Emergence of polarization in a voter model with personalized information

Giordano De Marzo ●

Dipartimento di Fisica Università "Sapienza" P. le A. Moro, 2, I-00185 Rome, Italy

Andrea Zaccaria ●

Istituto dei Sistemi Complessi (ISC-CNR), UoS Sapienza, P. le A. Moro, 2, I-00185 Rome, Italy

Claudio Castellano ●*

Istituto dei Sistemi Complessi (ISC-CNR), via dei Taurini 19, I-00185 Rome, Italy

The flourishing of fake news is supported by recommendation algorithms of online social networks, which, based on previous user activity, provide content adapted to their preferences and so create filter bubbles. We introduce an analytically tractable voter model with personalized information, in which an external field tends to align the agent's opinion with the one she held more frequently in the past. Our model shows a surprisingly rich dynamics despite its simplicity. An analytical mean-field approach, confirmed by numerical simulations, allows us to build a phase diagram and to predict if and how consensus is reached. Remarkably, polarization can be avoided only for weak interaction with personalized information and if the number of agents is below a threshold. We compute analytically this critical size, which depends on the interaction probability in a strongly nonlinear way.

I. INTRODUCTION

The way that our opinions are formed and change over time, giving rise to emerging collective phenomena, is a topic that has attracted rapidly increasing interest. Apart from its clear relevance for fundamental issues, such as the stability of democracy and the preservation of individual liberties, opinion dynamics is a paradigmatic example of a social phenomenon that can be studied quantitatively using a combination of theoretical and empirical approaches [1–5].

Traditional models for opinion dynamics focus on the interaction among a large number of peers, often in the presence of external fields—possibly varying over time but equal for all agents [6]—describing the effect of conventional media, such as television and the press, acting in the same way on all individuals. The advent of the Internet, and the online social network (OSN) revolution in particular, have made the scenario more complex. Although in principle the Internet allows users to access an unprecedented diversity of news and viewpoints, this abundance is overwhelming and pushes users to rely on automated recommendation systems to cope with information overload. To provide successful recommendations, web sites and social networks constantly track our online activity, and based on it, they propose personalized "information," which varies not only over time but also from person to person. Examples of this phenomenon are posts appearing on Facebook's news feed, which are chosen and ordered by the social network on the basis of previous interactions with other posts, or Google's Personalized PageRank. The dependence on past user behavior coupled with the human inclination to favor sources confirming one's own preferences [7] creates a feedback loop that drastically reduces the huge diversity of available content.

The selective exposure of individuals to a biased representation of the world (so that they remain confined within their "filter-bubble") is thought to play a crucial role in shaping opinions at all scales [8]. A most worrying aspect of this mechanism is the possible reinforcement of personal biases with the consequence of favoring radicalization phenomena [9]. An astonishing example in this sense is the recent revelation that 64% of people who join extremist groups on Facebook are recommended to do so by the Facebook algorithm itself [10].

It is therefore crucial to properly understand the effect of personalized information or advertising on the dynamics of opinions. Some modeling efforts in this direction have already been made. Perra and Rocha [11] studied a binary opinion dynamics where each user updates her opinion based on the opinion of others filtered in various ways. In the framework of continuous opinion dynamics with bounded confidence [2], in Ref. [12] the effect of recommender algorithms in OSN is mimicked by enhancing the probability to interact with individuals having close opinions. An increased tendency toward fragmentation (no consensus) and polarization (clusters with

distant opinions) is observed. A model for polarization in continuous dynamics has been recently presented and compared with empirical results by Baumann *et al.* [13].

In this paper, we consider arguably the simplest possible type of opinion dynamics, namely the voter model, and we study in detail its behavior in the presence of external personalized information, modeled, in its turn, in an extremely simple way. The fundamental questions we want to answer are whether, in this simple setting, selective exposure prevents the reaching of consensus and how this comes about. Is a minimal amount of personalized information sufficient to lead to a polarized state? Is consensus possible for a very strong influence of personalized information on opinions? On which timescale is consensus, if any, reached? What is the role of the system size? By means of an analytical approach, corroborated by numerical simulations, we fully understand the model behavior and thus get complete answers to all these questions.

The rest of the paper is organized as follows. In Sec. II we introduce the voter model modified by the presence of personalized information. Its analytical investigation is described in Sec. III, divided into three subsections, dealing with different values of the parameter c. The final section summarizes and discusses the results and presents some perspectives. Several Appendixes contain details of the analytical calculations.

II. THE VOTER MODEL WITH PERSONALIZED INFORMATION

In voter dynamics [14,15], individuals are endowed with a binary opinion (spin) $s_i = \pm 1$; at each time step a spin is randomly extracted and its value is replaced with the value of one of the spins to which it is connected. In other words, an individual becomes equal to a randomly chosen neighbor. The model has been studied extensively, both in regular lattices and on complex networks [2,16–21]. It is well known that this dynamics always leads in finite systems to full consensus—i.e., all spins get aligned after a certain amount of time—and that this is due only to stochastic fluctuations [22].

To understand the most basic effects of personalized information on opinion dynamics, we couple the voter model to a simple source of personalized information, which feeds back on each individual a signal depending on the past evolution of her own opinion. Other recent works have investigated the effect of opposing (but fixed) sources of external information on voter dynamics [23,24]. Our model is also similar to a voter dynamics with aging, recently introduced by Peralta *et al.* [25]. The main difference between that model and the present one is that the effective memory in our model is never erased, while it is, when a spin flips, in the model of Ref. [25].

Let us consider N agents distributed over the nodes i of a network. Each agent can assume two states $s_i = \pm 1$ that correspond to two different opinions, and each agent interacts with the agents to which she is connected. We define the adjacency matrix A_{ij} so that $A_{ij} = 1$ if spin i and j are linked and $A_{ij} = 0$ otherwise. With this convention, the number of agents a given spin s_i is connected to is simply $k_i = \sum_j A_{ij}$. The evolution of each individual depends also on another variable, a "personalized external information" e_i. This last quantity is a random variable $e_i = \pm 1$ assuming a positive
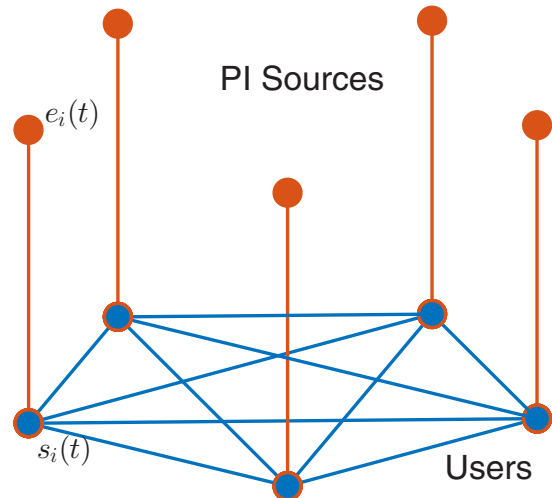

FIG. 1. Voter model with personalized information. Graphical representation of the voter model with personalized information. Blue circles represent the agents s_i that interact following the usual voter dynamics, while red circles are the external agents, carrying personalized information. Note that the external agent e_i influences only the corresponding voter spin, namely s_i.

value with a probability $P[e_i(t) = +1]$ that changes over time depending on the history of the agent's opinion.

The dynamics takes place as follows. Initially each spin is set to $s_i = \pm 1$ with equal probability. At each time step, a given individual i is selected at random and, with probability $1 - \lambda$, she follows the usual voter dynamics: Her state s_i is made equal to the state of a randomly selected neighbor s_j. With complementary probability λ, the individual copies the state of the external source:

$$s_i(t + \delta t) = \begin{cases} e_i(t) & \text{with probability } \lambda, \\ s_j(t) & \text{with probability } \frac{1-\lambda}{k_i}, \end{cases} \quad (1)$$

where $\delta t = 1/N$, j is one of the neighbors of i (i.e., $A_{ij} = 1$), and N_i is the total number of such neighbors. Pictorially, we are adding another layer of "external agents" $e_i(t)$, each of them coupled only to the original agent $s_i(t)$ and influencing her in the same way as the other neighbors, except for a different probability of interaction. See Fig. 1 for a graphical representation of the model.

To mimic the reinforcing effect of personalized information, we assume that, whenever the spin i is selected for update, then $P[e_i(t) = +1]$ changes, increasing the probability that in the future e_i will be equal to the current state of the agent, s_i.

More precisely, at each time step the update of the probability occurs after the update of the opinion variable. In other words, one first updates the opinion variable (which may or may not change) and afterwards increases by a factor c the ratio between $P[e_i = s_i(t + \delta t)]$ and $P[e_i = -s_i(t + \delta t)]$, where $s_i(t + \delta t)$ is the opinion variable after the update:

$$\frac{P[e_i = s_i(t + \delta t)]}{P[e_i = -s_i(t + \delta t)]} \rightarrow c \frac{P[e_i = s_i(t + \delta t)]}{P[e_i = -s_i(t + \delta t)]}.$$
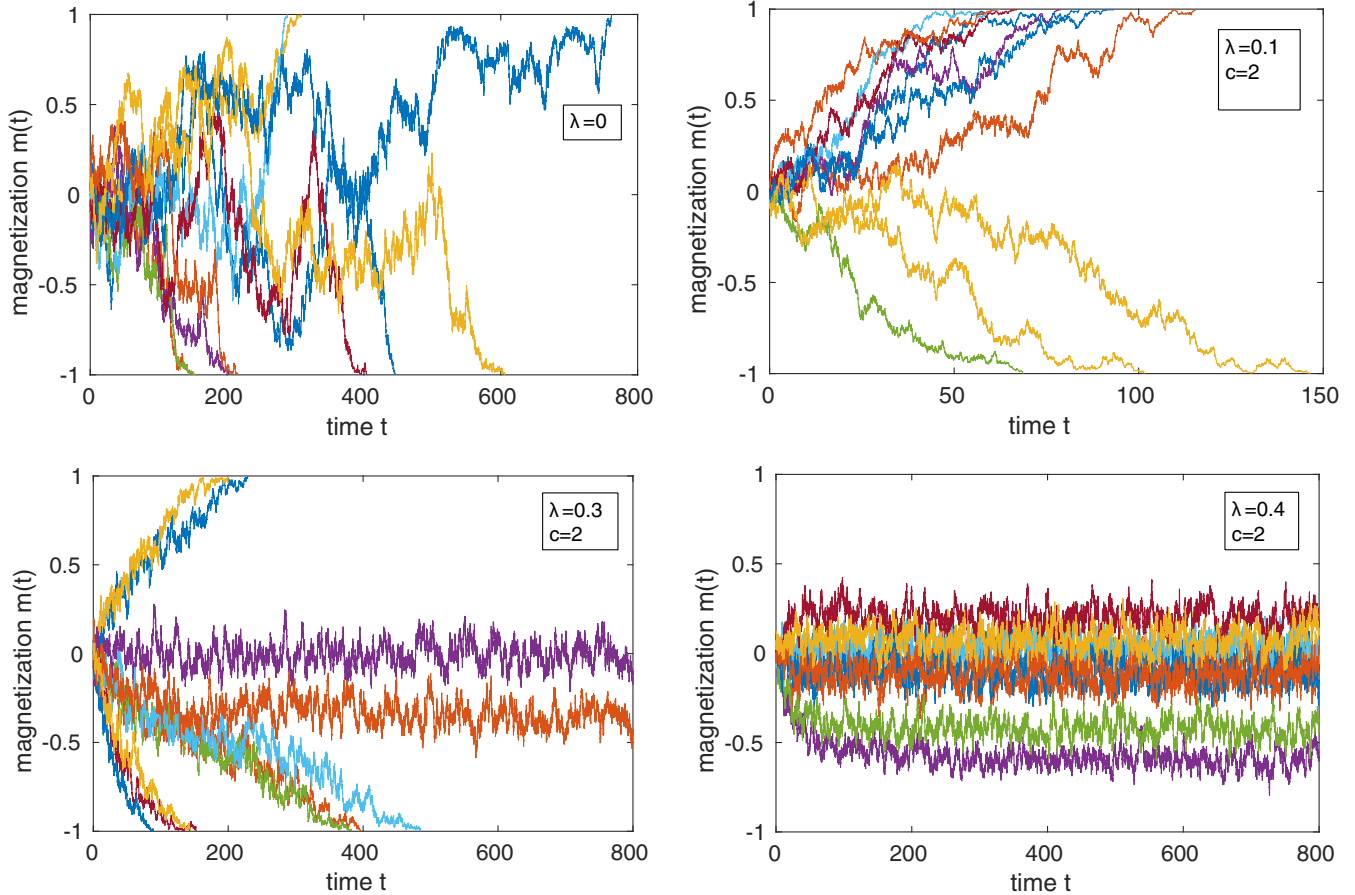
FIG. 2. From consensus to polarization. Temporal evolution of the magnetization $m(t)$ for different values of the probability λ, starting from $\lambda = 0$ (pure voter dynamics). Agents form a complete graph of size $N = 500$. The trajectories of 10 different runs are displayed.

In this way, the time-depending probability $P(e_i = +1)$ keeps track of the history of the spin s_i. For example, if agent i stays in state $s_i = +1$ for a long time, then $P(e_i = +1)$ tends to grow toward 1 and this makes it more likely that opinion $s_i = +1$ is maintained. The polarizing effect of this personalized source of information is clear. The parameter c determines the speed at which the balance between the two alternatives is disrupted. Notice that the change for the probability $P(e_i = +1)$ occurs at each update of agent i, even if the latter does not actually change opinion (because the agent interacts with a neighbor already sharing the same state).

It is useful to define the quantity

$$n_i(t) = \sum_{t'=1}^{t} s_i(t'),$$

which keeps the memory of the evolution of agent i's opinion. Assuming that initially no knowledge about the agent's preferences is available and therefore external information is fully random, $P[e_i(0) = +1] = 1/2$, we can write

$$P[e_i(t) = 1] = P[n_i(t)] = \frac{c^{n_i(t)}}{1 + c^{n_i(t)}}. \tag{2}$$

Hence, a positive (negative) value of n_i implies that personalized information is more probably equal to $e_i = +1$ ($e_i = -1$).

To give a qualitative idea of the model behavior, we report in Fig. 2 the temporal evolution of the magnetization $m(t) = \sum_i s_i/N$ for different values of the probability λ of interaction with the personalized information. For $\lambda = 0$ the system is exactly the usual voter dynamics and it reaches consensus $m = \pm 1$ because of random diffusive fluctuations, in a time of order N [22]. As personalized information is turned on, consensus is still reached, but surprisingly over *shorter* time intervals, and it is clear that drift plays now a relevant role. Increasing λ further, we observe that some runs do not reach consensus anymore and magnetization fluctuates around some constant value. Finally, for large λ consensus is never reached and all runs remain stuck in a disordered state. In the next section, through a mean-field analytical approach, we understand when and how consensus is reached, depending on the values of the parameters c and λ.

III. ANALYTICAL RESULTS

The evolution of a system following Eqs. (1) and (2) depends on the topology of the network defining the interactions among spins. In the following we will focus on complete graphs, for which each pair of spins is equally likely to interact, meaning that $A_{ij} = 1 - \delta_{ij}$ for any (i, j). This corresponds, in the absence of external information, to the mean-field limit of the voter model. We denote by $N_\uparrow$ the

number of spins in state $+1$, while $N_\downarrow = N - N_\uparrow$ is the number of spins in the opposite state. In these terms, the updating process is

$$s_i(t + \delta t) = \begin{cases} e_i(t) \text{ with probability } \lambda, \\ +1 \text{ with probability } (1 - \lambda)\frac{N_\uparrow}{N}, \\ -1 \text{ with probability } (1 - \lambda)\frac{N_\downarrow}{N}, \end{cases} \quad (3)$$

where the probability of positive external information is given by Eq. (2). Recalling that the magnetization can be written as

$$m = \frac{N_\uparrow - N_\downarrow}{N},$$

we obtain from Eq. (3) that each time a node is selected, its value s_i evolves according to

$$s_i(t) \rightarrow s_i(t + \delta t) = \begin{cases} +1 \text{ with prob: } (1-\lambda)\left(\frac{1+m}{2}\right) + \lambda P(n_i), \\ -1 \text{ with prob: } (1-\lambda)\left(\frac{1-m}{2}\right) + \lambda[1 - P(n_i)], \end{cases} \quad (4)$$

and, immediately after, n_i is updated as follows:

$$n_i(t) \rightarrow n_i(t + \delta t) = \begin{cases} n_i(t) + 1 \text{ if } s_i(t + \delta t) = 1, \\ n_i(t) - 1 \text{ if } s_i(t + \delta t) = -1. \end{cases} \quad (5)$$

Thus the state of each node is defined by the pair (s_i, n_i), and therefore the evolution of the system depends on the set $\{(s_i, n_i)\}_{i=1}^N$. In Appendix A we calculate the drift and diffusion coefficients [22] for the magnetization and for the average value of n_i,

$$n = \frac{1}{N}\sum_i n_i. \quad (6)$$

We obtain the magnetization drift

$$v^m = 2\frac{\lambda}{N}\sum_i \left\{ \left(\frac{1-s_i}{2}\right)P(n_i) - \left(\frac{1+s_i}{2}\right)[1 - P(n_i)] \right\} \quad (7)$$

and the magnetization diffusion coefficient

$$D^m = \frac{1-\lambda}{N}(1 - m^2)$$
$$+ \frac{2\lambda}{N^2}\sum_i \left\{ \left(\frac{1-s_i}{2}\right)P(n_i) + \left(\frac{1+s_i}{2}\right)[1 - P(n_i)] \right\}. \quad (8)$$

For the quantity n, the drift coefficient reads

$$v^n = (1 - \lambda)m + \frac{\lambda}{N}\sum_i [2P(n_i) - 1] \quad (9)$$

while the diffusion coefficient is

$$D^n = \frac{1}{N}. \quad (10)$$

These expressions contain $P(n_i)$ and therefore they are different depending on the value of c.

A. The case $c = 1$

Let us first discuss the case $c = 1$, which is equivalent to the noisy voter model or the Kirman model [26]. In this case,

Eq. (2) reduces to

$$P(n_i) = \frac{1}{2} \; \forall \; i, t$$

and therefore the variables n_i do not play any role. Setting $c = 1$, Eqs. (7) and (8) reduce to

$$\begin{cases} v_{c=1}^m = -\lambda m, \\ D_{c=1}^m = \frac{1}{N}[(1 - \lambda)(1 - m^2) + 2\lambda]. \end{cases} \quad (11)$$

Differently from the standard voter model, there is a nonzero drift term, driving the system toward the disordered symmetric configuration $m = 0$. However, depending on the value of λ, the system may still spend most of its time in the consensus state $m = \pm 1$, which is no longer absorbing. See Refs. [27–29] for a detailed analysis of the noisy voter model.

B. The behavior for $c \gtrsim 1$

We now study the behavior for $c > 1$, considering separately two cases. We set $c = 1 + \delta$ and first take $\delta \ll 1$. Under this hypothesis and focusing on short times, we can expand $P(n_i)$ to first order in $n_i\delta$, obtaining

$$P(n_i) = \frac{c^{n_i}}{1 + c^{n_i}} = \frac{(1 + \delta)^{n_i}}{1 + (1 + \delta)^{n_i}} \approx \frac{1}{2} + \frac{n_i\delta}{4}. \quad (12)$$

Inserting this expression into Eq. (7), we get

$$v^m \approx \lambda\left(\frac{n\delta}{2} - m\right), \quad (13)$$

and analogously for the diffusion coefficient

$$D^m \approx \frac{(1-\lambda)}{N}(1 - m^2) + \frac{2\lambda}{N^2}\left(1 - \frac{\delta}{2N}\sum_i s_i n_i\right). \quad (14)$$

Inserting the expansion of $P(n_i)$ into Eq. (9), we obtain

$$v^n \approx (1 - \lambda)m + \frac{\lambda\delta}{2}n. \quad (15)$$

In summary, combining Eqs. (13) and (15), the evolution of the system is given, as long as the condition $|n_i|\delta \ll 1$ is satisfied for any i, by

$$\begin{cases} \dot{m} = -\lambda m + \frac{\lambda\delta}{2}n, \\ \dot{n} = (1 - \lambda)m + \frac{\lambda\delta}{2}n, \end{cases} \quad (16)$$

where fluctuations due to diffusion have been neglected. By integrating we find, under the assumption $\delta \ll \lambda$,

$$\begin{cases} m = C_1\frac{\delta}{2}e^{t\delta/2} - C_2\frac{\lambda}{1-\lambda}e^{-\lambda t}, \\ n = C_1 e^{t\delta/2} + C_2 e^{-\lambda t}, \end{cases} \quad (17)$$

where C_1 and C_2 are determined by the initial conditions. We note that the deterministic evolution described by Eq. (16) is preceded by a regime dominated by stochastic effects, where we can effectively assume $c = 1$. During such an interval, m fluctuates around zero due to the presence of the term $-\lambda m$ in its drift, with fluctuations of the order of $\pm m_0 = \pm\sqrt{D^m} \approx \pm\sqrt{(1-\lambda)/N}$. Conversely n grows diffusively, up to the time $\tau = 2/\delta$, after which the exponential growth becomes dominant. Moreover, after a short time of order $1/\lambda$ the terms proportional to C_2 in Eqs. (17) become negligible. As a consequence, we can use as the initial condition for n its
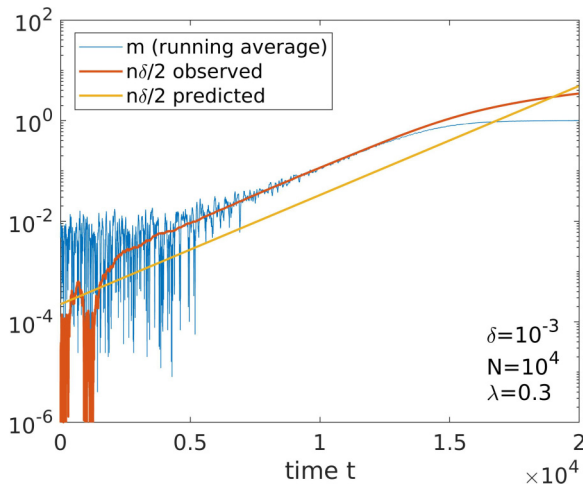
FIG. 3. Growth of n and m. Growth of m compared with the observed $n\delta/2$ and with the predicted trend, as defined by Eq. (18). n grows diffusively up to $\tau = 2/\delta$, the time at which the exponential contribution becomes dominant.

value at time τ, i.e., at the end of the diffusive regime, yielding $C_1 \approx \sqrt{\tau D_n} \approx \sqrt{\frac{2}{\delta N}}$,

$$\begin{cases} n \approx \pm\sqrt{\frac{2}{\delta N}}e^{t\delta/2}, & \text{for } t > \tau. \\ m = \frac{\delta}{2}n, \end{cases} \quad (18)$$

Note that the exponential growth of m actually begins only when $\frac{\delta}{2}|n| \sim m_0 \approx \sqrt{(1-\lambda)/N}$. Figure 3 shows that Eq. (18) describes well this stage of the temporal evolution of n and m.

The linear approximation remains valid until the time T_c when, for some i, $|n_i|$ becomes so large that the condition $|n_i|\delta \ll 1$ breaks down. This may happen in two different ways, depending on the shape of the probability distribution $Q(n_i)$ of the variables n_i. When the linear approximation breaks down, if the standard deviation of $Q(n_i)$ is small, the personalized information is approximately the same for all individuals. As a consequence, this source of information can be regarded as a constant field, and this results in the presence of a drift for the magnetization, which quickly reaches $m = \pm 1$. Conversely, if the standard deviation is much larger than the average $\langle n_i \rangle$, some of the spins are characterized by a negative n_i, while n_i is positive for the others. This implies that a fraction of the spins is influenced by a positive external field, while a negative field acts on the remaining part, thus leading to a polarized state. More detail, if $Q(n_i)$ is narrowly peaked around its mean value n over time, linearization breaks down for a time T such that $|n(T)|\delta \sim 1$. From Eq. (18),

$$T = \frac{1}{\delta}\ln\left(\frac{N}{2\delta}\right). \quad (19)$$

At this time, all n_i have the same sign, hence the drift (see Appendix A)

$$v_i = m(T)(1-\lambda) + \lambda[2P(n_i) - 1] \quad (20)$$

has the same sign for all individuals and thus consensus is rapidly reached.

Alternatively, if the distribution is very broad so that its standard deviation σ is much larger than the absolute mean value $|n|$, linearization starts to fail at a different time T^*. Assuming, for the sake of simplicity, $n > 0$, this occurs when

$$\delta[n + \sigma(T^*)] \approx \delta\sigma(T^*) = 1. \quad (21)$$

The calculation of the variance $\sigma^2(t)$ of the distribution $Q(n_i)$, reported in Appendix B, gives

$$\sigma^2(t) = \frac{1}{\lambda\delta}(e^{\lambda\delta t} - 1). \quad (22)$$

Imposing $\delta\sigma(T^*) = 1$, we then obtain

$$T^* = \frac{1}{\lambda\delta}\ln\left(1 + \frac{\lambda}{\delta}\right). \quad (23)$$

The time T_c when the linear approximation breaks down consequently is

$$T_c = \min(T, T^*).$$

Since T grows with N, while T^* does not depend on it, for small size $T < T^*$ and the opposite relationship $T > T^*$ is instead true for large N. Setting $T = T^*$ we can compute the crossover size

$$N^* = 2\delta\left(1 + \frac{\lambda}{\delta}\right)^{1/\lambda}. \quad (24)$$

For $N < N^*$ linearization breaks down due to the growth of n, and as a consequence consensus is always reached, all the drifts having the same sign. In contrast, if $N > N^*$ the end of the linear regime is caused by the growth of the variance. In this second case, at T^* most of the individuals have positive n_i and hence a positive drift, but some have negative n_i (see Fig. 4). Determining in this case whether consensus is reached or not is more involved, as discussed in the following.

Assuming $N \gg N^*$, so that the standard deviation is much larger than the mean value, the smallest of the negative values is

$$n_i \approx n - \sigma(T^*) \approx -\sigma(T^*) \approx -\frac{1}{\delta}, \quad (25)$$

and as a consequence the smallest drift is

$$v_i = m(T^*)(1-\lambda) + \lambda[2P(n_i) - 1] \approx m(T^*)(1-\lambda) - \lambda. \quad (26)$$

If this value is positive, the corresponding individual, which is the one whose external information is more negatively polarized, will be pushed toward positive values of n_i. It then follows that also in this case the system reaches consensus. More quantitatively, the condition for having consensus, $v_i > 0$, implies (considering also the symmetric case when $n < 0$)

$$|m(T^*)| > \frac{\lambda}{1-\lambda} = m_c. \quad (27)$$

Inserting Eq. (23) into Eq. (18), we obtain

$$|m(T^*)| = \sqrt{\frac{\delta}{2N}}\left(1 + \frac{\lambda}{\delta}\right)^{\frac{1}{2\lambda}}, \quad (28)$$
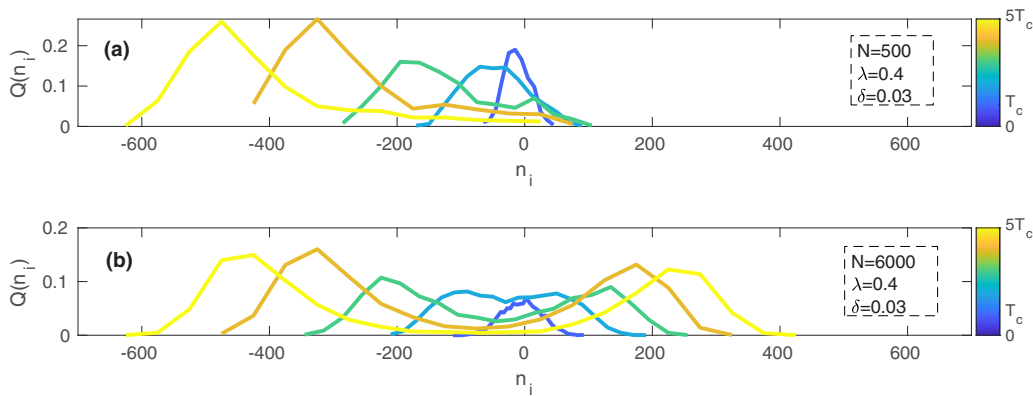
FIG. 4. Evolution of $Q(n_i)$. Evolution of $Q(n_i)$ for $N < N_c$ [panel (a)] and $N > N_c$ [panel (b)].

which, combined with the condition (27), implies that consensus is undoubtedly reached for $\lambda < 1/2$ if $N < \bar{N}$, with

$$\bar{N} = \frac{(1-\lambda)^2}{\lambda^2} \frac{\delta}{2} \left(1 + \frac{\lambda}{\delta}\right)^{\frac{1}{\lambda}}. \tag{29}$$

Note, however, that this is only a lower bound for the true N_c for $\lambda < 1/2$. Actually consensus may occur also for $N > \bar{N}$, provided that $\lambda < 1/2$. Indeed, even if the smallest drift (corresponding to the most negative n_i) is negative when linearization breaks down, it can become positive later on, thus making the system reach consensus. This may occur if only a few spins (among those with $n_i < 0$) have a negative drift. The others, moving toward positive n_i, produce an increase of the magnetization, which eventually overcomes the critical value $m_c = \frac{\lambda}{1-\lambda}$. The critical size N_c, determining if the system can reach consensus or not, is then larger than $\bar{N}$ for $\lambda < 1/2$. It is actually possible to improve on this result. In Appendix C a more refined argument is presented, allowing us to determine numerically a tighter lower bound $\hat{N}$ for N_c. This bound, as shown in Fig. 5, is in good agreement with simulations.
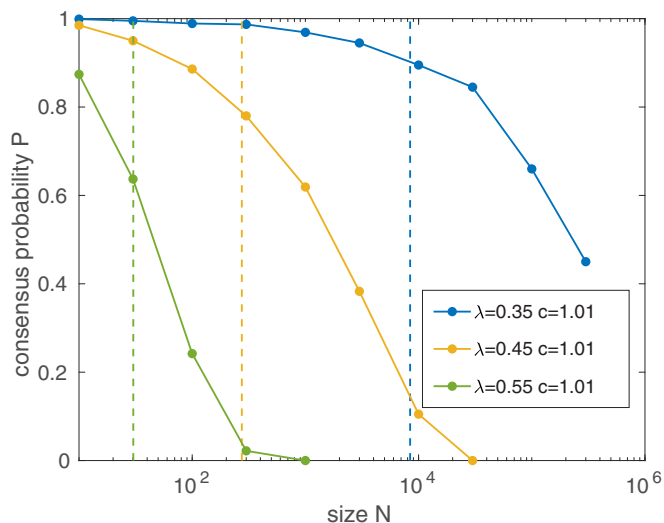

FIG. 5. Crossover between consensus and polarization for small c. Fraction of runs reaching consensus as a function of N, in 1000 realizations of the dynamics for $\delta = 0.01$ and various λ. The crossover values N_c predicted by Eq. (30) are marked by vertical lines.

The situation is different if $\lambda \geqslant 1/2$, because in this case the critical magnetization is larger than 1, and therefore, even if some spins move from negative n_i to positive n_i, the smallest drift remains negative for any magnetization. This implies that for $\lambda \geqslant 1/2$, consensus can be reached only if linearization breaks down due to the growth of n. Therefore, if $\lambda \geqslant 1/2$, the critical size coincides with N^*. In conclusion,

$$\begin{cases} N_c \geqslant \hat{N} & \text{for } \lambda < \frac{1}{2}, \\ N_c = N^* = 2\delta\left(1 + \frac{\lambda}{\delta}\right)^{1/\lambda} & \text{for } \lambda \geqslant \frac{1}{2}. \end{cases} \tag{30}$$

Of course, given the dependence of the argument on random fluctuations, these values are intended to indicate a crossover and not a sharp transition. Simulation results presented in Fig. 5 confirm that the probability of reaching consensus exhibits for various λ a crossover in reasonable agreement with Eq. (30).

C. The behavior for $c \gg 1$

As shown in the previous subsection, if c is close to 1, the system can be described in terms of the magnetization and of the first two moments of the distribution $Q(n_i)$: n and σ^2. This is possible because (see Fig. 4) this distribution is unimodal during the linear regime, which lasts for long times, up to $T_c = \min(T, T^*)$. Conversely, when c is large, the pair (n, σ^2) alone is not sufficient to describe the state of the system, even in the first stage of the dynamics. This can be seen by inspecting the local probabilities $P(n_i)$. If $c \gg 1$, we have

$$P(n_i) = \begin{cases} 1 - c^{-n_i} \approx 1 & \text{if } n_i > 0, \\ \frac{1}{2} & \text{if } n_i = 0, \\ c^n \approx 0 & \text{if } n_i < 0. \end{cases}$$

This means that, as soon as n_i becomes different from zero at the first update of node i, the external information almost certainly has the same sign of n_i and thus tends to increase its absolute value. The behavior of nodes with different signs of n_i is opposite, and this rapidly leads to the splitting of $Q(n_i)$ in two separate peaks. We then write the overall distribution as the sum of two unimodal distributions,

$$Q(n_i) = pQ^+(n_i) + (1-p)Q^-(n_i).$$

Here $Q^+(n_i)$ is the distribution of positive n_i, while $Q^-(n_i)$ is the distribution of the negative ones and the weight is $p = N^+/N$. By using the same formalism described above, in Appendix D we derive expressions for the moments of the
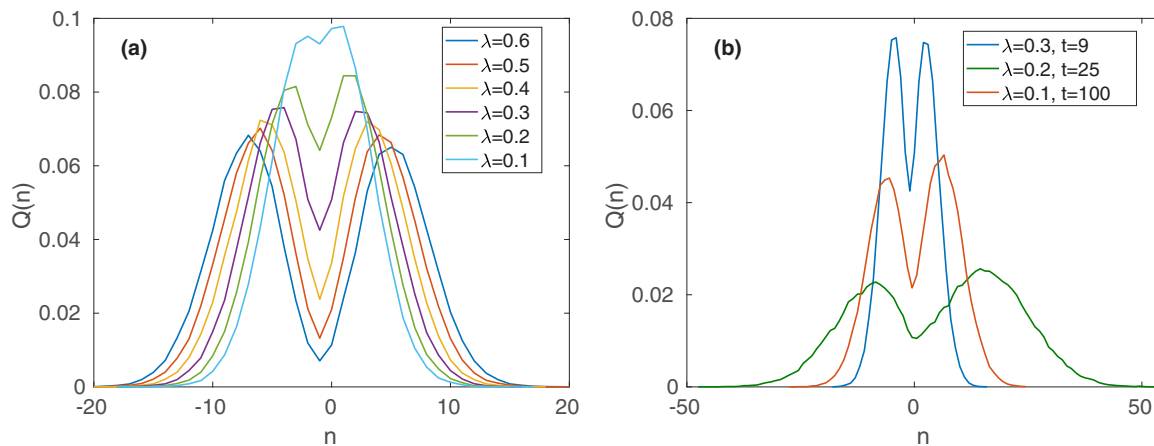
FIG. 6. Splitting of $Q(n_i)$. (a) Distribution $Q(n_i)$ at $t = 10$ for various values of λ. Note that, for $\lambda = 0.1$, $t \ll t_c = 100$ while for $\lambda = 0.6$, $t \gg t_c = 2.77$. (b) Distribution $Q(n_i)$ at $t \approx t_c = 1/\lambda^2$ for various values of λ.

two distributions,

$$v^{\pm} = \frac{d\langle n_i \rangle_{\pm}}{dt} = m(1 - \lambda) \pm \lambda, \tag{31}$$

$$\sigma_{\pm}^2 = t. \tag{32}$$

Since m is initially very small, Eq. (31) confirms that positive (negative) n_i tend to increase (decrease) and the overall distribution $Q(n_i)$ broadens and eventually splits. In the same Appendix we derive the expression for the variance $\sigma^2(t) = \langle n_i^2 \rangle - \langle n_i \rangle^2$ of the whole distribution $Q(n_i)$, finding

$$\sigma^2(t) = t + \lambda^2 t^2. \tag{33}$$

From Eq. (33) we can determine the time at which $Q(n_i)$ splits into two separate components. For small times $\sigma(t)$ is dominated by the diffusive widening of the central peak, while for larger times it is determined mainly by the ballistic distancing between the two peaks. Denoting by t_c the time at which the crossover occurs, it follows from Eq. (33) that $t_c \approx \lambda^2 t_c^2$, which yields

$$t_c \approx \frac{1}{\lambda^2}. $$

Figures 6(a) and 6(b) confirm numerically this prediction. Note that, at odds with the case $c \gtrsim 1$, the splitting always occurs, and over a much shorter temporal scale since $t_c = 1/\lambda^2 \ll \ln(1 + \lambda/\delta)/(\lambda\delta) = T^*$.

To understand whether the system reaches consensus or not, the argument is similar to the one presented for $c \gtrsim 1$, but in this case it provides the actual critical size N_c rather than a lower bound. If, when the distribution splits, the drift of the right peak is positive and that of the left peak is negative, consensus is not reached. From Eq. (31) this implies that consensus is reached only if

$$|m(t_c)| > m_c = \frac{\lambda}{1 - \lambda}. \tag{34}$$

This condition means, as before, that consensus for $\lambda > 1/2$ can occur only before the splitting, thus during the transient regime. It turns out numerically that, for $t < t_c$, the magnetization m grows as $m(t) \approx \xi t/\sqrt{N}$, where ξ is a random prefactor ranging between approximately -1 and $+1$.

Hence for $\lambda > 1/2$ the condition for consensus reads

$$|m(t)| = 1 \text{ for } t < t_c \rightarrow \frac{t_c}{\sqrt{N_c}} = 1 \rightarrow N_c = \frac{1}{\lambda^4}.$$

For $\lambda < 1/2$ instead, inserting the expression for $m(t)$ into Eq. (34) yields that consensus cannot be reached if the number N of individuals is larger than

$$N_c = \frac{(1 - \lambda)^2}{\lambda^6}. \tag{35}$$

For $N > N_c$ the system remains asymptotically disordered in a polarized state. In the opposite case instead consensus is rapidly reached after t_c, unless by chance the initial absolute value of ξ is particularly small.

In conclusion, recalling Eq. (35), the critical size satisfies

$$\begin{cases} N_c = \frac{(1-\lambda)^2}{\lambda^6} & \text{for } \lambda < \frac{1}{2}, \\ N_c = \frac{1}{\lambda^4} & \text{for } \lambda \geqslant \frac{1}{2}. \end{cases} \tag{36}$$

Note that N_c is continuous in $\lambda = 1/2$. Simulations presented in Fig. 7 show that the probability of reaching consensus exhibits a crossover at values well predicted by Eq. (36).

The long-time dynamics in the case $c \gg 1$ can be studied also in more detail by describing the evolution of the two peaks and their mutual interaction. As shown in Appendix E, we can write a closed integrodifferential equation for the magnetization $m(t)$. In particular, defining

$$y = \int_0^t m(t') dt'$$

we obtain the following second-order nonlinear ordinary differential equation (ODE):

$$\frac{d^2 y}{dt^2} = \sqrt{\frac{2}{\pi t (1 - \lambda)^2}} \left\{ \left(1 - \frac{dy}{dt}\right) \frac{1}{\frac{1}{t}y - m_c} \right.$$

$$\times \exp\left[-\frac{t(1-\lambda)^2 \left(\frac{1}{t}y - m_c\right)^2}{2} \right] - \left(1 + \frac{dy}{dt}\right) \frac{1}{\frac{1}{t}y + m_c}$$

$$\left. \times \exp\left[-\frac{t(1-\lambda)^2 \left(\frac{1}{t}y + m_c\right)^2}{2} \right] \right\}. \tag{37}$$
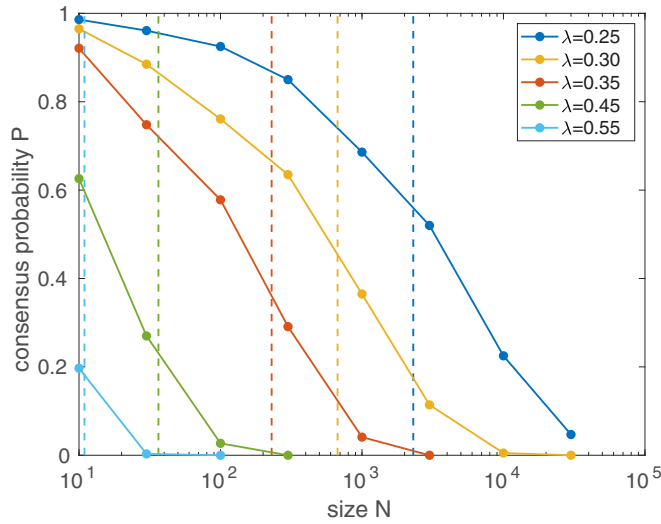
FIG. 7. Crossover between consensus and polarization for large c. Fraction of runs reaching consensus as a function of N, in 1000 realizations of the dynamics for $c = 100$ and several λ values. The crossover values N_c predicted by Eq. (36) are marked by vertical lines.

Note that the evolution of $m(t)$ is expressed in terms of y, which is a variable containing the past history of the system. This reflects the fact that personalized information keeps the memory of the preferences of the spin to which it is coupled. The fact that the evolution of m is governed by an integrodifferential equation is then a very natural consequence of the dynamics of the model. Solutions of this equation are reported in Fig. 8(b), where it is possible to see that states with $|m| < m_c$ remain disordered, while if $m < -m_c$ or $m > m_c$ consensus is reached, in good agreement with numerical simulations, as shown in Fig. 8(a).

Disordered states for $c > 1$ are completely different from the disordered states for $c = 1$. In the latter case, the constant magnetization is the effect of the external information being a random uncorrelated variable equal for all nodes so that all agents spend half of their time with $s_i = +1$ and half with $s_i = -1$. For $c > 1$ instead, the system is divided into two polarized clusters whose agents have a preferred spin value. A further characterization in terms of self-overlap is presented in Appendix F.

IV. DISCUSSION AND CONCLUSIONS

Let us summarize the results of our investigation. The evolution of the system is described by the magnetization m and the distribution of the local quantities n_i, describing the personalized information for each agent. Depending on whether δ ($c = 1 + \delta$) is much smaller or much larger than 1, the temporal evolution exhibits some variation.

In the first case, there are three temporal regimes in the evolution of the system. For short times up to $2/\delta$ stochastic effects dominate, the magnetization fluctuates around zero, and the distribution of the local n_i remains centered around zero. Later on, the symmetry between positive and negative external information breaks down because the mean value $n(t)$ of the single-peaked distribution $Q(n_i)$ starts drifting away from $n = 0$ exponentially in time, while also its width $\sigma(t)$ grows. At the same time also $|m|$ grows exponentially. This regime ends when linearization of the equations for m and n is no longer valid. The nonlinear subsequent evolution varies depending on the relative width of the peak. If the peak is narrow, all n_i have the same sign and consensus is quickly reached. If the peak is broad, then individuals with both $n_i > 0$ and $n_i < 0$ exist. If the magnetization in this moment is small enough, then the system gets trapped in a disordered (polarized) state, where disagreement persists and the magnetization maintains a constant value $|m| < m_c = \lambda/(1 - \lambda)$.

When $\delta \gg 1$ (i.e., $c \gg 1$) linearization is never valid, the $Q(n_i)$ distribution always splits into two components, and this happens much earlier, over a time scale equal to $t_c = 1/\lambda^2$. What happens next depends again on the value of the magnetization at t_c. If $|m(t_c)|$ is sufficiently large, the drift of the two components has the same sign. For example, if this sign is positive, it means that individuals with negative n_i have
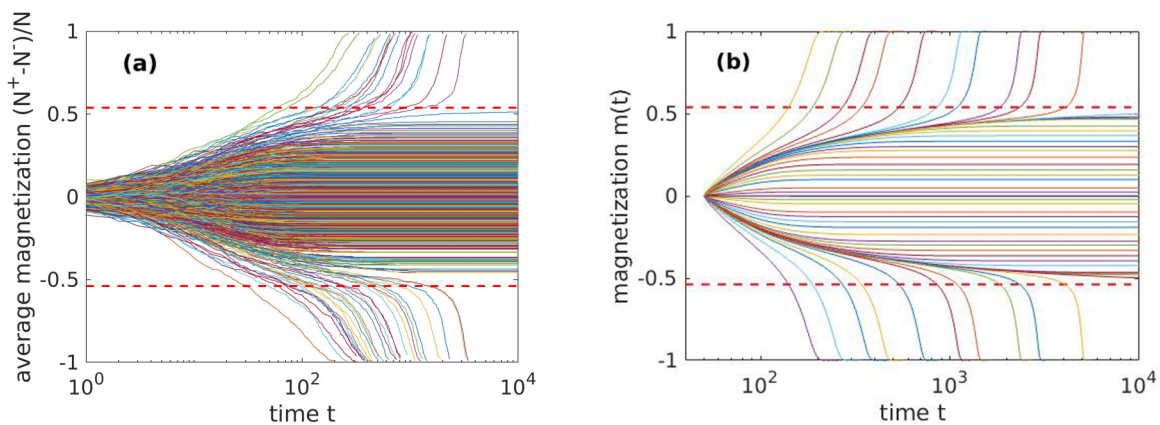

FIG. 8. Time evolution of the magnetization for $c = 100$. (a) 10^4 trajectories obtained simulating the model with $N = 10^3$, $c = 100$. and $\lambda = 0.35$. Red dashed lines represent $\pm m_c = \pm \lambda/(1 - \lambda)$; disordered states are found in the region $[-m_c; +m_c]$. (b) Solutions of Eq. (37) for $\lambda = 0.35$ and various m_0 in the range 10^{-5}–10^{-3}. The differential equation is valid only for times much larger than $t_c \approx 8$, therefore we set $t_0 = 50$. As the initial condition for the variable y_0, we used $y_0 = y(t_0) = m_0 t^2/2$, because the magnetization grows linearly during the initial regime. Red dashed lines represent $\pm m_c = \pm \lambda/(1 - \lambda)$, and in this case disordered states are found in the region $[-m_c; +m_c]$.
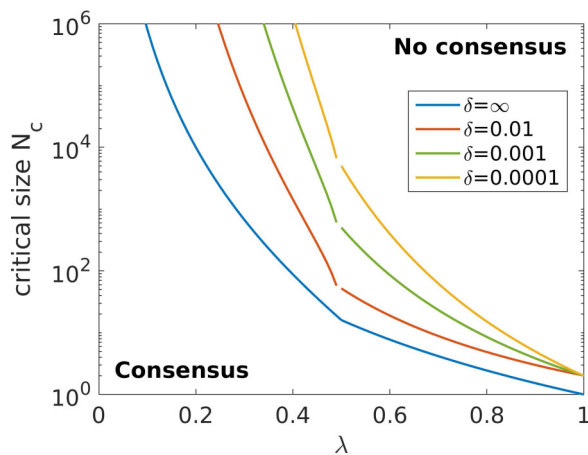
FIG. 9. Phase diagram for all c. The lines separate the region where consensus is reached from the region where it is not. Lines are obtained using Eq. (30) for $\delta \ll 1$ and Eq. (36) for $\delta = \infty$.

nevertheless an overall positive drift: The negative component of the $Q(n_i)$ distribution gets rapidly depleted and consensus is reached. Otherwise, the competition between the two opinions persists forever and the system gets stuck in the polarized state with $m < m_c$.

Although the detailed temporal evolution is rather different depending on whether $c = 1 + \delta$ is very close to 1 or much larger, the final overall phenomenology is similar. The parameter c sets the temporal scales of the dynamics and the details of the phase diagram, but not the qualitative features of the behavior: consensus for $N < N_c(\lambda)$, polarization otherwise. The dependence of N_c on λ is different depending on whether c is close to 1 [Eq. (30)] or large [Eq. (36)], so the boundaries between the two regions depend on the value of c. Figure 9 represents this phase diagram for several of these values, showing that, as expected, increasing the strength of personalized information makes consensus more difficult. However, we observe that increasing c reduces the consensus region but with a nontrivial limit: even in the $c \to \infty$ limit, consensus is possible in sufficiently small systems.

This rather rich phenomenology has been obtained in the simplest possible setting: an extremely simple binary opinion dynamics always leading to consensus, in a mean-field framework, augmented with an elementary form of personalized information. Clearly our results open some interesting issues for future research: What is the effect of a less trivial contact pattern among agents? What changes when different opinion dynamics models are considered? What happens if personalized information is parametrized differently, for example considering different functional forms for the probability $P(n_i)$? Finally, despite our oversimplified assumptions, some predictions derived in this work may be testable in empirical systems. In particular, the fact that disordered states cannot have magnetization larger than m_c or the existence of a maximum size N_c for the reaching of consensus could be observable in systems in which a community has to decide between two alternative options.

APPENDIX A: DRIFT AND DIFFUSION COEFFICIENTS

Let us define as R_i the probability that s_i switches from -1 to $+1$ in a single update. Equation (4) implies

$$R_i = \frac{1}{N}\left[\frac{1-\lambda}{N}\left(\frac{1-|,s_i}{2}\right)\underbrace{\sum_j\left(\frac{1+s_j}{2}\right)}_{N_\uparrow}+\lambda\left(\frac{1-s_i}{2}\right)P(n_i)\right],$$

where the prefactor stems from the fact that the ith spin is selected with probability $1/N$. Analogously L_i is the probability of the opposite transition, namely from $s_i = +1$ to -1:

$$L_i = \frac{1-\lambda}{N^2}\left(\frac{1+s_i}{2}\right)\underbrace{\sum_j\left(\frac{1-s_j}{2}\right)}_{N_\downarrow}+\frac{\lambda}{N}\left(\frac{1+s_i}{2}\right)[1-P(n_i)].$$

We can obtain the corresponding transition probabilities for the magnetization m by summing over all spins,

$$R^m = (1-\lambda)\frac{1-m}{2}\frac{1+m}{2} + \frac{\lambda}{N}\sum_i\left(\frac{1-s_i}{2}\right)P(n_i),$$

$$L^m = (1-\lambda)\frac{1+m}{2}\frac{1-m}{2} + \frac{\lambda}{N}\sum_i\left(\frac{1+s_i}{2}\right)[1-P(n_i)].$$

In a single update, occurring in a time $\delta t = 1/N$, the variation of the magnetization is $\delta m = 2/N$, hence its drift is

$$v^m = \frac{\delta m}{\delta t}[R^m - L^m]$$

$$= 2\frac{\lambda}{N}\sum_i\left\{\left(\frac{1-s_i}{2}\right)P(n_i) - \left(\frac{1+s_i}{2}\right)[1-P(n_i)]\right\}.$$

Analogously the diffusion coefficient is

$$D^m = \frac{(\delta m)^2}{\delta t}[R^m + L^m] = \frac{1-\lambda}{N}(1-m^2)$$

$$+ \frac{2\lambda}{N^2}\sum_i\left\{\left(\frac{1-s_i}{2}\right)P(n_i) + \left(\frac{1+s_i}{2}\right)[1-P(n_i)]\right\}.$$

As expected, the evolution of the magnetization depends also on the variables n_i, encoding the probability of the personalized information. For this reason, it is necessary to consider in detail their evolution. Their transition probabilities are

$$R^{n_i} = \frac{1}{N}\left[(1-\lambda)\left(\frac{1+m}{2}\right) + \lambda P(n_i)\right],$$

$$L^{n_i} = \frac{1}{N}\left[(1-\lambda)\left(\frac{1-m}{2}\right) + \lambda[1-P(n_i)]\right]. \quad (A1)$$

The difference between these probabilities and those for s_i stems from the fact that n_i is updated at each step, not only when s_i changes its state. Using these expressions, we can determine the drift for individual variables n_i,

$$v_i = \frac{\delta n_i}{\delta t}(R^{n_i} - L^{n_i}) = (1-\lambda)m + \lambda[2P(n_i) - 1]. \quad (A2)$$

Summing the transition probabilities over all nodes, we obtain the probabilities for n:

$$R^n = \sum_i R^{n_i} = (1-\lambda)\left(\frac{1+m}{2}\right) + \frac{\lambda}{N}\sum_i P(n_i),$$

$$L^n = \sum_i L^{n_i} = (1-\lambda)\left(\frac{1-m}{2}\right) + \frac{\lambda}{N}\sum_i [1 - P(n_i)].$$

In this case, the variation in a single update is $\delta n = 1/N$. As a consequence, the drift coefficient reads

$$v^n = \frac{\delta n}{\delta t}[R^n - L^n] = (1-\lambda)m + \frac{\lambda}{N}\sum_i [2P(n_i) - 1].$$
(A3)

The diffusion coefficient is instead

$$D^n = \frac{(\delta n)^2}{\delta t}[R^n + L^n] = \frac{1}{N}\left[(1-\lambda) + \frac{\lambda}{N}\sum_i 1\right] = \frac{1}{N}.$$
(A4)

APPENDIX B: VARIANCE OF $Q(n_i)$ FOR $c \gtrsim 1$

The master equation governing the evolution of $Q(n_i)$ is

$$Q(n_i, t+\delta t) = R^{n_i-1}Q(n_i-1, t) + L^{n_i+1}Q(n_i+1, t) + (1 - R^n - L^n)Q(n_i, t),$$

where $1 - R^n - L^n = (1-1/N)$ is the probability that the spin i is not selected and thus not updated. Expanding for small δt, we obtain

$$\frac{d}{dt}Q(n_i) = \frac{1}{\delta t}[R^{n_i-1}Q(n_i-1, t) + L^{n_i+1}Q(n_i+1) - (R^n + L^n)Q(n_i)],$$
(B1)

where $\delta t = 1/N$ and the transition probabilities R^{n_i} are given by Eq. (A1). Noticing that $R^n + L^n = \frac{1}{N}$, we rewrite this equation as

$$\frac{d}{dt}Q(n_i) = NR^{n_i-1}Q(n_i-1, t) + (1 - NR^{n_i+1})Q(n_i+1) - Q(n_i).$$
(B2)

Using the condition $|n_i|\delta \ll 1$, we rewrite the transition probabilities as

$$R^{n_i} = \frac{1}{N}\left[(1-\lambda)\left(\frac{1+m}{2}\right) + \lambda\left(\frac{1}{2} + \frac{1}{4}\delta n_i\right)\right].$$
(B3)

Substituting this expression into Eq. (B2), multiplying by n_i^2, and summing over n_i, we obtain

$$\frac{d}{dt}\langle n_i^2\rangle = \sum_{n_i}[4NR^{n_i}n_iQ(n_i) + (1 - 2n_i)Q(n_i)]$$
$$= 2m(1-\lambda)\langle n_i\rangle + 1 + \lambda\delta\langle n_i^2\rangle,$$
(B4)

where $\langle n_i\rangle$ is the quantity n defined in Eq. (6). The variance of $Q(n_i)$ is $\sigma^2 = \langle n_i^2\rangle - \langle n_i\rangle^2$, hence using Eqs. (16) and (B4)

we obtain

$$\frac{d}{dt}\sigma^2 = \frac{d}{dt}\langle n_i^2\rangle - 2\langle n_i\rangle\frac{d}{dt}\langle n_i\rangle$$
$$= 2m(1-\lambda)\langle n_i\rangle + 1 + \lambda\delta\langle n_i^2\rangle$$
$$\quad - 2\langle n_i\rangle\left[(1-\lambda)m + \frac{\delta\lambda}{2}\langle n_i\rangle\right]$$
$$= 1 + \lambda\delta\sigma^2,$$

whose solution, with initial condition $\sigma^2(0) = 0$, is

$$\sigma^2(t) = \frac{1}{\lambda\delta}(e^{\lambda\delta t} - 1).$$
(B5)

APPENDIX C: A MORE REFINED LOWER BOUND FOR N_c

The drift of spin i is given by Eq. (26),

$$v_i = m(1-\lambda) + \lambda[2P(n_i) - 1].$$

For small $\delta = c - 1$, the probability $P(n_i)$ can be expanded, giving

$$P(n_i) = \frac{1}{2} + \frac{n_i\delta}{4},$$

therefore

$$v_i = m(1-\lambda) + \lambda\frac{n_i\delta}{2}.$$

As a consequence, the spins with negative drift are those whose n_i satisfies $n_i < \bar{n}$, with $\bar{n}$ defined by

$$m(1-\lambda) + \lambda\frac{\bar{n}\delta}{2} = 0 \rightarrow \bar{n} = -\frac{2m}{\delta}\frac{1-\lambda}{\lambda} = -\frac{2}{\delta}\frac{m}{m_c}.$$
(C1)

We assume that at time T^*, when linearization breaks down, all spins with $n_i < 0$ but positive drift $v_i(T^*) > 0$ will eventually end up with $n_i \rightarrow +\infty$, while those with $v_i(T^*) < 0$ will drift toward $-\infty$. The number of spins with $v_i(T^*) < 0$ is easily obtained by integrating the distribution $Q_t(n_i)$ and using Eq. (C1). More precisely, denoting this quantity by N_∞^-, it holds that

$$N_\infty^- = N\int_{-\infty}^{\bar{n}} Q(n_i)dn_i = N\int_{-\infty}^{-\frac{2}{\delta}\frac{m(T^*)}{m_c}} Q_{T^*}(n_i)dn_i.$$

The distribution $Q_t(n_i)$ can be approximated by a Gaussian with mean value $n(t)$ and variance $\sigma^2(t)$. At time T^* the variance satisfies by definition

$$\sigma^2(T^*) = \frac{1}{\delta^2},$$

while recalling that

$$n(t) = \sqrt{\frac{2}{\delta N}}e^{t\delta/2}, \quad T^* = \frac{1}{\lambda\delta}\ln\left(1 + \frac{\lambda}{\delta}\right),$$

we get

$$n(T^*) = \sqrt{\frac{2}{\delta N}}\left(1 + \frac{\lambda}{\delta}\right)^{\frac{1}{2\lambda}}.$$
(C2)

Asymptotically, the number of negative spin can then be written as

$$N_\infty^- = \frac{N}{\sqrt{2\pi\sigma^2(T^*)}} \int_{-\infty}^{-\frac{2}{\delta}\frac{m(T^*)}{m_c}} \exp\left[-\frac{(n(T^*)-n_i)^2}{2\sigma^2(T^*)}\right]dn_i$$

$$= \frac{N\delta}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{2}{\delta}\frac{m(T^*)}{m_c}}$$

$$\times \exp\left\{-\frac{\left[\sqrt{\frac{2}{\delta N}}\left(1+\frac{\lambda}{\delta}\right)^{\frac{1}{2\lambda}}-n_i\right]^2\delta^2}{2}\right\}dn_i.$$

The asymptotic magnetization m_∞ will consequently be

$$m_\infty = \frac{N_\infty^+ - N_\infty^-}{N} = \frac{N - 2N_\infty^-}{N}$$

$$= 1 - 2\frac{\delta}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{2}{\delta}\frac{m(T^*)}{m_c}}$$

$$\times \exp\left\{-\frac{\left[\sqrt{\frac{2}{\delta N}}\left(1+\frac{\lambda}{\delta}\right)^{\frac{1}{2\lambda}}-n_i\right]^2\delta^2}{2}\right\}dn_i.$$

The asymptotic state is disordered only if this magnetization is smaller than the critical magnetization, otherwise the drift of the negative peak is positive and therefore consensus is reached. The condition for consensus consequently reads

$$m_\infty > m_c = \frac{\lambda}{1-\lambda},$$

that is,

$$1 - 2\frac{\delta}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{2}{\delta}\frac{m(T^*)}{m_c}}$$

$$\times \exp\left\{-\frac{\left[\sqrt{\frac{2}{\delta N}}\left(1+\frac{\lambda}{\delta}\right)^{\frac{1}{2\lambda}}-n_i\right]^2\delta^2}{2}\right\}dn_i > m_c. \quad \text{(C3)}$$

Here, using Eq. (C2) and recalling that $m = \frac{\delta}{2}n$, $m(T^*)$ satisfies

$$m(T^*) = \sqrt{\frac{\delta}{2N}}\left(1+\frac{\lambda}{\delta}\right)^{\frac{1}{2\lambda}}.$$

By solving numerically Eq. (C3) for the value of N, we derive a numerical lower bound of the critical size N_c of systems reaching consensus.

APPENDIX D: VARIANCE OF $Q(n_i)$ FOR $c \gg 1$

We can study the evolution of the component $Q^+(n_i)$ using the master equation defined in Eq. (B2),

$$\frac{d}{dt}Q^+(n_i) = NR^{n_i-1}Q^+(n_i-1)$$

$$+ (1 - NR^{n_i+1})Q^+(n_i+1) - Q^+(n_i), \quad \text{(D1)}$$

where the transition probabilities R^{n_i} satisfy

$$R^{n_i} = \frac{1}{N}\left[(1-\lambda)\frac{1+m}{2} + \lambda P(n_i)\right].$$

The first moment n^+ varies as

$$v^+ = \frac{dn^+}{dt} = \sum_{n_i} \frac{d}{dt}Q^+(n_i)n_i$$

$$= \sum_{n_i} Q^+(n_i)[2NR^{n_i} - 1].$$

Assuming $Q^+(n_i \leq 0) \approx 0$, we can rewrite the transition probabilities as

$$NR^{n_i} = (1-\lambda)\frac{1+m}{2} + \lambda(1 - c^{-n_i}), \quad \text{(D2)}$$

and then obtain for the drift v^+

$$v^+ = \sum_{n_i} Q^+(n_i)[2\lambda + (1-\lambda)(1+m) - 2\lambda c^{-n_i} - 1]$$

$$= m(1-\lambda) + \lambda - 2\lambda c^{-n^+} \approx m(1-\lambda) + \lambda, \quad \text{(D3)}$$

where we have made the approximation $\langle f(n) \rangle = f(\langle n \rangle)$. Similarly

$$v^- = m(1-\lambda) - \lambda + 2\lambda c^{n^-} \approx m(1-\lambda) - \lambda. \quad \text{(D4)}$$

Let us now turn to the evolution of the variance. As shown in the derivation of Eq. (B4), it holds that

$$\frac{d\langle n^2 \rangle_+}{dt} = \sum_{n_i}[4NR^{n_i}Q^+(n_i)n_i + (1 - 2n_i)Q^+(n_i)].$$

Again, making the approximation $Q^+(n_i \leq 0) \approx 0$ we can use Eq. (D2),

$$\frac{d\langle n^2 \rangle_+}{dt} = \sum_{n_i} \left\{4\left[(1-\lambda)\frac{1+m}{2} + \lambda\right]Q^+(n_i)n_i\right.$$

$$\left. + (1 - 2n_i)Q^+(n_i)\right\},$$

hence

$$\frac{d\langle n^2 \rangle_+}{dt} = 2[\lambda + m(1-\lambda)]n^+ + 1.$$

This implies that the variance $\sigma_+^2(t)$ of $Q^+(n_i)$ evolves as

$$\frac{d\sigma_+^2}{dt} = \frac{d\langle n^2 \rangle_+}{dt} - 2n^+\frac{d}{dt}n^+ = \frac{d\langle n^2 \rangle_+}{dt} - 2n^+v^+$$

$$= 2[\lambda + m(1-\lambda)]n^+ + 1$$

$$- 2n^+[m(1-\lambda) + \lambda] = 1.$$

Assuming $\sigma_+^2(0) = 0$, we obtain

$$\sigma_+^2(t) = t. \quad \text{(D5)}$$

Similarly $\sigma_-^2(t) = t$. We then conclude that the two peaks widen as it would happen for an unbiased random walk.

We can now study the dynamics of the variance $\sigma^2(t)$ of $Q(n_i)$. For a bimodal distribution it holds that

$$\sigma^2 = p[\sigma_1^2 + (\mu_1 - \mu)^2] + (1-p)[\sigma_2^2 + (\mu_2 - \mu)^2],$$

which in our case becomes

$$\sigma^2 = \frac{N^+}{N}[\sigma_+^2 + (n_+ - n)^2] + \frac{N^-}{N}[\sigma_-^2 + (n_- - n)^2]$$

$$= \frac{N^+}{N}\left\{\sigma_+^2 + \left[\frac{N^-}{N}(n^+ - n^-)\right]^2\right\}$$

$$+ \frac{N^-}{N}\left\{\sigma_-^2 + \left[\frac{N^+}{N}(n^+ - n^-)\right]^2\right\}.$$

The quantities N^+ and N^- vary slowly, and we can then approximate them as constants and derive over time, obtaining

$$\frac{d\sigma^2}{dt} = \frac{N^+}{N}\left\{\frac{d\sigma_+^2}{dt} + \left(\frac{N^-}{N}\right)^2 2(n^+ - n^-)\frac{d(n^+ - n^-)}{dt}\right\}$$

$$+ \frac{N^-}{N}\left\{\frac{d\sigma_-^2}{dt} + \left(\frac{N^+}{N}\right)^2 2(n^+ - n^-)\frac{d(n^+ - n^-)}{dt}\right\}$$

$$= \frac{N^+}{N}\left[1 + \left(\frac{N^-}{N}\right)^2 2(v^+ - v^-)^2 t\right]$$

$$+ \frac{N^-}{N}\left[1 + \left(\frac{N^+}{N}\right)^2 2(v^+ - v^-)^2 t\right],$$

where we used Eq. (D5). Noting that Eqs. (D3) and (D4) imply $v^+ - v^- \approx 2\lambda$ and recalling that $c \gg 1$, we obtain

$$\frac{d\sigma^2}{dt} \approx 1 + 8\frac{N^+ N^-}{N^2}\lambda^2 t.$$

It is easy to show that N^+ and N^- can be approximated by $N_\uparrow$ and $N_\downarrow$. Indeed, using Eq. (7) we can write the drift for m as

$$v^m = \frac{\lambda}{N}\sum_i \left(\frac{c^{n_i} - 1}{c^{n_i} + 1}\right) - \lambda m \approx \lambda\left[\frac{N^+ - N^-}{N} - m\right],$$

which implies that, up to diffusive contributions, $m \approx \frac{N^+ - N^-}{N}$. We can then set $N^+ \approx N(1 + m)/2$ arriving at the following expression:

$$\frac{d\sigma^2}{dt} = 1 + 2(1 - m^2)\lambda^2 t. \tag{D6}$$

For small m we can neglect the term m^2, obtaining

$$\frac{d\sigma^2}{dt} \approx 1 + 2\lambda^2 t,$$

which, setting $\sigma^2(0) = 0$, yields

$$\sigma^2(t) = t + \lambda^2 t^2. \tag{D7}$$

APPENDIX E: INTEGRODIFFERENTIAL EQUATION FOR *m*

Up to $t = t_c$ the distribution $Q(n_i)$ is unimodal and can be described in terms of n and σ^2; to understand the subsequent evolution, we approximate the distribution $Q(n_i)$ as the sum of two Gaussian distributions $Q^+(n_i)$ and $Q^-(n_i)$ with variance $\sigma_\pm^2 = t$ [see Eq. (D5)] and mean values n^+ and n^-. Let us focus on the $Q^+(n_i)$ component. This peak moves with drift v^+ and widens linearly in time. This implies that, in the reference frame moving with the peak, each spin performs an unbiased random walk. As a consequence, neglecting temporal correlations, the probability that one of these spins during

the random walk reaches negative values of n_i can be obtained by integrating $Q^+(n_i)$ for all $n_i < 0$. The probability $P_{R\to L}$ that an individual performs a transition from $n_i > 0$ to $n_i < 0$ is then

$$P_{R\to L} = \int_{-\infty}^0 Q^+(n)dn \approx \frac{1}{\sqrt{2\pi\sigma_+^2}}\int_{-\infty}^0 \exp\left[-\frac{(n - n^+)^2}{2\sigma_+^2}\right]$$

$$= \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{n^+}{\sqrt{2\sigma_+^2}}\right)\right]. \tag{E1}$$

Focusing on $t \gg t_c$, we can assume that the two peaks are well separated and therefore we can expand the error function for a large argument,

$$\mathrm{erf}(x) \approx 1 - \frac{1}{\sqrt{\pi}x}\exp(-x^2),$$

yielding

$$P_{R\to L}(t) \approx \sqrt{\frac{2}{\pi}}\frac{\sigma_+(t)}{n^+(t)}\exp\left[-\frac{(n^+(t))^2}{2\sigma_+^2(t)}\right]. \tag{E2}$$

We can express the mean value $n^+(t)$ in terms of the drift $v^+(t)$ as

$$n^+(t) = \int_0^t v^+(t')dt' = \lambda t + (1 - \lambda)\int_0^t m(t')dt',$$

where we used Eq. (D3). Substituting this expression into Eq. (E2) we get

$$P_{R\to L}(t) \approx \sqrt{\frac{2}{\pi}}\frac{\sigma_+(t)}{\lambda t + (1 - \lambda)\int_0^t m(t')dt'}$$

$$\times \exp\left\{-\frac{\left[\lambda t + (1 - \lambda)\int_0^t m(t')dt'\right]^2}{2\sigma_+^2(t)}\right\}.$$

Recalling that $\sigma_+(t) = t$, defining $m_c = \frac{\lambda}{1-\lambda}$, and introducing the temporal average of the magnetization

$$\bar{m}(t) = \frac{1}{t}\int_0^t m(t')dt' \tag{E3}$$

we can rewrite this probability as

$$P_{R\to L}(t) \approx \sqrt{\frac{\tau^+(t)}{\pi t}}\exp\left(-\frac{t}{\tau^+(t)}\right), \tag{E4}$$

where we introduced the characteristic temporal scale $\tau^+(t)$, which satisfies

$$\tau^+(t) = \frac{2}{(1 - \lambda)^2}\frac{1}{[\bar{m}(t) + m_c]^2}. \tag{E5}$$

Of course perfectly analogous formulas hold for the negative peak:

$$P_{L\to R}(t) \approx \sqrt{\frac{\tau^-(t)}{\pi t}}\exp\left(-\frac{t}{\tau^-(t)}\right), \tag{E6}$$

where the characteristic time $\tau^-(t)$ satisfies

$$\tau^-(t) = \frac{2}{(1 - \lambda)^2}\frac{1}{[\bar{m}(t) - m_c]^2}. \tag{E7}$$

Equation (E4) implies that the transition probability is exponentially small for $t > \tau^+(t)$, while if $t < \tau^+(t)$ the spins can make a transition from $n_i > 0$ to $n_i < 0$ with nonvanishing probability. If $\tau^+ \to \infty$, the transition probability in Eq. (E4) becomes essentially constant. This is a manifestation of the fact that the expansion of the error function in Eq. (E1) cannot be performed as the positive peak is not narrow. Physically, this implies that the positive peak gets rapidly absorbed by the negative one leading to consensus. To determine if consensus is reached, we then have to study this characteristic temporal scale.

We first observe that if, starting from a given time, the magnetization is smaller than $-m_c$, then $\bar{m}(t)$, which starts from $\bar{m}(t) = 0$, gradually decreases, necessarily reaching at some point the value $-m_c$. Hence τ^+ diverges and consensus is rapidly reached. If instead m does not become smaller than $-m_c$, then $\bar{m}(t)$ always remains larger than $-m_c$, τ^+ remains finite, and consensus is never reached. Hence we can predict that asymptotically disordered configurations are possible only for magnetization in the interval $-m_c < m < m_c$, while if m exceeds these bounds consensus is reached. Figure 8 confirms this expectation.

We can use the transition probabilities for writing a closed-form equation for $m(t)$. The derivative of N^+ satisfies

$$\frac{dN^+}{dt} = N^-(t)P_{L \to R}(t) - N^+(t)P_{R \to L}(t).$$

Expressing N^+ and N^- in terms of m, we get

$$\frac{dm}{dt} = [1 - m(t)]P_{L \to R}(t) - [1 + m(t)]P_{R \to L}(t). \quad \text{(E8)}$$

Inserting Eqs. (E4)–(E7) into Eq. (E8), we finally obtain an integrodifferential equation for the magnetization $m(t)$,

$$\frac{dm}{dt} = [1 - m(t)]\sqrt{\frac{2}{\pi t \left[\frac{1}{t}\int_0^t m(t')dt' - m_c\right]^2 (1-\lambda)^2}}$$

$$\times \exp\left\{-\frac{t(1-\lambda)^2\left[\frac{1}{t}\int_0^t m(t')dt' - m_c\right]^2}{2}\right\}$$

$$- [1 + m(t)]\sqrt{\frac{2}{\pi t \left[\frac{1}{t}\int_0^t m(t')dt' + m_c\right]^2 (1-\lambda)^2}}$$

$$\times \exp\left\{-\frac{t(1-\lambda)^2\left[\frac{1}{t}\int_0^t m(t')dt' + m_c\right]^2}{2}\right\}.$$

APPENDIX F: DETECTING FILTER BUBBLES: SELF-OVERLAP

To properly characterize disordered configurations, characterized by $|m| < 1$, we consider the local magnetization $m_i = \langle s_i \rangle$, defined as

$$m_i(t) = P_+(t) - P_-(t),$$

where

$$P_+ = (1-\lambda)\frac{N_\uparrow(t)}{N} + \lambda P(n_i(t)),$$

$$P_- = (1-\lambda)\frac{N_\downarrow(t)}{N} + \lambda[1 - P(n_i(t))].$$

Noting that the mean magnetization $m(t) = 1/N \sum m_i(t)$ satisfies

$$m(t) = \frac{N_\uparrow(t) - N_\downarrow(t)}{N} \rightarrow N_\uparrow(t) = \frac{N[m(t) + 1]}{2},$$

we obtain

$$m_i(t) = (1-\lambda)\frac{1+m(t)}{2} + \lambda P(n_i(t))$$

$$- (1-\lambda)\frac{1-m(t)}{2} - \lambda[1 - P(n_i(t))].$$

Therefore, the magnetization of site i is

$$m_i(t) = m(t)(1-\lambda) + \lambda[2P(n_i(t)) - 1]. \quad \text{(F1)}$$

We can now introduce the self-overlap $q(t)$, which allows us to detect the presence of filter bubbles,

$$q(t) = \frac{1}{N}\sum_i^N m_i(t)^2. \quad \text{(F2)}$$

Indeed if the network is split into two bubbles, this quantity is expected to be different from zero even if the overall magnetization is null, because each node tends to be aligned to one of the two opinions. It is easy to show that the disordered state found for $c = 1$ does not contain separate filter bubbles. Indeed using Eq. (F1) and recalling that $P(n_i(t)) = 1/2$, we obtain

$$q_{c=1}(t) = \frac{1}{N}\sum_i m(t)^2 (1-\lambda)^2 = m(t)^2 (1-\lambda)^2 \sim \frac{1}{\sqrt{N}},$$

where we used the fact that the diffusion coefficient, defined by Eq. (11), scales as $1/N$. We thus see that in the limit of large systems, the self-overlap is null, meaning that each node randomly flips between $+1$ and -1: no polarization is present. This also implies that the variables n_i perform an unbiased random walk, and therefore the width of the probability distribution of the n_i, $Q(n_i)$, is expected to broaden as $\sqrt{t}$.

Let us turn now to the case $c > 1$. As we have shown, there are disordered states in which the external information, after an initial transient, completely polarizes. This implies that N^+ spins receive positive information, while negative information acts on the remaining N^- spins. Let us consider a positively polarized spin i, for which it holds that $P(n_i) = 1$: using expression (F1), we obtain for its magnetization

$$m_i^+(t) = m(t)(1-\lambda) + \lambda.$$

Similarly for a negatively polarized spin it holds that

$$m_i^-(t) = m(t)(1-\lambda) - \lambda.$$

Splitting the sum of relation (F2), we can write the overlap as

$$q_{c>1} = \frac{1}{N}\left[\sum_{i|n_i>0}[m(t)(1-\lambda) + \lambda]^2\right.$$

$$\left. + \sum_{i|n_i<0}[m(t)(1-\lambda) - \lambda]^2\right]$$

$$= m(t)^2(1-\lambda)^2 + \lambda^2 + 2m(t)^2\lambda(1-\lambda).$$

In this case, the overlap is non-null also if $m = 0$, meaning that the variables are partially frozen due to the personalized information. The disordered state is then profoundly different from the one characteristic of $c = 1$. Indeed, a null overlap implies that spins randomly flip, spending half of their time with a positive orientation and half with a negative one. Conversely, an overlap different from zero indicates the presence of two polarized clusters, whose spins tend to remain fixed.

---

[1] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, Science **323**, 721 (2009).

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

[3] D. Acemoglu and A. Ozdaglar, Dyn. Games Appl. **1**, 3 (2011).

[4] P. Sen and B. K. Chakrabarti, *Sociophysics: An Introduction* (Oxford University Press, Oxford, 2014).

[5] H. Noorazar, Eur. Phys. J. **135**, 521 (2020).

[6] Q. Michard and J.-P. Bouchaud, Eur. Phys. J. B **47**, 151 (2005).

[7] S. Iyengar and K. S. Hahn, J. Commun. **59**, 19 (2009).

[8] E. Bakshy, S. Messing, and L. A. Adamic, Science **348**, 1130 (2015).

[9] M. Maes and L. Bischofberger, *Will the Personalization of Online Social Networks Foster Opinion Polarization?* (2015). Available at SSRN: https://ssrn.com/abstract=2553436.

[10] J. Horwitz and D. Seetharaman, Wall Street J. (2020).

[11] N. Perra and L. E. Rocha, Sci. Rep. **9**, 7261 (2019).

[12] A. Sirbu, D. Pedreschi, F. Giannotti, and J. Kertesz, PLoS One **14**, 1 (2019).

[13] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Phys. Rev. Lett. **124**, 048301 (2020).

[14] P. Clifford and A. Sudbury, Biometrika **60**, 581 (1973).

[15] R. A. Holley and T. M. Liggett, Ann. Prob. **3**, 643 (1975).

[16] I. Dornic, H. Chaté, J. Chave, and H. Hinrichsen, Phys. Rev. Lett. **87**, 045701 (2001).

[17] V. Sood and S. Redner, Phys. Rev. Lett. **94**, 178701 (2005).

[18] E. Pugliese and C. Castellano, Europhys. Lett. **88**, 58004 (2009).

[19] K. Suchecki, V. M. Eguíluz, and M. S. Miguel, Phys. Rev. E **72**, 036132 (2005).

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

[21] A. Carro, R. Toral, and M. San Miguel, Sci. Rep. **6**, 24775 (2016).

[22] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, 2010).

[23] D. Bhat and S. Redner, Phys. Rev. E **100**, 050301(R) (2019).

[24] D. Bhat and S. Redner, J. Stat. Mech.: Theor. Expt. (2020) 013402.

[25] A. F. Peralta, N. Khalil, and R. Toral, Physica A **552**, 122475 (2020).

[26] A. Kirman, Q. J. Econ. **108**, 137 (1993).

[27] S. Alfarano, T. Lux, and F. Wagner, Comput. Econ. **26**, 19 (2005).

[28] O. Artime, A. F. Peralta, R. Toral, J. J. Ramasco, and M. San Miguel, Phys. Rev. E **98**, 032104 (2018).

[29] O. Artime, N. Khalil, R. Toral, and M. San Miguel, Phys. Rev. E **98**, 042143 (2018).