Stochastic Opinion Dynamics under Social Pressure in Arbitrary Networks

Jennifer Tang, Aviv Adler, Amir Ajorlou, and Ali Jadbabaie

***Abstract*— Social pressure is a key factor affecting the evolution of opinions on networks in many types of settings, pushing people to conform to their neighbors' opinions. To study this, the *interacting Pólya urn* model was introduced by Jadbabaie et al. [1], in which each agent has two kinds of opinion: *inherent beliefs*, which are hidden from the other agents and fixed; and *declared opinions*, which are randomly sampled at each step from a distribution which depends on the agent's inherent belief and her neighbors' past declared opinions (the social pressure component), and which is then communicated to their neighbors. Each agent also has a *bias parameter* denoting her level of resistance to social pressure. At every step, each agent updates her declared opinion (simultaneously with all other agents) according to her neighbors' aggregate past declared opinions, her inherent belief, and her bias parameter. We study the asymptotic behavior of this opinion dynamics model and show that agents' declaration probabilities converge almost surely in the limit using Lyapunov theory and stochastic approximation techniques. We also derive a sufficient condition for the agents to approach consensus on their declared opinions.**

I. Introduction and Related Work

Opinion dynamics – the modeling and study of how people's opinions change in a social setting (particularly through communication on a network, whether online or offline) – is an extremely useful tool for analyzing various social and political phenomena such as consensus and social learning [2] as well as for designing strategies for political, marketing and information campaigns, such as the effort to curb vaccine hesitancy [3]. It is generally assumed in such models that the agents report their opinions truthfully. In reality, however, there are many occasions in which people make declarations contrary to their real views in order to conform socially [4], a fact confirmed both by common sense and by psychological studies [5]. This can make it difficult to determine the true beliefs governing observed interactions.

In this work, we study an *interacting Pólya urn model* for opinion dynamics, originating from [1], that captures a system of agents who might be untruthful due to their local social interactions. This model consists of n agents on a fixed network communicating on an issue with two basic sides, 0 and 1. Each agent has an *inherent belief* (true and unchanging), which is either 0 or 1, and an honesty parameter $\tilde{\gamma}$. Then the agents communicate their *declared opinions* to their neighbors at discrete time steps: at each step $t = 1, 2, \ldots$, all the agents simultaneously declare one of the two opinions (i.e. either '0' or '1'), which is then observed by their neighbors; the declarations of all the agents at any given step are made at random and independently of each other but with probabilities determined by their inherent belief, honesty parameter, and the ratio of the two declared opinions observed by the agent up to the current time. This can represent scenarios where agents (say, people using social media) alter their statements to better fit in with the opinions they have observed from others in the past; it may also represent scenarios where the agents update their opinions according to the declared opinions of others, but retain a bias towards their original position. The goal of this model is to shed light on how opinions might evolve in the presence of social pressure.

Opinion dynamics originally grew from a need to mathematically understand psychological experiments on the behavior of individuals in group settings [6], [5], [7]. Notable among these is the DeGroot model [8], where agents in a network average their neighbors' opinion in an iterative manner. With this procedure, the entire group asymptotically approaches a state where they all share a single opinion, a phenomenon known as *consensus*. While the DeGroot model is highly influential, it is clear that consensus is not always approached in reality. This problematic aspect of the DeGroot model and other similar models inspired follow-up work aiming to account for disagreement among agents [9], [10], [11].

However, opinions are often influenced not only by others' opinions but by personal inclinations or beliefs; for instance, each agent in the Friedkin-Johnsen model [9] updates her opinion at each step by averaging her neighbors' opinions (as in the DeGroot model) and then averaging the result with her initial opinion, which represents her innate beliefs. Other models adjust whether interactions with neighbors cause an agent to conform or be unique [12], [13]. Dandekar et al. [14] look at bias assimilation, in which agents weigh the average of their neighbors' opinions by an additional bias factor. Another relevant line of research is the competitive contagion and product adoption in the marketing literature [15], [16], [17], [18], where individuals' choices of products and services are influenced both by personal tastes and desires and by others' choices, a phenomenon commonly known as the network effect. Authors in [19] use a threshold diffusion model to numerically study cascades of self-reinforcing support for a highly unpopular norm on social networks.

Besides the model in [1], several other models also include the feature that agents do not always update their initial beliefs [20], [21], [22]. Ye et al. [22] study a model in

The work was supported by ARO MURI W911 NF-19-1-0217 and a Vannevar Bush Fellowship from the Office of the Under Secretary of Defense.

which each agent has both a private and expressed opinion, which evolve differently. Agents' private opinions evolve using the same update as in the Friedkin-Johnsen model whereas agents' public opinion are an average of their own private opinion and the public average opinion. Both [19] and [22] are very similar to [1], since agents are willing to express opinions they do not actually believe in. However, unlike [1], [22] assumes opinions are precisely expressed on a continuous interval, which is unrealistic for certain applications. On the other hand [19] works with binary opinions like [1], but with an additional reinforcement step which adds complexity. The model in [1] captures the core idea of [19] in a different way that is more tractable for analysis. Additionally, the honesty parameter from [1] is similar to the conformity parameter in [23], which measures how likely an agent is to conform to others. Other types of interacting Pólya urn models have also been used by [24], [25] to study contagion networks.

A. Contributions

While [1] originally proposed an interacting Pólya urn model for opinion dynamics, they studied it only in the special case of a (unweighted) complete graph as the network, with agents that all have the same honesty parameter $\tilde{\gamma}$. In this work we remove these constraints and study this process on arbitrary undirected graphs with agents whose honesty parameters may differ. Our contributions are:

1) We establish that the behavior of agents (i.e. their probabilities of declaring each opinion) almost surely converges to a steady-state asymptotically.
2) We determine a sufficient condition for *consensus*. Due to the stochastic nature of our model we define consensus as a property of the *declared* opinions: the network *approaches consensus* if all agents declare the same opinion (either all 0 or all 1) with probability approaching 1 as time goes to infinity. This corresponds to cases where social pressure forces increasing conformity over time, and makes estimating the agents' inherent beliefs from their behavior difficult, as shown in [1].

We discuss more details of these contributions and their comparison with [1] below.

a) Convergence of Agent Declaration Probabilities: We use Lyapunov theory and stochastic approximation to determine convergence for the opinion dynamics model. We show that on undirected networks, the probability that each agent i declares 1 at the next step converges asymptotically to some (possibly random) value p_i, and the values $p_1, p_2, \ldots, p_n$ represent an *equilibrium point* of the process.

b) Conditions for Consensus: An interesting result from [1] is that if the proportion of agents (connected in a complete graph) with inherent beliefs 0 or 1 passes a certain threshold, then asymptotically the system almost surely converges to a behavior where $p_i = 0$ for all agents or $p_i = 1$ for all agents. In this work, we find an analogous result for general networks, determining a condition under which all agents in the network almost surely converge to consensus (declaring the same opinion with probability 1). The condition is derived by incorporating the structure of the network, the inherent beliefs of the agents and their honesty parameters.

c) Analysis of Simplified Community Network: We apply our convergence and consensus results to study in depth a simplified community network. In this model, there are two communities, a and b, which are represented as two agents (or two vertices). To model that each community is more connected to itself than to the other community, vertices a and b have self-loops of greater weight than the edge connecting them. This network is designed to capture *homophily*, a property of real and online communities where people with similar traits, opinions or interests tend to form communities with relatively dense in-community connections [14]. We show that whether or not all agents in the network converge to declaring the same opinion (i.e. approach consensus) depends on whether the ratio of the proportion of in-community edges of each community is greater than the honesty parameter.

II. Model Description

Our model is a slight generalization of the model from [1], with the addition that each edge in the network has a (nonnegative) weight denoting how much the two agents' declared opinions influence each other. As mentioned, we also extend the model by permitting agents to have different honesty parameters.

A. Graph Notation

Let (undirected) graph $G = (V, E)$ be a network of n agents (corresponding to the vertices) labeled $i = 1, 2, \ldots, n$, so $V = [n]$. The graph G can have self-loops. For each edge $(i, j) \in E$, there is a weight $a_{i,j} \geq 0$, where by convention we let $a_{i,j} = 0$ if $(i, j) \notin E$. We denote the matrix of these weights as $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, i.e. the weighted adjacency matrix of G; since G is undirected, $\boldsymbol{A}$ is symmetric.

The vector of degrees of all agents is denoted as

$$\boldsymbol{d} \triangleq [\mathsf{deg}(1), \mathsf{deg}(2), \ldots, \mathsf{deg}(n)] \tag{1}$$

and its diagonalization is denoted $\boldsymbol{D} = \mathsf{diag}(\boldsymbol{d})$, i.e. the diagonal matrix of the degrees. Let the *normalized adjacency matrix* be $\boldsymbol{W} = \boldsymbol{D}^{-1}\boldsymbol{A}$. The matrix $\boldsymbol{W}$ can be interpreted as the transition matrix for a random walk on G, where the probability of choosing an edge at a given step is proportional to its weight. We assume that $\boldsymbol{W}$ is irreducible (G is connected) and not bipartite.

B. Inherent Beliefs and Declared Opinions

Each agent i has an *inherent belief* $\phi_i \in \{0, 1\}$, which does not change. At each time step $t \in \mathbb{Z}_+$, each agent i (simultaneously) announces a *declared opinion* $\psi_{i,t} \in \{0, 1\}$. At time t, we denote by $\mathcal{H}_t$ the *history* of the process, consisting of all $\psi_{i,\tau}$ for $\tau \leq t$. The declarations $\psi_{i,t}$ are based on the following probabilistic rule:

$$\psi_{i,t} \triangleq \begin{cases} 1 & \text{with probability } p_{i,t-1} & \text{if } \phi_i = 1 \\ 0 & \text{with probability } 1 - p_{i,t-1} & \text{if } \phi_i = 1 \\ 1 & \text{with probability } q_{i,t-1} & \text{if } \phi_i = 0 \\ 0 & \text{with probability } 1 - q_{i,t-1} & \text{if } \phi_i = 0 \end{cases} \tag{2}$$

where the parameters $p_{i,t-1}$ and $q_{i,t-1}$ depend on the history $\mathcal{H}_{t-1}$ via an interacting Pólya urn process in the following way. Each agent i has *honesty parameter* $\tilde{\gamma}_i \geq 1$ (we permit heterogeneous honesty parameters, while $\tilde{\gamma}_1 = \ldots = \tilde{\gamma}_n = \gamma$ in [1]).

Then for $t \in \mathbb{Z}_+$ let

$$M_i^0(t) = m_i^0 + \sum_{\tau=2}^{t} \sum_{j=1}^{n} a_{i,j} \mathbb{I}[\psi_{j,\tau} = 0] \tag{3}$$

$$M_i^1(t) = m_i^1 + \sum_{\tau=2}^{t} \sum_{j=1}^{n} a_{i,j} \mathbb{I}[\psi_{j,\tau} = 1] \tag{4}$$

where $m_i^0, m_i^1 > 0$ represent the initial settings of the model. (Initial settings are used in place of declared opinions at time 1. Some requirements for the initial settings are given shortly.) The quantity $M_i^0(t)$ represents the (weighted) number of times agent i observed a neighbor declare opinion 0 up to step t (plus initial settings), and $M_i^1(t)$ represents the analogous total of observed 1's. If each $a_{i,j} \in \{0, 1\}$, then $M_i^0(t)$ and $M_i^1(t)$ represent counts of agent's neighbors' declarations (plus initial settings). The ratio of $M_i^1(t)$ to $M_i^0(t)$ can be viewed as the social pressure on agent i to declare opinion 1. Then for $t \geq 1$:

$$p_{i,t} = \frac{\tilde{\gamma}_i M_i^1(t)}{\tilde{\gamma}_i M_i^1(t) + M_i^0(t)} \tag{5}$$

$$q_{i,t} = \frac{M_i^1(t)}{M_i^1(t) + \tilde{\gamma}_i M_i^0(t)} \,. \tag{6}$$

C. Declaration Proportions

Let $M_i(t) \triangleq m_i^0 + m_i^1 + (t-1)\,\mathsf{deg}(i) = M_i^0(t) + M_i^1(t)$ and

$$\mu_i^0(t) \triangleq M_i^0(t)/M_i(t) \tag{7}$$

$$\mu_i^1(t) \triangleq M_i^1(t)/M_i(t) \,. \tag{8}$$

The parameter $\mu_i^1(t)$ is essentially the sufficient statistic that summarizes the proportion of declared opinions in the neighborhood of given agent i up to time t. Since $\mu_i^0(t) = 1 - \mu_i^1(t)$, we simplify the notation to $\mu_i(t) \triangleq \mu_i^1(t)$.

We also define a sufficient statistic that summarizes agent i's declarations. Let $b_i^0, b_i^1 > 0$ (the initialization) be such that $b_i^0 + b_i^1 = 1$ for each i and

$$m_i^0 = \sum_{j=1}^{n} a_{i,j} b_j^0 \quad \text{and} \quad m_i^1 = \sum_{j=1}^{n} a_{i,j} b_j^1 \,. \tag{9}$$

For $t \in \mathbb{Z}_+$, let

$$B_i^0(t) = b_i^0 + \sum_{\tau=2}^{t} (1 - \psi_{i,\tau}), \;\; B_i^1(t) = b_i^1 + \sum_{\tau=2}^{t} \psi_{i,\tau} \tag{10}$$

$$\beta_i^0(t) = \frac{b_i^0}{t} + \frac{1}{t} \sum_{\tau=2}^{t} (1 - \psi_{i,\tau}), \;\; \beta_i^1(t) = \frac{b_i^1}{t} + \frac{1}{t} \sum_{\tau=2}^{t} \psi_{i,\tau} \,. \tag{11}$$

These are counts and proportions of declarations of each opinion (or "time-averaged declarations") for each agent (plus initial conditions). We similarly use $\beta_i(t) \triangleq \beta_i^1(t)$. It then follows that

$$\mu_i(t) = \frac{1}{\mathsf{deg}(i)} \sum_{j=1}^{n} a_{i,j} \beta_j(t) \,. \tag{12}$$

Finally, we define the vectors of observed declared opinions and given declared opinions for each agent at time t as

$$\boldsymbol{\mu}(t) \triangleq [\mu_1(t), \ldots \mu_n(t)]^\top \tag{13}$$

$$\boldsymbol{\beta}(t) \triangleq [\beta_1(t), \ldots \beta_n(t)]^\top \,. \tag{14}$$

D. Bias Parameters

One simplification to the notation from [1] is to combine the inherent belief ϕ_i and honesty parameter $\tilde{\gamma}_i$ into a single parameter we call the *bias parameter* $\gamma_i > 0$:

$$\gamma_i = \begin{cases} \tilde{\gamma}_i & \text{if } \phi_i = 1 \\ 1/\tilde{\gamma}_i & \text{if } \phi_i = 0 \end{cases} \tag{15}$$

and $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_n]$ is the set of bias parameters.

Define the function (note that μ, γ are scalars)

$$f(\mu, \gamma) \triangleq \frac{\gamma \mu}{1 + (\gamma - 1)\mu} = \frac{1}{1 + \frac{1}{\gamma}\left(\frac{1}{\mu} - 1\right)} \tag{16}$$

which then satisfies

$$f(\mu_i(t), \gamma_i) = \begin{cases} p_{i,t} & \text{if } \phi_i = 1 \\ q_{i,t} & \text{if } \phi_i = 0 \end{cases} \tag{17}$$

so (2) can be rewritten as

$$\psi_{i,t+1} \triangleq \begin{cases} 1 & \text{with probability } f(\mu_i(t), \gamma_i) \\ 0 & \text{with probability } 1 - f(\mu_i(t), \gamma_i) \end{cases} . \tag{18}$$

Note that the bias parameter γ_i is always defined as agent i's bias towards opinion 1. However, the model is symmetric in the following way: a γ bias towards 1 is equivalent to a $1/\gamma$ bias towards 0, which is captured by the equation

$$f(\mu_i^1(t), \gamma) = 1 - f(\mu_i^0(t), 1/\gamma) \,. \tag{19}$$

Define the diagonal matrix with $\boldsymbol{\gamma}$ along the diagonal as

$$\boldsymbol{\Gamma} = \mathsf{diag}(\boldsymbol{\gamma}) \,. \tag{20}$$

We assume for this work that $\boldsymbol{\Gamma} \neq \boldsymbol{I}$. This parallels the assumption $\tilde{\gamma}_i > 1$ used in [1].

E. Stochastic and Deterministic Expected Dynamics

Using (10), the recursive equations that govern the count of declared opinions by agent i are:

$$B_i^0(t+1) = B_i^0(t) + (1 - \psi_{i,t+1}) \tag{21}$$
$$B_i^1(t+1) = B_i^1(t) + \psi_{i,t+1}. \tag{22}$$

To work with $\beta_i(t)$ instead of $B_i^1(t+1)$, we rewrite (21) as

$$\beta_i(t+1) = \frac{t}{t+1}\beta_i(t) + \frac{1}{t+1}\psi_{i,t+1}. \tag{23}$$

Conditioned on the history $\mathcal{H}_t$ (which contains all information declared up to and including time t), the expected value of $\beta_i(t+1)$ is

$$\mathbb{E}[\beta_i(t+1)|\mathcal{H}_t] = \frac{t}{t+1}\beta_i(t) + \frac{1}{t+1}f(\mu_i(t), \gamma_i). \tag{24}$$

We then put the dynamics in (24) together for all the agents in the network, to get

$$\mathbb{E}[\boldsymbol{\beta}(t+1)|\mathcal{H}_t] = \frac{t}{t+1}\boldsymbol{\beta}(t) + \frac{1}{t+1}\begin{bmatrix} f(\mu_1(t), \gamma_1) \\ f(\mu_2(t), \gamma_2) \\ \vdots \\ f(\mu_n(t), \gamma_n) \end{bmatrix} \tag{25}$$

$$= \frac{t}{t+1}\boldsymbol{\beta}(t) + \frac{1}{t+1}f(\boldsymbol{\mu}(t), \boldsymbol{\gamma}), \tag{26}$$

or alternatively

$$\mathbb{E}[\boldsymbol{\beta}(t+1) - \boldsymbol{\beta}(t)|\mathcal{H}_t] = \frac{1}{t+1}(f(\boldsymbol{\mu}(t), \boldsymbol{\gamma}) - \boldsymbol{\beta}(t)). \tag{27}$$

Definition 1. *The* deterministic expected dynamics *are*

$$\boldsymbol{\beta}(t+1) - \boldsymbol{\beta}(t) = \frac{1}{t+1}(F(\boldsymbol{\beta}(t), \boldsymbol{\gamma}) - \boldsymbol{\beta}(t)) \tag{28}$$

where $F(\boldsymbol{\beta}(t), \boldsymbol{\gamma}) = [F_1(\boldsymbol{\beta}(t), \boldsymbol{\gamma}), \ldots, F_n(\boldsymbol{\beta}(t), \boldsymbol{\gamma})]$ *and*

$$F_i(\boldsymbol{\beta}(t), \boldsymbol{\gamma}) = f(\mu_i(t), \gamma_i). \tag{29}$$

We refer to original dynamics governed by (18) and (23) as the *full stochastic dynamics*.

F. Intuition for Interacting Pólya Urn Model

In this section, we consider how the interacting Pólya urn model is meaningful for opinion dynamics with social pressure. (For this, we use the case when $a_{i,j}$ is either 0 or 1.) Typically, urn models start with some composition of balls of different colors in an urn. At each step, a ball is drawn (independent of previous draws given the urn composition) from the urn and additional balls are added based on the drawn ball according to some urn functions. In the interacting Pólya urn model, when a neighbor of agent i declares an opinion, this is modeled as agent i putting a corresponding ball (labeled 0 or 1) into her own urn.

Then, when agent i declares an opinion, it is modeled by the following: she draws a ball from her urn and declares the corresponding opinion; each ball corresponding with opinion 1 is γ_i times as likely to be drawn as one with opinion 0. Note that if $\gamma_i = 1$ then agent i is (stochastically) mimicking the opinions her neighbors have declared in the past (plus her initial state, which becomes asymptotically negligible). We remark that the bias parameter is similar to the initial opinions in the Friedkin-Johnsen model [9] since they both are fixed parameters that influence all steps; however, note that there is a significant difference as the bias parameter can be overwhelmed over time by social pressure, thus leading to consensus.

III. Convergence Analysis

A. Equilibria of the Expected Dynamics

Definition 2. *A vector* $\boldsymbol{\beta}$ *is an* equilibrium point *of the expected dynamics if* $F(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \boldsymbol{\beta}$.

Note that vector **1** and vector **0** are always equilibrium points. We call these *boundary equilibrium points*, while other equilibrium points are *interior equilibrium points*. Equivalently, an interior equilibrium point is an equilibrium point $\boldsymbol{\beta}$ where $0 < \beta_i < 1$ for all i (see Lemma 1).

Lemma 1. *Suppose that a finite network of agents is connected. Then, if* $\boldsymbol{\beta}^*$ *is an equilibrium point such that for some* i, $\beta_i^* = 0$ *(or* $\beta_i^* = 1$*), then it must be that for all* i, $\beta_i^* = 0$ *(or respectively for all* i, $\beta_i^* = 1$*).*

Proof. Suppose that $\beta_i^* = 0$. The only way for this to occur at an equilibrium point is for each of agent i's neighbors j to also have $\beta_j^* = 0$. If any $\beta_j^* > 0$, then $\beta_i^* > 0$ since β_i gets a positive contribution from β_j in its sum. We continue by inducting on the neighbors of neighbors, and it gives that all agents j in the connected network must have $\beta_j^* = 0$. □

Finding an exact analytic expression for the equilibrium points is unfortunately difficult in general. In Section V, we show how to find equilibrium points for the simplified community network, which is possible because it is a small example. However, the following results can be used to solve for them numerically.

Proposition 1. *The equilibrium points of the expected dynamics are given by* $\boldsymbol{\beta}$ *such that for all* i,

$$0 = (\gamma_i - 1)\beta_i\mu_i + \beta_i - \gamma_i\mu_i \tag{30}$$

Proof. This follows from the fact that at any equilibrium,

$$\beta_i = f(\mu_i, \gamma_i) = \frac{\gamma_i\mu_i}{1 + (\gamma_i - 1)\mu_i}. \tag{31}$$

□

B. Tools from Stochastic Approximation

In order to prove the convergence of the full stochastic dynamics to the equilibrium points of the expected dynamics, we use results on the long-term behavior of path-dependent stochastic processes. In summary, [26, Theorem 3.1] uses stochastic approximation to show that dynamics using generalized urn functions converge an equilibrium point if a Lyapunov function V can be found that satisfies a certain set of conditions. One important condition is that $V > 0$ needs to satisfy $\langle F(\boldsymbol{\beta}, \boldsymbol{\gamma}) - \boldsymbol{\beta}, \nabla V(\boldsymbol{\beta})\rangle < 0$ except in a small neighborhood of points around the equilibria.

C. Convergence for General Networks

One of the primary contributions of the present work is to show the convergence of the time-averaged declared opinions $\boldsymbol{\beta}(t)$ to an equilibrium point, under the stochastic dynamics of (18) and (23) in any network. To carry out this result, we take advantage of two key properties of $f(\mu,\gamma)$:

- $f:[0,1]\to[0,1]$ is bijective in μ (this can be shown by the fact that $f(f(\mu,1/\gamma),\gamma)=\mu$)
- f is monotonic in μ.

Theorem 1. *Let* $\mu_i=\frac{1}{\mathsf{deg}(i)}\sum_{j=1}^{n}a_{i,j}\beta_j$ *and*

$$F(\boldsymbol{\beta},\boldsymbol{\gamma})=\begin{bmatrix}h(\mu_1,\gamma_1)\\ \vdots\\ h(\mu_n,\gamma_n)\end{bmatrix} \tag{32}$$

Suppose that the network associated with the adjacency matrix $\boldsymbol{A}$ *is undirected. Then, there exists a Lyapunov function* V*, where* $V\geq 0$ *such that*

$$\langle F(\boldsymbol{\beta},\boldsymbol{\gamma})-\boldsymbol{\beta},\nabla V(\boldsymbol{\beta})\rangle\leq 0 \tag{33}$$

so long as

- $h(\cdot,\gamma)$ *is bijective from* $[0,1]$ *to* $[0,1]$
- $h(\cdot,\gamma)$ *is monotonic.*

Equality in (33) *holds iff* $F(\boldsymbol{\beta},\boldsymbol{\gamma})=\boldsymbol{\beta}$.

Proof. Because $h(\cdot,\gamma)$ is bijective from $[0,1]$ to $[0,1]$, there exists an inverse $h^{-1}(\cdot,\gamma)$. The function $h^{-1}(\cdot,\gamma)$ is also (strictly) monotonically increasing. We use the notation

$$H(\mu,\gamma)=\int_0^{\mu}h^{-1}(\nu,\gamma)d\nu\,. \tag{34}$$

Let

$$V(\boldsymbol{\beta})=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{i,j}\left(H(\beta_i,\gamma_i)-\frac{1}{2}\beta_i\beta_j\right)+C \tag{35}$$

(where C is a constant to make V positive). Taking the partial derivatives gives that

$$\frac{\partial V}{\partial\beta_i}=\mathsf{deg}(i)h^{-1}(\beta_i,\gamma_i)-\sum_{j=1}^{n}a_{i,j}\beta_j \tag{36}$$

$$=\mathsf{deg}(i)\left(h^{-1}(\beta_i,\gamma_i)-\frac{1}{\mathsf{deg}(i)}\sum_{j=1}^{n}a_{i,j}\beta_j\right) \tag{37}$$

$$=\mathsf{deg}(i)\left(h^{-1}(\beta_i,\gamma_i)-\mu_i\right)\,. \tag{38}$$

(The property that $\boldsymbol{A}$ is symmetric is necessary for (36).) We can write the ith entry in vector $F(\boldsymbol{\beta},\boldsymbol{\gamma})-\boldsymbol{\beta}$ as

$$(F(\boldsymbol{\beta},\boldsymbol{\gamma})-\boldsymbol{\beta})_i=h(\mu_i,\gamma_i)-\beta_i\,. \tag{39}$$

Then

$$\langle F(\boldsymbol{\beta},\boldsymbol{\gamma})-\boldsymbol{\beta},\nabla V(\boldsymbol{\beta})\rangle$$
$$=\sum_{i=1}^{n}\mathsf{deg}(i)\left(h^{-1}(\beta_i,\gamma_i)-\mu_i\right)\left(h(\mu_i,\gamma_i)-\beta_i\right)\,. \tag{40}$$

Suppose $h(\mu_i,\gamma_i)>\beta_i$. Then since h is (strictly) monotone,

$$h(\mu_i,\gamma_i)>\beta_i \tag{41}$$

$$\iff h^{-1}(h(\mu_i,\gamma_i),\gamma_i)>h^{-1}(\beta_i,\gamma_i) \tag{42}$$

$$\iff \mu_i>h^{-1}(\beta_i,\gamma_i)\,. \tag{43}$$

We can conclude that in the case of $h(\mu_i,\gamma_i)\neq\beta_i$ the sign of the terms $\left(h^{-1}(\beta_i,\gamma_i)-\mu_i\right)$ and $\left(h(\mu_i,\gamma_i)-\beta_i\right)$ are necessarily different. Hence, their product must be negative. When $h(\mu_i,\gamma_i)=\beta_i$, then the values of both $h(\mu_i,\gamma_i)-\beta_i$ and $h(\mu_i,\gamma_i)-\beta_i$ are zero.

Each term in the sum of (40) must be nonpositive and thus

$$\langle F(\boldsymbol{\beta},\boldsymbol{\gamma})-\boldsymbol{\beta},\nabla V(\boldsymbol{\beta})\rangle\leq 0\,. \tag{44}$$

Equality holds when all terms in the sum of (40) are zero, which only occurs when $h(\mu_i,\gamma_i)=\beta_i$ for all i. This means that (40) is zero if and only if $\boldsymbol{\beta}$ is an equilibrium. □

Note that Theorem 1 holds for all h satisfying the given conditions (not just the specific f defined in (16)), and hence is a general result showing the existence of Lyapunov functions for any dynamics satisfying the given conditions on undirected graphs (which do not need to be connected).

Theorem 2. *The time-averaged declared opinions* $\boldsymbol{\beta}(t)$ *under the stochastic opinion dynamics governed by* (18) *and* (23) *almost surely converges to an equilibrium point of the expected dynamics, that is a fixed point of* $F(\cdot,\boldsymbol{\gamma})$.

Proof. This follows from Theorem 1 and [26, Thm 3.1]. □

IV. Convergence to Consensus

The previous section showed that the opinion dynamics under social pressure almost surely converges to an equilibrium point, but does not specify which equilibrium point the system converges to. Since there are multiple equilibrium points (not all necessarily stable) in any opinion dynamics system, in this section we explore conditions under which the system asymptotically converges to a boundary equilibrium point or an interior equilibrium point. When the system converges to a boundary equilibrium point (both $\mathbf{0}$ and $\mathbf{1}$ are boundary equilibrium points of $F(\boldsymbol{\beta},\boldsymbol{\gamma})$), we say that the agents approach consensus. Consensus occurs when all agents (asymptotically) converge to declaring the same opinion with probability 1.

Definition 3. Consensus *is approached if*

$$\boldsymbol{\beta}(t)\to\mathbf{1}\ \textit{or}\ \boldsymbol{\beta}(t)\to\mathbf{0}\quad\textit{as}\quad t\to\infty\,. \tag{45}$$

In this section, we establish a sufficient condition for convergence to consensus. Recall that $\boldsymbol{\beta}$ is the vector (over the agents) of the fraction of declared opinion 1 over time (plus initial conditions). Definition 3 does not imply that any agent will always declare the same opinion, only that her ratio of declared opinions tends to 0 or 1.

Which equilibrium point $\boldsymbol{\beta}(t)$ converges to (either boundary and interior) is closely related to the Jacobian matrix of $F(\cdot,\boldsymbol{\gamma})$. To calculate the Jacobian $\frac{\partial}{\partial\boldsymbol{\beta}}F(\boldsymbol{\beta},\boldsymbol{\gamma})$, recall

$$F_i(\boldsymbol{\beta},\boldsymbol{\gamma})=f(\mu_i,\gamma_i)=\frac{\gamma_i\mu_i}{1+(\gamma_i-1)\mu_i} \tag{46}$$

where we denote $\mu_i = \frac{1}{\mathsf{deg}(i)} \sum_{j=1}^n a_{i,j}\beta_j$. As a result,

$$\frac{\partial}{\partial \mu_i} F_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{\gamma_i}{(1+(\gamma_i - 1)\mu_i)^2}. \tag{47}$$

Finally,

$$\frac{\partial}{\partial \boldsymbol{\beta}} F(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{\partial}{\partial \boldsymbol{\mu}} F(\boldsymbol{\beta}, \boldsymbol{\gamma}) \frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}} \tag{48}$$

$$= \mathsf{diag}\left(\begin{bmatrix} \frac{\gamma_1}{(1+(\gamma_1-1)\mu_1)^2} \\ \vdots \\ \frac{\gamma_n}{(1+(\gamma_n-1)\mu_n)^2} \end{bmatrix} \right) \boldsymbol{W}. \tag{49}$$

We define for each vector $\boldsymbol{x}$ where $x_i \in [0,1]$,

$$\boldsymbol{J_x} \triangleq \mathsf{diag}\left(\begin{bmatrix} \frac{\gamma_1}{(1+(\gamma_1-1)\frac{1}{\mathsf{deg}(1)}\sum_j a_{1,j}x_1)^2} \\ \vdots \\ \frac{\gamma_n}{(1+(\gamma_n-1)\frac{1}{\mathsf{deg}(n)}\sum_j a_{n,j}x_n)^2} \end{bmatrix} \right) \boldsymbol{W}. \tag{50}$$

Importantly, $\boldsymbol{J_x}$ has all real eigenvalues.

The Jacobian at boundary equilibrium points $\mathbf{0}$ and $\mathbf{1}$ are

$$\boldsymbol{J_1} = \frac{\partial}{\partial \boldsymbol{\beta}} F(\boldsymbol{\beta}, \boldsymbol{\gamma})|_{\boldsymbol{\beta}=\mathbf{1}} = \boldsymbol{\Gamma}^{-1}\boldsymbol{W} \tag{51}$$

$$\boldsymbol{J_0} = \frac{\partial}{\partial \boldsymbol{\beta}} F(\boldsymbol{\beta}, \boldsymbol{\gamma})|_{\boldsymbol{\beta}=\mathbf{0}} = \boldsymbol{\Gamma}\boldsymbol{W}, \tag{52}$$

where $\boldsymbol{\Gamma}$ is defined in (20).

We will prove that properties of $\boldsymbol{J_0}$ and $\boldsymbol{J_1}$ suffice to determine whether $\boldsymbol{\beta}(t)$ approaches a boundary equilibrium point or not.

Theorem 3. *Let $\boldsymbol{x}$ be a boundary equilibrium point (either $\mathbf{0}$ or $\mathbf{1}$). If $\lambda_{\max}(\boldsymbol{J_x}) \leq 1$, then $\boldsymbol{\beta}(t)$ converges to a boundary equilibrium point almost surely.*

Proof. We assume WLOG that $\boldsymbol{J_1}$ has all eigenvalues less than or equal to 1. Let $\lambda = \lambda_{\max}(\boldsymbol{J_1})$ and let $\boldsymbol{v}$ be the corresponding eigenvector. Let v_i be the ith element of $\boldsymbol{v}$. Scale $\boldsymbol{v}$ so that $\boldsymbol{v}^T\mathbf{1} = 1$. We will use the fact $v_i \geq 0$ (shown by Perron-Frobenius) and $\boldsymbol{v}^T\boldsymbol{J_1} = \lambda \boldsymbol{v}^T$.

Observe that (as $\frac{1}{\mathbb{E}[X]} \leq \mathbb{E}[1/X]$ by Jensen's Inequality)

$$\sum_{i=1}^n v_i \left(\frac{1}{f(\mu_i, \gamma_i)} - 1 \right) = \sum_{i=1}^n \frac{v_i}{\gamma_i} \left(\frac{1}{\mu_i} - 1 \right) \tag{53}$$

$$\leq \sum_{i=1}^n \frac{v_i}{\gamma_i \mathsf{deg}(i)} \sum_{j=1}^n a_{i,j} \left(\frac{1}{\beta_j} - 1 \right). \tag{54}$$

Then

$$\sum_{i=1}^n v_i \left(\frac{1}{f(\mu_i, \gamma_i)} - 1 \right) = \sum_{j=1}^n \left(\frac{1}{\beta_j} - 1 \right) \sum_{i=1}^n \frac{v_i}{\gamma_i \mathsf{deg}(i)} a_{i,j} \tag{55}$$

$$= \lambda \sum_{j=1}^n v_j \left(\frac{1}{\beta_j} - 1 \right) \leq \sum_{j=1}^n v_j \left(\frac{1}{\beta_j} - 1 \right) \tag{56}$$

$$\implies \sum_{i=1}^n \frac{v_i}{f(\mu_i, \gamma_i)} \leq \sum_{i=1}^n \frac{v_i}{\beta_i}. \tag{57}$$

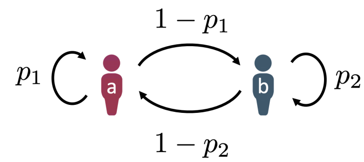

Fig. 1. The simplified community network used to study community structure. Agent a has bias parameter γ and agent b has bias parameter $1/\gamma$.

Interior equilibrium points must have that $\beta_i = f(\mu_i, \gamma_i)$ for all i. Inequality (57) is a strict inequality when $\lambda < 1$, in which case there must not exist any interior equilibrium points $\boldsymbol{\beta}$. When $\lambda = 1$, (57) can only be an equality if (54) is an equality. Since $1/x$ is a strictly convex function, equality in (54) only holds if all β_j's are equal for all j which is a neighbor of i. Since the graph is not bipartite and connected, this implies that all β_j are the same for each j. (We can see this since the non-bipartite property implies that there is a path with an even number of nodes connecting any node i to node j. The nodes at odd positions in the path will force the pair of two adjacent even position nodes to be the same.) However, the only way $\boldsymbol{\beta}$ can be an equilibrium point with this condition that β_j's are all equal is if $\boldsymbol{\beta} = \mathbf{1}$ or $\boldsymbol{\beta} = \mathbf{0}$, or if $\gamma_i = 1$ for all i. Thus, if $\lambda < 1$ or $\lambda = 1$, the only equilibrium points are $\mathbf{1}$ and $\mathbf{0}$.

Then by Theorem 2, $\boldsymbol{\beta}(t)$ must converge to one of the two boundary equilibrium points almost surely. □

Next we examine when consensus fails to occur; this is related to a similar condition on the eigenvalues of the Jacobian matrix for interior equilibria. (Proof omitted.)

Proposition 2. *For an interior equilibrium point $\boldsymbol{x}$, if all the eigenvalues of the Jacobian matrix $\frac{\partial}{\partial \boldsymbol{\beta}} F(\boldsymbol{\beta}, \boldsymbol{\gamma})|_{\boldsymbol{\beta}=\boldsymbol{x}}$ are less than 1, then $\mathbb{P}[\boldsymbol{\beta}(t) \to \boldsymbol{x}] > 0$.*

V. Community Network Example

In this section, we apply our results to get explicit results for the *simplified community network*, which is a two-agent network simulating the interaction of two communities.

The simplified community network has two vertices, agent a and agent b. Agent a has bias parameter γ where $\gamma > 1$ and agent b has bias parameter $1/\gamma$. The transition matrix for the edge weights between the two agents is given by

$$\boldsymbol{W} = \begin{bmatrix} p_1 & 1-p_1 \\ 1-p_2 & p_2 \end{bmatrix} \tag{58}$$

where $p_1, p_2 \in [0,1]$ and p_1 represents the proportion of in-community edges for agent a and p_2 represents the proportion of in-community edges for agent b. (See Figure 1 for a diagram.) The property that the agents have more in-community edges occurs when $p_1 > 1/2$ and $p_2 > 1/2$.

To analyze this network, we first find the equilibrium points.

Proposition 3. *The equilibrium points of the simplified community network are: $\mathbf{0} = (0,0)$; $\mathbf{1} = (1,1)$; and, when $\max\{\frac{p_1}{p_2}, \frac{p_2}{p_1}\} < \gamma$, the interior equilibrium point $\boldsymbol{\beta}^* =$*

(β_a^*, β_b^*) *where*

$$\beta_a^* = \frac{\gamma\left((\gamma+1)p_1p_2 - 2p_2 + \sqrt{p_1p_2}\Delta\right)}{(\gamma-1)((\gamma+1)p_1p_2 + \sqrt{p_1p_2}\Delta)} \tag{59}$$

$$\beta_b^* = \frac{2\gamma p_1 - (\gamma+1)p_1p_2 - \sqrt{p_1p_2}\Delta}{(\gamma-1)((\gamma+1)p_1p_2 + \sqrt{p_1p_2}\Delta)} \tag{60}$$

where $\Delta = \sqrt{4\gamma(1-p_1-p_2) + (\gamma+1)^2 p_1 p_2}$.

The calculations are omitted. To apply Theorem 1, we need the underlying adjacency matrix $\boldsymbol{A} = \boldsymbol{D}\boldsymbol{W}$ to be symmetric. We choose

$$\boldsymbol{D} = \begin{bmatrix} (1-p_2) & 0 \\ 0 & (1-p_1) \end{bmatrix} \tag{61}$$

which results in a symmetric $\boldsymbol{A}$.

Then we apply Theorem 1 as desired to show that asymptotically the dynamics on the simplified community network almost surely converges to one of the equilibrium points. Next we determine under what conditions the dynamics asymptotically approaches consensus.

Proposition 4. *For the simplified community network,*

$$\lim_{t\to\infty} \boldsymbol{\beta}(t) = \begin{cases} \boldsymbol{\beta}^* & \text{if } \max\{\frac{p_1}{p_2}, \frac{p_2}{p_1}\} < \gamma \\ \mathbf{1} & \text{if } \gamma \leq \frac{p_1}{p_2} \\ \mathbf{0} & \text{if } \gamma \leq \frac{p_2}{p_1} \end{cases} \tag{62}$$

almost surely where $\boldsymbol{\beta}^*$ *is given by Proposition 3.*

(The proof is omitted.) By Theorem 3, when the conditions $\gamma \leq \frac{p_1}{p_2}$ or $\gamma \leq \frac{p_2}{p_1}$ do not hold, there are in fact no interior equilibrium points. This matches the conclusion of Proposition 3.

VI. Conclusion

In this work, we studied the interacting Pólya urn model of opinion dynamics under social pressure. We expanded upon [1] by showing results for arbitrary networks and general bias parameters. To show that the probability of declared opinions converges asymptotically, we used an appropriate Lyapunov function and applied stochastic approximation, thus guaranteeing that in arbitrary networks, the behavior of agents almost surely converges. We also gave an easily-computable sufficient condition, for when the dynamics approach consensus. Our results provide insight as to how and when social pressure can force conformity of (expressed) opinions even against the true beliefs of some individuals.

A possible direction for further work is to find what consequences our techniques have for other models. Other directions include finding the interior equilibrium points for arbitrary networks and determining whether it is possible to infer inherent opinions in arbitrary networks.

References

[1] Ali Jadbabaie, Anuran Makur, Elchanan Mossel, and Rabih Salhab, "Inference in opinion dynamics under social pressure," *IEEE Transactions on Automatic Control*, vol. 68, no. 6, pp. 3377–3392, 2023.

[2] Hossein Noorazar, "Recent advances in opinion propagation dynamics: a 2020 survey," *The European Physical Journal Plus*, vol. 135, no. 6, pp. 521, 2020.

[3] Camilla Ancona, Francesco Lo Iudice, Franco Garofalo, and Pietro De Lellis, "A model-based opinion dynamics approach to tackle vaccine hesitancy," *Scientific Reports*, vol. 12, no. 1, pp. 11835, 2022.

[4] C. Miłosz, *The Captive Mind*, Knopf, 1953.

[5] Solomon E. Asch, "Studies of independence and conformity: I. a minority of one against a unanimous majority," *Psychological monographs*, vol. 70, no. 9, pp. 1–70, 1956.

[6] Muzafer Sherif, "A study of some social factors in perception.," *Archives of Psychology (Columbia University)*, 1935.

[7] Jr. French, John R. P., "A formal theory of social power," *Psychological review*, vol. 63, no. 3, pp. 181–194, 05 1956.

[8] Morris H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[9] Noah E Friedkin and Eugene C Johnsen, "Social influence and opinions," *Journal of Mathematical Sociology*, vol. 15, no. 3-4, pp. 193–206, 1990.

[10] Rainer Hegselmann and Ulrich Krause, "Opinion dynamics and bounded confidence: models, analysis and simulation," *J. Artif. Soc. Soc. Simul.*, vol. 5, no. 3, 2002.

[11] Guillaume Deffuant, Frédéric Amblard, Gérard Weisbuch, and Thierry Faure, "How can extremism prevail? a study based on the relative agreement interaction model," *Journal of artificial societies and social simulation*, vol. 5, no. 4, 2002.

[12] Claudio Altafini, "Consensus problems on networks with antagonistic interactions," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 935–946, 2013.

[13] Peter Duggins, "A psychologically-motivated model of opinion change with applications to american politics," *Journal of Artificial Societies and Social Simulation*, vol. 20, no. 1, pp. 13, 2017.

[14] Pranav Dandekar, Ashish Goel, and David T. Lee, "Biased assimilation, homophily, and the dynamics of polarization," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5791–5796, 2013.

[15] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137–146.

[16] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," *Internet and Network Economics*, vol. 4858, pp. 306–311, 2007.

[17] Arastoo Fazeli, Amir Ajorlou, and Ali Jadbabaie, "Competitive diffusion in social networks: Quality or seeding?," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 3, pp. 665–675, 2017.

[18] H. Amini, M. Draief, and M. Lelarge, "Marketing in a random network," *Network Control and Optimization*, vol. 5425, pp. 17–25, 2009.

[19] Damon Centola, Robb Willer, and Michael Macy, "The emperor's dilemma: A computational model of self-enforcing norms," *American Journal of Sociology*, vol. 110, no. 4, pp. 1009–1040, 2005.

[20] Daron Acemoğlu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Mathematics of Operations Research*, vol. 38, no. 1, pp. 1–27, 2013.

[21] Jason Gaitonde, Jon Kleinberg, and Eva Tardos, "Adversarial perturbations of opinion dynamics in networks," in *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020, pp. 471–472.

[22] Mengbin Ye, Yuzhen Qin, Alain Govaert, Brian D.O. Anderson, and Ming Cao, "An influence network model to study discrepancies in expressed and private opinions," *Automatica*, vol. 107, pp. 371–381, 2019.

[23] Abhimanyu Das, Sreenivas Gollapudi, Arindham Khan, and Renato Paes Leme, "Role of conformity in opinion dynamics in social networks," in *Proceedings of the second ACM conference on Online social networks*, 2014, pp. 25–36.

[24] Mikhail Hayhoe, Fady Alajaji, and Bahman Gharesifard, "Curing epidemics on networks using a polya contagion model," *IEEE/ACM Transactions on Networking*, vol. 27, no. 5, pp. 2085–2097, 2019.

[25] Somya Singh, Fady Alajaji, and Bahman Gharesifard, "A finite memory interacting polya contagion network and its approximating dynamical systems," *SIAM Journal on Control and Optimization*, vol. 60, no. 2, pp. S347–S369, 2022.

[26] W. Brian Arthur, Yu. M. Ermoliev, and Yu. M. Kaniovski, "Strong laws for a class of path-dependent stochastic processes with applications," in *Stochastic Optimization*, Vadim I. Arkin, A. Shiraev, and R. Wets, Eds., Berlin, Heidelberg, 1986, pp. 287–300, Springer Berlin Heidelberg.