Modularity, asymmetry, and polarization shape consensus speed in the voter model

Madi Yerlanov,* Zachary P. Kilpatrick,† and Nancy Rodriguez‡
Department of Applied Mathematics, University of Colorado Boulder, 1111 Engineering Center, Boulder, Colorado 80309, USA

(Received 30 March 2026; revised 25 June 2026; accepted 29 September 2026; published 13 October 2026)

In populations with community structure, the formation of consensus requires both alignment within and diffusion of beliefs across groups, processes that evolve on distinct timescales. How do modularity, asymmetry, and polarization shape this process? We study a variant of the voter model in which a population is divided into two cliques. At each time step, a pair of nodes is selected; if their binary opinions differ, then each agent adopts the opinion of the other with a certain probability. The type of pair selected depends on a network structure parameter—*modularity* that is central in this paper. We analyze how this coupling strength, population imbalance, and initial polarization jointly determine the time to consensus. Formation of consensus generally starts with interclique interactions rapidly synchronizing the two cliques' opinion fractions, after which consensus is reached through a slower diffusion along the synchronized curve. This slow stage is largely insensitive to modularity except when the cliques are nearly disconnected. To analyze these dynamics, we derive time-continuous approximations in the large-population limit and assess their accuracy against the discrete model. While modularity primarily affects the synchronization stage, initially polarized and asymmetric populations exhibit nontrivial effects, including regimes in which an intermediate level coupling minimizes consensus time. A small-clique scaling analysis reveals that this optimum arises from a competition between fast alignment drift and noise amplification in the smaller group.

DOI: 10.1103/br97-fw3p

I. INTRODUCTION

Many systems that display collective decision-making are inherently modular, with interactions occurring more frequently within communities than across them [1,2]. In social and political opinion dynamics, individuals are often organized into geographical regions or interest-based communities, where internal interactions are common while cross-community interactions occur at lower rates, sometimes facilitated by social media [3,4]. Indeed, experiments have shown that clustered network structure can accelerate the spread of new behaviors [5], yet the mechanisms linking modularity to consensus speed remain poorly understood analytically. Classical voter models, which assume uniform mixing, do not capture how such modularity influences consensus formation or the persistence of disagreement [6–8]. Extensions to structured networks have begun to relax these assumptions by incorporating heterogeneous connectivity [9–11], community structure [12,13], and opinion-dependent interaction rules [14,15], but a general understanding of how weak coupling between subgroups shapes fixation remains incomplete.

A parallel and well-developed tradition models opinion formation through kinetic equations of Boltzmann and Fokker-Planck type, originating with [16] and developed in a substantial body of subsequent work [17], including recent extensions that incorporate network structure through graphons [18]. These models typically describe continuous opinions evolving through binary interactions, with the Fokker-Planck description arising from a grazing-collision (quasi-invariant) limit. Our setting differs in that opinions are binary and the dynamics are those of the discrete voter model, for which the Fokker-Planck equation instead emerges as a large-population diffusion limit; nonetheless, both approaches share the goal of characterizing consensus via continuum descriptions of interacting-agent dynamics.

Many ecological populations are divided into subgroups that interact more frequently within their own habitat than with others [19]. Occasionally, individuals move between habitats or observe and copy behaviors from other groups, thereby enabling the spread of behavioral traits, such as foraging strategies or antipredator responses, across an entire population [20]. Over time, one behavior may become *fixed*, meaning that nearly all individuals in both groups adopt the same strategy [11,21]. The interplay between population structure and fixation dynamics has been studied extensively in evolutionary graph theory [22,23], where timescale separation between local equilibration and global fixation plays a central role [24,25].

At a more abstract level, similar dynamics arise in distributed computing systems, where many independent units must coordinate through local communication [26,27]. In such systems, nodes often interact in a modular manner, contacting nearby or related nodes more frequently than most nodes in the network. A binary opinion can represent a local decision, protocol choice, or stored value, and pairwise interactions model information exchange between nodes. In this

*Contact author: madi.yerlanov@colorado.edu
†Contact author: zpkilpat@colorado.edu
‡Contact author: rodrign@colorado.edu

 ©2026 American Physical Society

context, consensus corresponds to global agreement across the network, and the strength of coupling between modules plays a critical role in determining how rapidly such agreement is reached [28].

In this work, we study a tractable, modular extension of the classic voter model [13,29] to quantify how network modularity impacts consensus formation. We consider two connected communities where inter- and intragroup interactions occur at rates governed by a modularity parameter, individuals hold one of two states, and opinion changes arise through random pairwise interactions with a tunable alignment parameter. Voter models on modular and multiclique networks have been studied previously, including effects of external sources [30], the interplay between population size and interclique connectivity [13], and consensus times in structured populations [10–12], with reviews in Refs. [4,29]. Building on this literature, we vary the relative strength of within- and between-community interactions to examine how connectivity and population size shape consensus and fixation times. We analyze these dynamics through agent-based simulations alongside stochastic and continuum approximations [25], systematically relating modular structure to the mechanisms governing consensus.

We begin by introducing a discrete, two-clique voter model and specifying the opinion-update rules that govern interactions within and between communities. From this formulation, we compute the mean and variance of opinion changes and use these quantities to derive a stochastic differential equation approximation in the large-population limit, followed by a fully continuous description in the form of a Fokker-Planck equation. This continuum formulation provides a natural setting for asymptotic analysis, which we use to uncover the mechanisms underlying consensus formation in large systems. In particular, the analysis reveals a separation of timescales, with rapid alignment between cliques followed by slower diffusion toward global consensus. To assess the validity and limitations of these approximations, we complement the analysis with numerical simulations of the discrete, stochastic, and continuum models, comparing their predictions across a range of parameters. These results highlight how modularity, population imbalance, and interaction strength shape consensus times, and motivate further investigation of data-driven and heterogeneous extensions of voter models.

A key take-home message is that modular structure matters most when the system begins far from symmetry. In near-symmetric settings in which the clique sizes and initial opinion fractions are balanced, the well-mixed model is often the fastest route to consensus, and introducing modularity tends to add an alignment step without improving the slow diffusion to fixation. By contrast, strongly polarized initial conditions expose regimes in which nontrivial modularity can be advantageous: intermediate interclique coupling can accelerate the initial equalization of opinion fractions while still allowing efficient drift toward consensus. These effects are especially pronounced in mismatched systems, where a moderately coupled two-module network can enable the larger clique to quickly absorb the smaller one, reducing overall consensus times relative to both extremes of near-disconnection and near-complete mixing. The analysis and simulations that follow make these regimes precise, showing that modularity is neither universally beneficial nor universally detrimental, but instead acts as a context-dependent control parameter whose impact depends critically on initial polarization and population imbalance.

II. MODULAR VOTER MODEL AND MEAN-FIELD LIMITS

In this section, we introduce a modular extension of the classical voter model and specify the discrete interaction rules governing opinion updates within and between communities. Starting from this stochastic description, we derive diffusion and Fokker-Planck approximations that allow us to analyze consensus formation across different coupling regimes.

A. Model definition and interaction rules

We consider a network with N nodes partitioned into two cliques of sizes N_1 and N_2, with $N_1 + N_2 = N$. Each node holds a binary opinion, denoted as -1 or $+1$ (equivalently, A and B), and we use them interchangeably throughout this paper. Let $x_1 := x_1(t)$ and $x_2 := x_2(t)$ represent the proportions of nodes holding the opinion -1 in cliques 1 and 2. We denote the corresponding initial conditions by $y_1 := x_1(0)$ and $y_2 := x_2(0)$. The dynamics proceed in discrete time steps of duration $\Delta t = 2/N$. At each step, an edge is selected uniformly at random from the network, with the relative frequencies of intra- and interclique edges governed by a modularity parameter $\alpha \in [0, 1]$. This construction should be understood as a mean-field interaction rule: all node pairs are in principle possible, but intra- and interclique pairs are sampled with different weights determined by α, rather than through a fixed adjacency structure. Let E_1, E_2, and E_3 denote the (expected) numbers of edges within clique 1, within clique 2, and between cliques,

$$E_1 = \frac{\alpha}{2} N_1(N_1 - 1),$$
$$E_2 = \frac{\alpha}{2} N_2(N_2 - 1),$$
$$E_3 = (1 - \alpha) N_1 N_2.$$

Let $E = E_1 + E_2 + E_3$, and define $\gamma_i = E_i/E$, the probability that a randomly selected edge is of type i. If the selected edge connects two nodes holding opposite opinions, then each node independently switches its opinion with probability p; if both nodes hold the same opinion, then no update occurs. We refer to this parameter as *flip/switch probability*. Pair selection is governed by the *modularity parameter* α, which controls the relative frequency of within- versus between-clique interactions. The flip probability p reflects that an interaction is a mutual exchange and probabilistic update. Both agents may revise their opinions, in contrast to standard dynamics in which a single updating node adopts a sampled neighbor's opinion [31]. Realistically, an agent neither always nor never adopts the belief of a neighbor it disagrees with, so the plausible regime is intermediate, $0 < p < 1$, with the deterministic endpoints as limiting special cases. Such probabilistic, partial revision is the natural assumption for social, animal, and artificial agents and is standard in opinion-dependent update rules [14,15,29]. The two endpoints also behave degenerately, since

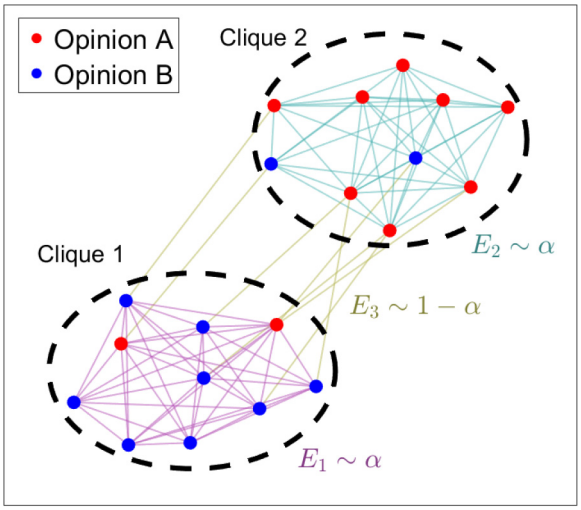


FIG. 1. Example of a two-clique network. Each clique has a fixed number of nodes (10 in this case), whose opinions are updated at discrete time steps. Intra- and interclique connections were sampled with different weights governed by a modularity parameter, $\alpha = 0.9$. We define consensus as a state in which (almost) all nodes have the same color (i.e., almost all red or all blue).

at $p = 1$ on a heterogeneous pair the opinions simply swap and the aggregate fraction of each is conserved within the interaction, leaving interclique exchange as the only source of net change. We assume p is the same for all agents, though expect future work could treat the heterogeneous case.

Concretely, per-pair uniformity over the full N-complete graph requires equal weight on intra- and interclique pairs, hence $\alpha = 1/2$; but the available intraclique pairs scale as $N_i(N_i - 1)$ in each clique against N_1N_2 cross pairs, so for $N_1 \neq N_2$ the two cliques retain unequal within- versus between-clique interaction frequencies that no single α can equalize. Conditional on selecting a node in clique 1, the probability that its interaction partner lies in clique 1 versus clique 2 is proportional to $\alpha(N_1 - 1)$ and $(1-\alpha)N_2$, respectively. In Fig. 1, intraclique interactions occur more frequently than interclique interactions, corresponding to a large value of α. Our model is equivalent to that studied in Ref. [13], which is parameterized by the number of interclique connections X and the fraction r of nodes in the first clique. In our notation, $X = (1-\alpha)N_1N_2$ and $r = N_1/N$. While the interaction rules coincide, our analysis proceeds along different lines.

B. Stochastic approximation

To analyze the discrete dynamics, we formulate a diffusion (stochastic differential equation, SDE) approximation for x_1 and x_2 by computing first and second moments of their increments over a single update. Specifically, we compute the joint transition probabilities $p_{i,j}$ for the changes in (x_1, x_2) over a time step $\Delta t = 2/N$, where $i, j \in \{-1, 0, 1\}$. Here, $p_{i,j}$ denotes the probability that x_1 and x_2 change by i/N_1 and j/N_2, respectively, during one update:

$$p_{i,j} := \mathbb{P}\left(x_1 \mapsto x_1 + \frac{i}{N_1},\ x_2 \mapsto x_2 + \frac{j}{N_2}\right).$$

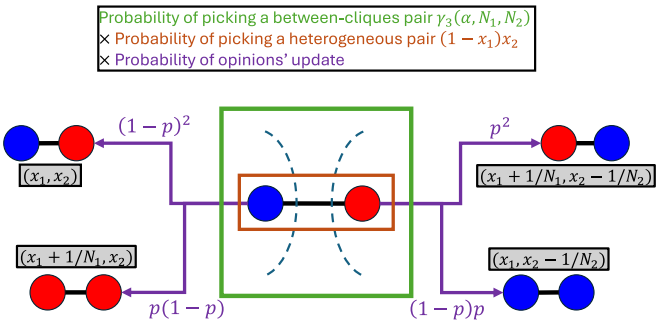


FIG. 2. Example of an update event. We illustrate the update of a heterogeneous $(+1, -1)$ pair (shown as blue and red nodes in cliques 1 and 2, respectively), starting from state (x_1, x_2). An edge is first selected (green text and box) with dashed arcs indicating a between-clique interaction. A specific pair type is then chosen (brown text and box) with edges represented by black lines between nodes. Finally, opinions are updated probabilistically (indicated by purple text and an arrow), yielding the new state, which is highlighted in gray.

Each transition probability is obtained by conditioning on the type of interacting pair selected at the update step. We first illustrate this procedure with representative update events, characterized by changes (i, j) in the number of -1 opinions in each clique.

Since both cliques cannot simultaneously lose or gain a -1 opinion, we have $p_{\pm1,\pm1} = 0$. The transition $(i, j) = (-1, 0)$ occurs when a heterogeneous edge of type $(-1, +1)$ is selected within clique 1 or between cliques and only the -1 node flips and the $+1$ opinion remains unchanged, so that

$$p_{-1,0} = 2\gamma_1 p(1-p)x_1(1-x_1) + \gamma_3 p(1-p)x_1(1-x_2).$$

The transition $(i, j) = (-1, +1)$ occurs when a heterogeneous interclique edge of type $(-1, +1)$ is selected and both opinions flip, yielding

$$p_{-1,+1} = \gamma_3 x_1(1-x_2)p^2.$$

The remaining transition probabilities follow from analogous considerations; all nine are collected in Table I. Figure 2 provides a representative visual illustration of how specific entries arise.

Using the joint transition probabilities $p_{i,j}$, we compute the moments of the increments $\Delta x_i := x_i(t + \Delta t) - x_i(t)$. The expected change in x_1 over one update is

$$\mathbb{E}(\Delta x_1) = \frac{1}{N_1}\sum_{i,j} i p_{i,j} = \frac{\gamma_3 p}{N_1}(x_2 - x_1),$$

and similarly,

$$\mathbb{E}(\Delta x_2) = \frac{1}{N_2}\sum_{i,j} j p_{i,j} = \frac{\gamma_3 p}{N_2}(x_1 - x_2).$$

Moreover, the second moments satisfy

$$\mathbb{E}(\Delta x_1^2) = \frac{1}{N_1^2}\sum_{i,j} i^2\, p_{i,j}$$

$$= \frac{1}{N_1^2}(\gamma_1 4x_1(1-x_1)p(1-p) + \gamma_3 p(x_1 + x_2 - 2x_1x_2)),$$

TABLE I. Joint transition probabilities for a single update. Rows: change in x_1; columns: change in x_2. "Complement" denotes $p_{0,0} = 1 - \sum_{(i,j)\neq(0,0)} p_{i,j}$. The time step is $\Delta t = 2/N$.

Update	$x_2 \to x_2 - 1/N_2$	$x_2 \to x_2$	$x_2 \to x_2 + 1/N_2$
$x_1 \to x_1 - 1/N_1$	0	$\gamma_1 2x_1(1-x_1)p(1-p)$ $+ \gamma_3 x_1(1-x_2)p(1-p)$	$\gamma_3 x_1(1-x_2)p^2$
$x_1 \to x_1$	$\gamma_2 2x_2(1-x_2)p(1-p)$ $+ \gamma_3(1-x_1)x_2 p(1-p)$	Complement	$\gamma_2 2x_2(1-x_2)p(1-p)$ $+ \gamma_3 x_1(1-x_2)p(1-p)$
$x_1 \to x_1 + 1/N_1$	$\gamma_3(1-x_1)x_2 p^2$	$\gamma_1 2x_1(1-x_1)p(1-p)$ $+ \gamma_3(1-x_1)x_2 p(1-p)$	0

and

$$\mathbb{E}(\Delta x_2^2) = \frac{1}{N_2^2}\sum_{i,j} j^2\, p_{i,j} = \frac{1}{N_2^2}(\gamma_2 4x_2(1-x_2)p(1-p) + \gamma_3 p(x_1 + x_2 - 2x_1x_2)),$$

and the mixed second moment

$$\mathbb{E}(\Delta x_1 \Delta x_2) = \frac{1}{N_1 N_2}\sum_{i,j} ij\, p_{i,j} = -\frac{1}{N_1 N_2}(\gamma_3 p^2(x_1 + x_2 - 2x_1x_2)).$$

Finally, we define the variances and covariance of the increments via

$$\mathrm{var}(\Delta x_i) := \mathbb{E}(\Delta x_i^2) - \mathbb{E}(\Delta x_i)^2,$$
$$\mathrm{cov}(\Delta x_1, \Delta x_2) := \mathbb{E}(\Delta x_1 \Delta x_2) - \mathbb{E}(\Delta x_1)\mathbb{E}(\Delta x_2).$$

Using the expressions derived above, this yields

$$\mathrm{var}(\Delta x_1) = \frac{1}{N_1^2}(\gamma_1 4x_1(1-x_1)p(1-p) + \gamma_3 p(x_1 + x_2 - 2x_1x_2) - \gamma_3^2 p^2(x_2 - x_1)^2),$$

$$\mathrm{var}(\Delta x_2) = \frac{1}{N_2^2}(\gamma_2 4x_2(1-x_2)p(1-p) + \gamma_3 p(x_1 + x_2 - 2x_1x_2) - \gamma_3^2 p^2(x_1 - x_2)^2),$$

and

$$\mathrm{cov}(\Delta x_1, \Delta x_2) = \frac{1}{N_1 N_2}(-\gamma_3 p^2(x_1 + x_2 - 2x_1x_2) + \gamma_3^2 p^2(x_2 - x_1)^2).$$

Let $\Sigma(\mathbf{x})$ denote the covariance matrix of the increment $\Delta\mathbf{x} := (\Delta x_1, \Delta x_2)^\top$, with entries defined above. We introduce a diffusion matrix $\sigma(\mathbf{x})$ satisfying $\Sigma = \sigma\sigma^\top$, noting that $\sigma(\mathbf{x})$ is not unique.

Over a macroscopic time interval $\Delta t = 1$, the dynamics consist of $N/2$ update events, so that on average each agent participates in one interaction. Rescaling time accordingly and retaining the leading-order drift and diffusion terms yields an SDE approximation of the discrete voter dynamics according to a two-dimensional Ornstein-Uhlenbeck process [32–34]:

$$dx_1 = \frac{N}{2}\mathbb{E}(\Delta x_1)\,dt + \sqrt{\frac{N}{2}}\,(\sigma_{11}\,dW_1 + \sigma_{12}\,dW_2),$$
$$dx_2 = \frac{N}{2}\mathbb{E}(\Delta x_2)\,dt + \sqrt{\frac{N}{2}}\,(\sigma_{21}\,dW_1 + \sigma_{22}\,dW_2), \qquad (1)$$

where (W_1, W_2) is a two-dimensional Wiener process. The state variables are constrained to the unit interval, $0 \leqslant x_i \leqslant 1$, for $i = 1, 2$. The drift term is linear in both x_i's, hence, when expanded around an equilibrium, no higher order term is present. The diffusion term is quadratic and scales $\approx 1/N$, hence the higher order terms are negligible as $N \to \infty$. Thus, the SDE approximation is valid for a sufficiently large population. See Kramers-Moyal expansion in Refs. [32,34] for further information on the convergence.

A major consequence of the diffusion approximation is the need to redefine consensus. In the general continuous-time approximation, there are two main types of boundary conditions: absorbing and reflective, corresponding to homogeneous Dirichlet and no-flux, respectively. Absorbing boundary conditions signify where the process stops. In our setting, this will correspond to consensus. The reflecting boundary conditions are imposed to prevent the mass from exiting or the process from stopping otherwise. For the associated stochastic process to exit the domain with positive probability, the absorbing portion of the boundary must have positive measure [35]. Consequently, consensus cannot be represented solely by the two corner points, (0,0) and (1,1), corresponding to complete agreement in having opinion -1 or $+1$, respectively. We therefore enlarge the absorbing boundary by introducing a tolerance parameter $\theta > 0$. We say that a *consensus* is reached if either

$$\|(N_1 x_1, N_2 x_2)\| \leqslant \theta N \quad \text{OR} \quad \|(N_1(1-x_1), N_2(1-x_2))\| \leqslant \theta N,$$

where $\|\cdot\|$ denotes a chosen norm. Throughout this paper, we take $\|\cdot\| = \|\cdot\|_1$. This definition corresponds to a state in which at most a fraction θ of dissenters disagree with the majority. The larger θ is, the faster mass leaks through the absorbing boundaries.

For numerical simulations, we use $0.01 \leqslant \theta \leqslant 0.025$. In real voting settings, social rules requiring more than half of participants to agree are referred to as *majority rule*, while larger thresholds (e.g., three-quarters) are termed

supermajority rules [36]. The values of θ considered here correspond to thresholds much closer to unanimity; informally, they could be described as an *uber-/ultramajority*. For consistency, we refer to this condition as consensus throughout.

C. Continuum approximation

Applying the functional central limit theorem yields the PDE formulation. Let $P(\mathbf{x}, t)$ represent the probability of finding the process at a given state $\mathbf{x} = (x_1, x_2)$ at time t. More precisely, P is a probability density function of an associated random variable $(\mathbf{x}, t)$. The system (1) thus has a Fokker-Planck representation [35]:

$$\frac{\partial P}{\partial t}(\mathbf{x}, t) = -\nabla \cdot \big(\boldsymbol{\mu}(\mathbf{x})\, P(\mathbf{x}, t)\big) + \nabla \cdot \nabla \cdot (D(\mathbf{x})\, P(\mathbf{x}, t)), \tag{2}$$

where $\boldsymbol{\mu}(\mathbf{x}) = (\mu_1, \mu_2)$ is the drift vector with

$$\mu_i(\mathbf{x}) := \mathbb{E}(\Delta x_i)\,\frac{N}{2},$$

and $D(\mathbf{x})$ is the diffusion matrix defined by

$$D(\mathbf{x}) := \sigma(\mathbf{x})\sigma(\mathbf{x})^{\top}\,\frac{N}{4} = \mathbf{Cov}(\Delta x_1, \Delta x_2)\,\frac{N}{4}.$$

The spatial domain is $[0, 1]^2$ with the consensus corners removed, and we denote it by Ω^θ. By the Cauchy-Schwarz inequality applied to the variance–covariance relation, the operator in Eq. (2) is parabolic in the interior of Ω^θ. Imposing absorbing conditions means that the mass can exit the system, hence it is not preserved.

We assume a point initial condition for the SDE $(x_1(0), x_2(0)) = (y_1, y_2) \equiv \mathbf{y}$, as described by

$$P_{\text{in}}(\mathbf{x}) := P(\mathbf{x}, 0) = \delta(\mathbf{x} - \mathbf{y}),$$

which will allow us to specify the associated backward equation for the first-passage-time statistics. It is convenient to write Eq. (2) in conservation (divergence) form,

$$\frac{\partial P}{\partial t} + \nabla \cdot \mathbf{J} = 0,$$
$$\mathbf{J}(\mathbf{x}, t) := \boldsymbol{\mu}(\mathbf{x})\, P(\mathbf{x}, t) - \nabla \cdot (D(\mathbf{x})\, P(\mathbf{x}, t)), \tag{3}$$

where

$$(\nabla \cdot (DP))_i := \sum_{j=1}^{2} \frac{\partial}{\partial x_j}(D_{ij}(\mathbf{x})\, P(\mathbf{x}, t)), \quad i = 1, 2.$$

We decompose the boundary of Ω^θ as

$$\partial\Omega^\theta = \partial\Omega^\theta_{\text{ref}} \cup \partial\Omega^\theta_{\text{abs}},$$
$$\partial\Omega^\theta_{\text{abs}} := \Gamma^\theta_0 \cup \Gamma^\theta_1,$$
$$\partial\Omega^\theta_{\text{ref}} := \partial\Omega^\theta \setminus \partial\Omega^\theta_{\text{abs}},$$

with absorbing (homogeneous Dirichlet) boundary segments

$$\Gamma^\theta_0 := \left\{ (x_1, x_2) \in [0, 1]^2 : \sum_{i=1,2} N_i x_i = \theta N \right\},$$

$$\Gamma^\theta_1 := \left\{ (x_1, x_2) \in [0, 1]^2 : \sum_{i=1,2} N_i (1 - x_i) = \theta N \right\}.$$

We impose reflecting (no-flux) boundary conditions on $\partial\Omega^\theta_{\text{ref}}$,

$$\mathbf{J}(\mathbf{x}, t) \cdot \mathbf{n} = 0 \quad \text{for } \mathbf{x} \in \partial\Omega^\theta_{\text{ref}},$$

and absorbing boundary conditions on $\partial\Omega^\theta_{\text{abs}}$,

$$P(\mathbf{x}, t) = 0 \quad \text{for } \mathbf{x} \in \partial\Omega^\theta_{\text{abs}},$$

where $\mathbf{n}$ denotes the outward unit normal on $\partial\Omega^\theta$.

From this forward problem one derives, via the adjoint (backward Kolmogorov) equation and standard first-passage arguments, a boundary-value problem for the mean consensus time $T(\mathbf{y})$, namely Eq. (A3) with the no-flux boundary conditions inherited from Eq. (A1). The derivation can be found in Appendix A.

Having established the backward problem for the mean consensus time (Appendix A) and clarified the associated boundary conditions, we now turn to the forward equation (2) to understand how probability mass is transported through the state space. In particular, we exploit the separation between $\mathcal{O}(1)$ drift terms and $\mathcal{O}(1/N)$ diffusive terms to identify the dominant dynamics and the relevant boundary-layer scalings in the large-population limit.

III. MULTISCALE ASYMPTOTICS OF CONSENSUS FORMATION

In this section, we analyze Eq. (2) using asymptotic methods in the limit of large populations $N \to \infty$, to identify the mechanisms governing consensus formation across distinct timescales. Extreme values of other parameters (e.g., $p = 0$ or $p = 1$) lead to degenerate dynamics and singular behavior; these cases are examined numerically in the next section, while their analytical treatment is left for future work. Throughout this section, we assume equal population sizes, $N_1 = N_2 = N/2$. This symmetry simplifies both the equations and the domain, making the analysis more tractable. Nevertheless, the same approach can be extended in the case that $N_1 \neq N_2$, provided both $N_i \to \infty$.

First, we introduce the change of variables

$$w = \frac{x_1 + x_2 - 1}{\sqrt{2}} \text{ and } z = \frac{x_2 - x_1}{\sqrt{2}}.$$

The parabolicity of the associated operator and smoothness of drift and diffusion terms imply that P is regular, hence the change of variables is justified. This transformation centers the domain at (0,0) and aligns the coordinate axes with the diagonal and antidiagonal directions, mapping the original diagonal $x_1 = x_2$, corresponding to alignment of the belief fractions in each module, onto the horizontal axis. The variable w runs from $-(1 - 2\theta)/\sqrt{2}$ to $(1 - 2\theta)/\sqrt{2}$, while z runs from $-1/\sqrt{2}$ to $1/\sqrt{2}$. Recall that the domain is a truncated diamond whose left- and right-hand corners have been removed. In the new variables, the drift and diffusion terms naturally separate along the diagonal (w) and transverse (z)

directions, so that Eq. (2) becomes

$$\begin{aligned}\frac{\partial P}{\partial t} &= 2p(1-\alpha)\frac{\partial}{\partial z}[zP] + \frac{p(1-p)}{2N}\left[\frac{\partial^2}{\partial w^2} + \frac{\partial^2}{\partial z^2}\right] \\ &\times [(1-2w^2+2(1-2\alpha)z^2)P] \\ &+ \frac{p^2(1-\alpha)}{N}\frac{\partial^2}{\partial z^2}[(1-2w^2-2(1-2\alpha)z^2)P] \\ &- \frac{4\alpha p(1-p)}{N}\frac{\partial^2}{\partial w \partial z}[wzP], \end{aligned} \tag{4}$$

where we have expanded the coefficients to make their dependence on w and z explicit.

In the singular limit $N \to \infty$, all second-derivative terms vanish and Eq. (4) reduces to the first-order transport equation

$$\frac{\partial P^0}{\partial t}(w,z,t) = 2p(1-\alpha)\frac{\partial}{\partial z}[zP^0(w,z,t)]. \tag{5}$$

Along characteristics, w is constant and $z(t) = z_0 e^{-2p(1-\alpha)t}$. Hence, if $P^0(w,z,0) = P_{\text{in}}(w,z)$, then

$$\begin{aligned} P^0(w,z,t) &= e^{2p(1-\alpha)t} P_{\text{in}}(w, ze^{2p(1-\alpha)t}) \\ &= \delta(w-w_0)\delta(z - z_0 e^{-2p(1-\alpha)t}). \end{aligned} \tag{6}$$

This implies exponential contraction toward the diagonal $z = 0$, with deviations transverse to alignment decaying at rate $2p(1-\alpha)$ (equivalently, on the characteristic timescale $[2p(1-\alpha)]^{-1}$).

As a result, probability mass rapidly concentrates near $z = 0$ on $\mathcal{O}(1)$ timescales. Diffusive effects, which are suppressed by a factor of $1/N$ relative to the drift, become relevant only within a narrow boundary layer around $z = 0$. This separation of scales motivates a multiscale expansion in $\epsilon = 1/N$,

$$P(w,z,t) = P^0(w,z,t) + \epsilon P^1(w,z,t) + \epsilon^2 P^2(w,z,t) + \cdots,$$

where P^0 captures the rapid drift toward alignment and higher-order terms encode finite-size diffusive corrections. This procedure is methodologically analogous to the Chapman-Enskog expansion in kinetic theory [37] and to the grazing-collision limit of kinetic opinion models [16], with the finite-size parameter $\epsilon = 1/N$ playing the role of the small scale and the first-order solvability condition projecting onto the slow dynamics.

To describe the subsequent evolution once the distribution has concentrated near the alignment manifold $z = 0$, we examine an inner region where diffusive effects compete with the vanishing drift. This motivates the introduction of rescaled spatial and temporal variables that capture the dynamics within a boundary layer around $z = 0$.

Boundary-layer dynamics near the alignment manifold

When z becomes small ($z = \mathcal{O}(\sqrt{\epsilon})$), we must resolve dynamics by introducing a stretched variable $s = z/\sqrt{\epsilon}$. In the inner region, the probability density therefore depends on (w, s, t, T). At leading order, the dynamics are confined to the z direction, with no evolution in w; evolution in w occurs on a slower timescale.

Let $s_0 = \mathcal{O}(1)$. We first shift time by

$$t_{\text{in}} := \frac{1}{2p(1-\alpha)} \ln \frac{z_0}{s_0\sqrt{\epsilon}},$$

which corresponds to the time at which $z = s_0\sqrt{\epsilon}$. We then introduce the slow timescale $T = \epsilon t$, treated as an independent variable in the inner region. Accordingly, in the inner region, the density is written as $P(w, s, t, T)$. With these rescalings, the leading-order term satisfies

$$P^0_t = 2p(1-\alpha)(sP^0)_s + \frac{p + p^2 - 2\alpha p^2}{2}(1-2w^2)P^0_{ss}. \tag{7}$$

Equation (7) differs from the outer transport equation (5) by the presence of an additional w-dependent diffusion term. The initial condition is

$$P(w,s,0,0) = \delta(w-w_0)\,\delta(\sqrt{\epsilon}(s-s_0)).$$

Here "initial" refers to the shifted variables $t' = t - t_{\text{in}}$ and $T' = \epsilon(t - t_{\text{in}})$, so that $t' = 0$ describes the moment the solution enters the boundary layer; the initial condition is imposed at this moment. Solving Eq. (7) with this initial condition yields

$$\begin{aligned} P^0 &= \frac{A(w,T')}{\sqrt{4\pi\epsilon\frac{p+p^2-2\alpha p^2}{8p(1-\alpha)}(1-e^{-4p(1-\alpha)t'})(1-2w^2)}} \\ &\times \exp\left[-\frac{(z - s_0\sqrt{\epsilon}e^{-2p(1-\alpha)t'})^2}{4\epsilon\frac{p+p^2-2\alpha p^2}{8p(1-\alpha)}(1-e^{-4p(1-\alpha)t'})(1-2w^2)}\right] \\ &\equiv A(w,T')B(s,t'), \end{aligned} \tag{8}$$

where $B(s,t')$ denotes the Gaussian factor, with mean $s_0\sqrt{\epsilon}\,e^{-2p(1-\alpha)t'}$ and variance $2\epsilon\frac{p+p^2-2\alpha p^2}{8p(1-\alpha)}(1 - e^{-4p(1-\alpha)t'})(1-2w^2)$. Equation (8) corresponds to a modified Ornstein-Uhlenbeck-type solution [38]: compared with Eq. (6), it introduces diffusive smoothing, though the overall dynamics are unchanged. Probability mass approaches the alignment manifold $z = 0$, advection diminishes, and diffusion takes over. P^0 is the product of $A(w,T')$ and the following Gaussian density in z:

$$\begin{aligned} &\mathcal{N}\Bigg(s_0\sqrt{\epsilon}e^{-2p(1-\alpha)t'}, \epsilon\frac{p+p^2-2\alpha p^2}{4p(1-\alpha)} \\ &\times (1-e^{-4p(1-\alpha)t'})(1-2w^2)\Bigg). \end{aligned} \tag{9}$$

Equation (9) is true locally only, and it is, in fact, not a true Gaussian, but rather a truncated version of it. This shows that the dynamics in the z direction follow a Gaussian density centered at $s_0\sqrt{\epsilon}\,e^{-2p(1-\alpha)t'}$, which decays to zero as $t' \to \infty$. Its width grows from zero and saturates at a level proportional to $\sqrt{1-2w^2}$, vanishing as $w \to \pm 1/\sqrt{2}$. Note that since we "cut" the corners, we do not run into singularities. The speed at which mass concentrates at $z = 0$ is still dependent on $p(1-\alpha)$, similar to Eq. (6). The presence of ϵ is an artifact of the initial condition that changed when $z \to s$. As ϵ appears in both the mean and variance, this signifies how population size affects the spread: the larger the population, the closer to zero and the thinner the Gaussian "bell" in z. The simulation of the

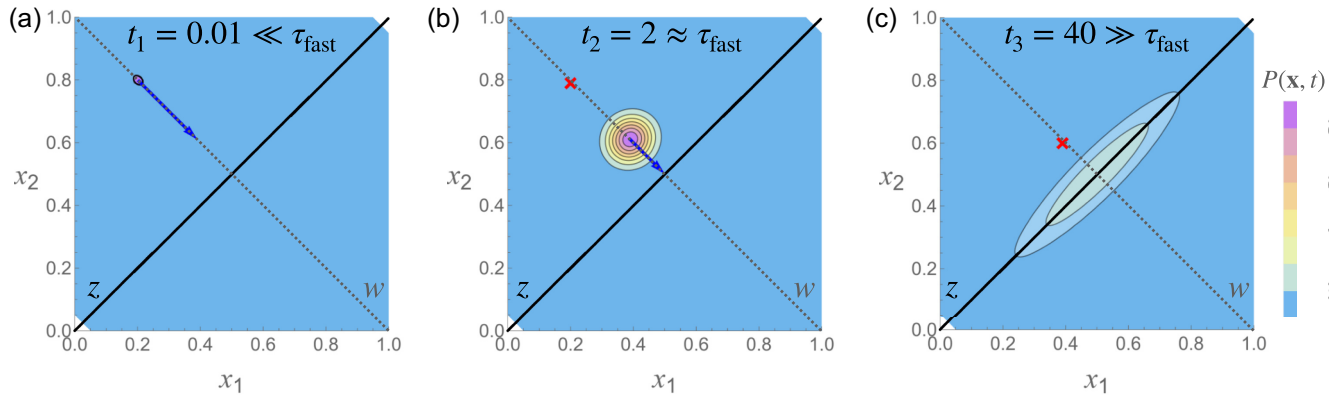


FIG. 3. Separation of fast and slow dynamics in the voter model. The solid line is the diagonal $z = 0$; dashed lines show the rotated coordinates w, z. Red crosses and blue arrows track successive centres of mass. (a) For $t \ll \tau_{\text{fast}} = [2p(1-\alpha)]^{-1}$, the density drifts toward the diagonal as predicted by Eq. (6). (b) At $t \approx \tau_{\text{fast}}$, diffusion competes with the weakening drift, producing the Gaussian profile (8). (c) For $t \gg \tau_{\text{fast}}$, the density is narrow in z (width $\propto \sqrt{\epsilon(1-2w^2)}$) and elongated along the diagonal, reflecting slow w-dynamics on the $T = \epsilon t$ timescale. Peak reduction from panel (b) to panel (c) reflects absorbing-boundary mass loss. Parameters: $N_1 = N_2 = 100$, $\alpha = 0.5$, $p = 0.5$, $\theta = 0.025$; initial condition $\mathbf{x}_0 = (1/5,\ 4/5)$.

original equation (2) is shown in Fig. 3. The numerical results, even for relatively small populations, confirm our preliminary findings on the dynamics. In Fig. 3(a), after a relatively small time, we notice that the initial mass spreads as it approaches the main diagonal along the antidiagonal. In Fig. 3(b), after two time units, the movement toward the diagonal continues. The spread is more apparent and appears to be equal in both directions. Finally, in Fig. 3(c), the spread primarily happens along the main diagonal, while it remains thin along the antidiagonal. The slow dynamics take effect. The solution "leaks" through the absorbing boundaries, hence the depletion of the maximum values and overall mass.

Figure 4 showcases how the asymptotic result agrees with numerical simulations of the actual PDE (2) across various parameter combinations. We start with a point far from the diagonal and track how long it takes for the point to reach the diagonal. We observe that the agreement holds until z gets considerably small, with the threshold varying with α. Hence, increasing α extends the window of agreement. This can be explained by the appearance of α in the variance of Eq. (9): the larger α is, the narrower the Gaussian. In other words, the solution preserves its shape, dictated by the initial conditions, longer with larger α. Beyond the threshold, the two functions diverge: the numerical z coordinate still approaches zero, but at a sub-exponential rate, visible as the red curve bending above the asymptotic straight line in the ln plot. The exact divergence threshold is not known analytically and likely depends on the interplay between boundary-layer diffusion and the absorbing boundaries. Nevertheless, Figure 4 confirms that the leading-order asymptotic expansion accurately captures the fast dynamics across a range of α values.

Proceeding to the next order in ϵ, the equation for P^1 is solvable only if its right-hand side satisfies a Fredholm condition, orthogonality to the constants spanning the null space of the adjoint s-operator [38]. Carrying out this projection eliminates the fast variable and yields the evolution equation for the slow amplitude $A(w, T)$ derived in Appendix B,

$$\frac{\partial A}{\partial T} = C_2\left[-4A - 8w\frac{\partial A}{\partial w} + (1-2w^2)\frac{\partial^2 A}{\partial w^2}\right], \tag{10}$$

where $C_2 = p(1-p)/2$. Separating variables as $\tau(T)\mathcal{W}(w)$ shows the temporal part satisfies $\tau(T) = e^{\lambda C_2 T}$. Since C_2 only depends on $p(1-p)$, the consensus time in this regime is minimized when $p = 1/2$, which we verify numerically in the next section. Moreover, α is absent from Eq. (10), so modularity influences the fast alignment stage but not the slow dynamics along the aligned manifold. Introducing

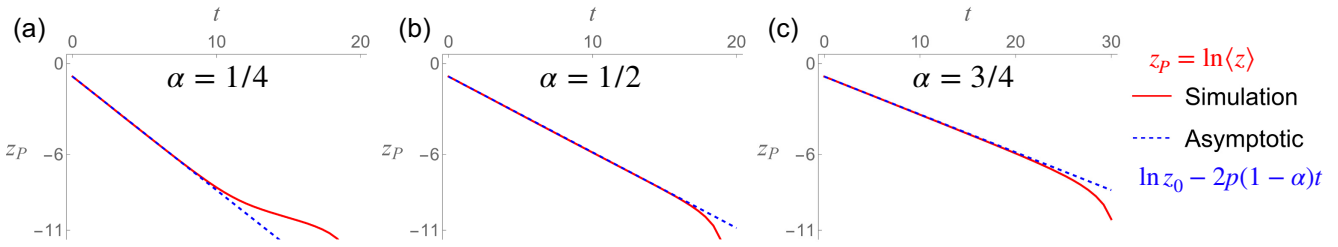


FIG. 4. Accuracy of asymptotic approximation for the fast dynamics. The red curve shows $z_P = \ln\langle z\rangle$, the natural log of the mean z coordinate of P, computed from numerical simulation of (2) using a center-of-mass formula based on 10^5 sample points. The blue dashed curve is the asymptotic prediction $\ln z_0 - 2p(1-\alpha)t$ from Eq. (6). The two agree well during the fast-drift regime and diverge once boundary-layer diffusion and absorbing-boundary effects become significant. Panels differ in α: (a) $\alpha = 1/4$, (b) $\alpha = 1/2$, (c) $\alpha = 3/4$. Remaining parameters: $y_1 = 1/5$, $y_2 = 4/5$, $N_1 = N_2 = 100$, $p = 1/2$, $\theta = 0.025$.

$u = (1 + \sqrt{2}w)/2$, the spatial part reduces to

$$u(1-u)\frac{d^2\mathcal{W}}{du^2} + 2(1-2u)\frac{d\mathcal{W}}{du} + \mu\mathcal{W} = 0,$$

with $\mu = -\frac{4+\lambda}{2}$. Equivalently, setting $v = \sqrt{2}w \in [-1, 1]$ yields

$$(1-v^2)\frac{d^2\mathcal{W}}{dv^2} - 4v\frac{d\mathcal{W}}{dv} + \mu\mathcal{W} = 0.$$

This is the Jacobi differential equation with parameters $(\alpha, \beta) = (1, 1)$, whose eigenfunctions are the Jacobi polynomials $\mathcal{P}_n^{(1,1)}(v)$ and eigenvalues $\mu_n = n(n+3)$, $n \in \mathbb{N} \cup \{0\}$. Therefore,

$$\lambda_n = -2n(n+3) - 4 = -2n^2 - 6n - 4.$$

Since all eigenvalues are negative, every mode decays in time, confirming that consensus is reached with probability one. Thus, once the cliques have aligned, the slow dynamics are insensitive to modularity and reduce to the same one-dimensional neutral diffusion structure that underlies the well-mixed model.

To summarize: the multiscale analysis reveals two distinct stages of consensus formation. On $\mathcal{O}(1)$ timescales, interclique coupling drives exponential alignment of opinion fractions at rate $2p(1-\alpha)$. On the slow $T = t/N$ timescale, the system evolves along the alignment manifold via a diffusion whose rate depends on $p(1-p)$ but is independent of the modularity parameter α. In the next section, we validate these predictions with numerical simulations.

IV. CONSENSUS TIMES ACROSS PARAMETER REGIMES

The asymptotic analysis of the previous section predicts that consensus evolves first by a fast alignment stage governed by α, followed by a slow diffusion stage governed by $p(1-p)$. However, it is not clear how these stages interact across parameter regimes, particularly when the symmetry assumptions underlying the analysis are relaxed. In this section, we address this question numerically. We begin by verifying that the discrete, SDE, and PDE formulations agree across a range of parameters, then turn to the central question: is there an optimal level of interclique coupling that minimizes the time to reach consensus? We will show that the answer depends critically on population imbalance and initial polarization, with nontrivial optima emerging precisely when the system is far from symmetric. The discrete and SDE simulations use Euler-Maruyama integration, while the PDE solutions are obtained via a finite element method. The implementation details and the source code are available in Ref. [39]. To compare the three formulations on equal footing, we require θN to be an integer in the discrete model and replace θ by $\theta' := (\theta N + \frac{1}{2})/N$ in the SDE and PDE to account for the half-step offset; all stated thresholds refer to the original θ.

A. Single parameter and pairwise comparisons

Here we verify that the discrete, SDE, and PDE models produce consistent predictions by performing single-parameter experiments. Specifically, we examine how each of the four parameters N, α, p, and the relative clique size $r := N_1/N$ affects the system by varying them one at a time. We begin with a symmetric baseline where $p = \alpha = r = y_1 = y_2 = 1/2$, and then move to an asymmetric setting with $p = \alpha = y_2 = 3/4$ and $r = y_1 = 1/4$. Both cases are presented in Fig. 5. Across both settings, the three models are in good agreement, with deviations appearing only at extreme parameter values, where either the diffusion coefficients degenerate ($\alpha \to 1$, $p \to 0, 1$) or finite-size effects limit the validity of the continuum approximation ($r \to 0, 1$). This confirms that the PDE analysis of the previous section captures the behavior of the discrete and SDE dynamics, and justifies our use of the PDE for the parameter explorations that follow.

We observe an approximately linear relation between N and the mean consensus time T [Fig. 5(a)]. Asymmetry increases the slope: in Fig. 5(a), the symmetric $T(N = 1000) \approx 2500$, whereas asymmetric $T(N = 1000) \approx 3250$. By contrast, α has little effect on consensus time except near the disconnected limit $\alpha \to 1$, where T increases sharply. In this regime, the cliques evolve nearly independently and may reach opposing consensus states; since heterogeneous interactions are absent, the system becomes trapped near the (0,1) or (1,0) states. Correspondingly, these points are singular for Eq. (A3). In the symmetric case, T increases in α but is nonmonotonic and convex in the asymmetric setting [Fig. 5(b)], indicating a stronger sensitivity to modularity away from symmetry. Asymmetry increases T as well. Figure 5(c) shows that the dependence on p is symmetric about $p = 1/2$, with divergence as $p \to 0$ or $p \to 1$. Thus, the primary driver of the consensus time is p: the farther p is from $1/2$, the larger T becomes. This is consistent with the asymptotic analysis, where the slow diffusive timescale is proportional to $[p(1-p)]^{-1}$. Symmetry or asymmetry does not seem to affect the time, as two PDE solutions almost coincide.

Our final comparison examines how the relative sizes of the cliques affect the mean consensus time, as shown in Fig. 5(d). Since the total population size is fixed, r also reflects the ratio N_1/N_2. Once again, the three models agree well, with deviations only at extreme values of r, where the approximation breaks down for very small subpopulations (e.g., $N_1 = 10$). The curves are approximately symmetric about $r = 1/2$, but the shape changes qualitatively between settings: in the symmetric case, T is minimized at $r = 1/2$, whereas in the asymmetric setting the minimum shifts toward r closer to 0 or 1. In fact, in the asymmetric case, the maximum consensus time occurs at $r = 1/2$. This reversal suggests that the role of equal clique sizes depends sensitively on the overall parameter regime. More broadly, the dependence of T on r interacts nontrivially with the other parameters, a point we return to in the next subsection.

The single-parameter comparisons above suggest that p acts largely as a uniform scaling factor, while α and r interact in less obvious ways. To make this precise, we solve Eq. (A3) over each parameter pair (α, p), (α, r), and (p, r), holding the third fixed along with $N = 500$, $\theta = 0.01$, and $y_1 = y_2 = 1/2$ (Fig. 6). Note that N was not considered, as it is expected to affect the consensus time only linearly. Thus, we do not expect any new insights for sufficiently larger populations, $N \geqslant 50$. The contour plots confirm the scaling role of p: in Figs. 6(a) and 6(c), contours run nearly horizontal and vertical, respectively, so p shifts T uniformly without

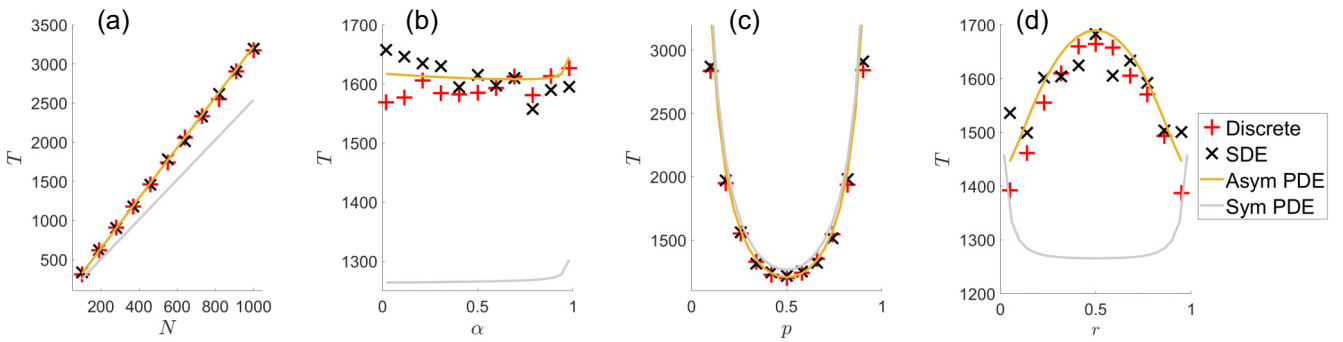


FIG. 5. Comparison between discrete, SDE, and PDE models across parameter sweeps. Each panel overlays results from a symmetric baseline (shown in grey) and an asymmetric setting (shown in color): (a) total population size, N, (b) modularity parameter, α, (c) flip probability, p, (d) relative sizes, r. The symmetric case uses $y_1 = y_2 = p = \alpha = r = 1/2$, while the asymmetric case uses $y_1 = 1/4$, $y_2 = 3/4$, $p = \alpha = 3/4$, and $r = 1/4$. In all panels, parameters not being varied are held fixed at these values, with $N = 500$ and $\theta = 0.01$.

reshaping its dependence on α or r; slight deviations appear only near $\alpha \to 1$ and $r \to 0$. The more revealing panel is Fig. 6(b): as r decreases, T becomes increasingly convex in α, with a minimum at $r = 1/2$ and a maximum in the corner of small r and small α—likely reflecting the low probability of selecting heterogeneous pairs in that regime. These results confirm that modularity matters most when clique sizes are highly unequal, motivating the detailed exploration of α in the next subsection.

The single-parameter results (Fig. 5) clearly indicate a hierarchy of influence. The simplest dependence is on N: consensus time grows linearly with population size, as one would expect from the diffusive timescale identified in the asymptotic analysis. The flip probability p is the dominant control parameter—T is symmetric about $p = 1/2$, where it is minimized, and diverges toward both $p \to 0$ and $p \to 1$, in direct agreement with the slow-stage diffusion rate $p(1-p)$. Neither of these dependencies changes qualitatively between symmetric and asymmetric settings. The modularity parameter α tells a different story: in the symmetric case, T is nearly flat across a wide range of α before diverging as the cliques decouple ($\alpha \to 1$), but once asymmetry is introduced the curve becomes convex, suggesting that an interior optimum may exist. The richest behavior belongs to the relative clique size r. In symmetric settings, equal-sized cliques ($r = 1/2$) minimize consensus time; in asymmetric settings, this relationship reverses, and $r = 1/2$ becomes a maximizer. This reversal, together with the emergent convexity in α, points to a nontrivial interplay between modularity and population imbalance that we explore in detail in the next subsection.

B. When can modularity accelerate consensus?

The previous subsection established that modularity and population imbalance interact nontrivially, while p acts as a uniform scaling factor. We now ask a sharper question: can a two-clique network reach consensus faster than a well-mixed population, and if so, is there an optimal level of interclique coupling? To expose these effects, we adopt polarized initial conditions ($y_1 = 1$, $y_2 = 0$), modeling groups that are initially sorted by opinion [40–42], and rely on the PDE (A3) throughout. As a baseline, we compare against a well-mixed model on the complete graph with the same update rules. The corresponding mean consensus time admits an explicit solution [29,43]:

$$T_{1D} := -\frac{N}{p(1-p)}\big[x_0 \ln(x_0) + (1-x_0)\ln(1-x_0) - \theta \ln(\theta) - (1-\theta)\ln(1-\theta)\big], \tag{11}$$

where $x_0 := (y_1 N_1 + y_2 N_2)/N$ is the initial fraction of opinion-A holders. The linear prefactor N sets the overall timescale, in rescaled units of one interaction per agent, while the dependence on the initial composition enters only through the fraction x_0; this $\mathcal{O}(N)$ scaling is the standard voter-model

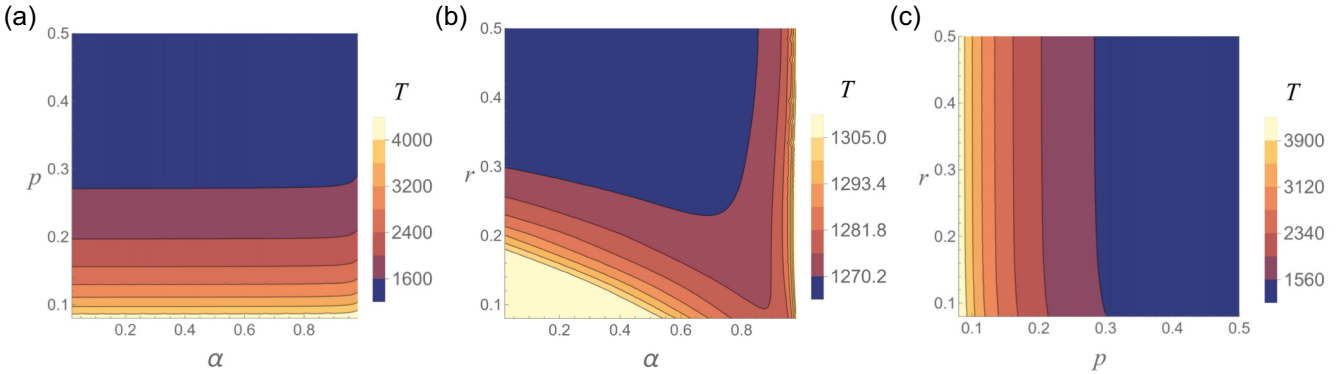


FIG. 6. Mean consensus time as a function of parameter pairs. Contours of the PDE solution (A3) over (a) (α, p) with $r = 1/2$, (b) (α, r) with $p = 1/2$, and (c) (p, r) with $\alpha = 1/2$. Remaining parameters $y_1 = 1/2$, $y_2 = 1/2$, $N = 500$, and $\theta = 0.01$.

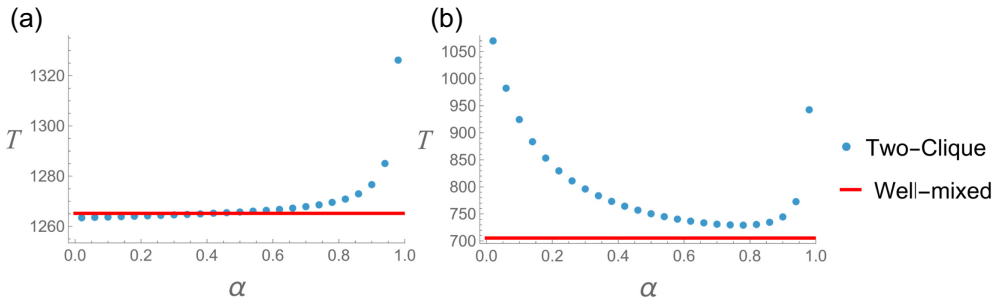


FIG. 7. Mean consensus time: two-clique versus well-mixed model. (a) Symmetric case, $N_1 = N_2 = 250$, $p = 1/2$. (b) Asymmetric case: $N_1 = 50$, $N_2 = 450$, $p = 1/4$. Remaining parameters: $\theta = 0.01$, $y_1 = 1$, $y_2 = 0$. The red line is the well-mixed solution (11); blue dots are the two-clique PDE solution (A3).

result and the diffusion approximation becomes more accurate as $N \to \infty$.

It is also natural to ask which consensus is reached, quantified by the exit probability $\mathcal{E}_A$, the probability of absorbing into the opinion-A region. The size-weighted fraction $M = (N_1x_1 + N_2x_2)/N$ is linear, and since the drifts cancel under this weighting,

$$r\,\mathbb{E}(\Delta x_1) + (1-r)\mathbb{E}(\Delta x_2) = \tfrac{\gamma_3 p}{N}[(x_2 - x_1) + (x_1 - x_2)] = 0,$$

with $r = N_1/N$, M is an exact martingale of the discrete dynamics. For $\alpha < 1$, where consensus is reached almost surely, optional stopping with the absorbing bands at $M = 1-\theta$ and $M = \theta$ yields

$$\mathcal{E}_A(y_1, y_2) = \frac{x_0 - \theta}{1 - 2\theta}, \qquad x_0 = \frac{N_1 y_1 + N_2 y_2}{N} = r y_1 + (1-r) y_2. \tag{12}$$

The conserved quantity is set by the clique sizes alone, independent of α and p, which control the relaxation rate but not the outcome. Hence $\mathcal{E}_A = 1/2$ under symmetry and deviates toward the larger or more strongly A-leaning clique under population imbalance and polarized initial conditions, paralleling the consensus-time analysis in which modularity reshapes the timescales but not the final state. The same neutrality underlies the degree-weighted conservation law for voter models on heterogeneous graphs [10].

In the symmetric case [Fig. 7(a)], the two-clique model is marginally faster at small α where interclique edges preferentially select heterogeneous pairs under polarized initial conditions; the advantage is confined to the fast alignment stage and amounts to less than 1% of the total consensus time. The asymmetric case [Fig. 7(b)] is more revealing: $T(\alpha)$ becomes convex with a clear interior minimum, indicating an optimal level of interclique coupling exists.

The U-shape can be understood through a small-r scaling argument. Let $r := N_1/N \ll 1$. In this limit, the edge fractions simplify to

$$\gamma_1 = \mathcal{O}(r^2), \quad \gamma_2 \approx 1 - \frac{2(1-\alpha)r}{\alpha}, \quad \gamma_3 \approx \frac{2(1-\alpha)r}{\alpha}.$$

Since $\gamma_3 = \mathcal{O}(r)$, interclique interactions become rare. Substituting into the drift expressions from Sec. II B and rescaling by $N/2$ gives

$$\mu_1 \approx \frac{(1-\alpha)p}{\alpha}(x_2 - x_1), \quad \mu_2 \approx \frac{(1-\alpha)p\,r}{\alpha}(x_1 - x_2).$$

The small clique tracks the large one at rate $(1-\alpha)p/\alpha$, while the large clique is slower by a factor of r. The alignment timescale is the reciprocal of this rate,

$$\tau_{\text{fast}} \sim \frac{\alpha}{(1-\alpha)p},$$

which favors small α. The leading-order diffusion coefficient for x_1 scales as

$$D_{11} \sim \frac{(1-\alpha)\,p}{\alpha\, r\, N}(x_1(1-x_2) + x_2(1-x_1)),$$

which diverges as $\alpha \to 0$: reducing modularity amplifies fluctuations in the small clique. On the alignment manifold ($x_1 \approx x_2 \approx x$), the parenthetical factor reduces to $2x(1-x)$, recovering the expression used below.

After the fast stage, the system evolves along the alignment manifold $x_1 \approx x_2 =: x$ with effective initial fraction $x_0 \approx r$ (under polarized initial conditions $y_1 = 1$, $y_2 = 0$). At leading order, the slow-stage dynamics reduce to the well-mixed model on the large clique, giving a total consensus time of approximately

$$T \approx \underbrace{\frac{\alpha}{(1-\alpha)p}}_{\text{fast stage}} \underbrace{- \frac{N}{p(1-p)}[r\ln r + (1-r)\ln(1-r) - \theta\ln\theta - (1-\theta)(1-\theta)]}_{\text{slow stage}}.$$

The first term decreases with α; the second is independent of α and typically dominates.

However, a correction to the slow stage restores α-dependence. Evaluating the small-r diffusion components

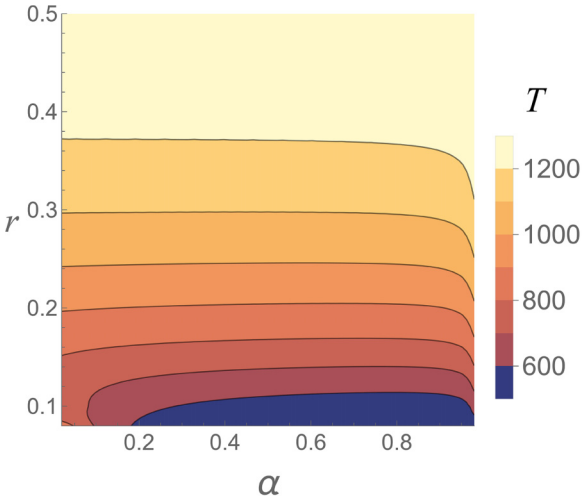


FIG. 8. Mean consensus time as a function of α and r under polarized initial conditions. Contours of the PDE solution (A3) with $y_1 = 1$, $y_2 = 0$, $N = 500$, $p = 1/2$, $\theta = 0.01$. Compare with Fig. 6(b), which uses balanced initial conditions.

on the alignment manifold and forming the effective scalar diffusion $D_{\text{eff}} = r^2 D_{11} + 2r(1-r)D_{12} + (1-r)^2 D_{22}$ gives Eq. (C1) of Appendix C. Replacing the well-mixed diffusion in the slow stage with it yields the corrected estimate

$$T \approx \frac{\alpha}{(1-\alpha)p} - \frac{N}{p[(1-p) + \frac{r(1-\alpha)}{\alpha}]} \times [r \ln r + (1-r)\ln(1-r) - \theta \ln \theta - (1-\theta)\ln(1-\theta)]. \quad (13)$$

Equation (13) makes the competing roles of α explicit. As α decreases from 1, the fast-stage cost drops and the enhanced diffusion D_{eff} (see Eq. (C1) in Appendix C) shortens the slow stage, so T decreases, consistent with the right branch of the U observed in simulations. For large α the correction term $r(1-\alpha)/\alpha$ vanishes, the slow stage reverts to the well-mixed value, and the fast-stage cost diverges as $\alpha \to 1$. The observed optima ($\alpha^* \approx 0.7$–0.9) lie in the transitional regime where both contributions are comparable.

Figure 8 confirms this picture under polarized initial conditions ($y_1 = 1$, $y_2 = 0$): as r decreases, $T(\alpha)$ becomes increasingly convex, with both the depth of the minimum and the overall range of T growing. Compared with the balanced initial conditions of Fig. 6(b), polarization shifts the minimum to smaller α and amplifies the penalty for near-disconnection. As in the previous subsection, varying p shifts T uniformly without altering this structure.

We next examine how initial polarity affects the optimal α and the corresponding T. Fixing $r = 1/10$ (Fig. 8), both the optimal α and the corresponding minimal T decrease as $|y_1 - y_2|$ increases. The small-r scaling argument above provides intuition: stronger polarization means the system starts further from alignment, increasing the relative contribution of the fast stage to the consensus time. To compensate, the optimum shifts toward smaller α, which accelerates drift at the cost of larger fluctuations in the small clique. The decrease in minimal T is more straightforward: as the minority shrinks, fewer agents must be converted to obtain consensus.

A natural follow-up question is whether this acceleration persists when the total number of minority-opinion holders is conserved, rather than allowing initial fractions to vary freely. A more realistic scenario fixes the total number of minority-opinion holders M and asks how they should be distributed: concentrated in the small clique, spread across both, or somewhere in between? To model this, we define the *constrained polarity* parameter $m := y_1 N_1 / M$, so that $y_1 = mM/N_1$ and $y_2 = (1-m)M/N_2$ [Figs. 9(c) and 9(d)].

Under this constraint, the optimal α increases roughly linearly with m, and consensus generally takes longer than in the unconstrained case. A clear minimum in T emerges at an intermediate value $m \approx 0.05$, which can be understood through the competing timescales identified in Sec. III: small m concentrates the minority in the small clique, maximizing the opinion gap $|y_1 - y_2|$ and thereby shortening the fast alignment stage ($\tau_{\text{fast}} \sim \alpha/[(1-\alpha)p]$), while simultaneously pushing the post-alignment fraction $x_0 \approx M/N$ toward zero, where the slow diffusion to consensus is fastest. Below this optimum, fluctuation costs in the small clique ($D_{11} \sim (1-\alpha)/(\alpha\, r\, N)$) dominate and consensus slows; above it, the minority is diluted across both cliques, weakening the alignment advantage without compensating gains in the slow stage. In practical terms, the minimum suggests that concentrating most, but not all, of the minority in the smaller clique best balances rapid alignment against efficient diffusion toward consensus. As clique sizes become more comparable, these trends weaken and eventually reverse, with the optimal α dropping and the minimum in T shifting toward $m \approx 1$. An explicit expression for the optimal α under constrained polarity remains an open problem. The small-r scaling correctly predicts the qualitative trends, but the observed optima ($\alpha^* \approx 0.7$–0.9) lie closer to the disconnection limit than the leading-order balance suggests, indicating that the slow-stage cost and the $\alpha \to 1$ divergence play a significant role in setting the precise optimum.

To summarize: modularity has little effect on consensus time in symmetric settings, where the well-mixed model is nearly optimal. Its influence becomes significant under population imbalance and polarized initial conditions, where an optimal level of interclique coupling emerges from the competition between fast alignment drift and small-clique fluctuations.

V. DISCUSSION

In this paper, we studied a modular voter model on two cliques, analyzing how modularity, population imbalance, and initial polarization shape the time to consensus. We derived SDE and Fokker-Planck approximations of the discrete process and confirmed their agreement numerically across a range of parameter regimes. The multiscale asymptotic analysis of the Fokker-Planck equation revealed two distinct stages: a fast alignment phase in which interclique coupling drives exponential synchronization of opinion fractions at rate $2p(1-\alpha)$, followed by a slow diffusion along the alignment manifold governed by $p(1-p)$ and independent of α. For unequal clique sizes, a complementary small-r scaling

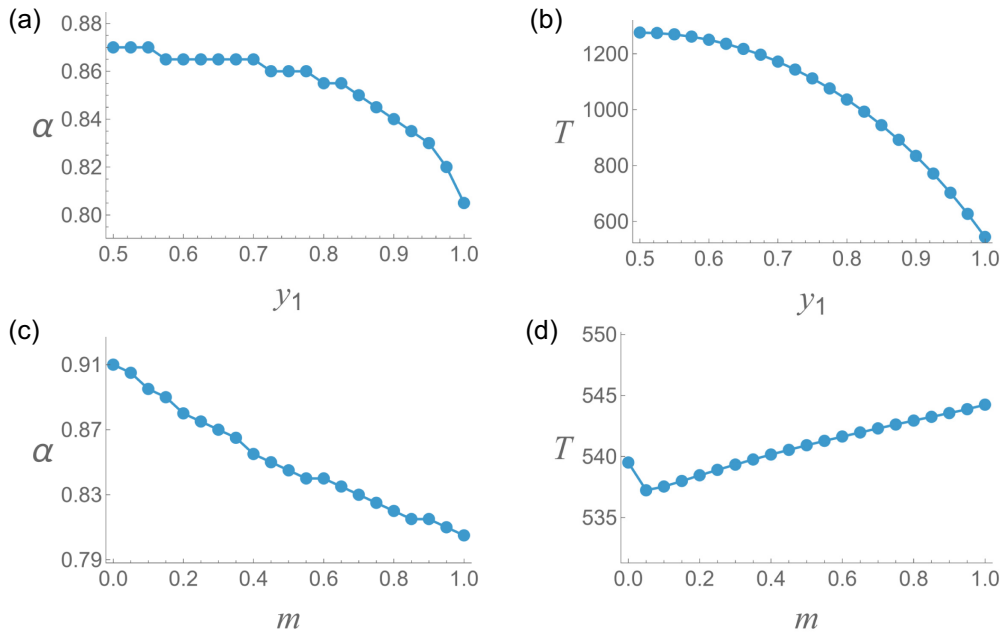


FIG. 9. Effect of initial polarization on optimal coupling and consensus time. (a), (b) Unconstrained polarization: $y_2 = 1 - y_1$ varies freely. (c), (d) Constrained polarization: the total number of minority-opinion holders is fixed at $M = 50$ and redistributed between cliques via $m := y_1 M/N_1$. In all panels, $N_1 = 50$, $N_2 = 450$, $p = 1/2$, $\theta = 0.01$. As clique sizes become more comparable, the constrained-polarity trends weaken and eventually reverse.

analysis identified the competing mechanisms that give rise to an optimal modularity: reducing α accelerates alignment ($\tau_{\text{fast}} \sim \alpha/[(1-\alpha)p]$) but amplifies fluctuations in the small clique ($D_{11} \sim (1-\alpha)/(\alpha\, r\, N)$), producing the U-shaped dependence of consensus time on α observed in simulations. The corrected decomposition (13) captures the qualitative structure of the U-shape, with the observed optima ($\alpha^* \approx 0.7$–0.9) lying in the transitional regime where fast-stage and slow-stage contributions are comparable.

Numerical exploration confirmed that p acts primarily as a uniform scaling factor, while the interaction between α and the relative clique size r produces the richest behavior. Under polarized initial conditions, an optimal α emerges that decreases with increasing polarization. Under constrained polarity, where a fixed minority is redistributed across cliques, an interior optimum in the distribution parameter m reflects the balance between fast alignment and slow diffusion. The observed optima ($\alpha^* \approx 0.7$–0.9) lie closer to the disconnection limit than leading-order scaling predicts, indicating that higher-order corrections are needed for quantitative agreement.

Several directions remain open. The small-r scaling could be extended to obtain sharper bounds on the optimal α, potentially through matched asymptotic expansions that connect the fast and slow regimes. Well-posedness of the mean consensus time PDE (A3) near the singular points (0,1) and (1,0) deserves rigorous treatment. The tolerance parameter θ, which we have not analyzed in detail, may itself influence the optimal coupling in ways worth exploring. Extensions to more than two cliques, nonbinary opinions, and heterogeneous flip probabilities would bring the model closer to realistic settings. Most importantly, calibrating the model against empirical data on opinion dynamics or behavioral spread in structured populations would test whether the mechanisms identified here operate in practice. More broadly, the drift-noise competition identified here for two cliques may generalize to hierarchical or multilayer networks, where similar timescale separations arise between local and global consensus [12], and the introduction of zealots or committed minorities could further enrich the dynamics [44].

ACKNOWLEDGMENTS

This work was partially funded by Grant No. NSF-DMS-2042413. This material is based upon work supported by the Air Force Office of Scientific Research under Awards No. FA9550-22-1-0380, No. NSF-DMS-2207700, and No. NSF-DMS-2527338.

DATA AVAILABILITY

The data (specifically, the software code that generates the data) supporting this article's findings are openly available in the repository [39], embargo periods may apply.

APPENDIX A: BACKWARD EQUATION AND THE MEAN CONSENSUS TIME PROBLEM

Given the PDE formulation (2), we now derive the mean consensus time. For this purpose, we regard the solution $P(\mathbf{x}, t)$ as a function of the initial state $\mathbf{y} = (y_1, y_2)$ through the point initial condition $P(\mathbf{x}, 0) = \delta(\mathbf{x} - \mathbf{y})$, and write

$$P(\mathbf{x}, t) = P(\mathbf{x}, t \mid \mathbf{y}, 0).$$

In this formulation, $\mathbf{y}$ serves as the state variable in the associated backward problem for first-passage time statistics. We

define a corresponding survival probability

$$G(\mathbf{y},t) := \int_{\Omega^\theta} P(\mathbf{x},t \mid \mathbf{y},0)\,d\mathbf{x},$$

which represents the probability that the process has not yet reached the absorbing boundary by time t, given the initial state $\mathbf{y}$.

Let $T(\mathbf{y})$ denote the first exit time of the process from Ω^θ. Then,

$$\mathbb{P}(T(\mathbf{y}) \geqslant t) = G(\mathbf{y},t).$$

The adjoint of the operator associated with Eq. (2) defines the backward Kolmogorov equation for functions of the initial state $\mathbf{y} = (y_1, y_2)$,

$$\frac{\partial P}{\partial t}(\mathbf{y},t) = \boldsymbol{\mu}(\mathbf{y}) \cdot \nabla_{\mathbf{y}} P(\mathbf{y},t) + \nabla_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} \cdot (D(\mathbf{y})\,P(\mathbf{y},t)), \tag{A1}$$

where the drift and diffusion coefficients are defined as in Eq. (2), but evaluated at $\mathbf{y}$ instead of $\mathbf{x}$. Integrating Eq. (A1) over the domain Ω^θ yields the corresponding backward equation for the survival probability G,

$$\frac{\partial G}{\partial t}(\mathbf{y},t) = \boldsymbol{\mu}(\mathbf{y}) \cdot \nabla_{\mathbf{y}} G(\mathbf{y},t) + \nabla_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} \cdot (D(\mathbf{y})\,G(\mathbf{y},t)). \tag{A2}$$

The function $G(\mathbf{y},t)$ is the survival probability associated with the first exit (consensus) time. Its negative derivative gives the probability density function of the first exit time. Using standard first-passage arguments [32] together with the no-flux boundary conditions (3), the mean consensus time satisfies

$$T(\mathbf{y}) = \int_0^\infty G(\mathbf{y},t)dt.$$

Since consensus is always reached, $G(\mathbf{y},t) \to 0$ as $t \to \infty$ and $G(\mathbf{y},0) = 1$, hence $\int_0^\infty \partial_t G(\mathbf{y},t)\,dt = G(\mathbf{y},\infty) - G(\mathbf{y},0) = -1$. Integrating Eq. (A2) with respect to time then yields

$$-1 = \mu_1(\mathbf{y})\frac{\partial T}{\partial y_1} + \mu_2(\mathbf{y})\frac{\partial T}{\partial y_2} + D_{11}(\mathbf{y})\frac{\partial^2 T}{\partial y_1^2} + 2D_{12}(\mathbf{y})\frac{\partial^2 T}{\partial y_1 \partial y_2} + D_{22}(\mathbf{y})\frac{\partial^2 T}{\partial y_2^2}. \tag{A3}$$

Now we comment on the boundary conditions. We still have no-flux boundary conditions dictated by the derivation of Eq. (A1). However, there is a caveat regarding the form of the flux, as seen in the following divergence form of Eq. (A3):

$$1 + \nabla \cdot [D(\mathbf{y})\nabla T] + \begin{bmatrix} -\frac{\partial D_{11}}{\partial y_1} - \frac{\partial D_{12}}{\partial y_2} + \mu_1(\mathbf{y}) \\ -\frac{\partial D_{12}}{\partial y_1} - \frac{\partial D_{22}}{\partial y_2} + \mu_2(\mathbf{y}) \end{bmatrix} \cdot \nabla T = 0. \tag{A4}$$

The term after $\nabla\cdot$ in Eq. (A4) is the "backward" flux for Eq. (A3). When comparing fluxes in Eqs. (3) and (A4), in addition to sign changes, the drift terms (those that include μ_i) are absorbed into the first-order (reaction) term. Instead, the drift terms and higher derivatives of diffusion rates are contained in the nondivergence contribution to the operator.

APPENDIX B: FIRST-ORDER SOLVABILITY AND THE SLOW-STAGE EQUATION

Here we carry out the first-order (P^1) calculation whose solvability condition yields the slow-stage equation (10). We proceed to the next order in the expansion. Dropping primes from the shifted temporal variables gives

$$\begin{aligned} P^1_t - 2p(1-\alpha)(sP^1)_s - \frac{p+p^2-2\alpha p^2}{2}(1-2w^2)P^1_{ss} \\ = [C_4 + 4C_2(L-1) - 4C_3L]P^0 + [-8C_2 + C_4]P^0_w \\ + [s(C_4 + 8L(C_2 - C_3))]P^0_s + C_2(1-2w^2)P^0_{ww} \\ + 2Ls^2(C_2 - C_3)P^0_{ss} + C_4 ws\,P^0_{ws} - P^0_T, \end{aligned} \tag{B1}$$

where

$$\begin{aligned} C_2 &= p(1-p)/2, \\ C_3 &= p^2(1-\alpha), \\ C_4 &= -4\alpha p(1-p), \quad L = 2(1-2\alpha). \end{aligned} \tag{B2}$$

Note, the coefficients in Eq. (B2) coincide with those of the diffusion operator in Eq. (2); we retain them in this form since simplifications will follow. Equation (B1) has a solution in P^1 only if the right-hand side satisfies a Fredholm solvability condition: it must be orthogonal to the null space of the adjoint of the s-operator on the left-hand side, which is spanned by constants [38]. Hence, we integrate the right-hand side over the boundary-layer domain Ω^θ_s in the stretched variable s (equivalently Ω^θ_z for z) and set the result to zero

$$\begin{aligned} \int_{\Omega^\theta_s} P^0_T\,ds = \int_{\Omega^\theta_s} \{[C_4 + 4C_2(L-1) - 4C_3L]P^0 + [-8C_2 + C_4]P^0_w + [s(C_4 + 8L(C_2 - C_3))]P^0_s \\ + C_2(1-2w^2)P^0_{ww} + 2Ls^2(C_2 - C_3)P^0_{ss} + C_4 ws\,P^0_{ws}\}ds. \end{aligned}$$

Since $P^0 = A(w,T)\,B(s,t)$, integrating over s eliminates the fast variable and yields an evolution equation for $A(w,T)$. To evaluate the resulting integrals, we impose

$$\int_{\Omega^\theta_s} B(s,t) = 1 \quad \text{and} \quad s^n B(s,t)\big|_{\partial\Omega^\theta_s} = 0 \quad \text{for } n = 0, 1.$$

The first condition normalizes B as a density of s; the second ensures that boundary terms vanish after integration by parts. Both are justified by the fact that B is a narrow Gaussian in s for fixed w, as established in Eq. (9). With these assumptions and after a few integrations by parts, one obtains the slow-stage equation (10) of the main text.

APPENDIX C: SMALL-CLIQUE SCALING OF THE EFFECTIVE DIFFUSION

Here we provide the small-r scaling algebra leading to the effective diffusion coefficient (C1). On the alignment manifold, $x_1 + x_2 - 2x_1x_2 \approx 2x(1-x)$ and $x_i(1-x_i) \approx x(1-x)$, so the small-$r$ diffusion components from Sec. II C reduce to

$$D_{11} = \frac{(1-\alpha)\,p\,x(1-x)}{\alpha\,r\,N},$$
$$D_{22} \approx \frac{p(1-p)\,x(1-x)}{N},$$
$$D_{12} = -\frac{(1-\alpha)\,p^2\,x(1-x)}{\alpha\,N}.$$

Since $x = rx_1 + (1-r)x_2$ is a linear combination, $D_{\text{eff}} = r^2D_{11} + 2r(1-r)D_{12} + (1-r)^2D_{22}$. The dominant contribution is $(1-r)^2D_{22} \approx p(1-p)\,x(1-x)/N$, the standard well-mixed diffusion. The term $r^2D_{11} = r(1-\alpha)\,p\,x(1-x)/(\alpha N)$ provides an $\mathcal{O}(r)$ correction (the r^2 cancels one power of r in D_{11}), while $2r(1-r)D_{12}$ is also $\mathcal{O}(r)$ but carries p^2 rather than $p(1-p)$ and is dropped at leading order. This gives

$$D_{\text{eff}} \approx \frac{p\,x(1-x)}{N}[(1-p) + \frac{r(1-\alpha)}{\alpha}]. \tag{C1}$$

[1] M. Girvan and M. E. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).

[2] S. Fortunato, Community detection in graphs, Phys. Rep. **486**, 75 (2010).

[3] D. Centola, *How Behavior Spreads: The Science of Complex Contagions* (Princeton University Press, Princeton, NJ, 2018), Vol. 3.

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

[5] D. Centola, The spread of behavior in an online social network experiment, Science **329**, 1194 (2010).

[6] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes* (Springer Science & Business Media, Berlin, 2013), Vol. 324.

[7] P. Clifford and A. Sudbury, A model for spatial conflict, Biometrika **60**, 581 (1973).

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

[9] V. Sood and S. Redner, Voter model on heterogeneous graphs, Phys. Rev. Lett. **94**, 178701 (2005).

[10] G. J. Baxter, R. A. Blythe, and A. J. McKane, Fixation and consensus times on a network: A unified approach, Phys. Rev. Lett. **101**, 258701 (2008).

[11] T. Antal, S. Redner, and V. Sood, Evolutionary dynamics on degree-heterogeneous graphs, Phys. Rev. Lett. **96**, 188104 (2006).

[12] N. Masuda, Voter model on the two-clique graph, Phys. Rev. E **90**, 012802 (2014).

[13] M. T. Gastner and K. Ishida, Voter model on networks partitioned into two cliques of arbitrary sizes, J. Phys. A: Math. Theor. **52**, 505701 (2019).

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

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

[16] G. Toscani, Kinetic models of opinion formation, Commun. Math. Sci. **4**, 481 (2006).

[17] G. Albi, L. Pareschi, G. Toscani, and M. Zanella, Recent advances in opinion modeling: Control and social influence, in *Active Particles*, Volume 1, edited by N. Bellomo, P. Degond, and E. Tadmor, Modeling and Simulation in Science, Engineering and Technology (Birkhäuser, Cham, 2017), pp. 49–98.

[18] B. Düring, J. Franceschi, M.-T. Wolfram, and M. Zanella, Breaking consensus in kinetic opinion formation models on graphons, J. Nonlinear Sci. **34**, 79 (2024).

[19] J. Grilli, T. Rogers, and S. Allesina, Modularity and stability in ecological communities, Nat. Commun. **7**, 12031 (2016).

[20] É. Danchin, L.-A. Giraldeau, T. J. Valone, and R. H. Wagner, Public information: From nosy neighbors to cultural evolution, Science **305**, 487 (2004).

[21] T. C. Bergstrom, Evolution of social behavior: Individual and group selection, J. Econ. Perspect. **16**, 67 (2002).

[22] B. Allen, G. Lippner, Y.-T. Chen, B. Fotouhi, N. Momeni, S.-T. Yau, and M. A. Nowak, Evolutionary dynamics on any population structure, Nature (London) **544**, 227 (2017).

[23] D. Hathcock and S. H. Strogatz, Fitness dependence of the fixation-time distribution for evolutionary dynamics on graphs, Phys. Rev. E **100**, 012408 (2019).

[24] G. W. A. Constable, T. Rogers, A. J. McKane, and C. E. Tarnita, Demographic noise can reverse the direction of deterministic selection, Proc. Natl. Acad. Sci. USA **113**, E4745 (2016).

[25] G. W. A. Constable and A. J. McKane, Exploiting fast-variables to understand population dynamics and evolution, J. Stat. Phys. **172**, 3 (2018).

[26] S. Ghosh, *Distributed Systems: An Algorithmic Approach* (Chapman and Hall/CRC, Boca Raton, FL, 2006).

[27] R. Olfati-Saber and R. M. Murray, Consensus problems in networks of agents with switching topology and time-delays, IEEE Trans. Autom. Control **49**, 1520 (2004).

[28] A. Olshevsky and J. N. Tsitsiklis, Convergence speed in distributed consensus and averaging, SIAM J. Control Optim. **48**, 33 (2009).

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

[30] D. Bhat and S. Redner, Polarization and consensus by opposing external sources, J. Stat. Mech. (2020) 013402.

[31] L. Becchetti, V. Bonifaci, E. Cruciani, and F. Pasquale, in *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI)*, edited by E. Elkind (International Joint Conferences on Artificial Intelligence, 2023), pp. 38–45.

[32] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, 2nd ed.,

Springer Series in Synergetics, Vol. 13 (Springer, Berlin, 1985).

[33] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence* (John Wiley & Sons, New York, NY, 1986).

[34] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, Amsterdam, 2007).

[35] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, Berlin, 1989).

[36] A. J. McGann, The tyranny of the supermajority, J. Theor. Politics **16**, 53 (2004).

[37] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-uniform Gases*, 3rd ed. (Cambridge University Press, Cambridge, UK, 1970).

[38] G. A. Pavliotis, *Stochastic Processes and Applications* (Springer, Berlin, 2014), Vol. 60.

[39] M. Yerlanov, Voter model code, 2025, https://github.com/MaYeatCo/Voter_Model.

[40] P. Trubowitz and N. Mellow, "Going bipartisan": Politics by other means, Political Sci. Q. **120**, 433 (2005).

[41] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, Exposure to opposing views on social media can increase political polarization, Proc. Natl. Acad. Sci. USA **115**, 9216 (2018).

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

[43] W. J. Ewens, *Mathematical Population Genetics: Theoretical Introduction* (Springer, Berlin, 2004), Vol. 27.

[44] M. Mobilia, Does a single zealot affect an infinite group of voters? Phys. Rev. Lett. **91**, 028701 (2003).