Polarization in increasingly connected societies

Tuan Minh Pham,[1,2,3,4,*] Sidney Redner,[5] Lourens Waldorp,[1,6] Jay Armas,[1,3,4,7] and Han L. J. van der Maas[1,6]
[1]*Dutch Institute for Emergent Phenomena, 1090 GE Amsterdam, The Netherlands*
[2]*Complexity Science Hub Vienna, Metternichgasse 8, A-1030 Vienna, Austria*
[3]*Institute for Advanced Study, Oude Turfmarkt 147, 1012 GC Amsterdam, The Netherlands*
[4]*Institute of Physics, University of Amsterdam, Science Park 904, Amsterdam, The Netherlands*
[5]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA*
[6]*Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, Amsterdam 1018 NP, The Netherlands*
[7]*The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100, Denmark*

(Received 29 September 2025; accepted 21 April 2026; published 11 May 2026)

Explanations of societal polarization often rely on one of three mechanisms: homophily, bounded confidence, and community-based interactions. Opinion dynamics models based on these mechanisms consider the lack of interactions as the main cause of polarization. Given the increasing connectivity in modern society, this explanation of polarization may be insufficient. To understand how society becomes more polarized as its connectedness increases, we propose a voter-type model (called I-voter) that incorporates involvement as a key mechanism in opinion formation and study its dependence on the network connectivity. We describe the steady-state behavior of the model analytically, at the mean-field and the moment-hierarchy levels, and stress the generality of our findings by considering various extensions and different network topologies.

DOI: 10.1103/5lwq-2n8m

I. INTRODUCTION

Polarization is a complex social phenomenon with likely multiple causes. While its study traces back to the 19th century [1], the field has grown rapidly over the past 30 years, driven by the availability of large-scale (social media) data and theoretical developments [2]. In particular, substantial progress has been made by numerous mathematical works that have rigorously established the emergence of polarization in coevolving social networks [3–5]. While not the only mechanism (see Ref. [6] for a review of various explanations [7]), many models emphasize the pivotal role of limited interactions, isolation, and, generally, low connectivity in driving polarization.

For instance, in classical consensus models, such as DeGroot's model [17], Abelson's model [18], and the voter model [19], polarization arises when subgroups become disconnected, preventing the formation of a unified consensus [20]. Also the model of Axelrod [21–23] predicts that, due to homophily, small societies fragment into cultural groups at a critical number of alternative traits per feature. In this latter model, individuals are more likely to interact with "similar" neighbors than with dissimilar ones, and they become more similar after every interaction. Other models show how fragmentation results from the presence of "boundedly confident" agents [24–26] who only interact with those not farther away than a given distance in opinion space. In threshold-type models, agents only adopt a view once the fraction of neighbors supporting the same view exceeds their own threshold drawn from a predefined distribution of adoption thresholds [27–29]. Polarization can also emerge from rearrangements of social ties in coevolving networks to form sparsely connected [30–33] or even antagonistic clusters of individuals [34–37].

In most of these models (apart from Ref. [38], which shows that, depending on the response rule, increasing out-group contacts may either leave polarization unchanged or even strengthen it), polarization typically arises from a lack of connectivity among agents holding opposing or distant opinions. However, in many respects, connectivity has increased in modern societies. Urbanization, globalization, international mobility, the rise of social media platforms, and cross-cultural marriages [39,40] have all contributed to humans being far more interconnected today than a century ago. One of the many consequences of this increasing connectivity is an increase in polarization, as suggested by Refs. [36,41].

In this paper, we explore an alternative explanation for polarization in increasingly interconnected societies, emphasizing the pivotal role of involvement in the process of opinion formation. The role of involvement (defined as sustained attention) has been extensively studied in psychology [42]. It can be measured using self-report questionnaires [43], behavioral indices, and psychophysiological measures [44]. Involvement has been used to explain why attitudes and opinions sometimes behave like dimensions and sometimes act as categories [45]. In network models of attitudes [46], involvement plays a similar role. Hoffstadt *et al.* [47] show in several datasets that involvement increases polarization. Central aspects of involvement in opinion or attitude formation include the following: low-involvement attitudes are more situationally influenced and less stable [48]; when people feel highly involved, their attitudes are less sensitive to persuasion

*Contact author: m.t.pham@uva.nl

[49]; and involvement weakens over time when not reinforced [50] but sharply increases due to interactions.

Building on these psychological models and empirical findings, we propose incorporating involvement into models of polarization. An earlier model developed along this line—the hierarchical Ising opinion model (HIOM) [51]—yielded several unexpected results, suggesting that polarization can emerge in highly connected networks even when all agents, on average, receive the same information from their social milieux. This finding of the HIOM, however, were based solely on simulations and possibly confounded by other factors. In this paper, to study the principal effect of involvement on polarization analytically, we develop an analytically tractable model whose relation to HIOM and other models is discussed in Appendix A.

We formulate our model based on generalizing the well-studied constrained three-state voter model [52,53]. In Ref. [52], agents can be in one of three states—leftist, centrist, and rightist—and can only switch to neighboring states (e.g., leftist to centrist or rightist to centrist) but not directly between the extremes (e.g., leftist to rightist) [54]. Here we generalize this model by assuming that extreme agents (either leftists or rightists), (a) when in isolation, can turn into neutral ones with a nonvanishing rate and, (b) when engaging in discussion with neutral agents, are less susceptible to the argument of the latter and are more likely to persuade the latter to become extreme. The resulting model is what we refer to as the I-voter model. In comparison with the constrained three-state voter model, the decaying effect of involvement (that always results in a nonzero fraction of centrists) makes the I-voter different from the latter model whose outcome can be either consensus or polarization with a *frozen* mixture of leftists and rightists but without any centrists. Such a polarized configuration is no longer an absorbing state in the I-voter model. The I-voter can thus be viewed as a "minimal" extension of the constrained three-state voter model that retains neutral opinions at stationarity without invoking any mechanisms hindering consensus like other variants of the voter model (VM), such as individual stubbornness, partisanship, nonlinear update, zealots, or individual and social heterogeneity [55–64].

In the present paper, we formulate an analytical framework for a general network topology that allows for a mean-field treatment of the model's behavior and then demonstrate the validity of our approach on different network topologies. We will show that on sparse networks, the I-voter model yields increased polarization with a growing number of interaction partners. We first describe and present results regarding the baseline model of three states and then provide a generalization to the case of an arbitrary, but odd, number of states.

II. THE MODEL

In the model, N agents, each residing at a node in a social network, hold an opinion $x_i \in \{-1, 0, 1\}$ that stands for "leftist," "centrist," and "rightist," respectively. When a leftist (rightist) and a centrist are in contact, the latter becomes left (right) with probability (per unit time) p, while an extremist turns centrist with probability $1-p$. The probability p represents combined effects of interactions leading to persuasion, while $1-p$ represents reading or hearing other opinions that

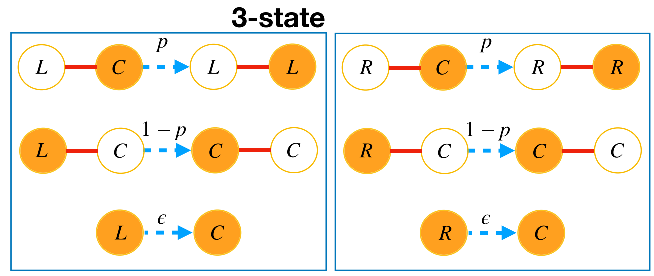

FIG. 1. Illustration of the three-state I-voter dynamics. Circles with the labels C, L, and R denote the centrists, leftists, and rightists, respectively. Lines indicate the interactions between two connected agents, while dashed arrows depict how the *highlighted* agent changes their opinion upon interaction. The updates that are independent of the agent interactions include the decay of leftist (rightist) to centrist. The parameters p, $1-p$, and ϵ are the respective rates of opinion updates.

will make someone's opinion more neutral. As centrists are expected to be more easily influenced, their transition rate is necessarily larger than that of extremists, i.e., $p > 1-p$. We thus only consider the case of $p > 0.5$. Furthermore, involvement (essentially attention) is a limited resource, meaning extremists may lose interest in the discussed issue and gradually become centrists, or there could be individual-specific effects, such as memory or disturbances (forgetting), or interference with other memories [65,66]. Thus the extremist, either left or right, can decay toward the center with probability (per unit time) ϵ. While real-world agents are always subjected to stimulation, especially on topics that tend to polarize, allowing involvement to decay results in a nontrivial opinion formation process that is not purely driven by transitions toward more extreme opinions. Figure 1 illustrates the dynamics of our I-voter model.

To implement the opinion formation process, we employ asynchronous updating, in which each agent is assigned its own independent Poisson clock, all with the same unit rate. If the agent is in state 0, then when its Poisson clock rings it changes its state from 0 to $1(-1)$ with probability p if the state of a randomly selected neighbor is right (left); otherwise, it remains unchanged. Similarly, if the agent is in the state 1 (-1), it changes its state to 0 with probability ϵ, regardless of its neighbor's state, and with probability $1-p$ if the state of a randomly selected neighbor is center. We simulate this model by the Gillespie algorithm [67] with implementation provided in Ref. [68].

III. RESULTS

A. The steady-state fraction of centrists

Let ρ_+, ρ_-, and ρ_0 denote the densities of rightists, leftists, and centrists, respectively. In Appendix B, we show that, for a *fully connected* network and in the limit of *infinite* system size $N \to \infty$, if $a := 2p - \epsilon - 1 > 0$, then the population reaches a steady state where the fraction ρ_0 of centrists is given by

$$\rho_0^* = \frac{\epsilon}{2p-1}. \quad (1)$$

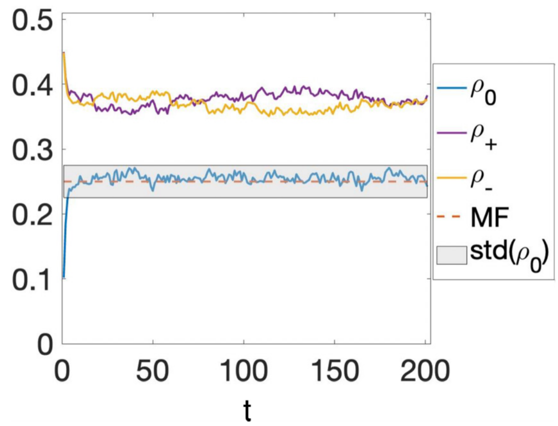

FIG. 2. The density of opinions as function of time. The mean-field (MF) prediction ρ_0^* is depicted by the dashed red line. Stochastic trajectories are generated by the Gillespie algorithm for $N = 100$, and then averaged over 100 independent runs on an all-to-all network for $\epsilon = 0.1$ and $p = 0.7$. The fraction of centrists converges to $\rho_0^* = \epsilon/(2p-1) = 0.25$. The fractions of centrists, rightists, and leftists are denoted by ρ_0, ρ_+, and ρ_- respectively. The shaded gray area depicts the standard deviation derived from the mean-field approximation of the full dynamics as given in Eq. (2).

We stress that a full parameter scanning over all combinations of p and ϵ will result in two phases, $\rho_0^* = \epsilon/(2p-1)$ for $a > 0$ and $\rho_0^* = 1$ (i.e., a society consisting of only centrists) for $a < 0$. Since we are not interested in the latter phase without any extremists, throughout the paper, we only consider the case of $a > 0$, representing the case that the probability to move to the center is much smaller than the probability to become more extreme. Next, for a system of finite size N, where finite-size fluctuations need to be taken into account, we first represent the model as a chemical reaction network and then use a continuous-time Markov Chain to describe the evolution of the distribution of different opinions considered as chemical species. Our approximation is based on a truncation of the moment hierarchy associated with this distribution up to the second order. This yields $\rho_0^* = \langle\rho_0\rangle_*$, where $\langle\cdot\rangle_*$ denotes averaging taken by the stationary distribution and with slight abuse of notation, and ρ_0 denotes the fraction of centrists in a single realization of the model dynamics. Using the same approximation scheme in Appendix C, we also obtain the variance of ρ_0 in the steady state:

$$\mathrm{Var}(\rho_0) := \langle\rho_0^2\rangle_* - \langle\rho_0\rangle_*^2 = \frac{\epsilon}{2(2p-1)}\frac{1}{N}. \tag{2}$$

This shows how fluctuations due to finite size alter the mean-field prediction. In particular, either a high decay rate toward the neutral state or a low persuasion results in increased variance.

Figure 2 demonstrates typical random trajectories of the I-voter model. Here, in agreement with the mean-field prediction, we find $\langle\rho_0\rangle_* = 0.25$ for $(p, \epsilon) = (0.7, 0.1)$. Note that because of the symmetry in the dynamical laws for leftists and rightists, we always obtain a statistical equality between the fraction of leftists and that of rightists if started from an unbiased initial condition with the same number. This is observed in Fig. 2 for ρ_+ and ρ_-. In addition, we also verify Eq. (2), where fluctuations are observed to be within the shaded area bounded by two bands $\langle\rho_0\rangle_* \pm \mathrm{std}(\rho_0)$, i.e., within one standard deviation of the mean-field solution [Eq. (1)].

B. The role of network connectivity

In a social network $\mathcal{G}$ with adjacency matrix $A_{ij} \in \{0, 1\}$, two agents i and j are connected if $A_{ij} = 1$, and they do not interact if $A_{ij} = 0$. Let $\mathcal{V}$ denote the set of nodes in $\mathcal{G}$. For every node $i \in \mathcal{V}$, we consider its local neighborhood $\partial_i := \{j \in \mathcal{V} : A_{ij} = 1\}$ consisting of its nearest neighbors only. A node i's degree then is given by the number of its neighbors, $\kappa_i := \sum_{j\in\partial_i} A_{ij}$. The level of connectedness in society is quantified by the average number of connections per node: $\kappa = N^{-1}\sum_i \kappa_i$. To study the effect of network connectivity on the opinion distribution, we consider N agents; each has a probability of flipping its opinion depending on the states of its nearest neighbors in $\mathcal{G}$.

Let $\mathbb{P}(\mathbf{x}, t)$ denote the joint distribution to observe a configuration $\mathbf{x} := (x_1, x_2, \dots, x_N)$ at time t. Appendix D provides details of how this distribution evolves according to a master equation [Eq. (D7)], whose transition rates $\mathbf{W}(\mathbf{x}'|\mathbf{x})$ from $\mathbf{x}$ to $\mathbf{x}'$ between t and $t + dt$ are given in Eqs. (D3)–(D6). This master equation is not solvable in general, so to construct a mean-field theory for our model, we introduce the averaged dynamical variable $\sigma_i(t)$, defined as the probability that node i is *not* a centrist at time t,

$$\sigma_i(t) := \sum_{\{\mathbf{x}\}} \mathbb{P}(\mathbf{x}, t)\big[\delta_{x_i,1} + \delta_{x_i,-1}\big], \tag{3}$$

and the probability $\rho_i^{(0)}(t)$ that a node i is a centrist at time t,

$$\rho_i^{(0)}(t) := \mathbb{E}\big[\delta_{x_i,0}\big] = \sum_{\{\mathbf{x}\}} \mathbb{P}(\mathbf{x}, t)\delta_{x_i,0} = 1 - \sigma_i(t), \tag{4}$$

where $\delta_{x,y}$ is Kronecker's delta and the sum $\sum_{\{\mathbf{x}\}}$ is carried over the entire phase space of 3^N configurations. In Appendix D, we derive from Eq. (D7) the following set of N *approximate* mean-field equations for σ_i, which measure i's averaged extremeness:

$$\frac{d\sigma_i}{dt} = -\epsilon\sigma_i(t) + \frac{p\rho_i^{(0)}}{\kappa_i}\sum_{j\in\partial_i}\sigma_j(t) - \frac{1-p}{\kappa_i}\sigma_i(t)\sum_{j\in\partial_i}\rho_j^{(0)}. \tag{5}$$

To quantify the level of polarization, we introduce the following measure:

$$\mathcal{P} = 1 - \frac{1}{N}\sum_i \rho_i^{(0)}(t) - \left(\frac{1}{N}\sum_i \mu_i\right)^2, \tag{6}$$

where

$$\mu_i(t) := \sum_{\{\mathbf{x}\}} \mathbb{P}(\mathbf{x}, t)\big[\delta_{x_i,1} - \delta_{x_i,-1}\big]. \tag{7}$$

This measure is in line with the idea that polarized societies typically lack a neutral attitude as common ground for global consensus and have a high variance of opinions [69]. If the probability of being centrist for any individual is low (for instance, a small fraction of respondents who chose the middle

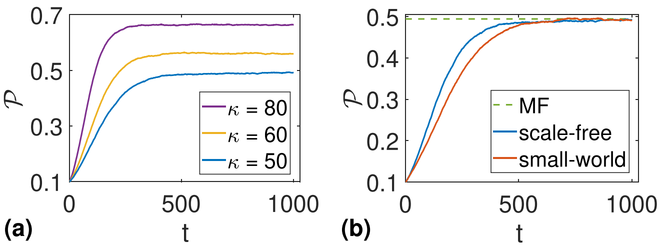

FIG. 3. (a) The polarization measure $\mathcal{P}$ in scale-free graphs with the exponent 2.8 and various mean degree $\kappa = 50, 60, 80$. (b) $\mathcal{P}$ in social networks with fixed $\kappa = 50$ but different topologies. Continuous lines are stochastic trajectories generated by the Gillespie algorithm for $N = 2000$ and then averaged over 100 independent runs at fixed $\epsilon = 0.1$ and $p = 0.7$. The initial fractions of leftists and rightists are equal to 0.05 in all these runs. The dashed line depicts the steady-state value obtained from the MF solutions of Eqs. (4) and (5) with suitable initial conditions.

category in an opinion poll), and it is equally likely to be either left or right, $\mathcal{P}$ will have a high value. So for $\mathcal{P} \in [0, 1]$, $\mathcal{P} = 0$ means no polarization and $\mathcal{P} = 1$ indicates the highest polarization level—this latter case corresponds to a population containing, on average, as many rightists as leftists. We remark that $\mathcal{P}$ can be considered a discrete version of the well-established *polarization index* introduced in Ref. [70] when the distribution of positive and that of negative opinions collapse into δ functions centered at 1 and -1, respectively.

Real social networks are often small worlds and structurally heterogeneous with an abundance of well-connected nodes (hubs) [71–75]. To capture the effect of such heterogeneity, in Fig. 3(a) we fix $p = 0.7$ and $\epsilon = 0.1$ and investigate scale-free networks of $N = 2000$ with the exponent 2.8 at varying mean degree κ generated by the Chung-Lu random graph model [76]. In Fig. 3(b) we next test if the particular network structure has an influence on the results by fixing $\kappa = 50$ and comparing scale-free with small-world topology gererated using the Watts-Strogatz parameter r—the probability to reconnect a link from any node to any other node in the network [71] ($r = 0$ corresponds to ring networks, and $r = 1$ to Erdős-Rényi graphs). Apart from the speed of convergence, there is no marked difference in terms of the steady-state behavior between these different topologies. This steady state can be decently predicted by our MF approximation in Eq. (5) as shown by the dashed line in Fig. 3(b). We note that the mean-field solutions can depend strongly on the initial conditions. We discuss this point in detail in Appendix D.

In Fig. 4, we compare our mean-field predictions with simulations on networks of varying average degrees κ and rewiring probability r. We obtain good agreement for dense networks (i.e., $\kappa = O(N)$), but deviations as the network becomes sparser. This can already be observed for $\kappa = 60$. Overall, both simulations and mean-field predictions show that as κ increases, $\mathcal{P}$ increases, indicating that polarization level rises with increasing connectivity. To check whether this behavior remains robust with variations in p and ϵ, provided that $a = 2p - \epsilon - 1 > 0$, we compute the phase diagram of $\mathcal{P}$ by means of Gillespie simulation in Fig. 5(a) and by our MF approximation in Eq. (5) in Fig. 5(b). We

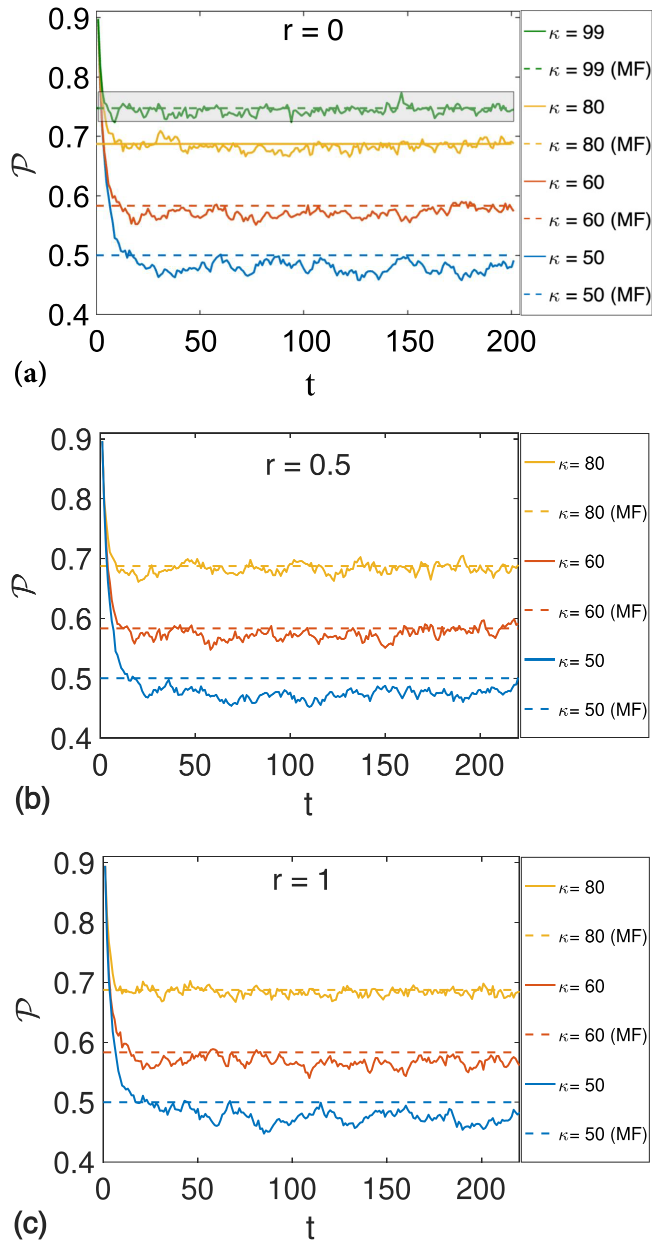

FIG. 4. The polarization measure $\mathcal{P}$ for (a) a social network with a ring topology ($r = 0$); (b) a Watts-Strogatz small-world network ($r = 0.5$); and (c) an Erdős–Rényi random graph ($r = 1$), where r is the rewiring probability [71], all with various degrees κ. $\mathcal{P}$ increases with increasing κ, showing polarization level rises up in more connected social networks. Dashed lines depict the MF prediction according to Eqs. (4) and (5). Continuous lines are stochastic trajectories generated by the Gillespie algorithm for $N = 100$, and then averaged over 100 independent runs. The shaded gray area depicts the standard deviation derived from the mean-field approximation of the full dynamics as given in Eq. (2). Here $\epsilon = 0.1$, $p = 0.7$; the initial fractions of leftists and rightists are equal to 0.45.

propose to use the ratio $\epsilon/(2p - 1)$ as an effective parameter controlling the level of involvement that intuitively decreases with increasing $\epsilon/(2p - 1)$. Polarization is more likely to occur in a society with highly involved agents: $\mathcal{P}$ vanishes as this ratio increases beyond a critical value, and the faster decay

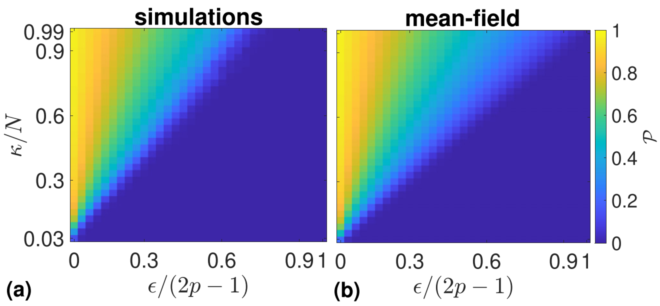

FIG. 5. The polarization measure $\mathcal{P}$ (coded by color: yellow for $\mathcal{P} = 1$ and dark blue for $\mathcal{P} = 0$) (a) for a social network with a ring topology and various degrees computed from simulations and (b) from the MF solution to Eqs. (4) and (5). In both panels, $\mathcal{P}$ is shown as a function of κ/N, the average degree scaled by the system size on the y axis, and $\epsilon/(2p-1)$, the effective parameter controlling the level of involvement on the x axis. We fixed $p = 0.7$ and increase ϵ, while keeping $\epsilon/(2p-1) \in [0, 1]$. The level of involvement decreases as this ratio increases. Here $N = 100$ and the initial fractions of leftists and rightists are equal to 0.45 in Gillespie simulations.

of the individual involvement, the lower $\mathcal{P}$ is. Apart from the special case of $\epsilon \to 0$, for a given level of involvement, a high level of polarization, $\mathcal{P} \simeq 1$, can only be achieved at sufficiently large degree κ. We note that our approximation qualitatively reproduces the boundary between polarized and nonpolarized phases, but it becomes more inaccurate as $\epsilon/(2p-1)$ increases. We remark that our results remain robust with respect to the inclusion of noise as shown in Appendix F. Moreover, we show in Appendix G that our key result of increased polarization with the mean degree is robust with respect to changing the measure of polarization. Specifically, when using the measure considered in Ref. [38] we find a qualitatively similar phase diagram to the one presented in Fig. 5.

C. *n*-state model

A natural extension of the three-state I-voter model is to include two extra states $x_i = +2$ and $x_i = -2$ that we call the five-state I-voter model. Here (i) decay means that an agent moves to an opinion state that is one level less extreme with probability (per unit time) ϵ; (ii) persuasion can happen only when $|x_i - x_j| = 1$ so that (without loss of generality, we consider $|x_i| < |x_j|$) either x_i goes one level more extreme with probability p or x_j goes one level less extreme with probability $1 - p$; and (iii) the reinforcement of extreme opinions can only occur between similar agents following their interaction, so if $x_i = x_j = \pm 1$, then both become one level more extreme with probability γ. The parameter γ describes an increased likelihood of moving toward more extreme opinions when individuals engage in discussion with like-minded others. This phenomenon is known as group polarization [77–81]. For example, in the so-called French jury study [82], French participants who already had a favorable attitude toward then-president Charles de Gaulle were asked to discuss their opinions in small groups. After the group discussion, their positive opinions became even more positive. Similarly, participants who disliked American foreign policy became even more negative about it after discussing it with like-minded others. Other evidence of this mechanism has been reported recently in online platforms, such as Reddit and Gab [83]. Therefore, we note that the implementation of the γ-based mechanism requires a two-body interaction, whereas that based on ϵ is a one-body effect. As a result, the effectiveness of the former is determined by the mean number of connections, κ, while the latter is independent of κ. Adding pairs of states $x_i = \pm 3$, $x_i = \pm 4$, and so on, while using the same rules for the five-state model, results in the seven-state model, nine-state model, and so on. Figure 6(a) illustrates the five-state I-voter model with $x_i = -2$ denoted by L_2 and $x_i = 2$ denoted by R_2.

In Fig. 6(b) we observe that while the steady-state fraction of centrist is invariant with respect to the introduction of γ and two extra states, the underlying dynamics change in comparison to the three-state I-voter model as shown in the inset. Here, ρ_+ and ρ_- both relax to values close to zero (but strictly positive as long as $\epsilon > 0$), while the densities of R_2 and L_2, denoted ρ_{2+} and ρ_{2-}, respectively, reach significantly higher values, indicating the emergence of more extreme opinions under the strong influence of γ. In Appendix E we derive the independence of ρ_0^* on γ within the mean-field description as well as by truncating at the second order in the moment hierarchy. Next, we generalize the use of the polarization measure $\mathcal{P}$ proposed in Eq. (6) to the n-state model. To this end, we modify the expressions for $\rho_i^{(0)}$, σ_i, and μ_i as follows: $\rho_i^{(0)}(t) := \mathbb{E}[\delta_{x_i,0}] = 1 - \sigma_i(t)$ and

$$\sigma_i(t) := \sum_{\{\mathbf{x}\}} \mathbb{P}(\mathbf{x}, t)\left\{\delta_{x_i,\,|x_i|} + \delta_{x_i,-|x_i|}\right\},$$

$$\mu_i(t) := \sum_{\{\mathbf{x}\}} \mathbb{P}(\mathbf{x}, t)\left\{\delta_{x_i,|x_i|} - \delta_{x_i,-|x_i|}\right\}. \quad (8)$$

In Fig. 6(c) we confirm a similar increase of $\mathcal{P}$ with increasing κ in this case. Given that the measure $\mathcal{P}$ depends on the joint distribution of all agents across five distinct states, it is nontrivial to see how an invariant fraction of centrists alone can lead to the same increase of polarization with the average degree κ. For now, we only remark that at the mean-field level, ρ_0^* can be shown to be independent of γ for any n-state I-voter. This suggests our result on polarization is general and is expected to go well beyond the three-state and five-state cases.

IV. DISCUSSION

Modern societies are characterized by unprecedented interconnectedness due to technological advancements, particularly social media platforms. These platforms facilitate rapid and widespread dissemination of information and opinions, significantly altering the dynamics of social interactions. In this context, the level of individual engagement, or involvement, becomes a critical factor in understanding how opinions are formed and sustained. In this regard, we proposed a way to understand the rise of opinion polarization in increasingly connected societies as a consequence of the joint effect of involvement characterized by (p, ϵ) and the mean degree κ. Here p quantifies the propensity to extremism due to the

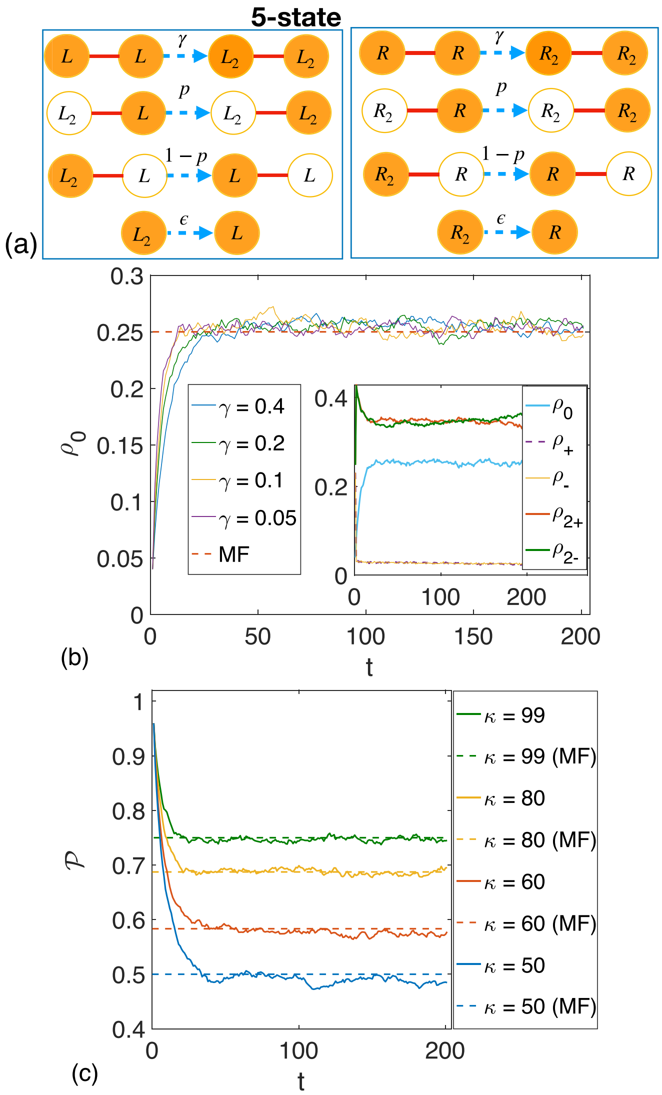

FIG. 6. (a) Illustration of the five-state I-voter dynamics. Circles with the legend L_2 and R_2 denote the state with $x_i = -2$ and $x_i = 2$, respectively. Lines indicate the interactions between two connected agents, while dashed arrows depict how the *highlighted* agent changes their opinion upon interactions. The updates that are independent of the agent interactions include the decay of an L_2 (R_2) agent to leftist (rightist). (b) The fraction of centrists ρ_0 in the five-state model in an all-to-all graph for $\gamma = 0.05, 0.1, 0.2, 0.4$, where "MF" denotes the mean-field prediction $\rho_0^* = \epsilon/(2p-1) = 0.25$ and is depicted by the dashed red line. Inset: The density of different opinions for $\gamma = 0.2$. The fraction of rightists (leftists) and that of $x_i = +2$ ($x_i = -2$) are denoted by ρ_+ (ρ_-) and ρ_{2+} (ρ_{2-}), respectively. (c) The polarization measure $\mathcal{P}$ of the five-state model as defined in Eq. (6) but with ($\rho_i^{(0)}, \mu_i$) given in Eq. (8), for varying degrees κ with fixed $\gamma = 0.2$. In (b) and (c), stochastic trajectories are generated by the Gillespie algorithm for $N = 100$, and then averaged over 100 independent runs for $\epsilon = 0.1$, $p = 0.7$.

interaction between an extremist and a centrist, while ϵ is the rate of decay of an extremist toward centrism.

We found that, for fixed values of p and ϵ, denser networks exhibit higher levels of polarization. This is shown to be the case in both the three-state and five-state I-voter models but is expected to hold for n-state dynamics with $\gamma > 0$ capturing a tendency of extremists to become even more extreme after discussion with like-minded others. These results are in qualitative agreement with recent empirical findings [39,40]. A consequence of these findings is that an increase in social relations, either in person or virtual, may lead to polarization while a decrease in social relations may lead to depolarization. For future work, we would investigate intervention strategies to shift the system between polarized and neutral states. Reducing p and increasing ϵ can decrease polarization. Centrists should be more resistant to extremist arguments, and involvement of extremisms should diminish more rapidly.

We note the following limitations. While the assumptions and predictions of our model align with a significant portion of the empirical literature (see main text for references), it does not yet offer quantitative predictions. A first step would be to estimate model parameters from a real dataset [14,84,85]. For the sake of analytical treatment, we have studied only homogeneous populations of agents with the same parameters p and ϵ, neglecting possible important effects of heterogeneity in the model parameters. A natural step then is to consider the case where each agent is characterized by individual values of p_i and ϵ_i. In this case, there might exist multiple stable steady states induced by individual heterogeneity. For this, one would compute the mean first-passage (convergence) time to reach a given steady state and the attractor-switching time.

Next, it is worth exploring the effect of antagonistic ties [86], which have been shown to play an important role in mitigating ideal polarization within village networks [87]. A reduction of opinion polarization by incidental similarities, i.e., shared personal traits between those individuals who hold different opinions on a political issue, has recently been found in Ref. [88]. Therefore, it would be interesting to include demographic and biographical features, such as age, gender, language, nationality, and personal interests, into our model and study how these features affect the ideological dimension. This will facilitate comparisons with the large-scale experiment of Ref. [88] and the Axelrod model's prediction [21].

Finally, we note that the psychological theory underlying the I-voter model has been recently extended in Ref. [89], where a new parameter, caution, which allows for involved neutral positions, is introduced. This involved neutral stance opens new depolarization pathways even in highly connected networks, suggesting that extending the I-voter model to incorporate this controlling parameter is a promising direction for future work.

ACKNOWLEDGMENTS

We thank Ben Meylahn and Wout Merbis for helpful comments. This work was supported in part by the Dutch Institute for Emergent Phenomena (DIEP) cluster at the University of Amsterdam under the Research Priority Area "Emergent Phenomena in Society: Polarisation, Segregation and Inequality" and the programme Foundations and Applications of Emergence (FAEME).

DATA AVAILABILITY

The data that support the findings of this article are openly available [68], embargo periods may apply.

APPENDIX A: RELATED MODELS

The hierarchical Ising opinion model (HIOM) [51] is a complex cascading transition model that captures the interplay between individual dynamics and polarization across individuals. The HIOM conceptualizes an agent's individual attitude as a network of beliefs, feelings, and behaviors toward an issue [46,90]. The alignment of nodes in an individual's attitude network depends on involvement. In lowly involved agents attitudes are weak and inconsistent, while highly involved agents develop extreme opinions. Changes in information (the external field) can lead to sudden jumps and hysteresis. In the HIOM, involvement plays a double role. First, agents with high involvement initiate more interactions and are more persuasive than less-involved ones. Second, involvement generally decays but increases due to interactions. Therefore, similar to the I-voter model, polarization increases in highly connected societies. However, due to the complexity of the setup, an analytical treatment of this effect is not feasible.

The constrained three-state voter model on all-to-all graphs [52] features a steady state, in which either no neutral opinion exists or a consensus on only one of the three opinions is reached. This means that the first kind of steady states of this model can be considered as the $\epsilon \to 0^+$ limit of the I-voter dynamics which also relaxes to a stationary mixture of leftists and rightists, but without any centrists. However, due to the decaying effect of involvement that turns extreme agents to neutral ones at a rate $\epsilon > 0$, configurations in the I-voter model always include some fraction of centrists, making it different from the constrained three-state voter model even in the mean-field limit. Another variant of the constrained voter model [91] features a "multi-opinion" phase in the mean-field limit similar to ours, but this phase does not persist in finite populations due to demographic fluctuations.

APPENDIX B: DERIVATION OF EQ. (1)

The I-voter model with three states is described by a set of two ordinary differential equations (ODEs) for ρ_- and ρ_+ (as $\rho_+ + \rho_- + \rho_0 = 1$) according to mass-action kinetics:

$$\begin{aligned}\dot{\rho}_+ &= (2p-1)\rho_+\rho_0 - \epsilon\rho_+\\ \dot{\rho}_- &= (2p-1)\rho_-\rho_0 - \epsilon\rho_-.\end{aligned} \tag{B1}$$

By introducing $y = \rho_+ + \rho_-$ and $a = 2p - 1 - \epsilon$, we get

$$\dot{y} = ay\Big(1 - \frac{y}{K}\Big), \quad K = \frac{2p-1-\epsilon}{2p-1}. \tag{B2}$$

This takes the same form as the logistic equation that describes the growth of a species with density $y(t)$ at a rate ay and decay ay^2/K with the *rescaled* carrying capacity in a given neighborhood $K < 1$. When $a > 0$, the stable fixed point of the above dynamics is $y_* = K$. Since we are only interested in physical solutions with positive value, we consider only those pairs of (p, ϵ) that satisfy $2p - 1 > \epsilon$. From the conservation law $\rho_0^* + y_* = 1$, we obtain

$$\rho_0^* = \frac{\epsilon}{2p-1}.$$

APPENDIX C: DERIVATION OF EQ. (2)

We remark that by mapping the dynamical rules of I-voter updates onto a chemical reaction network scheme with three chemical species L (leftist), R (rightist), and C (centrist), Eqs. (B1) can be derived as the $N \to \infty$ limit of the underlying master equation describing the evolution of the reactant concentrations. The set of reactions for the model reads

$$\begin{aligned}&C + L \xrightarrow{k_1} L + L\\ &L + C \xrightarrow{k_2} C + C\\ &L \xrightarrow{k_3} C\\ &C + R \xrightarrow{k_4} R + R\\ &R + C \xrightarrow{k_5} C + C\\ &R \xrightarrow{k_6} C,\end{aligned} \tag{C1}$$

where $k_1 = k_4 = p$, $k_2 = k_5 = 1 - p$, and $k_3 = k_6 = \epsilon$. Our chemical reaction network formulation of the opinion dynamics is inspired by Ref. [92] and similar in spirit to Ref. [93]. We start by writing the quasi-Hamiltonian H (i.e., the infinitesimal-time generator) for the master equation $\partial_t \mathcal{P} = H\mathcal{P}$, where for a *discrete* probability distribution $w_{\mathbf{n}}(t) := w(n_L(t), n_R(t), n_C(t))$ we introduce the associated generating function $\mathcal{P}(t, \mathbf{z}) = \sum_{\mathbf{n}} w_{\mathbf{n}}(t) z_L^{n_L} z_R^{n_R} z_C^{n_C}$, with the shorthand notations $\mathbf{n} := (n_L, n_R, n_C)$ and $\mathbf{z} := (z_L, z_R, z_C)$, for n_L, n_R, and n_C—the number of leftists, rightists, and centrists, respectively. Following Refs. [94,95], this Hamiltonian reads

$$\begin{aligned}H =\,& k_2\{[(a_C^\dagger)^2 - a_L^\dagger a_C^\dagger]a_L a_C + [(a_C^\dagger)^2 - a_R^\dagger a_C^\dagger]a_R a_C\}\\ &+ p[(a_L^\dagger)^2 - a_L^\dagger a_C^\dagger]a_L a_C + p[(a_R^\dagger)^2 - a_R^\dagger a_C^\dagger]a_R a_C\\ &+ \epsilon[a_C^\dagger - a_L^\dagger]a_L + \epsilon[a_C^\dagger - a_R^\dagger]a_R,\end{aligned} \tag{C2}$$

where we have introduced the creation and annihilation operators for the leftists, $a_L^\dagger$ and a_L, as well as their counterparts $a_R^\dagger$ ($a_C^\dagger$) and a_R (a_C) for the rightists (centrists). Now let us introduce the number operators $\hat{N}_L = a_L^\dagger a_L$, $\hat{N}_R = a_R^\dagger a_R$, and $\hat{N}_C = a_C^\dagger a_C$. Taking the derivatives of the generating functions, we can evaluate the averaged number of leftists as

$$\frac{d}{dt}\langle n_L\rangle = \frac{d}{dt}(\hat{N}_L \mathcal{P}|_{\mathbf{z}=\mathbf{1}}) \tag{C3}$$

and similarly for the average number of rightists and centrists. The time derivatives of these averages are then given by

$$\begin{aligned}&\frac{d}{dt}\langle n_L\rangle = (2p-1)\langle n_L n_C\rangle - \epsilon\langle n_L\rangle,\\ &\frac{d}{dt}\langle n_R\rangle = (2p-1)\langle n_R n_C\rangle - \epsilon\langle n_R\rangle,\\ &\frac{d}{dt}\langle n_C\rangle = -(2p-1)\langle (n_L + n_R)n_C\rangle + \epsilon\langle (n_L + n_R)\rangle.\end{aligned} \tag{C4}$$

This set of unclosed equations is an example of the typical "moment closure" problem encountered in numerous fields [96], where we need to know $\langle n_L n_C\rangle$ for determining the evolution, for instance, of $\langle n_L\rangle$. One can easily check that the second-order moments depend on the third-order moments, and so on and so forth. For instance, we have for $\langle n_L n_C\rangle$ the following:

$$\frac{d}{dt}\langle n_L n_C\rangle = (2p-1)\big[\big\langle n_L n_C^2\big\rangle - \big\langle n_L^2 n_C\big\rangle - \langle n_L n_R n_C\rangle\big] + \epsilon\langle n_L(n_L-1)\rangle + \epsilon\langle n_L n_R\rangle - \epsilon\langle n_L n_C\rangle. \quad \text{(C5)}$$

Since the total number of agents, $N = n_L + n_R + n_C$, is conserved in this case the last two terms can be expressed as

$$\epsilon\langle n_L n_R\rangle = N\epsilon\langle n_L\rangle - \epsilon\langle n_L n_C\rangle - \epsilon\big\langle n_L^2\big\rangle,$$
$$-\langle n_L n_R n_C\rangle = -N\langle n_L n_C\rangle + \big\langle n_L n_C^2\big\rangle + \big\langle n_L^2 n_C\big\rangle.$$

Substituting these expressions into Eq. (C5), rescaling $p \to p/N$, $1-p \to (1-p)/N$ and introducing the densities $\rho_+ = n_L/N$, $\rho_- = n_R/N$, and $\rho_0 = n_C/N$, we arrive at

$$\frac{d}{dt}\langle \rho_+\rangle = (2p-1)\langle \rho_+\rho_0\rangle - \epsilon\langle \rho_+\rangle$$
$$\frac{d}{dt}\langle \rho_+\rho_0\rangle = \Gamma\langle \rho_+\rho_0\rangle + \Lambda\big\langle \rho_+\rho_0^2\big\rangle + \epsilon\left(1-\frac{1}{N}\right)\langle \rho_+\rangle, \quad \text{(C6)}$$

where $\Gamma := -[2\epsilon + (2p-1)]$ and $\Lambda := 2(2p-1)$. The mean-field limit for the evolution of the *averaged* fraction of rightists $\langle \rho_+\rangle$ in Eq. (B1) is recovered by assuming statistical independence of the densities ρ_+ and ρ_0, resulting in

$$\frac{d}{dt}\langle \rho_+\rangle = (2p-1)\langle \rho_+\rangle\langle \rho_0\rangle - \epsilon\langle \rho_+\rangle, \quad \text{(C7)}$$

from which the mean-field fixed point in Eq. (1) with $\langle \rho_L\rangle_* > 0$ is obtained:

$$\langle \rho_0\rangle_* = \frac{\epsilon}{2p-1}. \quad \text{(C8)}$$

This assumption of statistical independence also allows us to obtain the stationary value of the second moment $\langle \rho_0^2\rangle$ from the second of Eqs. (C6). Indeed, $\langle \rho_0^2\rangle$ satisfies

$$-[2\epsilon + (2p-1)]\langle \rho_0\rangle_* + 2(2p-1)\big\langle \rho_0^2\big\rangle_* + \epsilon\left(1-\frac{1}{N}\right) = 0.$$

Hence,

$$\mathrm{Var}(\rho_0) = \frac{\epsilon}{2(2p-1)}\left[\frac{2\epsilon}{2p-1} + \frac{1}{N}\right] - \frac{\epsilon^2}{(2p-1)^2} = \frac{\epsilon}{2(2p-1)}\frac{1}{N}. \quad \text{(C9)}$$

APPENDIX D: DERIVATION OF EQ. (5)

We here show how Eq. (5) can be derived from a master equation for $\mathbb{P}(\mathbf{x}, t)$ that represents the distribution of chemical species reacting according to the set of reactions in Eqs. (C1). For every node i, we introduce its local fields:

$$h_i^{(0)} := \sum_{j\in\partial_i}\delta_{x_j,0}, \quad h_i^{(+)} := \sum_{j\in\partial_i}\delta_{x_j,1}, \quad h_i^{(-)} := \sum_{j\in\partial_i}\delta_{x_j,-1}. \quad \text{(D1)}$$

Thus, if i has κ_i neighbors, then $h_i^{(0)} + h_i^{(+)} + h_i^{(-)} = \kappa_i$. The master equation for $\mathbb{P}(\mathbf{x}, t)$ reads

$$\frac{1}{N}\frac{d}{dt}\mathbb{P}(\mathbf{x}', t) = \sum_{\{\mathbf{x}\}}\mathbf{W}(\mathbf{x}'|\mathbf{x})\,\mathbb{P}(\mathbf{x}, t) - \mathbb{P}(\mathbf{x}', t), \quad \text{(D2)}$$

where, as we consider that only one agent can change its state at any moment in time, the transition rate $\mathbf{W}(\mathbf{x}'|\mathbf{x})$ from $\mathbf{x} := (x_1, x_2, \ldots, x_i, \ldots, x_N)$ to $\mathbf{x}' := (x_1, x_2, \ldots, x_i', \ldots, x_N)$ is given by

$$\mathbf{W}(\mathbf{x}'|\mathbf{x}) := \frac{1}{N}\sum_{i=1}^{N}\left[\prod_{j=1(\neq i)}^{N}\delta_{x_j,x_j'}\right]\mathcal{F}(x_i'|x_i) \quad \text{(D3)}$$

with the individual rate matrix $\mathcal{F}(x_i'|\{x_i, \mathbf{x}_{\partial_i}\}) \equiv \mathcal{F}(x_i'|x_i)$,

$$\mathcal{F}\big(x_i'|x_i\big) = \begin{pmatrix} (0|0) & (0|1) & (0|-1) \\ (1|0) & (1|1) & 0 \\ (-1|0) & 0 & (-1|-1) \end{pmatrix}, \quad \text{(D4)}$$

subject to a normalization constraint

$$\sum_{x_i'}\mathcal{F}(x_i'|x_i) = \mathcal{F}(x_i|x_i) + \sum_{x_i'(\neq x_i)}\mathcal{F}(x_i'|x_i) = 1 \quad \text{(D5)}$$

and specified explicitly as

$$\mathcal{F}(-1|1) = 0, \quad \mathcal{F}(1|-1) = 0,$$
$$\mathcal{F}(0|0) = 1 - \frac{k_1 h_i^{(+)} + k_4 h_i^{(-)}}{\kappa_i},$$
$$\mathcal{F}(1|0) = \frac{k_1 h_i^{(+)}}{\kappa_i}, \quad \mathcal{F}(-1|0) = \frac{k_4 h_i^{(-)}}{\kappa_i},$$
$$\mathcal{F}(0|1) = \frac{k_2 h_i^{(0)}}{\kappa_i} + \epsilon, \quad \mathcal{F}(1|1) = 1 - \epsilon - \frac{k_2 h_i^{(0)}}{\kappa_i},$$
$$\mathcal{F}(0|-1) = \frac{k_5 h_i^{(0)}}{\kappa_i} + \epsilon, \quad \mathcal{F}(-1|-1) = 1 - \epsilon - \frac{k_5 h_i^{(0)}}{\kappa_i}, \quad \text{(D6)}$$

where $k_1 = k_4 = p$ and $k_2 = k_5 = 1-p$. Denoting the vector of all nodes' states apart from i as $\mathbf{x}_{\setminus i}$, according to Eq. (D3) we have $\mathbf{x}_{\setminus i} = \mathbf{x}'_{\setminus i}$. Now substituting Eq. (D3) into Eq. (D2), we obtain the following explicit form of the master equation:

$$\begin{aligned}\frac{d}{dt}\mathbb{P}(\mathbf{x}', t) &= \sum_{i=1}^{N}\sum_{x_i\in\{-1,0,1\}}[\mathcal{F}(x_i'|x_i)\mathbb{P}(\mathbf{x}'_{\setminus i}, x_i, t)] - N\mathbb{P}(\mathbf{x}', t) \\ &= \sum_{i=1}^{N}\mathcal{F}(x_i'|x_i')\mathbb{P}(\mathbf{x}', t) + \sum_{i=1}^{N}\sum_{\substack{x_i\in\{-1,0,1\}\\ x_i\neq x_i'}}\mathcal{F}(x_i'|x_i) \\ &\quad\times\mathbb{P}(\mathbf{x}'_{\setminus i}, x_i, t) - N\mathbb{P}(\mathbf{x}', t) \\ &= \sum_{i=1}^{N}\sum_{x_i(\neq x_i')}[\mathcal{F}(x_i'|x_i)\mathbb{P}(\mathbf{x}'_{\setminus i}, x_i, t) \\ &\quad - \mathcal{F}(x_i|x_i')\mathbb{P}(\mathbf{x}', t)].\end{aligned} \quad \text{(D7)}$$

Multiplying both sides of this equation by $[\delta_{x_i',1} + \delta_{x_i',-1}]$ and following Eqs. (D6), $\mathcal{F}(1|x_i) \neq 0$ and $\mathcal{F}(-1|x_i) \neq 0$ for $x_i = 0$, and $\mathcal{F}(x_i|1) \neq 0$ and $\mathcal{F}(x_i|-1) \neq 0$ for $x_i = 0$, and we arrive at

$$\frac{d\sigma_i}{dt} = \sum_{\{\mathbf{x}'_{\setminus i}\}}\{[\mathcal{F}(1|0)+\mathcal{F}(-1|0)]\mathbb{P}(\mathbf{x}'_{\setminus i},0,t)-[\mathcal{F}(0|1)\mathbb{P}(\mathbf{x}'_{\setminus i},1,t)+\mathcal{F}(0|-1)\mathbb{P}(\mathbf{x}'_{\setminus i},-1,t)]\}$$
$$= \sum_{\{\mathbf{x}'_{\setminus i}\}}\left\{p\,\frac{h_i^{(+)}+h_i^{(-)}}{\kappa_i}\mathbb{P}(\mathbf{x}'_{\setminus i},0,t)-\left[\epsilon+(1-p)\,\frac{h_i^{(0)}}{\kappa_i}\right][\mathbb{P}(\mathbf{x}'_{\setminus i},1,t)+\mathbb{P}(\mathbf{x}'_{\setminus i},-1,t)]\right\}. \quad \text{(D8)}$$

Equation (D8) is exact but unsolvable. This is because, for example, the first term depends on the correlation between x_i and x_j,

$$\sum_{\{\mathbf{x}'_{\setminus i}\}}[h_i^{(+)}+h_i^{(-)}]\mathbb{P}(\mathbf{x}'_{\setminus i},0,t) = \sum_{\{\mathbf{x}'\}}\mathbb{P}(\mathbf{x}',t)\sum_{j\in\partial_i}\left[\delta_{x'_i,0}\delta_{x'_j,1}+\delta_{x'_i,0}\delta_{x'_j,-1}\right]. \quad \text{(D9)}$$

Assuming statistical independence so that $\mathbb{P}(\mathbf{x}'_{\setminus i},0,t)\simeq\mathbb{P}(\mathbf{x}'_{\setminus i},t)P_i(x_i=0,t)$,

$$\sum_{\{\mathbf{x}'_{\setminus i}\}}[h_i^{(+)}+h_i^{(-)}]\mathbb{P}(\mathbf{x}'_{\setminus i},0,t)\simeq P_i(x_i=0,t)\sum_{\{\mathbf{x}'_{\setminus i}\}}\sum_{j\in\partial_i}\left[\delta_{x'_j,1}+\delta_{x'_j,-1}\right]\mathbb{P}(\mathbf{x}'_{\setminus i},t)=\rho_i^{(0)}(t)\sum_{j\in\partial_i}\sigma_j(t). \quad \text{(D10)}$$

Similarly, assuming $\mathbb{P}(\mathbf{x}'_{\setminus i},1,t)\simeq\mathbb{P}(\mathbf{x}'_{\setminus i},t)P_i(x_i=1,t)$ and $\mathbb{P}(\mathbf{x}'_{\setminus i},-1,t)\simeq\mathbb{P}(\mathbf{x}'_{\setminus i},t)P_i(x_i=-1,t)$ leads to

$$\sum_{\{\mathbf{x}'_{\setminus i}\}}[\mathbb{P}(\mathbf{x}'_{\setminus i},1,t)+\mathbb{P}(\mathbf{x}'_{\setminus i},-1,t)]=\sigma_i(t),$$
$$\sum_{\{\mathbf{x}'_{\setminus i}\}}h_i^{(0)}[\mathbb{P}(\mathbf{x}'_{\setminus i},1,t)+\mathbb{P}(\mathbf{x}'_{\setminus i},-1,t)]=\sigma_i(t)\sum_{j\in\partial_i}\rho_j^{(0)}. \quad \text{(D11)}$$

Using all the above expression allows us to arrive at Eq. (5):

$$\frac{d\sigma_i}{dt}=-\epsilon\sigma_i(t)+\frac{p\rho_i^{(0)}}{\kappa_i}\sum_{j\in\partial_i}\sigma_j(t)-\frac{1-p}{\kappa_i}\sigma_i(t)\sum_{j\in\partial_i}\rho_j^{(0)}. \quad \text{(D12)}$$

Note that without assuming statistical independence, the equation for $\sigma_i(t)$ becomes exactly the same as that for $\langle\rho_+ + \rho_-\rangle$ in Eqs. (C6) in the limit of making all particles indistinguishable (for arbitrarily finite N) as $\delta_{x_i,0}\to\rho_0$, $\delta_{x_i,1}\to\rho_+$, and $\delta_{x_i,-1}\to\rho_-$. As in this case, we would have

$$\frac{d\sigma_i}{dt}=-\epsilon\sigma_i(t)+\frac{p}{\kappa_i}\sum_{j\in\partial_i}\left[\langle\delta_{x_i,0}\delta_{x_j,1}\rangle+\langle\delta_{x_i,0}\delta_{x_j,-1}\rangle\right]$$
$$-\frac{1-p}{\kappa_i}\sum_{j\in\partial_i}\left[\langle\delta_{x_i,1}\delta_{x_j,0}\rangle+\langle\delta_{x_i,-1}\delta_{x_j,0}\rangle\right]. \quad \text{(D13)}$$

We first show in Fig. 7 that solutions to Eq. (5) strongly depend on the initial condition. Specifically, we distinguish between three families of initial conditions. In the first one, all the nodes have the same initial state, whose value is generated at random. In the second one, each of the nodes has its own random initial value. For both cases, we can control the ensemble average over the population $\bar{\sigma}(0)=\sum_i\sigma_i(t=0)/N$ by keeping it at a particular value. In the last family of initial conditions, we fixed this latter value $\bar{\sigma}(0)$ for a fraction of nodes, while varying the fraction of initial centrist nodes. Results of this mean-field approach indicate that when the dynamics starts with all nodes having the same values then the system relaxes quickly to a single steady state. In other cases, heterogeneity in the initial condition can give rise to multiple stationary states. It would be interesting to investigate in future work what is the probability to end up in any of these final steady states when starting from a given initial condition.

APPENDIX E: THE n-STATE I-VOTER MODEL

Let ρ_+, ρ_{2+}, ρ_-, ρ_{2-}, and ρ_0 denote the densities of voters whose states are $x_i=1, x_i=2, x_i=-1, x_i=-2$, and $x_i=0$, respectively. The five-state model (p,ϵ,γ) is given by four extra ODEs:

$$\dot{\rho}_+=(2p-1)[\rho_+\rho_0-\rho_+\rho_{2+}-\gamma\rho_+^2]+\epsilon[\rho_{2+}-\rho_+]$$
$$\dot{\rho}_{2+}=(2p-1)\rho_+\rho_{2+}+\gamma\rho_+^2-\epsilon\rho_{2+}$$
$$\dot{\rho}_-=(2p-1)[\rho_-\rho_0-\rho_-\rho_{2-}]-\gamma\rho_-^2+\epsilon[\rho_{2-}-\rho_-]$$
$$\dot{\rho}_{2-}=(2p-1)\rho_-\rho_{2-}+\gamma\rho_-^2-\epsilon\rho_{2-}. \quad \text{(E1)}$$

Therefore,

$$d(\rho_++\rho_{2+})/dt=(2p-1)\rho_+\rho_0-\epsilon\rho_+,$$
$$d(\rho_-+\rho_{2-})/dt=(2p-1)\rho_-\rho_0-\epsilon\rho_-. \quad \text{(E2)}$$

The fixed point of the dynamics for ρ_0 in the five-state model,

$$\dot{\rho}_0=-\tilde{y}[(2p-1)\rho_0-\epsilon],\quad \tilde{y}=\rho_++\rho_-, \quad \text{(E3)}$$

is the same as in Eq. (1) for $p>1/2$ and is independent of γ. For the five-state model with the two additional states $+2$ and -2, the set of reactions includes the following additional reactions with $k_7=k_8=\gamma$:

$$L_2+L\xrightarrow{k_1}L_2+L_2$$
$$L_2+L\xrightarrow{k_2}L+L$$
$$L_2\xrightarrow{k_3}L$$
$$R_2+R\xrightarrow{k_4}R_2+R_2$$
$$R_2+R\xrightarrow{k_5}R+R$$
$$R_2\xrightarrow{k_6}R$$
$$R+R\xrightarrow{k_7}R_2+R_2$$
$$L+L\xrightarrow{k_8}L_2+L_2. \quad \text{(E4)}$$

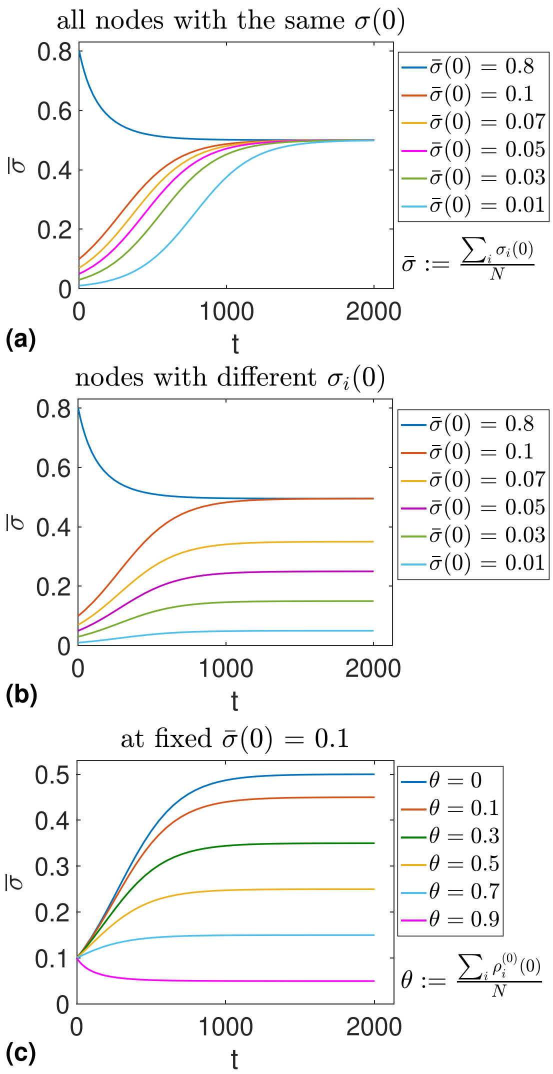

FIG. 7. The averaged extremeness of the population $\bar\sigma(t) := N^{-1}\sum_i \sigma_i(t)$ computed from Eq. (5) for various initial conditions. Here $N = 2000$, $\kappa = 50$, $\epsilon = 0.1$, and $p = 0.7$.

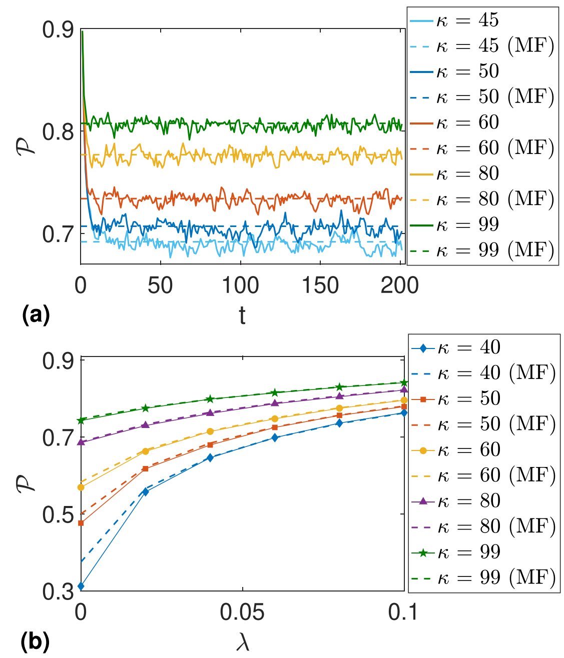

FIG. 8. The polarization measure $\mathcal{P}$ in a social network of varying degrees κ with random flipping of a centrist to either leftist or rightist (a) at rate $\lambda = 0.05$ and (b) at different λ. Continuous lines are stochastic trajectories generated from the Gillespie algorithm for $N = 100$, and then averaging over 100 independent runs. Dashed lines depict the MF prediction according to Eqs. (4) and (5). Here $\epsilon = 0.1$ and $p = 0.7$; the initial fractions of leftists and rightists are equal to 0.45 in Gillespie simulations.

The Hamiltonian in this case reads

$$\begin{aligned} H = {} & p[(a_L^\dagger)^2 - a_L^\dagger a_C^\dagger]a_L a_C + k_2[(a_C^\dagger)^2 - a_L^\dagger a_C^\dagger]a_L a_C \\ & + p\{[(a_R^\dagger)^2 - a_R^\dagger a_C^\dagger] + k_2[(a_C^\dagger)^2 - a_R^\dagger a_C^\dagger]\}a_R a_C \\ & + p\{[(a_{L_2}^\dagger)^2 - a_L^\dagger a_{L_2}^\dagger] + k_2[(a_L^\dagger)^2 - a_L^\dagger a_{L_2}^\dagger]\}a_L a_{L_2} \\ & + p\{[(a_{R_2}^\dagger)^2 - a_R^\dagger a_{R_2}^\dagger] + k_2[(a_R^\dagger)^2 - a_R^\dagger a_{R_2}^\dagger]\}a_R a_{R_2} \\ & + \gamma[(a_{R_2}^\dagger)^2 - (a_R^\dagger)^2](a_R)^2 + \gamma[(a_{L_2}^\dagger)^2 - (a_L^\dagger)^2](a_L)^2 \\ & + \epsilon[a_C^\dagger - a_L^\dagger]a_L + \epsilon[a_C^\dagger - a_R^\dagger]a_R \\ & + \epsilon[a_L^\dagger - a_{L_2}^\dagger]a_{L_2} + \epsilon[a_R^\dagger - a_{R_2}^\dagger]a_{R_2}, \end{aligned} \tag{E5}$$

from which, we can obtain the equations of motion for $\langle n_L\rangle$, $\langle n_{L_2}\rangle$, and $\langle n_L n_C\rangle$:

$$\begin{aligned} \frac{d}{dt}\langle n_L\rangle = {} & (2p-1)(\langle n_L n_C\rangle - \langle n_L n_{L_2}\rangle) + \epsilon(\langle n_{L_2}\rangle - \langle n_L\rangle) \\ & - 2\gamma\langle n_L(n_L-1)\rangle, \\ \frac{d}{dt}\langle n_{L_2}\rangle = {} & (2p-1)\langle n_L n_{L_2}\rangle - \epsilon\langle n_{L_2}\rangle + 2\gamma\langle n_L^2\rangle - 2\gamma\langle n_L\rangle, \\ \frac{d}{dt}\langle n_L n_C\rangle = {} & \epsilon(\langle n_L n_R\rangle + \langle n_C n_{L_2}\rangle) - (1+\epsilon-2\gamma)\langle n_L n_C\rangle \\ & + (2p-1)[\langle n_L n_C^2\rangle - \langle n_L n_R n_C\rangle - \langle n_L n_{L_2} n_C\rangle] \\ & - (2p+2\gamma-1)\langle n_L^2 n_C\rangle + \epsilon\langle n_L(n_L-1)\rangle. \end{aligned} \tag{E6}$$

Following similar calculations to what was used after Eq. (E1) yields

$$\frac{d}{dt}\langle n_C\rangle = -(2p-1)\langle (n_L+n_R)n_C\rangle + \epsilon\langle (n_L+n_R)\rangle. \tag{E7}$$

Dividing both sides by N, as well as assuming statistical independence between n_C, n_L, and n_R, we arrive at the same Eq. (E3) for $\langle\rho_0\rangle = \langle n_C\rangle/N$. This means that the steady-state fraction of centrists, $\langle\rho_0\rangle_*$, is independent of γ and equal to that given in Eq. (C8) if the set of moment equations is closed at the second order. A similar line of analysis shows that this also holds for the n-state model in the mean-field limit.

APPENDIX F: THE ROLE OF NOISE

To test the robustness of our results reported in the main text we introduce a random flip of centrist to either leftist or rightist with probability (per unit time) λ. So λ represents the effect of noise in the system as long as $\lambda \ll \epsilon$. Differently from the noisy voter model [97], we exclude the spontaneous changes from left to right and vice versa. Such noise can arise

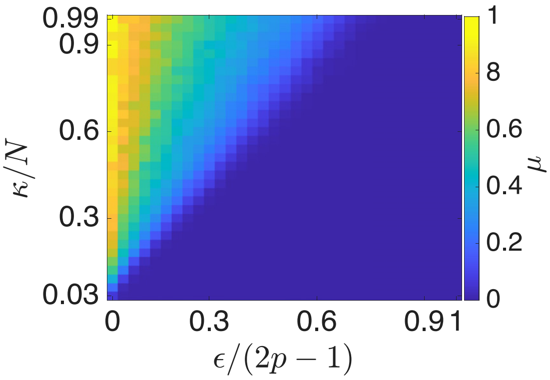

FIG. 9. The polarization measure μ as defined in Eq. (G1) in a social network with a ring topology and various degrees computed from simulations as a function of κ/N, the average degree scaled by the system size on the y axis, and $\epsilon/(2p-1)$, the effective parameter controlling the level of involvement on the x axis. We fix $p = 0.7$ and increase ϵ, while keeping $\epsilon/(2p-1) \in [0, 1]$. The level of involvement decreases as this ratio increases. Here $N = 100$ and the initial fraction of leftists and rightists is equal to 0.45 in simulations.

from many different factors that lead to a random flip of an individual's opinion, regardless of the state of its neighbors. The inclusion of $\lambda > 0$ also prevents the system from reaching an absorbing state of all agents being neutral. The individual rate matrix given in Eqs. (D6) gets modified as follows:

$$\mathcal{F}(0|0) = 1 - \frac{k_1 h_i^{(+)} + k_4 h_i^{(-)}}{\kappa_i} - 2\lambda,$$

$$\mathcal{F}(1|0) = \frac{k_1 h_i^{(+)}}{\kappa_i} + \lambda, \quad \mathcal{F}(-1|0) = \frac{k_4 h_i^{(-)}}{\kappa_i} + \lambda. \tag{F1}$$

Results for fixed $\lambda = 0.05$ on networks of $N = 100$ with various values of κ are presented in Fig. 8(a). Here we confirm that our main result for $\lambda = 0$ (increased polarization in more connected social networks) is robust with respect to the inclusion of $\lambda > 0$. Next, we test the quality of the MF solution for various λ in Fig. 8(b) and find better agreement with the simulations as λ increases. This is expected due to the underlying assumption of the MF approximation which embodies statistical independence. For higher values of λ, the correlations between different agents become weaker, thus justifying the MF assumption. All curves corresponding to different κ merge at high enough λ, when the effect of noise dominates the I-voter dynamics.

APPENDIX G: ANOTHER MEASURE OF POLARIZATION IN THE THREE-STATE MODEL

In order to confirm the robustness of our result with respect to the choice of polarization measure, in Fig. 9 we plot the following measure introduced by Ref. [38]:

$$\mu = \frac{1 - |\rho_+ - \rho_-|}{2}\left(\frac{\rho_+}{\rho_+ + \rho_0} + \frac{\rho_-}{\rho_- + \rho_0}\right). \tag{G1}$$

This alternative measure qualitatively agrees with our measure $\mathcal{P}$ presented in Fig. 5 (with only a minor shift toward the left of the highly polarized region), confirming that our result of increased polarization with increasing degree is robust.

[1] É. Durkheim, *De La Division Du Travail Social: Étude Sur L'organisation Des Sociétés Supérieures* (Alcan, Paris, 1893).

[2] E. Dimant and E. O. Kimbrough, Polarization in multidisciplinary perspective, PNAS Nexus **3**, pgae425 (2024).

[3] R. Basu and A. Sly, Evolving voter model on dense random graphs, Ann. Appl. Probab. **27**, 1235 (2017).

[4] S. Baldassarri, P. Braunsteins, F. den Hollander, and M. Mandjes, Opinion dynamics on dense dynamic random graphs, arXiv:2410.14618.

[5] S. Athreya, F. den Hollander, and A. Röllin, Co-evolving vertex and edge dynamics in dense graphs, arXiv:2504.06493.

[6] A. F. Peralta, J. Kertész, and G. Iñiguez, Opinion dynamics in social networks: From models to data, *Handbook of Computational Social Science* (Edward Elgar, CHeltenham, UK, 2025).

[7] Other models of polarization aim to capture phenomena like opinion distancing [8], persuasive arguments [9], social feedback between internal and expressed opinions [10,11], biased assimilation [12], algorithmic mediation [13], multidimensional (i.e., context-dependent) polarization [14,15], and conflict [16].

[8] W. Jager and F. Amblard, Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change, Comput. Math. Organ. Theor. **10**, 295 (2005).

[9] M. Mäs and A. Flache, Differentiation without distancing. Explaining bi-polarization of opinions without negative influence, PLoS ONE **8**, e74516 (2013).

[10] S. Banisch and E. O. and, Opinion polarization by learning from social feedback, J. Math. Soc. **43**, 76 (2019).

[11] B. V. Meylahn and J. M. Meylahn, How social reinforcement learning can lead to metastable polarisation and the voter model, PLoS ONE **19**, e0313951 (2024).

[12] P. Dandekar, A. Goel, and D. T. Lee, Biased assimilation, homophily, and the dynamics of polarization, Proc. Natl. Acad. Sci. USA **110**, 5791 (2013).

[13] A. F. Peralta, M. Neri, J. Kertész, and G. Iñiguez, Effect of algorithmic bias and network structure on coexistence, consensus, and polarization of opinions, Phys. Rev. E **104**, 044312 (2021).

[14] A. F. Peralta, P. Ramaciotti, J. Kertész, and G. Iñiguez, Multidimensional political polarization in online social networks, Phys. Rev. Res. **6**, 013170 (2024).

[15] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Emergence of polarized ideological opinions in multidimensional topic spaces, Phys. Rev. X **11**, 011012 (2021).

[16] P. Törnberg, E. Olbrich, and J. Uitermark, Editorial: The computational analysis of cultural conflicts, Front. Big Data **5**, 840584 (2022).

[17] M. H. DeGroot, Reaching a consensus, J. Am. Stat. Assoc. **69**, 118 (1974).

[18] R. P. Abelson, Mathematical models of the distribution of attitudes under controversy, in *Contributions to Mathematical Psychology*, edited by N. Frederiksen and H. Gulliksen (Holt, Rinehart and Winston, New York, 1964).

[19] R. A. Holley and T. M. Liggett, Ergodic theorems for weakly interacting infinite systems and the voter model, Ann. Probab. **3**, 643 (1975).

[20] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. **81**, 591 (2009).

[21] R. Axelrod, The dissemination of culture: A model with local convergence and global polarization, J. Conflict Resolut. **41**, 203 (1997).

[22] A. Flache and M. W. Macy, Local convergence and global diversity: From interpersonal to social influence, J. Conflict Resolut. **55**, 970 (2011).

[23] N. Lanchier, The Axelrod model for the dissemination of culture revisited, Ann. Appl. Probab. **22**, 860 (2012).

[24] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, Mixing beliefs among interacting agents, Adv. Complex Syst. **03**, 87 (2000).

[25] R. Hegselmann and U. Krause, Opinion dynamics and bounded confidence models, analysis and simulation, J. Artif. Soc. Soc. Simul. **5**, 1 (2002).

[26] C. Bernardo, C. Altafini, A. Proskurnikov, and F. Vasca, Bounded confidence opinion dynamics: A survey, Automatica **159**, 111302 (2024).

[27] M. Granovetter, Threshold models of collective behavior, Am. J. Sociol. **83**, 1420 (1978).

[28] D. J. Watts, A simple model of global cascades on random networks, Proc. Natl. Acad. Sci. USA **99**, 5766 (2002).

[29] D. Centola and M. Macy, Complex contagions and the weakness of long ties, Am. J. Sociol. **113**, 702 (2007).

[30] P. Holme and M. E. J. Newman, Nonequilibrium phase transition in the coevolution of networks and opinions, Phys. Rev. E **74**, 056108 (2006).

[31] F. Vazquez, V. M. Eguíluz, and M. S. Miguel, Generic absorbing transition in coevolution dynamics, Phys. Rev. Lett. **100**, 108702 (2008).

[32] R. Durrett, J. P. Gleeson, A. L. Lloyd, P. J. Mucha, F. Shi, D. Sivakoff, J. E. S. Socolar, and C. Varghese, Graph fission in an evolving voter model, Proc. Natl. Acad. Sci. USA **109**, 3682 (2012).

[33] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Modeling echo chambers and polarization dynamics in social networks, Phys. Rev. Lett. **124**, 048301 (2020).

[34] C. Altafini, Consensus problems on networks with antagonistic interactions, IEEE Trans. Autom. Control **58**, 935 (2013).

[35] A. Flache and M. W. Macy, Small worlds and cultural polarization, J. Math. Soc. **35**, 146 (2011).

[36] M. T. Pham, I. Kondor, R. Hanel, and S. Thurner, The effect of social balance on social fragmentation, J. R. Soc. Interface **17**, 20200752 (2020).

[37] P. J. Górski, C. Atkisson, and J. A. Hołyst, A general model for how attributes can reduce polarization in social groups, Net. Sci. **11**, 536 (2023).

[38] A. Lipiecki and K. Sznajd-Weron, Depolarizing power of anticonformity, Expert Syst. Appl. **285**, 127879 (2025).

[39] L. G. E. Smith, E. F. Thomas, A.-M. Bliuc, and C. McGarty, Polarization is the psychological foundation of collective engagement, Commun. Psychol. **2**, 41 (2024).

[40] Y. Kazmina, E. M. Heemskerk, E. Bokányi, and F. W. Takes, From contact to threat: A social network perspective on perceptions of immigration, arXiv:2407.06820.

[41] S. Thurner, M. Hofer, and J. Korbel, Why more social interactions lead to more polarization in societies, Proc. Natl. Acad. Sci. USA **122**, e2517530122 (2025).

[42] B. T. Johnson and A. H. Eagly, Effects of involvement on persuasion: A meta-analysis, Psychol. Bull. **106**, 290 (1989).

[43] J. L. Zaichkowsky, Measuring the involvement construct, J. Consum. Res. **12**, 341 (1985).

[44] M. Lalmas, H. O'Brien, and E. Yom-Tov, *Measuring User Engagement* (Springer Nature, Switzerland, 2022).

[45] B. Latané and A. Nowak, Attitudes as catastrophes: From dimensions to categories with increasing involvement, *Dynamical Systems in Social Psychology* (Academic Press, San Diego, CA, 1994), pp. 219–249.

[46] J. Dalege, D. Borsboom, F. van Harreveld, and H. L. J. van der Maas, The attitudinal entropy (AE) framework as a general theory of individual attitudes, Psychol. Inquiry **29**, 175 (2018).

[47] M. Hoffstadt, I. Smal, H. van der Maas, and J. Garcia-Bernardo, Involvement as a polarizing factor?—a comprehensive multimethod analysis across representative datasets, Eur. J. Soc. Psychol. **55**, 193 (2024).

[48] R. E. Petty and J. T. Cacioppo, The elaboration likelihood model of persuasion, *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* (Springer, New York, 1986), pp. 1–24.

[49] R. E. Petty and J. A. Krosnick, Attitude strength: An overview, *Attitude Strength: Antecedents and Consequences* (Psychology Press, New York, 1995), pp. 1–24.

[50] M. L. Richins and P. H. Bloch, After the new wears off: The temporal context of product involvement, J. Consum. Res. **13**, 280 (1986).

[51] H. L. J. van der Maas, J. Dalege, and L. Waldorp, The polarization within and across individuals: The hierarchical Ising opinion model, J. Complex Netw. **8**, cnaa010 (2020).

[52] F. Vazquez and S. Redner, Ultimate fate of constrained voters, J. Phys. A: Math. Gen. **37**, 8479 (2004).

[53] M. Mobilia, Fixation and polarization in a three-species opinion dynamics model, Europhys. Lett. **95**, 50002 (2011).

[54] This latter assumption can be considered as the restriction of a more general scenario, where, at high network connectivity, leftists and rightists can frequently interact with each other, but rarely update their respective opinions.

[55] A. Jędrzejewski and K. Sznajd-Weron, Statistical physics of opinion formation: Is it a SPOOF? C. R. Phys. **20**, 244 (2019).

[56] S. Redner, Reality-inspired voter models: A mini-review, C. R. Phys. **20**, 275 (2019).

[57] M. Mobilia, A. Petersen, and S. Redner, On the role of zealotry in the voter model, J. Stat. Mech. (2007) P08029.

[58] P. G. Meyer and R. Metzler, Time scales in the dynamics of political opinions and the voter model, New J. Phys. **26**, 023040 (2024).

[59] N. Khalil and T. Galla, Zealots in multistate noisy voter models, Phys. Rev. E **103**, 012311 (2021).

[60] G. De Marzo, A. Zaccaria, and C. Castellano, Emergence of polarization in a voter model with personalized information, Phys. Rev. Res. **2**, 043117 (2020).
[61] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, Ordering dynamics of the multi-state voter model, J. Stat. Mech. (2012) P10027.
[62] C. Castellano, M. A. Muñoz, and R. Pastor-Satorras, Nonlinear q-voter model, Phys. Rev. E **80**, 041129 (2009).
[63] P. Nyczka, K. Sznajd-Weron, and J. Cisło, Phase transitions in the q-voter model with two types of stochastic driving, Phys. Rev. E **86**, 011105 (2012).
[64] B. Nowak and K. Sznajd-Weron, Switching from a continuous to a discontinuous phase transition under quenched disorder, Phys. Rev. E **106**, 014125 (2022).
[65] J. M. J. Murre and J. Dros, Replication and analysis of Ebbinghaus' forgetting curve, PLoS ONE **10**, e0120644 (2015).
[66] W. H. Edwards, *Motor Learning and Control: From Theory to Practice* (Cengage Learning, London, 2010).
[67] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. **22**, 403 (1976).
[68] T. Pham, PHYSICSIDEA/Ivoter-model: Code for Ivoter model, version 1.0.0, Zenodo (2026), https://doi.org/10.5281/zenodo.19715368.
[69] A. Bramson, P. Grim, D. J. Singer, S. Fisher, W. Berger, G. Sack, and C. Flocken, Disambiguation of social polarization concepts and measures, J. Math. Soc. **40**, 80 (2016).
[70] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito, Measuring political polarization: Twitter shows the two sides of Venezuela, Chaos **25**, 033114 (2015).
[71] D. Watts and S. Strogatz, Collective dynamics of 'small-world' networks, Nature (London) **393**, 440 (1998).
[72] H. T. Williams, J. R. McMurray, T. Kurz, and F. Hugo Lambert, Network analysis reveals open forums and echo chambers in social media discussions of climate change, Global Environ. Change **32**, 126 (2015).
[73] G. Iñiguez, S. Heydari, J. Kertész, and J. Saramäki, Universal patterns in egocentric communication networks, Nat. Commun. **14**, 5217 (2023).
[74] J. Borge-Holthoefer, A. Rivero, I. García, E. Cauhé, A. Ferrer, D. Ferrer, D. Francos, D. Iñiguez, M. P. Pérez, G. Ruiz, F. Sanz, F. Serrano, C. Viñas, A. Tarancón, and Y. Moreno, Structural and dynamical patterns on online social networks: The Spanish May 15th movement as a case study, PLoS ONE **6**, e23883 (2011).
[75] Y. Zhao, W. Bai, T. Qiao, and W. Wang, Modularity of online social networks acts as a reliable predictor of both whole-network and ego-network characteristics over time, Humanit. Soc. Sci. Commun. **12**, 839 (2025).
[76] D. Fasino, A. Tonetto, and F. Tudisco, Generating large scale-free networks with the Chung–Lu random graph model, Networks **78**, 174 (2021).
[77] S. Y. Lee and J.-H. Kim, What makes people more polarized? The effects of anonymity, being with like-minded others, and the moderating role of need for approval, Telematics Inf. **76**, 101922 (2023).
[78] K. Strandberg, S. Himmelroos, and K. Grönlund, Do discussions in like-minded groups necessarily lead to more extreme opinions? Deliberative democracy and group polarization, Int. Pol. Sci. Rev. **40**, 41 (2019).
[79] S. B. Hobolt, K. Lawall, and J. Tilley, The polarizing effect of partisan echo chambers, Am. Polit. Sci. Rev. **118**, 1464 (2024).
[80] X. Zheng, Y. Lu, J. K. Lee, and J. C. and, Social media news use and polarized partisan perceptions: Mediating roles of like-minded and cross-cutting discussion, J. Inf. Technol. Pol. **22**, 200 (2025).
[81] D. Kuhn, D. Floyd, P. Yaksick, M. Halpern, and W. R. and, How does discourse among like-minded individuals affect their thinking about a complex issue? Think. Reason. **25**, 365 (2019).
[82] S. Moscovici and M. Zavalloni, The group as a polarizer of attitudes, J. Personality Soc. Psychol. **12**, 125 (1969).
[83] M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, The echo chamber effect on social media, Proc. Natl. Acad. Sci. USA **118**, e2023301118 (2021).
[84] M. Galesic and D. Stein, Statistical physics models of belief dynamics: Theory and empirical tests, Physica A **519**, 275 (2019).
[85] A. Vendeville, Voter model can accurately predict individual opinions in online populations, Phys. Rev. E **111**, 064310 (2025).
[86] A. Castro, T. M. Pham, E. Ortega, and D. Machado, Heterophobic interactions hinder consensus formation in sparse random networks, Phys. Rev. E **112**, 064309 (2025).
[87] A. Ghasemian and N. A. Christakis, The structure and function of antagonistic ties in village social networks, Proc. Natl. Acad. Sci. USA **121**, e2401257121 (2024).
[88] S. Balietti, L. Getoor, D. G. Goldstein, and D. J. Watts, Reducing opinion polarization: Effects of exposure to similar people with differing political views, Proc. Natl. Acad. Sci. USA **118**, e2112552118 (2021).
[89] H. L. J. van der Maas, D. Borsboom, and L. Waldorp, The statistical physics of psychological networks: Zero matters, Psychol. Rev. (2026).
[90] J. Dalege, M. Galesic, and H. Olsson, Networks of beliefs: An integrative theory of individual- and social-level belief dynamics., Psychol. Rev. **132**, 253 (2024).
[91] M. Mobilia, Commitment versus persuasion in the three-party constrained voter model, J. Stat. Phys. **151**, 69 (2013).
[92] J. Armas, W. Merbis, J. M. Meylahn, S. Rafiee Rad, and M. J. del Razo, Risk aversion can promote cooperation, J. Phys.: Complexity **6**, 015010 (2025).
[93] J. Holehouse and H. Pollitt, Non-equilibrium time-dependent solution to discrete choice with social interactions, PLoS ONE **17**, e0267083 (2022).
[94] M. J. del Razo, T. Lamma, and W. Merbis, Field theories and quantum methods for stochastic reaction-diffusion systems, Rev. Mod. Phys. **98**, 015001 (2026).
[95] J. Baez and J. Biamonte, *Quantum Techniques in Stochastic Mechanics* (World Scientific, Singapore, 2018).
[96] C. Kuehn, Moment closure—a brief review, in *Control of Self-Organizing Nonlinear Systems*, edited by E. Schöll, S. H. L. Klapp, and P. Hövel (Springer International Publishing, Cham, 2016), pp. 253–271.
[97] B. L. Granovsky and N. Madras, The noisy voter model, Stochastic Process. Appl. **55**, 23 (1995).