Consensus formation times in anisotropic societies

Juan Neirotti

Department of Mathematics, Aston University,

The Aston Triangle, B4 7ET, Birmingham, UK

Abstract

We developed a statistical mechanics model to study the emergence of a consensus in societies of adapting, interacting agents constrained by a social rule B. In the mean field approximation we find that if the agents' interaction H_0 is weak, all agents adapt to the social rule B, with which they form a consensus; but if the interaction is sufficiently strong a consensus is built against the established *status quo*. We observed that, after a transient time α_t, agents asymptotically approach complete consensus by following a path whereby they neglect their neighbors' opinions on socially neutral issues (i.e. issues for which the society as a whole has no opinion). α_t is found to be finite for most values of the inter-agent interaction H_0 and temperature T, with exception of the values $H_0 = 1$, $T \to \infty$ and the region determined by the inequalities $\beta < 2$ and $2\beta H_0 < 1 + \beta - \sqrt{1 + 2\beta - \beta^2}$, for which consensus, with respect to B, is never reached.

I. INTRODUCTION

In this article we propose a statistical mechanics approach to study the emergence and consolidation of opinion consensus in a society of adaptive agents, in the presence of a social field B. The term *consensus* is understood to be the level of agreement amongst the agents in favor or against the predetermined socially accepted position delivered by B [1]. B represents the set of rules that determine what is socially acceptable. Such rules are the result of previous consensus-forming processes, typically observed in any functioning society [2, 3].

We developed our model from the assumption that the agents form their opinions on social issues based on partial information received at regular intervals during the process. The volume of information increases over time and, the agents being adaptive, they update their opinions accordingly.

The model we work with has been inspired on the model presented [4] and possesses the following characteristics:

1. There is a mechanism for the agents to assimilate information and update their opinions.

2. The model considers the existence of a set of rules B that determines what is socially acceptable.

3. The model considers the interaction of the agents with their neighbors [5, 6], with a strength proportional to the credibility, number and proximity of neighbors to the agent.

The topology induced by the proximity of neighbors and the adaptability of the agents are both sources of disorder that have not been considered simultaneously in previous opinion-formation models. We are convinced that this effort is worth pursuing and expect that the inclusion of these components will enhance the suitability of our model.

Opinions, considered to be the belief or attitude towards different positions on a given subject, can be conveniently modeled by continuous variables. Yet there is sufficient evidence in support of modeling opinions (on *important* issues) with binary variables [7]. Thus both the opinion of an agent a and the social position delivered by B on an issue codified into a binary string of length N, $\boldsymbol{\xi} \in \{\pm 1\}^N$ are respectively $\sigma_a(\boldsymbol{\xi})$, $\sigma_B(\boldsymbol{\xi}) \in \{\pm 1\}$. According to [4], representing a and B with perceptrons ensures the analytical tractability of the model. In this manner, the socially accepted position on $\boldsymbol{\xi}$ is $\sigma_B(\boldsymbol{\xi}) = \mathrm{sgn}(\mathbf{B} \cdot \boldsymbol{\xi})$ where $\mathbf{B} \in \mathbb{R}^N$ is the synaptic vector of B, $\mathrm{sgn}(x) = 1$ if $x > 0$, -1 if $x < 0$ and 0 otherwise and $\mathbf{B} \cdot \boldsymbol{\xi} = \sum_{j=1}^{N} B_j \xi_j$. It is clear from this formalism that the presence of B introduces a privileged direction $\mathbf{B}$ in space, which gives an anisotropic character to the opinion formation process. We associated to the agent a a perceptron with a synaptic vector $\mathbf{J}_a \in \mathbb{R}^N$, such that $\sigma_a(\boldsymbol{\xi}) = \mathrm{sgn}(\mathbf{J}_a \cdot \boldsymbol{\xi})$.

There is a body of evidence supporting the effect of social influence on opinion formation processes [8]; in consequence, to model the agents' interactions, we follow social impact theory [5, 6]. Following item 3 above, and to give a topological structure to the system, we consider a society with M agents

$1 \leq a \leq M$ linked by a set of social strengths $\mathscr{S} \equiv \{\eta_{a,c} | 0 \leq \eta_{a,c} \in \mathbb{R}\}$, where $\eta_{a,c}$ represents the influence agent c has on the opinion of agent a. Reciprocity is not assumed and, therefore, the relationship $\eta_{a,c} = \eta_{c,a}$ is not expected. We define the neighborhood of a by $\mathbb{N}_a = \{c | c \neq a \,\text{and}\, \eta_{a,c} > 0\}$ which is the set of agents *connected* to a. The opinion formation process itself is modeled by an on-line learning scenario [9], where a set of social issues $\mathscr{L}_P \equiv \big\{(\boldsymbol{\xi}_\mu, \sigma_B(\boldsymbol{\xi}_\mu)),\ \mu = 1, \ldots, P\big\}$ is used to define the energy of the society:

$$E(\{\mathbf{J}_a\}; \mathscr{L}_P, \mathscr{S}) \equiv \sum_{\mu=1}^{P} \sum_{a=1}^{M} \Theta(-\sigma_a(\boldsymbol{\xi}_\mu)\sigma_B(\boldsymbol{\xi}_\mu)) \left[1 - \sum_{c \in \mathbb{N}_a} \eta_{a,c} \Theta(-\sigma_c(\boldsymbol{\xi}_\mu)\sigma_B(\boldsymbol{\xi}_\mu))\right] \qquad (1)$$

where $\Theta(x) = 1$ if $x > 0$ and 0 otherwise. Observe that for independent agents ($\forall a, c\ \eta_{a,c} = 0$) the energy (1) is minimized to 0 when all agents develop the same opinion as B. If $\mathbb{N}_a \neq \emptyset$, then the $\mu-$th term in the RHS of (1) is 0 if $\sigma_a = \sigma_B$ or $1 - \eta_{a,c_1} - \cdots - \eta_{a,c_m}$, if a disagrees with B ($\sigma_a \neq \sigma_B$) and agrees with some of its neighbors $c_i \in \{c \in \mathbb{N}_a | \sigma_a = \sigma_c\}$. Observe that if a disagrees with B and the social strengths $\eta_{a,c}$ are large enough, the added effect of a's agreeing neighbors could make the energy grow negative. This model of the energy accounts for the effect observed in social experiments, where people tend to agree with peers that share their same opinions [10].

II. THE FREE ENERGY IN THE MEAN FIELD APPROXIMATION

The energetic formulation of the problem allows us to apply the techniques from the statistical mechanics of disordered systems to better understand the behavior of the society. There are two sources of disorder in the model described by (1), one introduced through the set of issues $\mathscr{L}_P$, and the second through the topology imposed by $\mathscr{S}$. As a valid first approach to the full treatment of the present formalism we present in this article a study on the emergence of consensus in a mean field approximation (i.e. for all index a, $\mathbb{N}_a = \{1, 2, \ldots, a-1, a+1, \ldots, M\}$ and $\eta_{a,c} = \eta_0$ for all pairs (a, c)).

We apply the replica trick [11] in order to compute the expectation of the logarithm of the partition function $\overline{\log Z} = \lim_{n\to 0} n^{-1}\left(\overline{Z^n} - 1\right)$. The average of the replicated partition function is

$$\overline{Z^n}(\beta, \eta_0) \equiv \left\langle \exp\left\{-\beta \sum_{\gamma,\mu,a} \Theta\left(-\mathbf{J}_a^\gamma \cdot \boldsymbol{\xi}_\mu \mathbf{B} \cdot \boldsymbol{\xi}_\mu\right) \left[1 - \eta_0 \sum_{c} \Theta\left(-\mathbf{J}_c^\gamma \cdot \boldsymbol{\xi}_\mu \mathbf{B} \cdot \boldsymbol{\xi}_\mu\right)\right]\right\}\right\rangle_{\{\boldsymbol{\xi}_\mu\},\mathbf{B},\{\mathbf{J}_a^\gamma\}} \qquad (2)$$

where β (the inverse of the temperature) is a parameter that gauges the fluctuations of energy and the angular brackets represent the expectation over the set of issues $\{\boldsymbol{\xi}_\mu\}$, the distribution of synaptic vectors of the social rule $\mathbf{B}$ and the set of replicated synaptic vectors of the agents $\{\mathbf{J}_a^\gamma\}$ (the details of the calculation are presented in Appendix A).

The calculation of the average over the disorder introduced through the social issues in $\mathscr{L}_P$, produces an expression for the replicated partition function $\overline{Z^n}$ that depends on the following distributed variables:

$$R_a^\gamma \equiv \frac{\mathbf{J}_a^\gamma \cdot \mathbf{B}}{N}, \qquad W_{a,b}^\gamma \equiv \frac{\mathbf{J}_a^\gamma \cdot \mathbf{J}_b^\gamma}{N},$$
$$q_a^{\gamma,\rho} \equiv \frac{\mathbf{J}_a^\gamma \cdot \mathbf{J}_a^\rho}{N}, \qquad t_{a,b}^{\gamma,\rho} \equiv \frac{\mathbf{J}_a^\gamma \cdot \mathbf{J}_b^\rho}{N}.$$

These overlaps are the cosines of the angles between synaptic vectors and they represent a level of agreement between the agents and the society (R_a^γ), between two different agents ($W_{a,b}^\gamma$ and $t_{a,b}^{\gamma,\rho}$) or between versions of the same agent in different replicas ($q_a^{\gamma,\rho}$). We impose a replica and site symmetric approximation, which entails consideration of the values of the overlaps as site and replica independent $R_a^\gamma = R$, $q_a^{\gamma,\rho} = q$, $W_{a,b}^\gamma = W$ and $t_{a,b}^{\gamma,\rho} = t$. It is possible to justify that the difference between W and t satisfies the scaling $\tau \equiv M(W-t) \sim O(1)$ (see reference [12], equation (3)) which simplifies the matrix representation of the interaction between replicated systems.

In this approximation, and assuming that the length of the issues N is sufficiently large and τ sufficiently small, the replicated partition function can be expressed as:

$$\overline{Z^n}(\alpha,\beta,H_0) = \underset{q,R,W}{\operatorname{extr}} \left\{ \exp\left(\frac{N}{2}\mathcal{G}_S(q,R) + \alpha N \mathcal{G}_E(q,R,W;\beta,H_0) \right) \right\}$$

where $\alpha \equiv P/N$ is a parameter that measures the volume of information provided to the agents. Such information is supplied at constant rate, thus α can be interpreted as a measure of time. The quantity $H_0 \equiv M\eta_0 \sim O(1)$ is a measure of the total interaction between an agent and its neighborhood. It must be an $O(1)$ quantity to ensure the extensivity of the energy (1); and:

$$\mathcal{G}_S(q,R) \equiv nM\left(\ln(1-q) + \frac{q-R^2}{1-q} \right)$$
$$\mathcal{G}_E(q,R,W;\beta,H_0) \equiv -2nM \int \mathrm{d}z\, \mathcal{N}\left(z \,\middle|\, 0, \frac{W}{1-q} \right) \mathcal{H}\left(-\sqrt{\frac{1-q}{W(W-R^2)}} Rz \right) \Phi(z;\beta,H_0),$$

where $\mathcal{N}(x|\mu,\sigma^2) = \exp[(x-\mu)^2/2\sigma^2]/\sqrt{2\pi\sigma^2}$ is a Gaussian distribution in x, centered at μ and with variance σ^2 and $\mathcal{H}(x) \equiv \int_x^\infty \mathrm{d}z\,\mathcal{N}(z|0,1)$ is the Gardner error function. The function $\Phi(z;\beta,H_0)$ carries the information of the averaged inter-agent interaction, weighted by the thermal coefficient:

$$\Phi(z;\beta,H_0) \equiv -\lim_{M\to\infty} \frac{1}{M} \log\left\{ \int \mathcal{D}x \left[\mathcal{H}(-z) + \exp\left(\sqrt{\frac{2\beta H_0}{M}} x - \beta \right) \mathcal{H}(z) \right]^M \right\}$$
$$= \min_{u\in[0,1]} \tilde{\Phi}(u,z;\beta,H_0), \tag{3}$$

with

$$\tilde{\Phi}(u,z;\beta,H_0) \equiv \frac{[u-\mathcal{H}(z)]^2}{2\mathcal{H}(z)\mathcal{H}(-z)} - u^2\beta H_0 + u\beta.$$

This expression is obtained through the application of Laplace's method under the assumption that the size of the population (M) is sufficiently large [28]. There are three possible results to the minimization problem (3), depending on the values of the variable z and the parameters β and H_0.

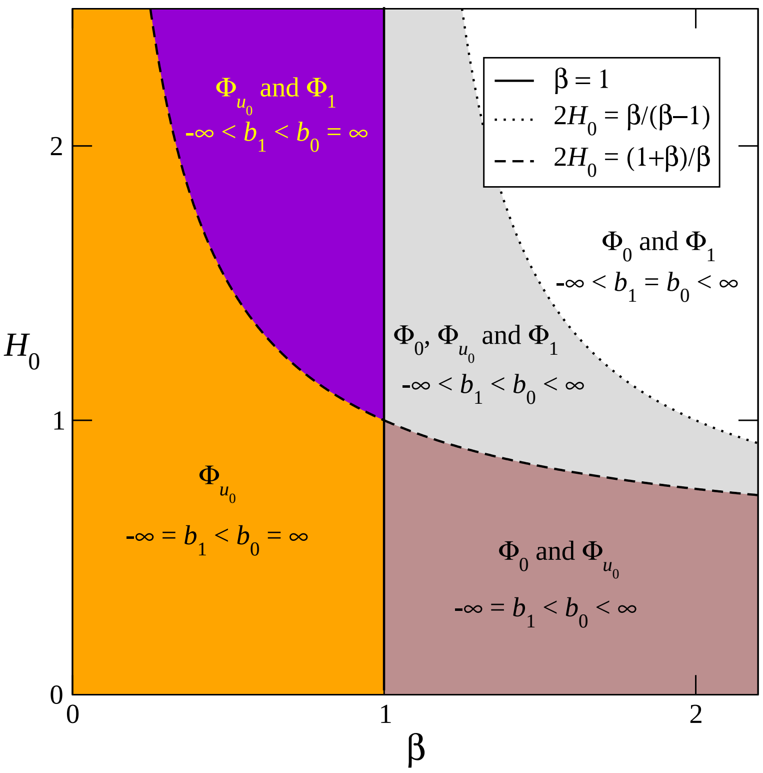

Figure 1: Distribution of the components (A9), with their correspondent boundaries b_0 (A7) and b_1 (A8), in the plane (β, H_0) (color on-line).

Given the functions $b_0(\beta, H_0)$ and $b_1(\beta, H_0)$ (equations (A7) and (A8) respectively), we observe that if $b_0 < z$, the minimum of (3) is at $u = 0$ and $\Phi(z) = \Phi_0(z) \equiv \tilde{\Phi}(0, z)$; if $b_1 < z < b_0$, the minimum is at $u = u_0$, where $0 < u_0 < 1$ is given by the equation (A3) and $\Phi(z) = \Phi_{u_0}(z) \equiv \tilde{\Phi}(u_0, z)$; and if $z < b_1$, the minimum is at $u = 1$ and $\Phi(z) = \Phi_1(z) \equiv \tilde{\Phi}(1, z)$. The explicit form of the components Φ_0, Φ_{u_0} and Φ_1 is given in expression (A9). Observe that the function Φ so defined is continuous but not differentiable at $z = b_0, b_1$. In figure 1 we present the distribution of the components $\Phi_{0,}\Phi_{u_0}$ and Φ_1 in the plane (β, H_0), which provides insight on the phase diagram of the system.

By defining the new parameters $w \equiv W/(1-q)$ and $r \equiv R/\sqrt{1-q}$ we have that:

$$\begin{aligned}\beta f(\alpha\beta, H_0) &\equiv -\lim_{n\to 0}\lim_{M,N\to\infty}\frac{\overline{Z^n}(\alpha,\beta,H_0)-1}{nNM}\\ &= \operatorname*{extr}_{q}\psi(q)\ +\operatorname*{extr}_{r,w}\phi(r,w;\alpha,\beta,H_0)\end{aligned} \tag{4}$$

where

$$\psi(q) \equiv -\frac{1}{2}\left(\ln(1-q) + \frac{q}{1-q}\right) \tag{5}$$

$$\phi(r,w;\alpha,\beta,H_0) \equiv \frac{r^2}{2} + 2\alpha\int \mathrm{d}z\,\mathcal{N}(z|0,w)\mathcal{H}\left(-\frac{rz}{\sqrt{w(w-r^2)}}\right)\Phi(z;\beta,H_0). \tag{6}$$

Observe that $\psi(q)$ is concave in q and its minimum is reached at $q = 1$. Given that ψ does not depend on the parameters α, β or H_0, we will consider the problem of optimizing the shifted free energy:

$$\beta f_0(\alpha, \beta, H_0) \equiv \operatorname*{extr}_{r,w} \phi(r, w; \alpha, \beta, H_0). \tag{7}$$

III. THE ROLE OF THE SOCIALLY NEUTRAL ISSUES

To better understand how the process of opinion formation evolves, we need to study what happens in the orthogonal hyper-space to $\mathbf{B}$. To this end we define as *socially neutral issues* all the binary strings $\mathbf{S}_0 \in \{\pm 1\}^N$ satisfying $\mathbf{B} \cdot \mathbf{S}_0 = 0$.Thus, a socially neutral issue is an issue for which there is no socially accepted position.

The optimization of the function ϕ with respect to the re-scaled parameters produces the equations $\partial_r \phi = \partial_w \phi = 0$, that are satisfied if:

$$r = -\sqrt{\frac{2}{\pi}}\alpha \int \mathrm{d}z \, \mathcal{N}(z|0, w - r^2) \, \frac{\partial \Phi(z; \beta, H_0)}{\partial z} \tag{8}$$

$$r^2 = 2\alpha \int \mathrm{d}z \, \mathcal{N}(z|0, w) \left(1 - \frac{z^2}{w}\right) \mathcal{H}\left(-\frac{rz}{\sqrt{w(w - r^2)}}\right) \Phi(z; \beta, H_0), \tag{9}$$

where $0 \leq r^2 \leq w$, which implies that $R^2 \leq W$. If two agents a and c have the same overlap with B, i.e. $R_a = R_c = R$, the relationship between R and W is $W = R^2 + (1 - R^2)\cos\varphi$, where φ is the angle between the components of $\mathbf{J}_a$ and $\mathbf{J}_c$ perpendicular to $\mathbf{B}$. In such a case, if $R^2 = W$, then $\varphi = \frac{\pi}{2}$ and the probability of both agents agreeing on any $\mathbf{S}_0$ is $\frac{1}{2}$ and no consensus can be built on socially neutral issues. If $R = 0$, then $0 < \cos\varphi = W$, indicating that there is no consensus in favor or against B but a level of agreement can be built on socially neutral issues.

A. $r^2 = w$ solution. Independence of opinion on socially neutral issues

Observe that equations (8) and (9) can be satisfied simultaneously with the condition $r^2 = w$ (implying $R^2 = W$) for a finite value of $\alpha = \alpha_t$ at a particular value of $r = r_t$ determined by the equations:

$$\alpha_t = -\sqrt{\frac{\pi}{2}} \frac{r_t}{\Phi^{(1)}(\beta, H_0)} \tag{10}$$

$$r_t = -\frac{\sqrt{2\pi}}{\Phi^{(1)}(\beta, H_0)} \int \mathrm{d}z \, \mathcal{N}(z|0, r_t^2) \left(1 - \frac{z^2}{r_t^2}\right) \Theta(r_t z) \, \Phi(z; \beta, H_0), \tag{11}$$

where

$$\Phi^{(n)}(\beta, H_0) \equiv \mathcal{A}_{u_0}(\beta, H_0) \left.\frac{\partial^n \Phi_{u_0}(z; \beta, H_0)}{\partial z^n}\right|_{z=0} + \mathcal{A}_0(\beta, H_0) \left.\frac{\partial^n \Phi_0(z; \beta, H_0)}{\partial z^n}\right|_{z=0} + \\ + \mathcal{A}_1(\beta, H_0) \left.\frac{\partial^n \Phi_1(z; \beta, H_0)}{\partial z^n}\right|_{z=0} \tag{12}$$

is the n-th derivative of Φ at $z = 0$ and $\mathcal{A}_1(\beta, H_0) \equiv \Theta(H_0 - 1)\Theta(2\beta H_0 - 2 - \beta)$, $\mathcal{A}_{u_0}(\beta, H_0) \equiv \Theta(1 - H_0)\Theta(2 - \beta) + \Theta(H_0 - 1)\Theta(2 + \beta - 2\beta H_0)$ and $\mathcal{A}_0(\beta, H_0) \equiv \Theta(1 - H_0)\Theta(\beta - 2)$ are signal functions such that $\mathcal{A}_\Gamma = 1$ if $z = 0$ is in the domain of Φ_Γ or 0 otherwise, with $\Gamma = 0, u_0, 1$. [29] In particular, the first derivative of Φ at 0 is given by:

$$\Phi^{(1)}(\beta, H_0) = \sqrt{\frac{2}{\pi}}\mathrm{sgn}(H_0 - 1)\left(\frac{\beta|H_0 - 1|}{2 - \beta H_0}\mathcal{A}_{u_0}(\beta, H_0) + \mathcal{A}_0(\beta, H_0) + \mathcal{A}_1(\beta, H_0)\right). \tag{13}$$

Observe that $\mathrm{sgn}(\Phi^{(1)}) = \mathrm{sgn}(H_0 - 1)$ and being $\alpha_t > 0$, through (10) the sign of r_t must be $\mathrm{sgn}(1 - H_0)$. Let us assume that $|r_t|$ is small enough, such that the error term:

$$\epsilon(\beta, H_0) \equiv \max_{z\in\mathbb{R},\gamma=0,1}\{|\Phi(z;\beta, H_0)|\}|b_\gamma|\mathcal{N}(b_\gamma|0, r_t^2) \tag{14}$$

is negligible, and that we are working in a region of the plane (β, H_0) such that the boundaries b_0 and b_1 are not zero. By using expressions (12) and (14) we can approximate (11) in the following way:

$$r_t \approx -\sqrt{2\pi}\sum_{n=0}^{\infty}\frac{r_t^n}{n!}\frac{\Phi^{(n)}(\beta, H_0)}{\Phi^{(1)}(\beta, H_0)}\int_0^\infty \mathcal{D}z\, z^n(1 - z^2) + O(\epsilon) \tag{15}$$

which implies that, keeping terms up to $O(r_t^4)$ in (15), we obtain:

$$r_t \approx \sqrt{\frac{\pi^3}{2}}\frac{-2\beta H_0 + 2(1+\beta)H_0 - \beta}{(1 - H_0)[(12-\pi)\beta H_0 + 2\pi]}\mathcal{A}_{u_0}(\beta, H_0) + \frac{\sqrt{2\pi^3}}{12 - \pi}\left[\mathcal{A}_0(\beta, H_0) - \mathcal{A}_1(\beta, H_0)\right], \tag{16}$$

and

$$\alpha_t \approx \frac{\pi^{5/2}}{2^{3/2}}\frac{(2 - \beta H_0)(-2\beta H_0 + 2(1+\beta)H_0 - \beta)}{\beta(1 - H_0)^2[(12-\pi)\beta H_0 + 2\pi]}\mathcal{A}_{u_0}(\beta, H_0) + \alpha_{0,1}\left[1 - \mathcal{A}_{u_0}(\beta, H_0)\right], \tag{17}$$

where

$$\alpha_{0,1} \equiv \frac{2^{1/2}\pi^{5/2}}{24 - 2\pi} \approx 1.396. \tag{18}$$

$\alpha_{0,1}$ is introduced as a measure of a typical time scale for most of the points of the (β, H_0) plane. Equation (16) is an approximation to the solution of (11) which is qualitatively suitable if $\mathrm{sgn}(r_t) = \mathrm{sgn}(1 - H_0)$. This is not the case for order pairs (β, H_0) satisfying:

$$\mathcal{B}(\beta, H_0) = \Theta(2 - \beta)\Theta\left(1 + \beta - \sqrt{1 + 2\beta - \beta^2} - 2\beta H_0\right). \tag{19}$$

In this region, the proposal $r_t^2 = w_t$ does not satisfy the saddle point equations (8) and (9). We will explore the behavior of the solution in this region in the next subsection. For almost all the region of the plane (β, H_0) determined by the equation $\mathcal{B}(\beta, H_0) = 0$, the solution $r^2 = w$ is stable (see Appendix B).

Most of the opinion formation process occurs for $\alpha > \alpha_t$. The effective energy for $\alpha > \alpha_t$ can be defined as

$$\phi_{\text{eff}}(r; \alpha, \beta, H_0) \equiv \frac{r^2}{2} + 2\alpha\int \mathrm{d}z\,\mathcal{N}\left(z\,\middle|0, r^2\right)\Theta(rz)\Phi(z;\beta, H_0). \tag{20}$$

The new saddle point equation is:

$$\partial_r\phi_{\text{eff}} = r - \frac{2\alpha}{|r|}\int \mathrm{d}z\,\mathcal{N}(z|0, r^2)\left(1 - \frac{z^2}{r^2}\right)\Theta(rz)\Phi(z;\beta, H_0)$$

which implies that for large values of α, $|r| \gg 1$,thus

$$r^3 \approx \frac{\mathrm{sgn}(1-H_0)}{2\pi}\alpha, \tag{21}$$

which implies that $|r| \sim O(\alpha^{1/3})$, and the second derivative is then $\partial^2_{r,r}\phi_{\text{eff}} \approx 1 + O(\alpha^{-1/3})$, which indicates that the solution (21) is stable.

Finally, observe that $r^2 \propto 1/(1-q)$, thus we expect for α sufficiently large to observe the asymptotic behavior $q \approx 1 - O(\alpha^{-2/3})$.

B. $r^2 < w$ solution. Consensus on socially neutral issues

The behavior $r^2 < w$ is observed for values of β and H_0 such that $\mathcal{B}(\beta, H_0) = 1$, indicating that the component of Φ that appears in (8) and (9) for these values of β and H_0 is Φ_{u_0}. Therefore, for small enough values of α we have that $w - r^2 \ll 1$ and $|r| \ll 1$, therefore:

$$r \approx -\sqrt{\frac{2}{\pi}}\alpha\Phi^{(1)}_{u_0}(\beta, H_0) \tag{22}$$

$$r^2 \approx 2\alpha \int_0^\infty \mathcal{D}z\,(1-z^2)\,\Phi_{u_0}(\sqrt{w}z;\beta,H_0) \tag{23}$$

where (22) and (13) indicate that $r > 0$ and in (23) we have use the approximation based on (14). By expanding $\Phi_{u_0}(z;\beta,H_0)$ around $z = 0$, we obtain an expression for r up to order one in w:

$$r \approx \sqrt{w} - \frac{2}{\pi}\frac{\beta H_0^2 - 2(\beta+1)H_0 + \beta}{(1-H_0)(2-\beta H_0)} w \tag{24}$$

where the factor of w in the second term of (24) is positive if $\mathcal{B}(\beta, H_0) = 1$.

For large values of α we suppose that $w > w - r^2 \gg 1$. Thus:

$$\begin{aligned} r &= -\sqrt{\frac{2}{\pi}}\frac{\alpha}{w-r^2}\int_{-\infty}^{\infty}\mathcal{D}z\,z\,\Phi_{u_0}\left(\sqrt{w-r^2}z;\beta,H_0\right) \\ &\approx \frac{\alpha\beta(1-H_0)}{\pi\sqrt{w-r^2}} \end{aligned} \tag{25}$$

$$\begin{aligned} r^2 &\approx \alpha\int_{-\infty}^{\infty}\mathcal{D}z\,(1-z^2)\,\Phi_{u_0}\left(\sqrt{w-r^2}z;\beta,H_0\right) \\ &\approx \frac{\alpha\beta(2-\beta)}{4\pi\sqrt{w}}. \end{aligned} \tag{26}$$

From (25) and (26) we obtain that $r \sim \frac{1}{4}(2-\beta)/(1-H_0)$ asymptotically, which does not depend on α. In a similar manner, we obtain the asymptotic behavior of $\sqrt{w} \sim \frac{4}{\pi}\alpha\beta(1-H_0)^2/(2-\beta)$ which indicates that $1 - q \sim O(\alpha^{-2})$. These results indicate that the overlap R approaches zero asymptotically $R \sim O(\alpha^{-1})$.

C. Phase diagram

We solved numerically the equations (10) and (11) and constructed the plot of the $\log(\alpha_t)$ as a function of β and H_0 presented in figure 2. α_t represents the transient period prior to the final stage

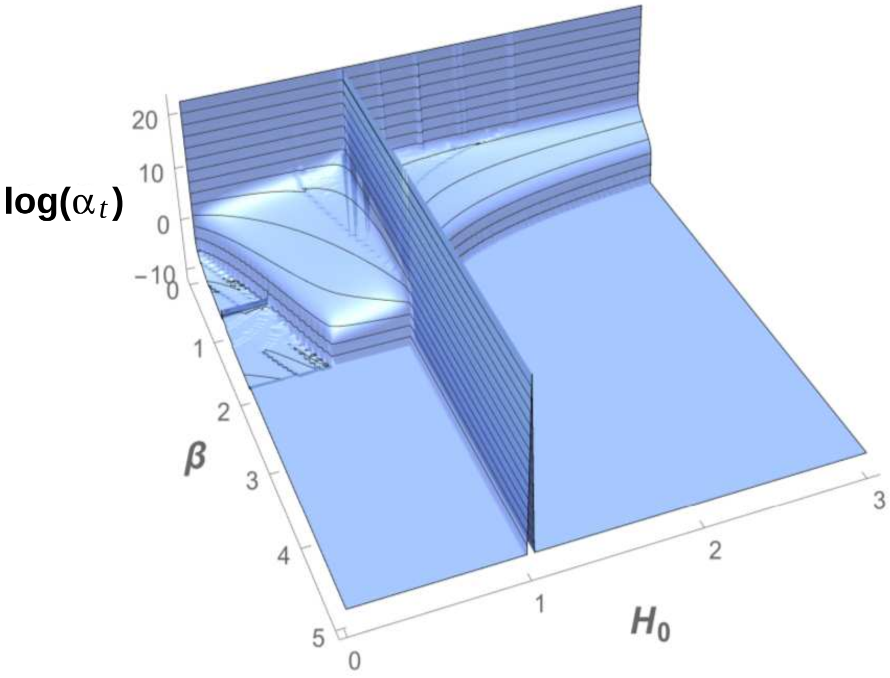

Figure 2: Logarithm of the transient time $\log(\alpha_t)$ as a function of β and H_0. (Color on-line)

of the opinion formation process, characterized by agents developing independent attitudes towards their peers' opinions on socially neutral issues. From figure 2 we observe that there is a sector of the (β, H_0) plane for which the system takes a relatively long time to reach the solution $r^2 = w$. This sector is formed by the order pairs (β, H_0) that make $\mathcal{A}_{u_0}(\beta, H_0) = 1$. In the triangular sector formed by order pairs (β, H_0) that make $\mathcal{B}(\beta, H_0) = 1$, no suitable numerical solution was found, as was expected.

In order to better understand the picture the system presents immediately after α_t and by considering the definitions of $\mathcal{A}_1$, $\mathcal{A}_{u_0}$, $\mathcal{A}_0$ and $\mathcal{B}$ with addition of the calculation of the instable region and the analysis of the signs of the solutions presented in (21) and (25), we constructed the diagram of figure 3. The areas marked A_{u_0} correspond to sectors of the (β, H_0) plane characterized by relatively long transient times $\alpha_t \gg \alpha_{0,1}$, whereas the areas marked A_0 and A_1 develop the solution $r^2 = w$ in relatively short transient times $\alpha_t = \alpha_{0,1}$.

With the asymptotic behavior of R inferred from the equations (21) and (25) we constructed the phase diagram of the system, presented in figure 4. Observe that for $H_0 > 1$ the asymptotic value of $R = -1$. At $H_0 = 1$ we have that $R = 0$ for all α, inside the sector with $\mathcal{B}(\beta, H_0) = 1$ R vanishes asymptotically and for order pairs (β, H_0) such that $H_0 < 1$ and $\mathcal{B}(\beta, H_0) = 0$ we have that $R = 1$. The transitions between the phases with $R = 0$ and $R = 1$, and between the phases with $R = 1$ and $R = -1$ are of the first order.

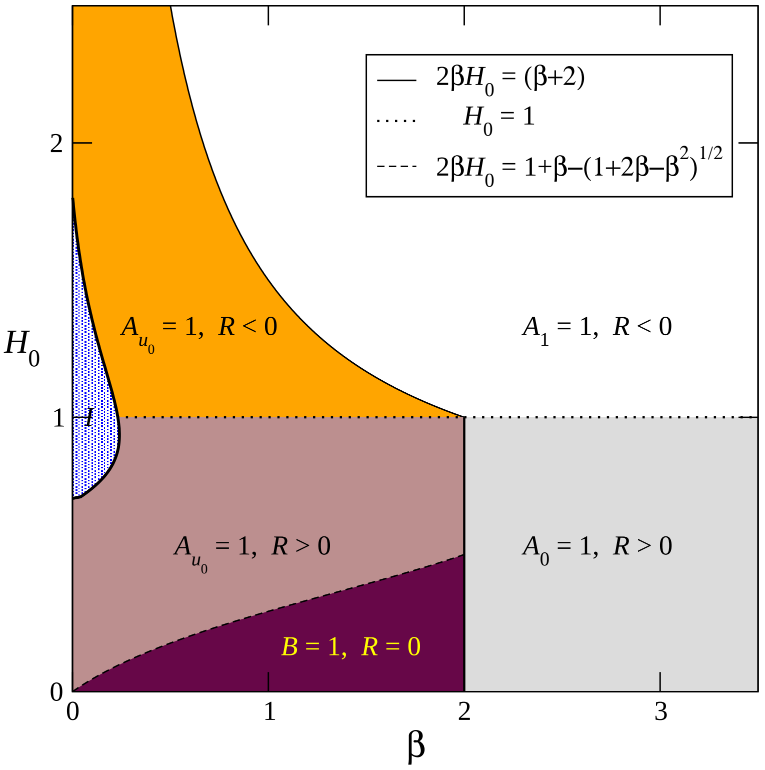

Figure 3: In this diagram we present a picture of the system at . We labeled the regions where the proposed solution $r_t^2 = w_t$ is stable by $\mathcal{A}_{u_0}$(where $\alpha_t \gg \alpha_{0,1}$), $\mathcal{A}_0$ and $\mathcal{A}_1$ (where, in both cases, $\alpha_t = \alpha_{0.1}$), by I where the proposed solution is instable and by $\mathcal{B}$ the region where $r^2 < w$ for all α. We also indicated the sign of R according to (16) and (25) (color on-line).

IV. DISCUSSION

We presented a model for the opinion formation process in a society of interacting agents, represented by binary perceptrons, in the presence of a social field B. The field is the result of many opinion formation processes prior to the current one; it provides the socially acceptable position on current issues and indicates a preferential direction in the space of issues given the anisotropic character to the system. The model, represented by equation (1), incorporates the interaction of two different sources of disorder, namely the topology of the interaction $\mathscr{S}$ and the training set $\mathscr{L}_P$ and, although our results have been obtained by considering a mean field approximation on the topology, we expect to tackle the complete model in a future work.

Our results are derived from the study of the shifted free energy (7), associated with the function ϕ (6) through an optimization procedure. The optimal solutions of the energy are obtained by solving the equations (8) and (9) for the reduced parameters $r \equiv R/\sqrt{1-q}$ and $w \equiv W/(1-q)$ respectively. For most of the values of β and H_0 (i.e. $\mathcal{B}(\beta, H_0)$=0), the solution $r_t^2 = w_t$ is reached after a transient time α_t. This transient is larger in the region determined by the values of β and H_0

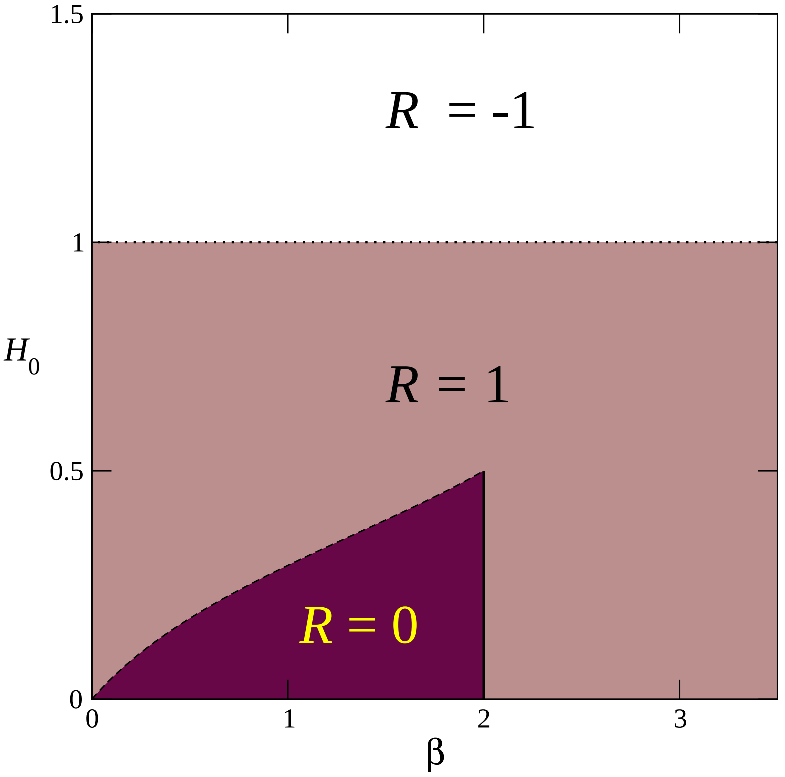

Figure 4: Phase diagram of the system in the limit of $\alpha \to \infty$. Transitions between any two phases are always of the first order (color on-line).

such that $\mathcal{A}_{u_0}(\beta, H_0) = 1$. This region is characterized by a high temperature ($\beta \to 0$) which is the cause of the long transients. The only region in the plane (β, H_0) for which the solutions found are not stable is located in the neighborhood of the point $\beta = 0$ and $H_0 = 1$, indicated in figure 3 by a label I.

We also constructed a phase diagram of the system by inferring the behavior of R for large values of α, presented in figure 4. For values of $H_0 > 1$ the consensus is always formed against B, i.e. $R = -1$. This is one of the effects studied within the context of moral foundation theory, which considers the cause of change in the society's *status quo* the frequent corroboration of opinion between equally minded voters [14, 15]. The conservative attitude of the agents ($R = 1$) interacting with low values of $H_0 < 1$ is consistent with previous studies done on a dynamical version of the model at zero temperature [4]. Inside the region $\mathcal{B}(\beta, H_0) = 1$ there is no consensus with respect to B ($R = 0$). The transitions between any two phases are of the first order in all the possible cases.

The fact that at α_t the overlaps become $R_t^2 = W_t$ indicates that the agents approach consensus disregarding the opinion of their peers on socially neutral issues (issues for which there is no definite socially accepted position). Given that the only anisotropy of energy (1) is due to the presence of the synaptic vector $\mathbf{B}$, it is reasonable to suppose that the agents evolve maximizing the diversity of opinions in the only region of the version space where there is no social reference, i.e. the hyper-space perpendicular to $\mathbf{B}$.

Consensus with respect to B is never formed for $\beta = 0$, $H_0 = 1$ and the values of β and H_0 satisfying $\mathcal{B}(\beta, H_0) = 1$. On the line $\beta = 0$, $\Phi(z)$ is zero, consensus is never achieved due to large energy fluctuations in the system, and $R = 0$ for all α. At $H_0 = 1$, $\Phi(z)$ is even and the solution to (8) is $R = 0$. This occurs because competing attitudes towards following either B or neighboring agents cancel each other out and consensus is never reached. At $\mathcal{B}(\beta, H_0)$ a consensus is initially built in favor of B ($R > 0$), but it vanishes asymptotically when more information is provided to the system ($R \to 0$ when $\alpha \to \infty$). The only consensus observed in this region is with respect to socially neutral issues which is an effect similar to the one observed when irrelevant events affect the opinion of voters on government performance [16].

A similar model, without the presence of B, has been studied in [17]. In this model the authors found the persistence of disagreement in a system compossed by consensus seekers. Apparently the lack of reference (B in our case) made impossible the formation of a consensus.

It is worth to mention that these results have been obtained assuming that the size of the population (M) is large enough. Although *large enough* in this context is equivalent to infinitely large, it may be interesting to explore the suitability of the results found as approximations to the behavior of finite sized communities

α is a time-like parameter, thus the reported α_t can be considered as characteristic times of the model, which, for a fully connected system, is expected to be shorter than the one obtained by other means than a mean field approximation [18, 19]. As is expected from a mean field approximation [20, 21], phenomena associated to the correlation length of the system (like the presence of clusters reported in [4, 22]), cannot be addressed within this framework. To do so we will need to consider more realistic graph topologies, particularly by introducing non-symmetric interaction (directed graphs) [23] and connectivity dynamics [24, 25] which facilitates the exchange of information between agents [26, 27].

Acknowledgments

The author would like to acknowledge discussions with Dr R. C. Alamino in the early stages of this work. The advice of L. E. Neirotti is also warmly appreciated.

Appendix A: Mean Field Approach

The average we need to compute is:

$$\overline{Z^n}(\beta, \eta_0) \equiv \left\langle \exp\left\{ -\beta \sum_{\gamma,\mu,a} \Theta\left(-\mathbf{J}_a^\gamma \cdot \boldsymbol{\xi}_\mu \mathbf{B} \cdot \boldsymbol{\xi}_\mu\right) \left[1 - \eta_0 \sum_c \Theta\left(-\mathbf{J}_c^\gamma \cdot \boldsymbol{\xi}_\mu \mathbf{B} \cdot \boldsymbol{\xi}_\mu\right)\right] \right\} \right\rangle_{\{\boldsymbol{\xi}_\mu\}, \mathbf{B}, \{\mathbf{J}_a^\gamma\}} .$$

We assumed that the components of the issues $\boldsymbol{\xi}$ are i.i.d variables drawn from $\mathcal{P}(\xi_i = \pm 1) = \frac{1}{2}$ (but any distribution with zero mean and unit variance would do). Any non-zero vector $\mathbf{B} \in \mathbb{R}^N$ could be used as the social rule's synaptic vector and so determine a privileged direction in space. For simplicity's sake we chose the vector with components $B_k = 1$, and thus $\mathcal{P}(\mathbf{B}) = \prod_k \delta(B_k - 1)$. The agents' synaptic vectors are uniformly distributed over the surface of a sphere in $\mathbb{R}^N$ centered at 0 and with radius $\sqrt{N}$, thus $\mathcal{P}(\mathbf{J}) \equiv \prod_{k=1}^N \delta\left(\sum_{k=1}^N J_k^2 - N\right) / \sqrt{2\pi \mathrm{e}}$.

In order to compute the partition function equation (2) we define the $O(1)$ variables $\lambda_{a,\mu}^\gamma \equiv \mathbf{J}_a^\gamma \cdot \boldsymbol{\xi}_\mu / \sqrt{N}$ and $u_\mu \equiv \mathbf{B} \cdot \boldsymbol{\xi}_\mu / \sqrt{N}$ and perform the average over the training set:

$$\overline{Z^n}(\beta) = \int \prod_{\gamma,\mu,a} \frac{\mathrm{d}\lambda_{a,\mu}^\gamma \mathrm{d}\hat{\lambda}_{a,\mu}^\gamma}{2\pi} \int \prod_\mu \frac{\mathrm{d}u_\mu \mathrm{d}\hat{u}_\mu}{2\pi} \exp\left(-i \sum_{\gamma,\mu,a} \hat{\lambda}_{a,\mu}^\gamma \lambda_{a,\mu}^\gamma - i \sum_\mu \hat{u}_\mu u_\mu\right)$$
$$\left\langle \prod_{\mu,k} \cos\left(\sum_{\gamma,a} \frac{\hat{\lambda}_{a,\mu}^\gamma J_{a,k}^\gamma}{\sqrt{N}} + \frac{\hat{u}_\mu B_k}{\sqrt{N}}\right)\right\rangle_{\mathbf{B},\{\mathbf{J}_a^\gamma\}}$$
$$\exp\left\{-\sum_{\gamma,\mu,a} \beta \Theta\left(-\lambda_{a,\mu}^\gamma u_\mu\right) \left[1 - \eta_0 \sum_c \Theta\left(-\lambda_{c,\mu}^\gamma u_\mu\right)\right]\right\}.$$

By applying a Gaussian approximation to the product of cosines, by introducing the overlaps:

$$R_a^\gamma \equiv \frac{\mathbf{J}_a^\gamma \cdot \mathbf{B}}{N}, \qquad W_{a,b}^\gamma \equiv \frac{\mathbf{J}_a^\gamma \cdot \mathbf{J}_b^\gamma}{N},$$
$$q_a^{\gamma,\rho} \equiv \frac{\mathbf{J}_a^\gamma \cdot \mathbf{J}_a^\rho}{N}, \qquad t_{a,b}^{\gamma,\rho} \equiv \frac{\mathbf{J}_a^\gamma \cdot \mathbf{J}_b^\rho}{N},$$

by defining the matrices:

$$[\hat{\boldsymbol{Q}}]_{a,b}^{\gamma,\rho} \equiv i\left\{\delta^{\gamma,\rho}\left(\delta_{a,b}\hat{\ell}_a^\gamma + (1-\delta_{a,b})\hat{W}_{a,b}^\gamma\right) + (1-\delta^{\gamma,\rho})\left(\delta_{a,b}\hat{q}_a^{\gamma,\rho} + (1-\delta_{a,b})\hat{t}_{a,b}^{\gamma,\rho}\right)\right\}$$

$$[\boldsymbol{Q}]_{a,b}^{\gamma,\rho} \equiv \delta^{\gamma,\rho}\left(\delta_{a,b} + (1-\delta_{a,b})W_{a,b}^\gamma\right) + (1-\delta^{\gamma,\rho})\left(\delta_{a,b}q_a^{\gamma,\rho} + (1-\delta_{a,b})t_{a,b}^{\gamma,\rho}\right)$$

and by integrating over the synaptic vectors we have that:

$$\overline{Z^n}(\alpha,\beta,\eta_0) = \mathscr{C}^{-1}\hat{\mathscr{C}}^{-1} \int \mathrm{d}\boldsymbol{Q}\, \mathrm{d}\boldsymbol{R}\, \mathrm{d}\hat{\boldsymbol{Q}}\, \mathrm{d}\hat{\boldsymbol{R}} \exp\left(N g_S(\boldsymbol{Q},\boldsymbol{R},\hat{\boldsymbol{Q}},\hat{\boldsymbol{R}})\right) \times$$
$$\left[\int \frac{\mathrm{d}\hat{\boldsymbol{\lambda}}\, \mathrm{d}\boldsymbol{\lambda}\, \mathrm{d}u}{(2\pi)^{nM+1/2}} \exp\left(g_E(\boldsymbol{Q},\boldsymbol{R},\hat{\boldsymbol{\lambda}},\boldsymbol{\lambda},u;\beta,\eta_0)\right)\right]^{\alpha N}$$

where $\mathscr{C}$ and $\hat{\mathscr{C}}$ are suitable normalization constants, $P = \alpha N$ and

$$g_S(\boldsymbol{Q},\boldsymbol{R},\hat{\boldsymbol{Q}},\hat{\boldsymbol{R}}) \equiv \frac{1}{2}\mathrm{tr}\boldsymbol{Q}\hat{\boldsymbol{Q}} - \frac{1}{2}\ln|\hat{\boldsymbol{Q}}| - \frac{1}{2}\sum_{a,b}\sum_{\gamma,\rho} \hat{R}_a^\gamma \left[\hat{\boldsymbol{Q}}^{-1}\right]_{a,b}^{\gamma,\rho} \hat{R}_b^\rho + i\sum_{\gamma,a} \hat{R}_a^\gamma R_a^\gamma - \frac{nM}{2}$$
$$g_E(\boldsymbol{Q},\boldsymbol{R},\hat{\boldsymbol{\lambda}},\boldsymbol{\lambda},u;\beta,\eta_0) \equiv -\frac{1}{2}\sum_{\gamma,a}\left(1-(R_a^\gamma)^2\right)\left(\hat{\lambda}_a^\gamma\right)^2 - \sum_{\gamma,a}\sum_{\gamma<\rho}\left(q_a^{\gamma,\rho} - R_a^\gamma R_a^\rho\right)\hat{\lambda}_a^\gamma \hat{\lambda}_a^\rho -$$
$$-\sum_{\gamma,a}\sum_{a<b}\left(W_{a,b}^\gamma - R_a^\gamma R_b^\gamma\right)\hat{\lambda}_a^\gamma\hat{\lambda}_b^\gamma - \sum_{\gamma,a}\sum_{\gamma\neq\rho}\sum_{a<b}\left(t_{a,b}^{\gamma,\rho} - R_a^\gamma R_b^\rho\right)\hat{\lambda}_a^\gamma\hat{\lambda}_b^\rho -$$
$$-\frac{u^2}{2} + i\sum_{\gamma,a}\hat{\lambda}_a^\gamma R_a^\gamma u - i\sum_{\gamma,a}\hat{\lambda}_a^\gamma\lambda_a^\gamma - \sum_{\gamma,a}\beta\Theta(-\lambda_a^\gamma u)\left(1 - \eta_0\sum_c \Theta(-\lambda_c^\gamma u)\right).$$

In the large N limit we can apply Laplace method to solve the integrals over $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{R}}$, thus obtaining:

$$\begin{aligned}\hat{R}^{\gamma}_{a} &= i\sum_{\rho,b}[\hat{\boldsymbol{Q}}]^{\gamma,\rho}_{a,b}R^{\rho}_{b}\\ \left[\hat{\boldsymbol{Q}}^{-1}\right]^{\gamma,\rho}_{a,b} &= [\boldsymbol{K}]^{\gamma,\rho}_{a,b}\equiv[\boldsymbol{Q}]^{\gamma,\rho}_{a,b}-R^{\gamma}_{a}R^{\rho}_{b},\end{aligned}$$

which produces that

$$\begin{aligned}\exp[N\mathcal{G}_S(\boldsymbol{K})] &\equiv \operatorname*{extr}_{\hat{\boldsymbol{Q}},\hat{\boldsymbol{R}}}\left\{\exp\left(Ng_S(\boldsymbol{Q},\boldsymbol{R},\hat{\boldsymbol{Q}},\hat{\boldsymbol{R}})\right)\right\}\\ &= |\boldsymbol{K}|^{N/2}\end{aligned}$$

thus

$$\overline{Z^n}(\alpha,\beta,\eta_0)=\operatorname*{extr}_{\boldsymbol{K}}\left\{\exp\left(\frac{N}{2}\ln|\boldsymbol{K}|+\alpha N\mathcal{G}_E(\boldsymbol{K};\beta,\eta_0)\right)\right\}$$

where:

$$\exp[\mathcal{G}_E(\boldsymbol{K};\beta,\eta_0)]\equiv\int\frac{\mathrm{d}\hat{\boldsymbol{\lambda}}\,\mathrm{d}\boldsymbol{\lambda}\,\mathrm{d}u}{(2\pi)^{nM+1/2}}\exp\left(g_E(\boldsymbol{Q},\boldsymbol{R},\hat{\boldsymbol{\lambda}},\boldsymbol{\lambda},u;\beta,\eta_0)\right).$$

By imposing the replica symmetric Ansatz and symmetry between agents, i.e. $R^{\gamma}_{a}=R$, $q^{\gamma,\rho}_{a}=q$, $W^{\gamma}_{a,b}=W$, and $t^{\gamma,\rho}_{a,b}=t$ with the assumption that the overlaps W and t satisfy the scaling $\tau\equiv M(W-t)\sim O(1)$ (see reference [12], equation (3)), the logarithm of the determinant of $\boldsymbol{K}$ is:

$$\ln|\boldsymbol{K}|=nM\left[\ln(1-q)+\frac{q-W}{1-q}+\frac{W-R^2}{1-q+\tau}+O(n)\right].\tag{A1}$$

By defining the function $B(x;\beta,\eta_0)\equiv\exp\left(\sqrt{2\beta\eta_0}x-\beta\right)$ and performing the integrals over the variables $\hat{\lambda}^{\gamma}_{a}$ and λ^{γ}_{a}, we have that

$$\begin{aligned}\exp[\mathcal{G}_E(\boldsymbol{K};\beta,\eta_0)] &= 2\int_0^{\infty}\mathcal{D}u\int\mathcal{D}w\int\prod_a\mathcal{D}w_a\left\{\int\mathcal{D}x\mathcal{D}s\prod_a[B+(1-B)\mathcal{H}(-y_a)]\right\}^n\\ &\approx 2\int_0^{\infty}\mathcal{D}u\int\mathcal{D}w\int\prod_a\mathcal{D}w_a\left\{\int\mathcal{D}x\mathcal{D}s\,[B+(1-B)\mathcal{H}(-\overline{y})]^M\right\}^n\\ &\approx\sqrt{\frac{2}{\pi}\frac{1-q}{W}}\int\mathrm{d}z\,\exp\left(-\frac{1-q}{W}\frac{z^2}{2}\right)\mathcal{H}\left(-\sqrt{\frac{1-q}{W(W-R^2)}}Rz\right)\\ &\quad\left\{\sqrt{\frac{M(1-q)}{2\pi\tau}}\int\mathcal{D}x\,\mathrm{d}\sigma\,\exp\left(-M\frac{1-q}{\tau}\frac{(\sigma-z)^2}{2}\right)[B+(1-B)\mathcal{H}(-\sigma)]^M\right\}^n\end{aligned}$$

where the intermediate step has used the average variable:

$$\overline{y}\equiv\frac{Ru+\sqrt{t-R^2}w+\sqrt{q-t}M^{-1}\sum_a w_a+\sqrt{W-t}s}{\sqrt{1-q+t-W}},$$

$\mathcal{D}x\equiv(2\pi)^{-1/2}\mathrm{d}x\,\mathrm{e}^{-x^2/2}$ is the Gaussian measure and $\mathcal{H}(x)\equiv\int_x^{\infty}\mathcal{D}y$ is the Gardner error function. In order to keep the extensivity of the energy (1) we will impose the scaling $H_0\equiv M\eta_0\sim O(1)$. For a large enough population size M we can use the Gaussian approximation for the Binomial factor, solve the integrals in σ and x by the Laplace's method and expand for small n:

$$\exp\left[\mathcal{G}_E(\boldsymbol{K};\beta,\eta_0)\right] \approx 1 - 2nM\sqrt{\frac{1-q}{W}}\int\frac{\mathrm{d}z}{\sqrt{2\pi}}\exp\left(-\frac{1-q}{W}\frac{z^2}{2}\right)\mathcal{H}\left(-\sqrt{\frac{1-q}{W(W-R^2)}}Rz\right)\times$$
$$\times\min_{u\in(0,1),\sigma\in\mathbb{R}}\left\{\frac{1-q}{\tau}\frac{(\sigma-z)^2}{2}+\frac{(u-\mathcal{H}(\sigma))^2}{2\mathcal{H}(\sigma)\mathcal{H}(-\sigma)}-u^2\beta H_0+u\beta\right\}+O(n^2). \tag{A2}$$

The factor between curly brackets at the RHS of (A2) emerges from the interaction between agents and is the responsible for the fragmentation of the phase space observed in the following. For sufficiently small values of τ the minimum of (A2) is achieved at $\sigma = z$. The remaining problem corresponds to the minimization of a quadratic polynomial in $u \in [0,1]$, for which the solution is either the minimum of the parabola:

$$u_0 = \frac{[1-\beta\mathcal{H}(-z)]\mathcal{H}(z)}{1-2\beta H_0\mathcal{H}(z)\mathcal{H}(-z)} \tag{A3}$$

if the factor of the quadratic component is positive, i.e. $1-2\beta H_0\mathcal{H}(z)\mathcal{H}(-z) > 0$ and if $0 < u_0 < 1$, or the border of the interval, i.e. $u = 0, 1$. Consider $\mathcal{H}^{-1}(x)$ the inverse of the Gardner error function. We found that, by defining the quantities:

$$a_1 \equiv -\mathcal{H}^{-1}\left(\Theta(2H_0-1)\max\left\{0,\frac{\beta(2H_0-1)-1}{\beta(2H_0-1)}\right\}\right) \tag{A4}$$
$$a_2 \equiv -\mathcal{H}^{-1}\left(\min\left\{1,\frac{1}{\beta}\right\}\right) \tag{A5}$$
$$a_3 \equiv -\mathcal{H}^{-1}\left(\frac{1}{2}-\frac{\sqrt{\beta^2(1-H_0)^2+1}-1}{2\beta(1-H_0)}\right) \tag{A6}$$
$$b_0 \equiv \Theta(a_2-a_1)a_2+\Theta(a_1-a_2)a_3 \tag{A7}$$
$$b_1 \equiv \Theta(a_2-a_1)a_1+\Theta(a_1-a_2)a_3 \tag{A8}$$

we observe that if $b_1 < z < b_0$ the minimum is achieved at $u = u_0$ (A3), if $b_0 < z$ the minimum is achieved at $u = 0$ and if $z < b_1$ the minimum is achieved at $u = 1$. The solution to the minimization problem, in zeroth order in τ, is then:

$$\Phi(z;\beta,H_0) \equiv \lim_{\tau\to 0}\min_{u\in(0,1),\sigma\in\mathbb{R}}\left\{\frac{1-q}{\tau}\frac{(\sigma-z)^2}{2}+\frac{(u-\mathcal{H}(\sigma))^2}{2\mathcal{H}(\sigma)\mathcal{H}(-\sigma)}-u^2\beta H_0+u\beta\right\}$$
$$= \begin{cases}\Phi_1 \equiv \frac{\mathcal{H}(-z)}{2\mathcal{H}(z)}+\beta(1-H_0) & z < b_1\\ \Phi_{u_0} \equiv \frac{\beta\mathcal{H}(z)[1-H_0\mathcal{H}(z)]}{1-2\beta H_0\mathcal{H}(z)\mathcal{H}(-z)}-\frac{\beta^2\mathcal{H}(z)\mathcal{H}(-z)}{2[1-2\beta H_0\mathcal{H}(z)\mathcal{H}(-z)]} & b_1 < z < b_0\\ \Phi_0 \equiv \frac{\mathcal{H}(z)}{2\mathcal{H}(-z)} & b_0 < z.\end{cases} \tag{A9}$$

$\Phi(z;\beta,H_0)$ is continuous in z but not differentiable at the boundaries defined in equations (A7) and (A8). In the plane defined by the independent parameters β and H_0 the components Φ_{z_0}, Φ_0 and Φ_1 cover the areas illustrated in figure 1. Observe that the component Φ_{z_0} appears in the sector

$\mathscr{S}_{z_0} \equiv \{(\beta, H_0)|\beta \leq 1 \,\mathrm{and}\, H_0 \geq 0\} \cup \{(\beta, H_0)|\beta > 1 \,\mathrm{and}\, 2H_0 < \beta/(\beta-1)\}$, the component Φ_1 appears in the sector $\mathscr{S}_1 \equiv \{(\beta, H_0)|\beta \geq 0 \,\mathrm{and}\, 2H_0 > (1+\beta)/\beta\}$ and the component Φ_0 appears in the sector $\mathscr{S}_0 \equiv \{(\beta, H_0)|\beta \geq 1 \,\mathrm{and}\, H_0 \geq 0\}$.

Appendix B: Stability of the solution $r^2 = w$

To explore the stability of the solution (16) we analyze the sign of the eigenvalues of the matrix of second derivatives $[\partial^2_{\gamma,\delta}\phi]$. The second derivatives of ϕ with respect to r and w are:

$$\begin{aligned}
\partial^2_{r,r}\phi &= 1 - \sqrt{\frac{2}{\pi}}\alpha r \int \mathrm{d}z\, \mathcal{N}(z|0, w-r^2)\frac{\partial^3\Phi(z;\beta,H_0)}{\partial z^3} \\
\partial^2_{r,w}\phi &= \frac{\alpha}{\sqrt{2\pi}} \int \mathrm{d}z\, \mathcal{N}(z|0, w-r^2)\frac{\partial^3\Phi(z;\beta,H_0)}{\partial z^3} \\
\partial^2_{w,w}\phi &= \frac{\alpha}{w^2} \int \mathrm{d}z\, \mathcal{N}(z|0, w)\left(\frac{3}{2} - \frac{3z^2}{w} + \frac{z^4}{2w^2}\right)\mathcal{H}(-\kappa z)\Phi(z;\beta,H_0) + \\
&\quad + \frac{\alpha}{2\sqrt{2\pi}}\frac{r(2w-r^2)}{w^3} \int \mathrm{d}z\, \mathcal{N}(z|0, w-r^2)\frac{\partial}{\partial z}\left[\left(1 - \frac{z^2}{w}\right)\Phi(z;\beta,H_0)\right] - \\
&\quad - \sqrt{\frac{2}{\pi}}\alpha\frac{r(r^2-w)}{w^3} \int \mathrm{d}z\, \mathcal{N}(z|0, w-r^2)\frac{\partial\Phi(z;\beta,H_0)}{\partial z} + \\
&\quad + \frac{\alpha}{2\sqrt{2\pi}}\frac{r(2w-r^2)}{w^2} \int \mathrm{d}z\, \frac{\mathcal{N}(z|0, w-r^2)}{w-r^2}\left(1 - \frac{z^2}{w-r^2}\right)\frac{\partial\Phi(z;\beta,H_0)}{\partial z}.
\end{aligned}$$

The evaluation of these derivatives at the solution (16) produces the entries of the Hessian matrix at the critical point:

$$h_{r,r} \approx 1 + \frac{\pi}{2}\frac{[\Phi^{(2)}(\beta,H_0)]^2}{\Phi^{(1)}(\beta,H_0)\Phi^{(3)}(\beta,H_0)} \tag{B1}$$

$$h_{r,w} = h_{w,r} \approx \sqrt{\frac{\pi}{8}}\frac{\Phi^{(2)}(\beta,H_0)}{\Phi^{(1)}(\beta,H_0)} \tag{B2}$$

$$h_{w,w} \approx \frac{\pi}{10}\frac{\Phi^{(2)}(\beta,H_0)\Phi^{(4)}(\beta,H_0)}{\Phi^{(1)}(\beta,H_0)\Phi^{(3)}(\beta,H_0)}. \tag{B3}$$

By numerical calculations we found that the Hessian matrix, with entries (B1), (B2) and (B3), possess two positive eigenvalues for all values of β and H_0 with the exception of a small neighborhood of the point $\beta = 0$, $H_0 = 1$, and inside the region described by $\mathcal{B}(\beta, H_0)$ (19), where the proposed solution $r_t^2 = w_t$ is not suitable.

[1] J. Torok, G. Iñiguez, T. Yasseri, M. San Miguel, K. Kaski and J. Kertesz, Phys. Rev. Lett. **110**, 088701 (2013).

[2] S. Galam, Y. Gefen and Y. Shapir, Mathematical Journal of Sociology **9**, 1(1982).

[3] S. Galam and S. Moscovici, European Journal of Social Psychology **21**, 49 (1991).

[4] J. Neirotti, Phys. Rev. E **94**, 012309 (2016).

[5] B. Latane, American Psychologist **36**, 343 (1981).

[6] M. Lewenstein, A. Nowak and B. Latane, Phys. Rev. A **45**, 763 (1992).

[7] K. Kacperski and J. A. Holyst, J. Stat. Phys. **84**, 169 (1996).

[8] J. Fernandez-Gracia, K. Suchecki, J. J. Ramasco, M. San Miguel and V. M. Eguiluz, Phys. Rev. Lett. **112**, 158701 (2014).

[9] A. Engel and C. Van den Broeck, Statistical mechanics of learning, Cambridge: CUP (2001).

[10] E. Gilbert, T. Bergstrom and K. Karahalios, Proceedings of the Hawaii International Conference on System Sciences, ed. R. J. Sprague, IEEE Computer Society, Washington DC, 1 (2009).

[11] M. Mezard, G. Parisi and M. A. Virasoro, Spin glasses theory and beyond, World Scientific (1987).

[12] J. P. Neirotti and L. Franco, J. Phys. A **43**, 445103 (2010).

[13] C. Castellano, S. Fortunato and V. Loretto, Rev. Mod. Phys. **81**, 591 (2009).

[14] R. Vicente, A. Susemihl, J. P. Jerico and N. Caticha, Physica A **400**, 124 (2014).

[15] J. Graham, J. Haidt and B. A. Nosek, J. of Personality and Social Psichology **96**, 1029 (2009).

[16] A. J. Healy, N. Malhotra and C. H. Mo, Proc. Natl. Acad. Sci. USA **107**, 12804 (2010).

[17] R. Vicente, A. C. R. Martins and N. Caticha, J. Stat. Mech. P03015 (2009).

[18] P. P. Li, D. F. Zheng and P. M. Hui, Phys. Rev. E **73**, 056128 (2006).

[19] G. J. Baxter, R. A. Blythe and A. J. McKane, Phys. Rev. Lett. **101**, 258701 (2008).

[20] P. L. Krapivsky and S. Redner, Phys. Rev. Lett. **90**, 238701 (2003).

[21] A. Soulier and T. Halpin-Healy, Phys. Rev. Lett. **90**, 258103 (2003).

[22] J. Shao, S. Havlin and H. E. Stanley, Phys. Rev. Lett. **103**, 018701 (2009).

[23] A. D. Sanchez, J. M. Lopez and M. A. Rodriguez, Phys. Rev. Lett. **88**, 048701 (2002).

[24] T. Gross, Carlos J. D. D'Lima and B. Blasius, Phys. Rev. Lett. **96**, 208701 (2006).

[25] C. Nardini, B. Kozma and A. Barrat, Phys. Rev. Lett. **100**, 158701 (2008).

[26] G. Toscani, Comm. Math. Sci. **4**, 481 (2006).

[27] S. Fortunato and C. Castellano, Phys. Rev. Lett. **99**, 138701 (2007).

[28] The relationship of the quantities that set the size of the replicated system (N, M and n) is as follows: $1 \ll M \ll N$ and $nNM \ll 1$.

[29] Observe that $\mathcal{A}_1 + \mathcal{A}_{u_0} + \mathcal{A}_0 = 1$ and $\mathcal{A}_1\mathcal{A}_{u_0}\mathcal{A}_0 = 0$ for all (β, H_0).