{\min})}$ is an integer for every $s \in S$, then (1) is satisfied if and only if $n_{y_j}^{(k_{\min})} \geq n_{x_j}^{(k_{\min})}$. Hence, for every $s \neq y_j$

$$n_{y_j}^{(k_{\min}+1)} > n_{y_j}^{(k_{\min})} \geq n_s^{(k_{\min})} \geq n_s^{(k_{\min}+1)},$$

thus showing that y_j is the unique strategy for which we have that $n_{y_j}^{(k_{\min}+1)} = \max_{s \in S} n_s^{(k_{\min}+1)}$.

Assume now that the claim is true for $k' - 1$; we prove it for k'. Let ℓ be the player moving at round k'. By inductive hypothesis, $s_{\max}$ is the one strategy played by more players in $\mathbf{x}^{(k'-1)}$. So if $x_\ell^{(k'-1)} \neq s_{\max}$, by the same argument of the base case, ℓ will switch to $s_{\max}$. Moreover, no ℓ such that $x_\ell^{(k'-1)} = s_{\max}$ will deviate from $s_{\max}$. In fact, as noted above, this would require ℓ to move to a strategy played by at least as many players playing $s_{\max}$ – this is impossible. $\square$

It is not hard to see that the two claims above yield the upper bound. Observe that the result holds no matter which agent is selected at Line 5 of the dynamics (if multiple choices are available). However, this choice influence $s_{\max}$, and, hence, the opinion on which the agents converge.

For the lower bound, we need to prove the following claim.

Claim 4. *Suppose that for all players i and strategies $s \in S$ we have $s - b_i = 1$ if $s \neq b_i$ and $s - b_i = 0$ otherwise, $c_i(x_i^{(0)}, \mathbf{x}_{-i}^{(0)}) - c_i(s, \mathbf{x}_{-i}^{(0)}) \leq 0$ and $\mathbf{x}^{(0)}$ is not a consensus. Then $\mathbf{x}^{(k+1)} = \mathbf{x}^{(0)}$ for every k such that $\rho^{(k)} \leq \frac{f(1)}{\delta g(1)}$, where $\delta = \max_{s:\, n_s^{(0)} > 0} \{n_{s'}^{(0)} - n_s^{(0)} + 1 \mid s' \in S\}$.*

Proof. Suppose that $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$ and $\mathbf{x}^{(k)}$ is not a consensus. Then for all players i and $s \in S$, $c_i(x_i^{(k)}, \mathbf{x}_{-i}^{(k)}) - c_i(s, \mathbf{x}_{-i}^{(k)}) \leq 0$. This yields

$$\rho^{(k)}g(1)(n_s^{(k)} - n_{x_i^{(k)}}^{(k)} + 1) \leq f(s - b_i) - f(x_i^{(k)} - b_i).$$

Note that the RHS of this inequality cannot be $-f(1)$. Indeed, if this is the case, since $\rho^{(k)}$ is non-negative and $g(1) > 0$, the inequality would be violated for $n_s^{(k)} \geq n_{x_i^{(k)}}^{(k)}$ – this always occurs since $\mathbf{x}^{(k)}$ is not a consensus. Therefore, we can safely substitute the RHS of the inequality with $f(1)$. Then, we have that $\rho^{(k)} \leq \frac{f(1)}{g(1)(n_s^{(k)} - n_{x_i^{(k)}}^{(k)} + 1)}$ for every i and every s.

Hence, given $\mathbf{x}^{(0)}$ and $\rho^{(k)}$ as from the hypothesis, the theorem follows by a simple inductive argument. $\square$

Now consider the following instance: for every $s, b \in S = B$ we have $s - b = 1$ if $s \neq b$ and $s - b = 0$ otherwise, $f(1) = g(1) = 1$ and every player has a different belief. In this case $\mathbf{x}^{(0)}$ is not a consensus profile, $\delta = 1$ and if $\rho^{(0)} \leq \frac{f(1)}{\delta g(1)}$, then $c_i(x_i^{(0)}, \mathbf{x}_{-i}^{(0)}) - c_i(s, \mathbf{x}_{-i}^{(0)}) \leq 0$ for every i and s. The lower bound then follows from the three claims. $\square$

3.1 Noisy Dynamics

For sake of presentation, let us consider now $S = B = \{0, 1\}$ (we emphasize again that our arguments do generalize to different settings). Let also $\ell(\mathbf{x})$ be the number of 1's in the opinion profile $\mathbf{x}$. By using the notation of the proof of Theorem 1, we note that if $\mathbf{x} = \mathbf{x}^{(k)}$ for some $k \geq 0$, then $\ell(\mathbf{x}) = n_1^{(k)}$.

Now we assume that players only have bounded rationality and update their opinion according to a *logit update rule*. I.e., after k steps, given that the current strategy profile is $\mathbf{x} = \mathbf{x}^{(k)}$, player i is selected at random for update and adopts opinion 1 with probability $P_i^k(\mathbf{x}) = \frac{e^{\beta c_i(\mathbf{x}_{-i}, 1)}}{e^{\beta c_i(\mathbf{x}_{-i}, 1)} + e^{\beta c_i(\mathbf{x}_{-i}, 0)}}$. By setting $\alpha(k) = e^{-\beta \rho^{(k)} g(1)}$ and $C = e^{-\beta f(1)}$, we have that

$$P_i^k(\mathbf{x}) = \begin{cases} \dfrac{\alpha(k)^{n - \ell(\mathbf{x}) - 1}}{\alpha(k)^{n - \ell(\mathbf{x}) - 1} + \alpha(k)^{\ell(\mathbf{x})} \cdot C^{-(1 - b_i)}}, & \text{if } x_i = 0; \\[2ex] \dfrac{\alpha(k)^{n - \ell(\mathbf{x})}}{\alpha(k)^{n - \ell(\mathbf{x})} + \alpha(k)^{\ell(\mathbf{x}) - 1} \cdot C^{-(1 - b_i)}}, & \text{if } x_i = 1. \end{cases}$$

We denote as τ the smallest integer for which $\mathbf{x}^{(\tau)}$ is such that $\ell(\mathbf{x}^{(\tau)}) \in \{0, n\}$. I.e., τ is the time the dynamics takes for reaching the consensus. Let $\mathbf{E}_\mathbf{x}[\tau]$ be the expectation of τ given that $\mathbf{x}^{(0)} = \mathbf{x}$. Fix $M = \max\left\{\frac{f(1)}{g(1)}, \frac{3}{\beta g(1)n}, \frac{2\log n}{\beta g(1)n}\right\}$. We prove the following theorem.

Theorem 5. *If G is a clique, $S = B = \{0, 1\}$, and $\rho^* > M$, then, for every $\mathbf{x}$, $\mathbf{E}_\mathbf{x}[\tau] \leq n^3 + k^*$, where $k^* = \min\{k \mid \rho^{(k)} > M\}$.*

To prove this theorem, let us consider an alternative process that works as follows: after k steps, given that the current strategy profile is $\mathbf{x} = \mathbf{x}^{(k)}$, a player i is selected at random and adopts opinion 1 with probability $Q_i^k(\mathbf{x}) = P_i^k(\mathbf{x})$ if $k < k^*$, otherwise $Q_i^k(\mathbf{x})$ is

$$\begin{cases} \dfrac{\alpha(k)^{n - \ell(\mathbf{x}) - 1}}{\alpha(k)^{n - \ell(\mathbf{x}) - 1} + \alpha(k)^{\ell(\mathbf{x})} \cdot C^{-1}}, & \text{if } x_i = 0 \text{ and } \ell(\mathbf{x}) > \frac{n-1}{2}; \\[2ex] \dfrac{\alpha(k)^{n - \ell(\mathbf{x}) - 1}}{\alpha(k)^{n - \ell(\mathbf{x}) - 1} + \alpha(k)^{\ell(\mathbf{x})} \cdot C}, & \text{if } x_i = 0 \text{ and } \ell(\mathbf{x}) \leq \frac{n-1}{2}; \\[2ex] \dfrac{\alpha(k)^{n - \ell(\mathbf{x})}}{\alpha(k)^{n - \ell(\mathbf{x})} + \alpha(k)^{\ell(\mathbf{x}) - 1} \cdot C^{-1}}, & \text{if } x_i = 1 \text{ and } \ell(\mathbf{x}) \geq \frac{n+1}{2}; \\[2ex] \dfrac{\alpha(k)^{n - \ell(\mathbf{x})}}{\alpha(k)^{n - \ell(\mathbf{x})} + \alpha(k)^{\ell(\mathbf{x}) - 1} \cdot C}, & \text{if } x_i = 1 \text{ and } \ell(\mathbf{x}) < \frac{n+1}{2}. \end{cases}$$

Roughly speaking, this new dynamics decreases its distance from consensus slower than the original dynamics. We next formally state this property.

Lemma 6. *Let τ' be as τ but with respect to the new process in place of the original dynamics. Then, $\mathbf{E_x}\left[\tau\right] \leq \mathbf{E_x}\left[\tau'\right]$.*

Proof Sketch. Let $\mathbf{x}^{(k)}$ be the strategy profile reached by the original dynamics after k steps, and $\mathbf{y}^{(k)}$ be the strategy profile reached by the new dynamics after the same number of steps. Let us also define $\delta(k) = \min\{\ell(\mathbf{x}^{(k)}), n - \ell(\mathbf{x}^{(k)})\}$, $\delta'(k) = \min\{\ell(\mathbf{y}^{(k)}), n-\ell(\mathbf{y}^{(k)})\}$, and $\Delta(k) = \delta'(k) - \delta(k)$.

The lemma follows by showing that the dynamics can be coupled so that, if $\Delta(0) \geq 0$, then $\Delta(k) \geq 0$ for each $k \geq 0$.

The idea behind the coupling is the following: first we try to match as many agents as possible in the original dynamics with agents in the new dynamics so that for both dynamics the agent's opinion agrees with the majority. Then the coupling picks a random agent i in the original dynamics and the matched agent in the new dynamics, and let these agents to both play the majority opinion or both play the minority opinion with the largest possible probability. It is easy then to see that with the remaining probability only the original dynamics moves towards consensus, as desired. □

Consider now a birth and death chain on $\{0, \ldots, n\}$ such that the probability p_ℓ of going from state ℓ to state $\ell+1$ is

$$p_\ell = \frac{n-\ell}{n} \cdot \begin{cases} \frac{\alpha(k^*)^{n-\ell-1}}{\alpha(k^*)^{n-\ell-1}+\alpha(k^*)^\ell \cdot C^{-1}}, & \text{if } \ell > \frac{n-1}{2}; \\ \frac{\alpha(k^*)^{n-\ell-1}}{\alpha(k^*)^{n-\ell-1}+\alpha(k^*)^\ell \cdot C}, & \text{otherwise.} \end{cases}$$

The probability q_ℓ of going from state ℓ to state $\ell-1$ is

$$q_\ell = \frac{\ell}{n} \cdot \begin{cases} 1 - \frac{\alpha(k^*)^{n-\ell}}{\alpha(k^*)^{n-\ell}+\alpha(k^*)^{\ell-1} \cdot C^{-1}}, & \text{if } \ell \geq \frac{n+1}{2}; \\ 1 - \frac{\alpha(k^*)^{n-\ell}}{\alpha(k^*)^{n-\ell}+\alpha(k^*)^{\ell-1} \cdot C}, & \text{otherwise.} \end{cases}$$

With the remaining probability the chain remains in ℓ. We denote as $s(k)$ the position of this chain after k steps and as $\tau_{0,n}$ the smallest integer such that $s(\tau_{0,n}) \in \{0, n\}$. Let $\mathbf{E}_\ell\left[\tau_{0,n}\right]$ be the expectation of $\tau_{0,n}$ given that $s(0) = \ell$. We have the following lemma, whose proof resembles the one for Lemma 6.

Lemma 7. $\mathbf{E_x}\left[\tau'\right] \leq k^* + \mathbf{E}_{\ell(\mathbf{x}^{(k^*)})}\left[\tau_{0,n}\right]$.

In order to bound $\mathbf{E}_\ell\left[\tau_{0,n}\right]$, let us consider an alternative birth and death chain defined on the set of states $\{\frac{n}{2}, \ldots, n\}$, such that the probability of a transition from state ℓ to state $\ell+1$ is $p_\ell^* = p_\ell$ for every $\ell > \frac{n}{2}$, and it is $p_\ell^* = p_\ell + q_\ell$ if $\ell = \frac{n}{2}$, whereas the probability of a transition from state ℓ to state $\ell-1$ is $q_\ell^* = q_\ell$ if $\ell > \frac{n}{2}$, and $q_\ell' = 0$ if $\ell = \frac{n}{2}$. It is not hard to see that, if $\ell \geq \frac{n}{2}$, then $\mathbf{E}_\ell\left[\tau_{0,n}\right] \leq \mathbf{E}_\ell\left[\tau_n^*\right]$, where τ_n^* is the first time step in which this new chain hits the state n. The case for $\ell \leq \frac{n}{2}$ is symmetric.

We next show that $\mathbf{E}_\ell\left[\tau_n^*\right] \leq n^3$, whenever $\ell \geq \frac{n}{2}$. Therefore, we can conclude that for every starting profile $\mathbf{x}$, the dynamics converges to consensus in expected time $\mathbf{E_x}\left[\tau\right] \leq \mathbf{E_x}\left[\tau'\right] \leq k^* + \mathbf{E}_{\ell(\mathbf{x})}\left[\tau_{0,n}\right] \leq k^* + \mathbf{E}_\ell\left[\tau_n^*\right] \leq k^* + n^3$.

The proof of this last step follows standard arguments (see, e.g., [Auletta *et al.*, 2012]). Indeed, it is well known

that $\mathbf{E}_\ell\left[\tau_n^*\right] = \sum_{i=\ell+1}^{n} \sum_{j=\frac{n}{2}}^{i-1} \frac{1}{p_j^*} \prod_{m=j+1}^{i-1} \frac{q_m^*}{p_m^*}$. Now observe that $\frac{q_m^*}{p_m^*} \leq \frac{1+\frac{h}{n}}{1-\frac{h}{n}} \cdot \alpha(k^*)^{h+1} C^{-1}$ for every $m \in \{\frac{n}{2} + 1, \ldots, n-1\}$, where $h = 2m-n \geq 2$ and $t = \alpha(k^*)^{h+1} C^{-1}$.

Observe that $\alpha(k^*)^{h+1} C^{-1} = e^{-\beta[\rho^{(k^*)}g(1)(h-1)+f(1)]}$. Since $\rho^{(k^*)} > \frac{f(1)}{g(1)}$, then $\alpha(k^*)^{h+1}C^{-1} \leq e^{-\beta\rho^{(k^*)}g(1)h}$. Moreover, if $m \leq \frac{3n}{4}$ (and, thus, $h \leq \frac{n}{2}$), then, since $e^{\frac{x}{1+x}} \leq 1 + x \leq e^x$, $\frac{1+\frac{h}{n}}{1-\frac{h}{n}} \leq e^{\frac{h}{n}} e^{\frac{h}{n-h}} \leq e^{\frac{3h}{n}}$. Thus, since $\rho^{(k^*)} > \frac{3}{\beta g(1)n}$, $\frac{q_m^*}{p_m^*} \leq e^{-h\left(\beta\rho^{(k^*)}g(1)-\frac{3}{n}\right)} \leq 1$. If $\frac{3n}{4} < m \leq n-1$ (and thus, $h > \frac{n}{2}$), then $\frac{m}{n-m} \leq n$. Then, $\frac{q_m^*}{p_m^*} \leq e^{-\left(\beta\rho^{(k^*)}g(1)h-\log n\right)} \leq e^{-\left(\frac{n}{2}\beta\rho^{(k^*)}g(1)-\log n\right)} \leq 1$, where in the last inequality we used that $\rho^{(k^*)} > 2\frac{\log n}{\beta g(1)n}$. Moreover, we have that $\frac{1}{p_j^*} = \frac{n}{n-j}\left(1 + \alpha(k^*)^{2j-n+1} \cdot C^{-1}\right)$. Since for every $j \in \{\frac{n}{2}, \ldots, n-1\}$ we have that $\frac{n}{n-j} \leq n$ and $\alpha(k^*)^{2j-n+1} \cdot C^{-1} \leq \alpha(k^*)^{2j-n} \leq 1$, then it follows that $\mathbf{E}_\ell\left[\tau_n^*\right] \leq \sum_{i=\ell+1}^{n} \sum_{j=\frac{n}{2}}^{i-1} n \leq n^3$.

4 Other Social Networks

We begin by characterizing the *stationary points* of our dynamics, i.e., profiles in which no player will have an incentive to change their opinion in any of the next steps of the dynamics. Given graph G, agent i and profile $\mathbf{x}$ we denote with $N_x^i(\mathbf{x})$, for $x \in S$, the number of neighbors of i in G playing strategy x in $\mathbf{x}$.

Lemma 8. *If $\rho^* > \frac{f(1)}{g(1)}$, then the profile $\mathbf{x}$ can be a stationary point for the dynamics on $G = (V, E)$ if and only if for all agents i and every opinion $y_i \neq x_i$, it holds that*

$$N_{x_i}^i(\mathbf{x}) - N_{y_i}^i(\mathbf{x}) \geq \begin{cases} 1 & \text{if } f(y_i - b_i) < f(x_i - b_i); \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Suppose that $\mathbf{x}^{(k)} = \mathbf{x}$ for some k such that $\rho^{(k)} > \frac{f(1)}{g(1)}$. We have that $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$ if and only if for all $i \in V$, and for every $y_i \neq x_i$, $c_i(\mathbf{x}^{(k)}) \leq c_i(y_i, \mathbf{x}_{-i}^{(k)})$. This is equivalent to

$$N_{x_i}^i(\mathbf{x}) - N_{y_i}^i(\mathbf{x}) \geq \frac{f(x_i - b_i) - f(y_i - b_i)}{\rho^{(k)}g(1)}.$$

If $f(y_i - b_i) \geq f(x_i - b_i)$, then the RHS of this inequality is in $\left[-\frac{f(1)}{\rho^{(k)}g(1)}, 0\right] \subseteq (-1, 0)$, since $\rho^{(k)} > \frac{f(1)}{g(1)}$. However, since $N_x^i(\mathbf{x}^{(k)})$ and $N_y^i(\mathbf{x}^{(k)})$ are integers, then this inequality is equivalent to the desired condition.

If $f(y_i - b_i) < f(x_i - b_i)$, then the RHS of the inequality is in $\left(0, \frac{f(1)}{\rho^{(k)}g(1)}\right] \subseteq (0, 1)$, and thus the inequality is still equivalent to the desired condition.

Moreover, since $\rho^{(k+1)} \geq \rho^{(k)}$, the inequality will hold also at round $k + 1$; then, by induction, $\mathbf{x}$ is stationary. □

Note that a consensus profile $\mathbf{x}$ is always a stationary point for the dynamics, no matter the graph G, as $N_y^i(\mathbf{x}) = 0$ for all

i and all $y \neq x_i$. However, for some graphs G, there might be additional stationary points that are different from consensus – in such cases, we say that the dynamics on G *diverges*.

Given $G = (V, E)$, for $v \in V$ and $A \subset V$, we let $N_v(A) = |\{j \in A \mid (v, j) \in E\}|$. We say that G is *well-partitioned* if V can be partitioned in sets $V_0, V_1, \ldots, V_{m-1}$, with $m > 1$, such that, for all $b, c \in \{0, 1, \ldots m-1\}$ with $b \neq c$ and $v \in V_b$, $N_v(V_b) \geq N_v(V_c)$.

Theorem 9. *If $\rho^* > \frac{f(1)}{g(1)}$, the dynamics on G diverges if and only if G is well-partitioned.*

Proof. Let us start by proving that whenever G is well-partitioned then there is an instance on which the dynamics on G diverges. Indeed, by hypothesis, there exists a partition of V in $V_0, V_1, \ldots, V_{m-1}$, with $m > 1$ such that for all $v \in V_b$, $N_v(V_b) \geq N_v(V_c)$, for every $b, c \in \{0, 1, \ldots, m-1\}$ with $b \neq c$. We can then naturally define $B = S = \{0, 1, \ldots, m-1\}$ and the belief of $i \in V_b$ to be b; moreover we can also set $\rho^{(0)} = \rho^*$ and $f(\alpha) = f(1)$ for every $\alpha > 0$. Therefore, by Lemma 8 we can conclude that $\mathbf{x}^{(0)}$ is a stationary point different from consensus.

For the other direction, let $\mathbf{x}$ be a stationary point of the dynamics on G different from consensus. Since $\rho^* > \frac{f(1)}{g(1)}$, Lemma 8 yields a partition of the vertices of the graph as requested, i.e., $i \in V_b$ iff $x_i = s_b$ where we rewrite w.l.o.g. the strategy set as $S = \{s_0, \ldots, s_{m-1}\}$. □

Basically, Theorem 9 proves that agents are content of having reached consensus within their own cluster/community, and they do not care for general consensus.

4.1 Understanding Well-Partitioned Graphs

It seems hard to give an explicit, topological characterization of graphs that are well-partitioned. The related literature [Gharan and Trevisan, 2014; Peng *et al.*, 2015; Kolev and Mehlhorn, 2016] on "clustering" of graphs[3], for example, focuses only on algorithmic characterizations (wherein, naturally, algorithms are intended to run in polynomial-time). We now follow a similar avenue, but only give a partial answer.

We next relate the problem of establishing whether a graph is well-partitioned to the existence of a *non-singleton min-cut*, i.e., a min-cut whose two sides are not singletons.

Theorem 10. *If there exists a non-singleton min-cut of a graph $G = (V, E)$, then G is well-partitioned.*

Proof. Let $(A, V \setminus A)$ be a non-singleton min-cut of G. We can set $m = 2$ and use $V_0 = A$ and $V_1 = V \setminus A$ as the partition needed by the definition of well-partitioned graphs.

Assume, by contradiction, that this is not the case. Then there exists $b \in \{0, 1\}$ and $v \in V_b$ such that $N_v(V_b) < N_v(V_{1-b})$. But then moving v from V_b to V_{1-b} would give rise to a new cut of smaller size; a contradiction with the fact that $(A, V \setminus A)$ is a min-cut. □

---

[3] In graph clustering, the nodes must have much more neighbors in their side of the partition; we only need at least as many neighbors.

It is not hard to see that there is a polynomial-time algorithm to establish whether a graph has a non-singleton min-cut. In fact, it is well known that we can use Karger's algorithm [Karger, 1993] to enumerate (with high probability) all of the (roughly $|V|^2/2$) min-cuts (of an undirected, unweighted graph) in polynomial-time; then we can simply check the size of both ends of the cut.

However, such a condition is not necessary for well-partitioned graphs already for $m = 2$, as we are going to discuss next. For an even number n, let G be the clique K_n with a perfect matching removed, i.e., $E = \{(i, j) : j \neq i\} \setminus \{(2i, 2i-1) : n/2 \geq i \geq 1\}$. Consider the partition in which even vertices are in V_0 and odd vertices in V_1. Each vertex has exactly half of its neighbors in V_b and exactly half in V_{1-b} for $b = 0, 1$. By Theorem 9, the dynamics diverges on G. However, G does not have a non-singleton min-cut, since all the min-cuts of G are of the kind $(\{v\}, V \setminus \{v\})$, $v \in V$.

Determining whether a graph is well-partitioned is, in fact, equivalent to the graph having a non-singleton cut which has local minimum weight (i.e., moving one vertex to the another set of the partition does not diminish the weight of the cut). To the best of our knowledge, the computational complexity of this problem is open.

5 Conclusions

In this work we initiated the study of opinion dynamics with a social pressure towards consensus. If for clique social networks, we have been able to give a complete picture of what happens both with fully rational players and bounded rational ones, much more is left to understand for different social network topologies: Is it possible to design a polynomial-time algorithm that on input a graph G is able to decide if the dynamics can diverge on G? Or, on the contrary, can we show that this problem is computationally hard? Can we bound the time that the dynamics takes in order to converge to consensus or to a generic stationary point? Does the convergence to consensus in non-clique graphs take a path similar to the one described by Claim 3, with the majority opinion becoming more and more supported as time goes on? Simulations may be an useful tool for answering this question.

In this work we focused on unweighted graphs. Naturally, it would be interesting to see to which extent our results hold on weighted networks. Note that, for example, it is not hard to see that for specific weight assignment of edges, the best response dynamics does not converge to consensus even if the underlying graph is a clique.

It would be interesting also to evaluate how bounded rationality influences the evolution of opinions in non-clique network topologies. Do noisy dynamics, such as logit dynamics, converge to the consensus whenever best response does?

References

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

[Auletta *et al.*, 2012] Vincenzo Auletta, Diodato Ferraioli, Francesco Pasquale, and Giuseppe Persiano. Metastability of logit dynamics for coordination games. In *SODA '12*, pages 1006–1024, 2012.

[Auletta *et al.*, 2013a] Vincenzo Auletta, Diodato Ferraioli, Francesco Pasquale, Paolo Penna, and Giuseppe Persiano. Logit dynamics with concurrent updates for local interaction games. In *ESA '13*, pages 73–84, 2013.

[Auletta *et al.*, 2013b] Vincenzo Auletta, Diodato Ferraioli, Francesco Pasquale, and Giuseppe Persiano. Mixing time and stationary expected social welfare of logit dynamics. *Theory of Computing Systems*, 53(1):3–40, 2013.

[Auletta *et al.*, 2015] Vincenzo Auletta, Ioannis Caragiannis, Diodato Ferraioli, Clemente Galdi, and Giuseppe Persiano. Minority becomes majority in social networks. In *WINE '15*, pages 74–88, 2015.

[Auletta *et al.*, 2016] Vincenzo Auletta, Ioannis Caragiannis, Diodato Ferraioli, Clemente Galdi, and Giuseppe Persiano. Generalized discrete preference games. In *IJCAI '16*, pages 53–59, 2016.

[Bhalgat *et al.*, 2010] Anandd Bhalgat, Tanmoy Chakraborty, and Sanjeev Khanna. Approximating pure Nash equilibrium in cut, party affiliation, and satisfiability games. In *EC '10*, pages 73–82, 2010.

[Bhawalkar *et al.*, 2013] Kshipra Bhawalkar, Sreenivas Gollapudi, and Kamesh Munagala. Coevolutionary opinion formation games. In *STOC '13*, pages 41–50, 2013.

[Bilò *et al.*, 2016] Vittorio Bilò, Angelo Fanelli, and Luca Moscardelli. Opinion formation games with dynamic social influences. In *WINE '16*, pages 444–458, 2016.

[Bindel *et al.*, 2011] David Bindel, Jon Kleinberg, and Sigal Oren. How bad is forming your own opinion? In *FOCS '11*, pages 55–66, 2011.

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

[Chierichetti *et al.*, 2013] Flavio Chierichetti, Jon Kleinberg, and Sigal Oren. On discrete preferences and coordination. In *EC '13*, pages 233–250, 2013.

[DeGroot, 1974] Morris H. DeGroot. Reaching a consensus. *J. American Statistical Association*, 69:118–121, 1974.

[Ferraioli *et al.*, 2016] Diodato Ferraioli, Paul W. Goldberg, and Carmine Ventre. Decentralized dynamics for finite opinion games. *Theoretical Computer Science*, 648:96 – 115, 2016.

[Friedkin and Johnsen, 1990] Noah E. Friedkin and Eugene C. Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[Gharan and Trevisan, 2014] Shayan Oveis Gharan and Luca Trevisan. Partitioning into expanders. In *SODA '14*, pages 1256–1266, 2014.

[Grandi *et al.*, 2015] Umberto Grandi, Emiliano Lorini, and Laurent Perrussel. Propositional opinion diffusion. In *AAMAS '15*, pages 989–997, 2015.

[Karger, 1993] David R. Karger. Global min-cuts in rnc, and other ramifications of a simple min-cut algorithm. In *SODA '93*, pages 21–30, 1993.

[Kolev and Mehlhorn, 2016] Pavel Kolev and Kurt Mehlhorn. A note on spectral clustering. In *ESA '16*, pages 57:1–57:14, 2016.

[McFadden, 1974] Daniel McFadden. *Conditional logit analysis of qualitative choice behavior*, pages 105–142. Number 2. 1974.

[Montanari and Saberi, 2009] Andrea Montanari and Amin Saberi. Convergence to equilibrium in local interaction games. In *FOCS '09*, pages 303–312, 2009.

[Mossel and Tamuz, 2014] Elchanan Mossel and Omer Tamuz. Opinion exchange dynamics. *arXiv preprint arXiv:1401.4770*, 2014.

[Peng *et al.*, 2015] Richard Peng, He Sun, and Luca Zanetti. Partitioning well-clustered graphs: Spectral clustering works! In *COLT '15*, pages 1423–1455, 2015.

[Peyton Young, 2006] Hobart Peyton Young. The diffusion of innovations in social networks. *The economy as an evolving complex system III: Current perspectives and future directions*, 267, 2006.

[Pryymak *et al.*, 2012] Oleksandr Pryymak, Alex Rogers, and Nicholas R. Jennings. Efficient opinion sharing in large decentralised teams. In *AAMAS '12*, pages 543–550, 2012.

[Rajtmajer *et al.*, 2016] Sarah Rajtmajer, Anna Squicciarini, Christopher Griffin, Sushama Karumanchi, and Alpana Tyagi. Constrained social energy minimization for multiparty sharing in online social networks. In *AAMAS '16*, pages 680–688, 2016.

[Schwind *et al.*, 2015] Nicolas Schwind, Katsumi Inoue, Gauvain Bourgne, Sebastien Konieczny, and Pierre Marquis. Belief revision games. In *AAAI'15*, pages 1590–1596, 2015.

[Tsang and Larson, 2014] Alan Tsang and Kate Larson. Opinion dynamics of skeptical agents. In *AAMAS '14*, pages 277–284, 2014.