

SUPPLEMENTARY INFORMATION: REPLICATOR–MUTATOR DYNAMICS ON NETWORKS WITH SPATIAL NONLOCALITY

PORIKSHIT MONDAL AND ERNESTO ESTRADA

S1. NUMERICAL METHODS AND SIMULATION DETAILS FOR THE GENERALIZED REPLICATOR-MUTATOR MODEL

In this section, we specifically focus on the impact of nonlocal interactions in the replicator-mutator model, realized on different artificial networks. For instance, all numerical experiments are performed using a common computational framework. Only the network topologies are varied within their respective structural parameters. Here, we consider three stochastic network models: Watts–Strogatz (WS), Erdős–Rényi (ER), and Barabási–Albert (BA) networks. All the outputs are obtained in both linear and logarithmic scales.

The numerical pipeline consists of the following steps:

- (1) generate an unweighted, undirected network and its adjacency matrix;
- (2) calculate the all-pairs shortest-path distance matrix using breadth-first search (BFS);
- (3) construct the Mellin-weighted nonlocal interaction matrix;
- (4) construct the row-normalized transition matrix and mutation matrix;
- (5) integrate the generalised nonlocal Replicator–Mutator differential equations;
- (6) remove small negative terminal components caused by numerical floating-point errors and renormalize the final state;
- (7) calculate the effective diversity $n_e = 1/\sum_i x_i^2$;
- (8) repeat the calculation over independent network realizations; and
- (9) average the resulting diversity over the prescribed graph realizations.

All three classes of networks use the same universal simulation framework, with the base parameters $n = 50$ and $a = 0.3$. Moreover, for the Watts–Strogatz network,

$$(S1.1) \quad k = 10, \quad p = 0.1,$$

where k is the number of nearest neighbours in the initial ring lattice and p is the rewiring probability.

For the Erdős–Rényi network,

$$(S1.2) \quad p_{\text{edge}} = 0.118,$$

where every unordered pair of nodes is independently connected with probability p_{edge} .

For the Barabási–Albert network,

$$(S1.3) \quad m = 3,$$

where every newly introduced node attaches to m existing nodes according to preferential attachment, starting from a fully connected core of $m + 1$ nodes.

S1.1. Independent graph realizations. The contour plots of different networks in the main manuscript are considered for a single realization (seed=123) with random initial conditions. Moreover, to reduce the dependence on a particular stochastic network realization, we consider $N_G = 20$ independent graph realizations for each network class.

To explore the behaviour of transitions for strong nonlocal interactions (small Mellin exponent s), we work with both (a) a single realization of the networks and (b) $N_G = 20$ independent graph realizations. The detailed explanations are given in section S2 of this Supplementary Information.

For multiple realizations with $r = 1, \dots, 20$, the network generator is called with a different seed. Here, the seeds are chosen as

$$(S1.4) \quad 123, 124, \dots, 142,$$

i.e.

$$(S1.5) \quad \text{seed}_r = 123 + (r - 1).$$

For each realization, the complete (μ, s) parameter grid is evaluated. The resulting diversity matrices are accumulated and then divided by $N_G = 20$. Thus, the reported quantity is

$$(S1.6) \quad \overline{n_e}(\mu, s) = \frac{1}{20} \sum_{r=1}^{20} n_e^{(r)}(\mu, s).$$

This procedure averages over network realizations, while using one random initial condition at each grid point for each realization. It therefore does not constitute an explicit ensemble average over multiple initial conditions for a fixed network and parameter point.

S1.2. Initial-condition distribution. For simulations, the initial state is considered independently from the symmetric Dirichlet distribution

$$(S1.7) \quad \mathbf{x}_0 \sim \text{Dirichlet}(1, \dots, 1).$$

This distribution is uniform over the probability simplex

$$(S1.8) \quad \Delta^{n-1} = \left\{ \mathbf{x} \in \mathbb{R}^n : x_i \geq 0, \sum_{i=1}^n x_i = 1 \right\}.$$

The standard grid calculation uses one such initial condition at each (μ, s) point for every graph realization. With 20 independent graph realizations, the full calculation therefore requires

$$(S1.9) \quad 20 \times 400 \times 400 = 3.2 \times 10^6$$

ODE integrations per network class and per grid type. The Dirichlet initial condition is not an ensemble average over initial conditions. It is one representative random interior state per grid point.

S1.3. Numerical ODE solver. Our differential equation is integrated using the `solve_ivp` routine from SciPy. The base implementations use the adaptive explicit Runge-Kutta method RK45.

Here, the integration interval is fixed at

$$(S1.10) \quad t \in [0, 5000].$$

This integration interval is chosen to allow the transient dynamics to decay before evaluating the terminal state. Additional numerical checks and validations of the terminal state are presented in section S4.

Moreover, the relative and absolute tolerances are

$$(S1.11) \quad \text{rtol} = 10^{-7}, \quad \text{atol} = 10^{-7}.$$

S1.4. Treatment of numerical negative components. Although the mathematical dynamics are formulated on the probability simplex, finite-precision numerical integration can produce extremely small negative components near zero (numerical checks and the corresponding validations are explained in section S4).

The base implementation treats such terminal numerical values by component-wise clipping:

$$(S1.12) \quad \tilde{x}_i = \max\{x_i(t_f), 0\}.$$

The clipped vector is then renormalized:

$$(S1.13) \quad x_i^{\text{final}} = \frac{\tilde{x}_i}{\sum_j \tilde{x}_j}.$$

The diversity is then calculated from this final normalized vector.

Note that the clipping is a numerical post-processing operation only, and it is not a part of the differential equation evaluated by the ODE solver.

TABLE S1.1. Numerical parameters used in the WS, ER, and BA network simulations.

Quantity	Setting
Network classes	WS, ER, BA
Network size	$n = 50$
WS parameters	$k = 10, p = 0.1$
ER parameter	$p_{\text{edge}} = 0.118$
BA parameter	$m = 3$
Interaction strength	$a = 0.3$
Graph realizations	$N_G = 20$
Graph seeds	123, ..., 142
Initial condition	Dirichlet(1, ..., 1)
Shortest-path algorithm	BFS
Mellin weight	d^{-s}
ODE solver	Adaptive RK45 (<code>solve_ivp</code>)
Integration interval	[0, 5000]
Relative tolerance	10^{-7}
Absolute tolerance	10^{-7}
Terminal negative values	Clipped to zero
Terminal state	Renormalized to unit sum
Diversity	$n_e = 1 / \sum_i x_i^2$
Linear μ -grid	[0, 10], 400 points
Linear s -grid	[0, 20], 400 points
Log μ -grid	$[10^{-4}, 10]$, 400 points
Log s -grid	$[10^{-2}, 10^{1.5}]$, 400 points
Grid size	400×400

S2. NONLOCALITY-INDUCED COLLAPSE AND ITS BEHAVIOUR OF TRANSITION

In the region of strong nonlocal interactions, corresponding to sufficiently small values of the Mellin exponent s , we observe an interesting phenomenon, namely nonlocality-induced collapse, in which a sharp transition from behavioural flocking to complete collapse occurs, as clearly visible in Fig. S2.1. However, a sharp crossover by itself is not sufficient to identify the transition as a first-order phase transition. Therefore, we perform a systematic three-step numerical verification: (i) system-size dependence, (ii) hysteresis under forward and backward mutation-rate sweeps, and (iii) initial-condition dependence and quantitative bistability.

These three tests are deliberately complementary. The system-size analysis asks whether the transition sharpens with increasing system size. The hysteresis analysis asks whether two distinct branches can persist over a finite interval of the control parameter. The bistability analysis directly examines whether, at the same parameter values, different initial conditions can converge to distinct attracting states.

All analyses are performed using extended-time integration to verify that the observed branches are not merely transient states.

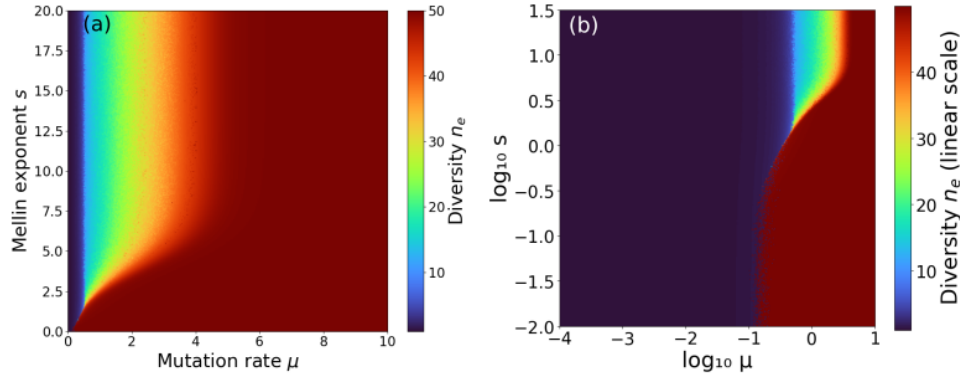


FIGURE S2.1. Contour plots for the Watts–Strogatz network, where the mutation rate μ and the Mellin exponent s are plotted against the effective diversity n_e . Subfigure (a) is shown on a linear scale, whereas subfigure (b) is shown on a logarithmic scale. Both correspond to rewiring probability $p = 0.1$. The common simulation parameters are $n = 50$, $k = 10$, and $a = 0.3$.

All three numerical experiments below use the same basic definition and implementation. We primarily consider Watts–Strogatz (WS) networks to present the three analyses in detail, with the common parameters

$$(S2.1) \quad k = 10, \quad p = 0.1, \quad a = 0.3.$$

The dynamical equation is integrated using SciPy’s `solve_ivp` with the explicit RK45 method. The solver tolerances used in the extended verification are

$$(S2.2) \quad \text{rtol} = 10^{-7}, \quad \text{atol} = 10^{-7}.$$

For all three analyses, the integration is performed with a maximum total integration time

$$(S2.3) \quad t_{\max} = 5000.$$

The terminal convergence criterion is defined using the Euclidean norm,

$$(S2.4) \quad \|\dot{\mathbf{x}}(t_f)\|_2 < \varepsilon_{\text{conv}}, \quad \varepsilon_{\text{conv}} = 10^{-6}.$$

The purpose of this residual criterion is to distinguish a genuinely stationary numerical state from a trajectory that merely evolves slowly at the end of a finite simulation window.

Before starting the analysis, we note that we have also performed the corresponding analyses for Erdős–Rényi and Barabási–Albert networks, using their corresponding 20 network realizations. However, to avoid unnecessary repetition, the three experiments below are discussed primarily using Watts–Strogatz networks. The corresponding qualitative behaviour for the other network classes is summarized later.

S2.1. System-size analysis. Here, we investigate whether the width of the transition decreases, or equivalently whether the maximum slope increases, as the number of behaviours/nodes is increased. This provides a finite-size analogue of asking whether the observed discontinuous response persists and sharpens with increasing system size.

First, we note that the different dynamical regimes are quantitatively characterized through the effective diversity n_e , which itself depends on the system size n . To compare systems with different values of n , we normalize the diversity by defining

$$(S2.5) \quad \eta = \frac{n_e - 1}{n - 1}.$$

Then,

$$(S2.6) \quad n_e = 1 \implies \eta = 0,$$

$$(S2.7) \quad n_e = n \implies \eta = 1.$$

Consequently,

$$(S2.8) \quad \eta \approx 0$$

represents behavioural flocking, while

$$(S2.9) \quad \eta \approx 1$$

represents complete collapse.

S2.1.1. Methodology. For WS networks with the fixed common parameters ($k = 10, p = 0.1, a = 0.3$), the system-size calculation considers three sufficiently small values of the Mellin exponent: (a) $s = 10^{-1}$, (b) $s = 10^{-2}$, and (c) $s = 10^{-3}$.

We then compute $\eta(\mu)$ for

$$(S2.10) \quad n = 25, 50, 100, 200, 400, 700, 1000.$$

The mutation rate is scanned logarithmically over

$$(S2.11) \quad 10^{-4} \leq \mu \leq 10.$$

For the numerical experiments, we use $M_\mu = 1000$ grid points within this μ range.

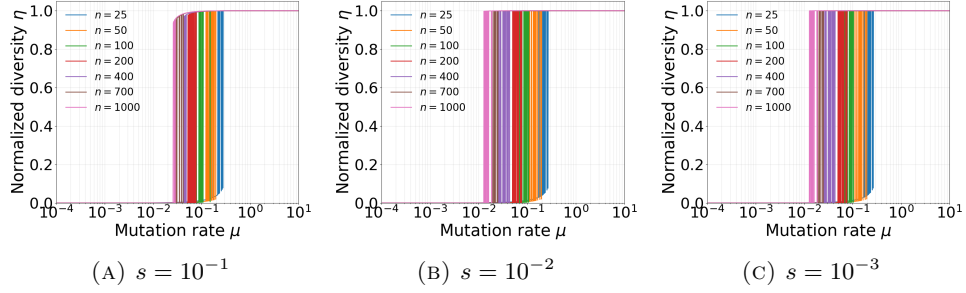


FIGURE S2.2. System-size analysis for the WS network ($\text{seed}=123$) with $k = 10$, $a = 0.3$, and $p = 0.1$ under varying initial conditions: (a) $s = 10^{-1}$, (b) $s = 10^{-2}$, and (c) $s = 10^{-3}$.

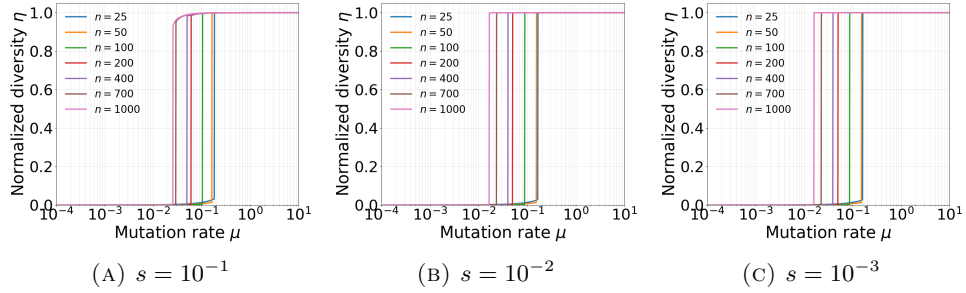


FIGURE S2.3. System-size analysis for the WS network ($\text{seed}=123$) with $k = 10$, $a = 0.3$, and $p = 0.1$ under fixed initial conditions: (a) $s = 10^{-1}$, (b) $s = 10^{-2}$, and (c) $s = 10^{-3}$.

Moreover, we consider two initial-condition protocols: (a) independently sampled random initial conditions (see Fig. S2.2); and (b) one fixed Dirichlet realization for each system size, reused across the entire μ scan for that system size (see Fig. S2.3).

The phrase “fixed initial condition” means that, for a given n , the same n -dimensional initial vector is used at every value of μ . It does *not* mean that exactly the same vector is used for different values of n , since the dimension of the state space changes.

The initial states are sampled from the symmetric Dirichlet distribution,

$$(S2.12) \quad \mathbf{x}_0 \sim \text{Dirichlet}(1, \dots, 1),$$

which is uniform over the simplex.

S2.1.2. Results. The resulting $\eta(\mu)$ curves become progressively sharper as n increases, as shown in Figs. S2.2 and S2.3. We observe that the response approaches a step-like transition between

$$(S2.13) \quad \eta \approx 0$$

and

$$(S2.14) \quad \eta \approx 1.$$

The random initial conditions can produce multiple apparent transition lines (Fig. S2.2), whereas fixing the initial condition produces a clean transition (Fig. S2.3). As discussed below, the subsequent bistability analysis provides evidence that these multiple lines can be associated with distinct coexisting attracting states. In contrast, the fixed-initial-condition curves effectively select one branch, while the hysteresis analysis reveals the coexistence region.

TABLE S2.1. Transition metrics for the Watts–Strogatz networks for different network sizes n , with $s = 10^{-2}$ and fixed initial conditions.

n	$\mu(\eta = 0.1)$	$\mu_c(\eta = 0.5)$	$\mu(\eta = 0.9)$	Width ($\Delta\mu_{0.1-0.9}$)	$\max_{\mu} \frac{d\eta}{d\mu}$
25	0.17522810	0.17606544	0.17690279	0.0016746924	240.22622
50	0.12839997	0.12900077	0.12960156	0.0012015934	334.80986
100	0.08983416	0.09025157	0.09066897	0.0008348164	481.90903
200	0.06071296	0.06099444	0.06127593	0.0005629687	714.61540
400	0.03040782	0.03054887	0.03068992	0.0002821051	1427.29953
700	0.02805116	0.02818129	0.02831142	0.0002602576	1547.50368
1000	0.02202154	0.02212402	0.02222651	0.0002049712	1969.28981

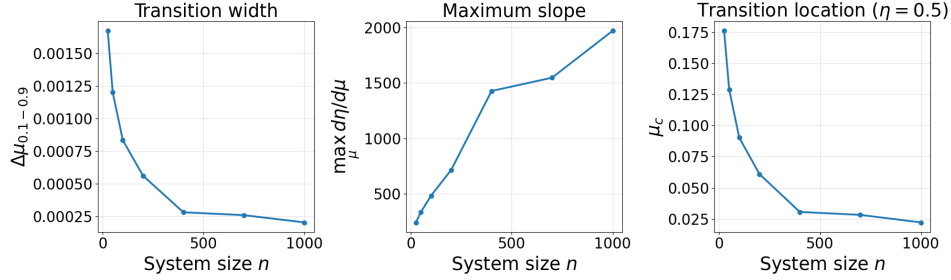


FIGURE S2.4. Quantification of the system-size sharpening for the WS network with $k = 10$, $a = 0.3$, $p = 0.1$, and $s = 10^{-2}$, under fixed initial conditions.

For the small value of the Mellin exponent $s = 10^{-2}$, the transition can be seen in Fig. S2.3(b). The corresponding system-size quantification is shown in Table S2.1, and its graphical representation is shown in Fig. S2.4.

For each value of n , we characterize the transition using three quantities. First, we define the transition width as

$$(S2.15) \quad \Delta\mu_{0.1-0.9} = \mu(\eta = 0.9) - \mu(\eta = 0.1),$$

which measures the interval over which the normalized diversity changes from 0.1 to 0.9. Second, we calculate the maximum slope

$$(S2.16) \quad \max_{\mu} \frac{d\eta}{d\mu}$$

as a measure of the steepness of the response. Finally, we define an operational transition location μ_c by

$$(S2.17) \quad \eta(\mu_c) = 0.5.$$

As shown in Fig. S2.4, the transition width decreases systematically with increasing system size, from 1.67×10^{-3} at $n = 25$ to 2.05×10^{-4} at $n = 1000$. Simultaneously, the maximum slope increases from approximately 240 to 1969. Thus, the diversity response becomes progressively sharper as the network size increases. The corresponding transition location also shifts toward smaller values of μ , from $\mu_c \simeq 0.176$ at $n = 25$ to $\mu_c \simeq 0.0221$ at $n = 1000$.

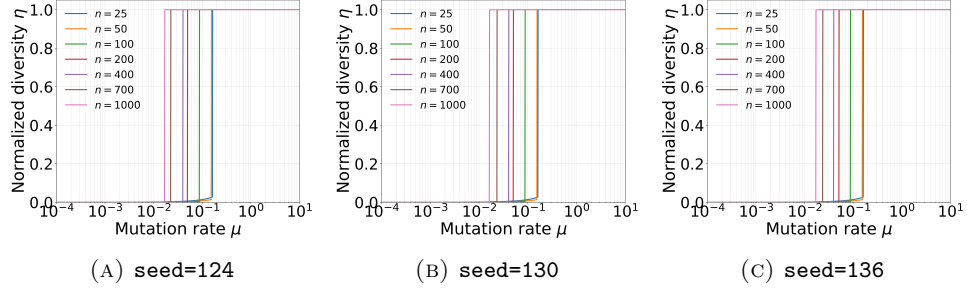


FIGURE S2.5. System-size analysis for the WS networks with $k = 10$, $a = 0.3$, $p = 0.1$, and $s = 10^{-2}$. The subfigures show different realizations of the WS network: (a) seed=124, (b) seed=130, and (c) seed=136.

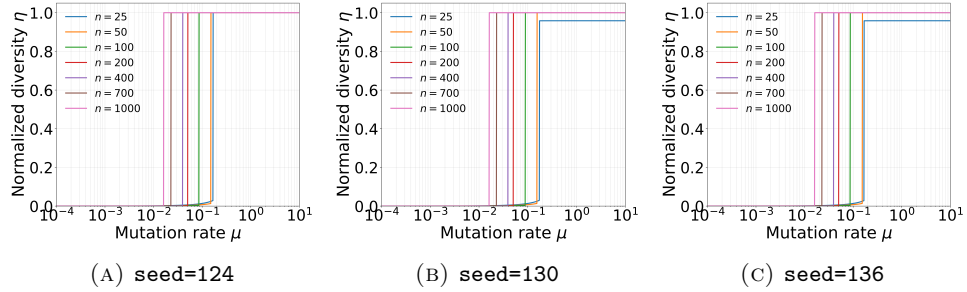


FIGURE S2.6. System-size analysis for the ER networks with $a = 0.3$, $p_{\text{edge}} = 0.118$, and $s = 10^{-2}$. The subfigures show different realizations of the ER network: (a) seed=124, (b) seed=130, and (c) seed=136.

These results provide evidence of finite-size sharpening of the transition. A similar analysis was performed for other sufficiently small values of the Mellin exponent s , using all 20 realizations of the WS network. We similarly performed the analysis for other network classes, including Erdős–Rényi and Barabási–Albert networks, using different network realizations. Representative outcomes are shown in Figs. S2.5, S2.6, and S2.7. We observed the same qualitative trend across the tested realizations, namely finite-size sharpening of the transition. Since finite-size sharpening alone does not establish the order of a transition, we complement this analysis with independent tests of hysteresis and bistability.

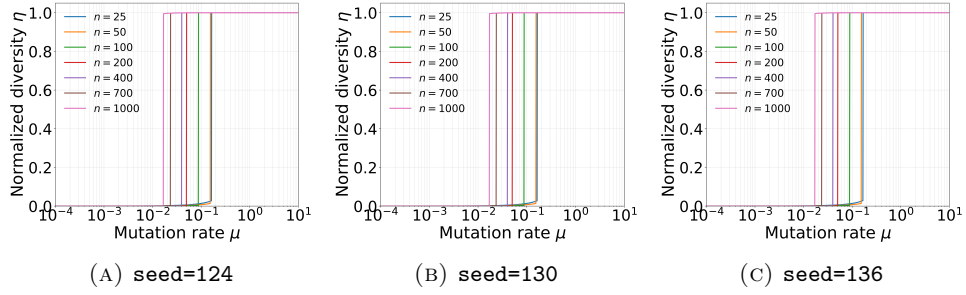


FIGURE S2.7. System-size analysis for the BA networks with $a = 0.3$, $m = 3$, and $s = 10^{-2}$. The subfigures show different realizations of the BA network: (a) seed=124, (b) seed=130, and (c) seed=136.

S2.2. Hysteresis. We investigate whether, at the same value of μ , the state obtained by increasing μ differs from the state obtained by decreasing μ . If, over a finite interval of μ , one observes

$$(S2.18) \quad \eta^\uparrow(\mu) \neq \eta^\downarrow(\mu),$$

this indicates hysteresis. Here, $\eta^\uparrow(\mu)$ denotes the response obtained while increasing μ , whereas $\eta^\downarrow(\mu)$ denotes the response obtained while decreasing μ . Thus, hysteresis means that the state obtained at a given control parameter μ depends on the direction in which the parameter was varied. A hysteresis loop occurs when

$$(S2.19) \quad \eta^\uparrow(\mu) \neq \eta^\downarrow(\mu)$$

over a finite interval of μ .

The hysteresis width is defined as

$$(S2.20) \quad \Delta\mu_{\text{hyst}} = \mu^\uparrow - \mu^\downarrow,$$

where $\mu^\uparrow$ and $\mu^\downarrow$ denote the transition points obtained from the forward and backward sweeps, respectively.

S2.2.1. Methodology. For the forward sweep, the calculation starts at a small value of μ . After integration at one value of μ , the final state is used as the initial condition for the next value:

$$(S2.21) \quad \mathbf{x}_0^{(j+1)} = \mathbf{x}_f^{(j)}.$$

After reaching the large- μ end of the parameter range, the procedure is reversed. In this case, the final state at one value of μ is used as the initial condition at the next lower value, thereby generating the backward branch.

The experiment is performed on a WS network with

$$(S2.22) \quad n = 50, \quad k = 10, \quad p = 0.1, \quad a = 0.3, \quad s = 10^{-2}.$$

S2.2.2. Numerical robustness tests. The hysteresis calculation was repeated using three continuation resolutions and three maximum integration times:

$$(S2.23) \quad M_\mu \in \{200, 500, 1000\}, \quad t_{\text{max}} \in \{1000, 2000, 5000\}.$$

The terminal convergence criterion was

$$(S2.24) \quad \|\dot{\mathbf{x}}(t_f)\|_2 < 10^{-6}.$$

The purpose of varying both the continuation resolution and the integration time is to distinguish a genuine hysteresis interval from a numerical artifact caused by coarse parameter sampling or insufficient relaxation time.

S2.2.3. Results. The complete robustness data are shown in Table S2.2. For $M_\mu = 200$, the forward sweep already converged at $t_{\max} = 1000$, whereas the backward sweep did not satisfy the residual criterion until a longer integration time was used. For $M_\mu = 500$, neither direction was fully converged at $t_{\max} = 1000$, whereas both converged at $t_{\max} = 2000$ and $t_{\max} = 5000$. For $M_\mu = 1000$, the backward branch required the longest integration time in one of the tested cases: at $t_{\max} = 2000$, its maximum residual was approximately 4.07×10^{-4} , whereas by $t_{\max} = 5000$ both branches satisfied the convergence criterion.

TABLE S2.2. Robustness of the hysteresis width with respect to the number of mutation-rate points M_μ and the maximum integration time $t_{\max}$. Here, $\mu^\uparrow$ and $\mu^\downarrow$ denote the forward and backward transition points, respectively, and $\Delta\mu_{\text{hyst}} = \mu^\uparrow - \mu^\downarrow$. “F” and “B” denote the forward and backward sweeps, respectively.

M_μ	$t_{\max}$	$\mu^\uparrow$	$\mu^\downarrow$	$\Delta\mu_{\text{hyst}}$	F conv.	B conv.	max residuals (F,B)
200	1000	0.301407	0.056419	0.244988	yes	no	9.47×10^{-7} , 4.71×10^{-6}
200	2000	0.301407	0.059776	0.241631	yes	yes	9.47×10^{-7} , 9.64×10^{-7}
200	5000	0.301407	0.059776	0.241631	yes	yes	9.47×10^{-7} , 9.64×10^{-7}
500	1000	0.310051	0.057617	0.252434	no	no	2.10×10^{-6} , 1.02×10^{-5}
500	2000	0.310051	0.058960	0.251091	yes	yes	9.94×10^{-7} , 9.91×10^{-7}
500	5000	0.310051	0.058960	0.251091	yes	yes	9.94×10^{-7} , 9.91×10^{-7}
1000	1000	0.309516	0.058921	0.250594	yes	no	9.95×10^{-7} , 1.87×10^{-6}
1000	2000	0.309516	0.058935	0.250580	yes	no	9.95×10^{-7} , 4.07×10^{-4}
1000	5000	0.309516	0.059604	0.249912	yes	yes	9.95×10^{-7} , 9.97×10^{-7}

The converged results at $t_{\max} = 5000$ are summarized in Table S2.3. The robust numerical result is therefore

$$(S2.25) \quad \mu^\uparrow \approx 0.30 \text{ to } 0.31, \quad \mu^\downarrow \approx 0.059 \text{ to } 0.060.$$

Consequently,

$$(S2.26) \quad \Delta\mu_{\text{hyst}} \approx 0.24 \text{ to } 0.25.$$

The persistence of a finite separation,

$$(S2.27) \quad \mu^\downarrow < \mu^\uparrow,$$

under refinement of the continuation grid is the central observation. The corresponding hysteresis loop is shown in Fig. S2.8, particularly for $s = 10^{-2}$. This observation provides numerical evidence for the existence of hysteresis in the system.

Hysteresis for the WS network at different sufficiently small values of the Mellin exponent s is visualized in Fig. S2.8. Moreover, we also performed the same analysis for other network classes using different network realizations. Representative outcomes are shown in Figs. S2.9, S2.10, and S2.11 for WS, ER, and BA networks, respectively. The qualitative trends are similar across the tested realizations, and hysteresis is observed in each case.

TABLE S2.3. Hysteresis robustness for the WS network at $s = 10^{-2}$. The reported values use the convergence criterion $\|\dot{\mathbf{x}}(t_f)\|_2 < 10^{-6}$.

M_μ	$t_{\max}$	$\mu_\uparrow$	$\mu_\downarrow$	$\Delta\mu_{\text{hyst}}$
200	5000	0.301407	0.059776	0.241631
500	5000	0.310051	0.058960	0.251091
1000	5000	0.309516	0.059604	0.249912

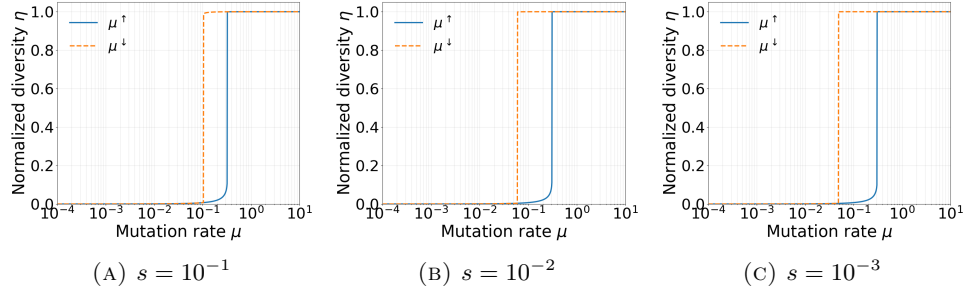


FIGURE S2.8. Hysteresis for the WS network (**seed=123**) with $n = 50$, $k = 10$, $a = 0.3$, and $p = 0.1$ under fixed initial conditions, for (a) $s = 10^{-1}$, (b) $s = 10^{-2}$, and (c) $s = 10^{-3}$.

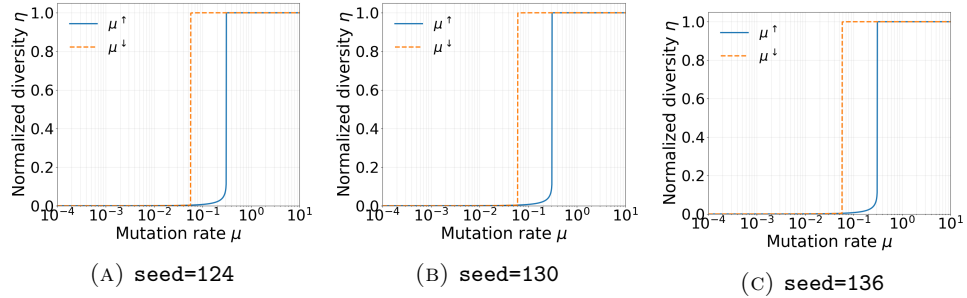


FIGURE S2.9. Hysteresis check for the WS networks with $k = 10$, $a = 0.3$, $p = 0.1$, and $s = 10^{-2}$. The subfigures show different realizations of the WS network: (a) **seed=124**, (b) **seed=130**, and (c) **seed=136**.

The hysteresis results establish a finite parameter interval in which the response depends on the continuation direction. The robustness analysis is particularly important because insufficient integration time can mimic or exaggerate hysteresis when the system relaxes slowly. The fact that both branches satisfy the same terminal residual criterion at $t_{\max} = 5000$, together with the approximate stability of the switching locations under continuation refinement, supports the conclusion that the observed hysteresis is dynamical rather than a simple finite-time numerical artifact.

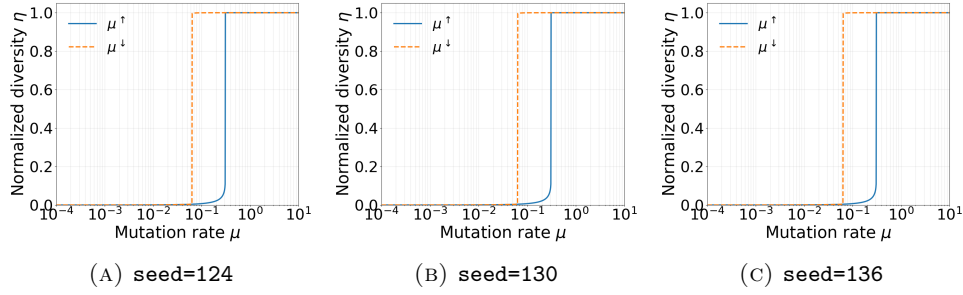


FIGURE S2.10. Hysteresis check for the ER networks with $a = 0.3$, $p_{\text{edge}} = 0.118$, and $s = 10^{-2}$. The subfigures show different realizations of the ER network: (a) seed=124, (b) seed=130, and (c) seed=136.

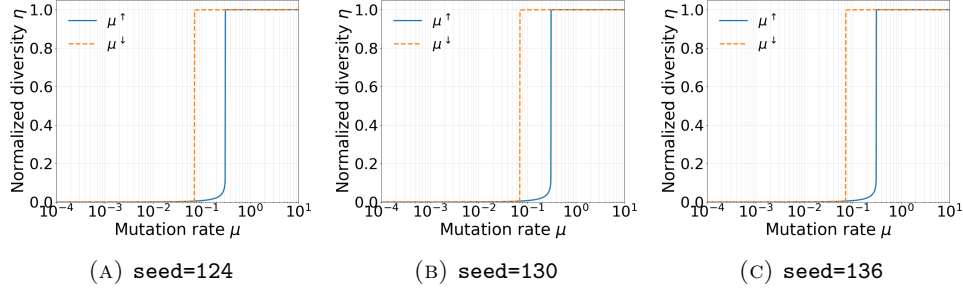


FIGURE S2.11. Hysteresis check for the BA networks with $a = 0.3$, $m = 3$, and $s = 10^{-2}$. The subfigures show different realizations of the BA network: (a) seed=124, (b) seed=130, and (c) seed=136.

S2.3. Bistability and initial-condition dependence. Here, we address the following question: for exactly the same (n, s, μ) and the same network realization, can different initial conditions converge to two distinct terminal states? If so, and if both groups of trajectories satisfy the terminal convergence criterion, this provides direct numerical evidence for the coexistence of distinct attracting states rather than merely slow transients.

Hysteresis establishes history dependence, but it does not by itself demonstrate that two attracting states coexist at a fixed parameter value. The direct test is therefore to hold the network and all model parameters fixed while varying only the initial condition.

In our case, the candidate states are (a) $\eta_{\text{low}} \approx 0$, corresponding to behavioural flocking, and (b) $\eta_{\text{high}} \approx 1$, corresponding to complete collapse.

Thus, the direct bistability test asks whether, for exactly the same (n, s, μ) , two different initial conditions can satisfy

$$(S2.28) \quad \mathbf{x}_0^{(1)} \longrightarrow \eta_{\text{low}},$$

while

$$(S2.29) \quad \mathbf{x}_0^{(2)} \longrightarrow \eta_{\text{high}}.$$

S2.3.1. *Methodology.* Here, the fixed network is a Watts–Strogatz (WS) realization with

$$(S2.30) \quad n = 50, \quad k = 10, \quad p = 0.1, \quad a = 0.3.$$

The tested mutation rates are

$$(S2.31) \quad \mu = 0.04, 0.10, 0.15, 0.20, 0.40.$$

For every fixed μ , 100 independent Dirichlet initial conditions are generated:

$$(S2.32) \quad \mathbf{x}_0^{(r)} \sim \text{Dirichlet}(1, \dots, 1), \quad r = 1, \dots, 100.$$

Each trajectory is integrated independently up to $t_{\max} = 5000$. The terminal state is then used to calculate the effective number of behaviours n_e and the normalized diversity η .

The quantitative calculation uses

$$(S2.33) \quad t_{\max} = 5000, \quad \varepsilon_{\text{conv}} = 10^{-6},$$

with the RK45 solver and $\text{rtol} = \text{atol} = 10^{-7}$. Convergence to a stationary state is assessed using the terminal residual

$$(S2.34) \quad \|\dot{\mathbf{x}}(t_f)\|_2 < \varepsilon_{\text{conv}},$$

where $\|\cdot\|_2$ denotes the Euclidean norm.

S2.3.2. *Results.* We observe in figure S2.12 that, for sufficiently small values of the Mellin exponent s , the terminal states can split into distinct branches for the suitable values of the mutation rate μ . Each trajectory corresponds to a different initial condition, while the colour identifies the mutation rate associated with the corresponding set of initial conditions. To quantify this behaviour, we introduce the following branch classification.

We define well-separated thresholds by

$$(S2.35) \quad \begin{cases} \eta < 0.1, & \text{low-diversity branch,} \\ \eta > 0.9, & \text{high-diversity branch,} \\ 0.1 \leq \eta \leq 0.9, & \text{intermediate state.} \end{cases}$$

Thus, a trajectory is counted as belonging to a clearly identified branch only when its terminal diversity is sufficiently far from the middle of the interval.

For $s = 10^{-2}$, shown in subfigure S2.12b, we further quantify the terminal states in table S2.4. The table shows the distribution of the final normalized diversity for the different values of the mutation rate μ . Here, M_{low} and M_{high} denote the numbers of trajectories ending in the low- and high-diversity states, respectively, while f_{low} and f_{high} denote their corresponding fractions. The quantities $\langle \eta \rangle$ and σ denote the mean and standard deviation of the normalized diversity within each branch.

We observe from table S2.4 that the number of trajectories classified as intermediate is zero for every tested mutation rate:

$$(S2.36) \quad M_{\text{intermediate}} = 0.$$

Furthermore,

$$(S2.37) \quad M_{\text{converged}} = M_{\text{low}} + M_{\text{high}} = 100$$

for every tested μ .

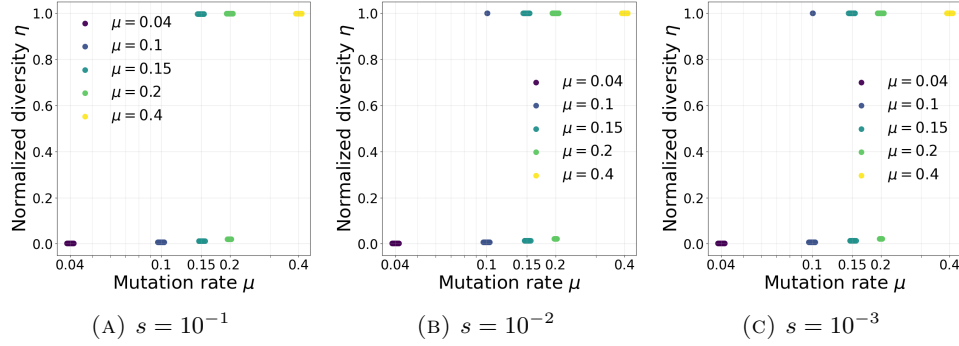


FIGURE S2.12. Bistability for the WS network (**seed=123**) with $n = 50$, $k = 10$, $a = 0.3$, and $p = 0.1$ under varying initial conditions, for (a) $s = 10^{-1}$, (b) $s = 10^{-2}$, and (c) $s = 10^{-3}$.

TABLE S2.4. Quantitative bistability verification. Branch statistics are based on the terminal normalized diversity η . The branch criterion is $\eta < 0.1$ for the low branch and $\eta > 0.9$ for the high branch. All 100 trajectories satisfy the convergence criterion $\|\dot{\mathbf{x}}(t_f)\|_2 < 10^{-6}$.

μ	M_{low}	f_{low}	$\langle \eta \rangle_{\text{low}}$	σ_{low}	M_{high}	f_{high}	$\langle \eta \rangle_{\text{high}}$	σ_{high}
0.04	100	1.00	0.002408	5.66×10^{-7}	0	0.00	—	—
0.10	98	0.98	0.007175	2.01×10^{-6}	2	0.02	0.999925	1.41×10^{-9}
0.15	44	0.44	0.012851	4.89×10^{-6}	56	0.56	0.999980	1.07×10^{-9}
0.20	5	0.05	0.021410	9.22×10^{-6}	95	0.95	0.999991	3.30×10^{-16}
0.40	0	0.00	—	—	100	1.00	0.999998	3.30×10^{-16}

The convergence statistics are summarized in table S2.5. All trajectories satisfy the convergence criterion $\|\dot{\mathbf{x}}(t_f)\|_2 < 10^{-6}$.

TABLE S2.5. Terminal residual statistics for the quantitative bistability experiment.

μ	M_{conv}	$\max \ \dot{\mathbf{x}}(t_f)\ _2$	$\overline{\ \dot{\mathbf{x}}(t_f)\ _2}$
0.04	100	9.4463×10^{-7}	2.9928×10^{-7}
0.10	100	9.6430×10^{-7}	3.2244×10^{-7}
0.15	100	8.6400×10^{-7}	1.5074×10^{-7}
0.20	100	2.2959×10^{-7}	6.5590×10^{-8}
0.40	100	2.2060×10^{-7}	5.9842×10^{-8}

We next compare the results with the hysteresis interval identified previously.

- (a) For the mutation rates $\mu = 0.04$ and $\mu = 0.40$, which are outside the hysteresis region (compare subfigures S2.12b and S2.8b, and recall that for $s = 10^{-2}$, the hysteresis analysis gave $\mu^\uparrow \approx 0.309516$, and $\mu^\downarrow \approx 0.059604$ for $M_\mu = 1000$ and $t_{\text{max}} = 5000$), the initial conditions select only one of the two observed branches.

At $\mu = 0.04$, all tested initial conditions converge to

$$(S2.38) \quad \eta \approx 0.$$

Thus, the low-diversity branch is selected for all tested initial conditions at this parameter value.

At $\mu = 0.40$, all tested initial conditions converge to

$$(S2.39) \quad \eta \approx 1.$$

Thus, the high-diversity branch is selected for all tested initial conditions at this parameter value.

- (b) For mutation rates inside the hysteresis region, the initial conditions can split into two distinct branches.

The clearest coexistence point is

$$(S2.40) \quad \mu = 0.15.$$

At this parameter value,

$$(S2.41) \quad M_{\text{low}} = 44, \quad M_{\text{high}} = 56,$$

so that

$$(S2.42) \quad f_{\text{low}} = 0.44, \quad f_{\text{high}} = 0.56.$$

The corresponding branch means are

$$(S2.43) \quad \langle \eta \rangle_{\text{low}} = 0.012851, \quad \sigma_{\text{low}} = 4.89 \times 10^{-6},$$

and

$$(S2.44) \quad \langle \eta \rangle_{\text{high}} = 0.999980, \quad \sigma_{\text{high}} = 1.07 \times 10^{-9}.$$

The separation between the two branch means is therefore

$$(S2.45) \quad 0.999980 - 0.012851 \approx 0.987129,$$

which is much larger than either within-branch standard deviation. This strong separation indicates that the two populations correspond to distinct terminal states rather than a single broad distribution of slowly relaxing trajectories.

The transition in the basin fractions is also visible across the tested parameter range. At $\mu = 0.10$, the low-diversity branch contains $M_{\text{low}} = 98$ of the 100 initial conditions, whereas at $\mu = 0.20$ it contains only $M_{\text{low}} = 5$. At $\mu = 0.15$, both branches are substantially populated. This behaviour is consistent with the bistable region identified from the hysteresis analysis.

We performed the same experiments for several sufficiently small values of the Mellin exponent s and for all 20 realizations of the WS, ER, and BA network ensembles considered in our analysis. Representative realizations of the WS, ER, and BA networks are shown in figures S2.13, S2.14, and S2.15, respectively. The qualitative and quantitative behaviours are similar across the tested realizations. In each tested realization, the numerical results provide evidence for the coexistence of two attracting dynamical states over the corresponding bistable parameter region.

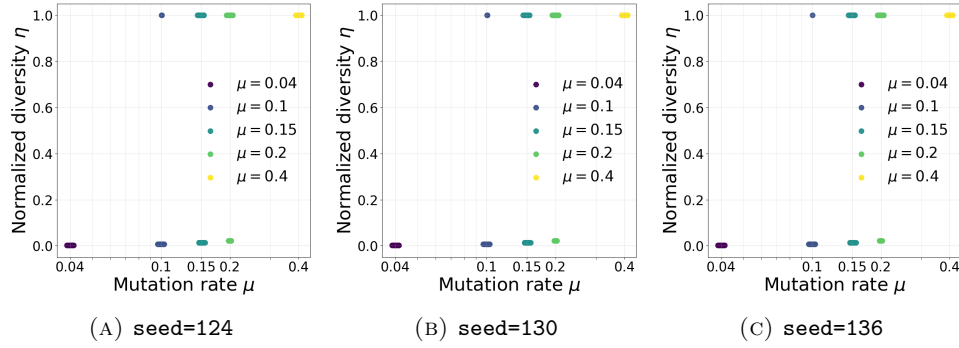


FIGURE S2.13. Bistability check for different WS network realizations with $k = 10$, $a = 0.3$, $p = 0.1$, and $s = 10^{-2}$. The subfigures show different WS network realizations: (a) seed=124, (b) seed=130, and (c) seed=136.

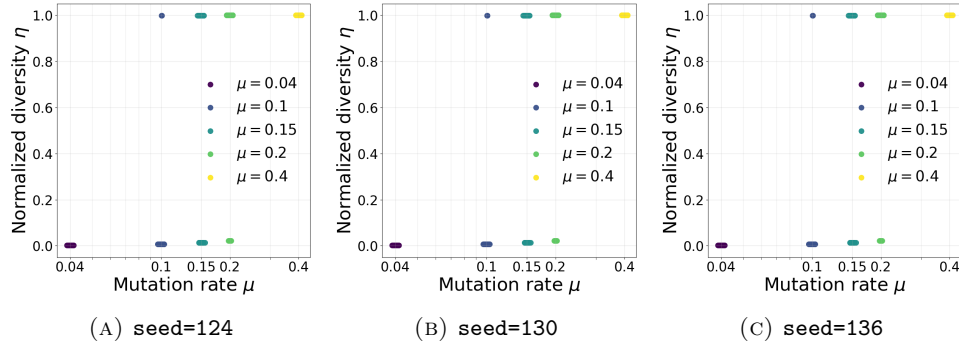


FIGURE S2.14. Bistability check for different ER network realizations with $a = 0.3$, $p_{\text{edge}} = 0.118$, and $s = 10^{-2}$. The subfigures show different ER network realizations: (a) seed=124, (b) seed=130, and (c) seed=136.

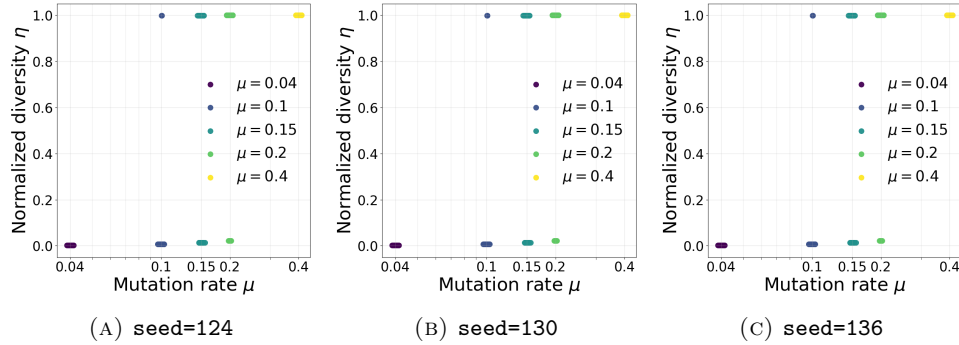


FIGURE S2.15. Bistability check for different BA network realizations with $a = 0.3$, $m = 3$, and $s = 10^{-2}$. The subfigures show different BA network realizations: (a) seed=124, (b) seed=130, and (c) seed=136.

S2.4. Multiple network realizations. The three numerical experiments, namely system-size analysis, hysteresis, and bistability, have been discussed quantitatively mainly for a WS network with a particular realization (`seed=123`). However, the same analyses were also performed for all 20 realizations considered for each network class, including Watts–Strogatz, Erdős–Rényi, and Barabási–Albert networks. Representative results for several realizations (`seed=124`, `130`, `136`) are shown in figures S2.5, S2.6, and S2.7 for the system-size analysis; in figures S2.9, S2.10, and S2.11 for the hysteresis analysis; and in figures S2.13, S2.14, and S2.15 for the bistability analysis.

Across the tested realizations, the qualitative behaviour is consistent. In particular, each realization exhibits the same three characteristic features: sharpening of the transition with increasing system size, hysteresis under forward and backward parameter continuation, and coexistence of low- and high-diversity terminal states under different initial conditions in the bistable regime. The precise location and extent of the transition, however, can vary between network realizations.

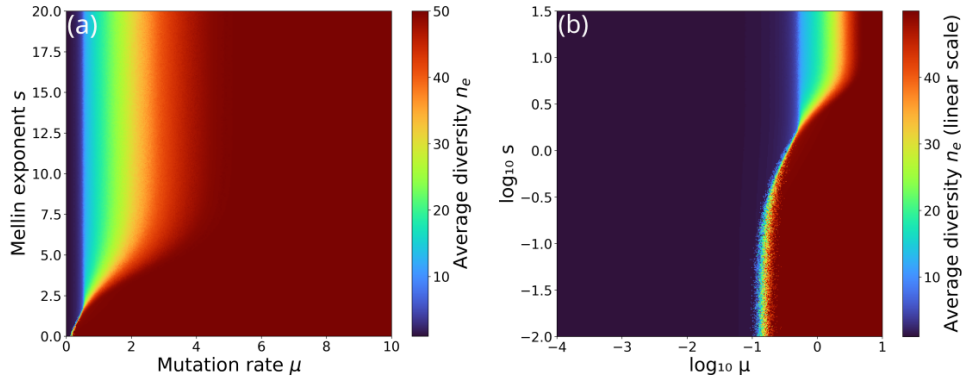


FIGURE S2.16. Contour plots for the Watts–Strogatz network obtained from a 20-realization ensemble average, showing the effective diversity n_e as a function of the mutation rate μ and the Mellin exponent s . Subfigure (a) is shown on a linear scale, whereas subfigure (b) is shown on a logarithmic scale. Both correspond to rewiring probability $p = 0.1$. The common simulation parameters are $n = 50$, $k = 10$, and $a = 0.3$.

It is important to distinguish the behaviour of individual network realizations from that of the ensemble-averaged response. When the results are averaged over all 20 network realizations, as shown in figure S2.16, the discontinuous transition observed for individual realizations is no longer visually sharp. Instead, the transition appears broadened or smeared over a finite range of μ , particularly for small values of s . A similar behaviour is observed for the other network classes, including Erdős–Rényi networks (figure S2.17) and Barabási–Albert networks (figure S2.18).

This broadening is expected because the transition point and the width of the bistable region are not identical across different network realizations. In addition, within the bistable regime, different initial conditions can select different attracting states. Consequently, averaging over network realizations and initial conditions

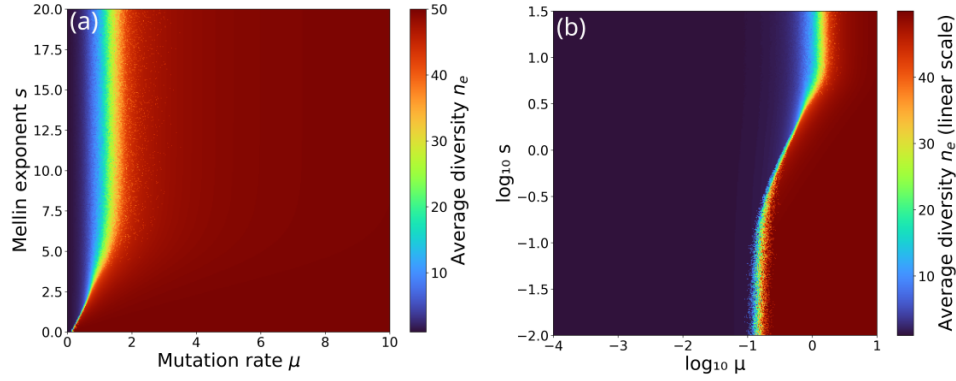


FIGURE S2.17. Contour plots for the Erdős-Rényi network obtained from a 20-realization ensemble average, showing the effective diversity n_e as a function of the mutation rate μ and the Mellin exponent s . Subfigure (a) is shown on a linear scale, whereas subfigure (b) is shown on a logarithmic scale. The common simulation parameters are $n = 50$, $a = 0.3$, and edge probability $p_{\text{edge}} = 0.118$.

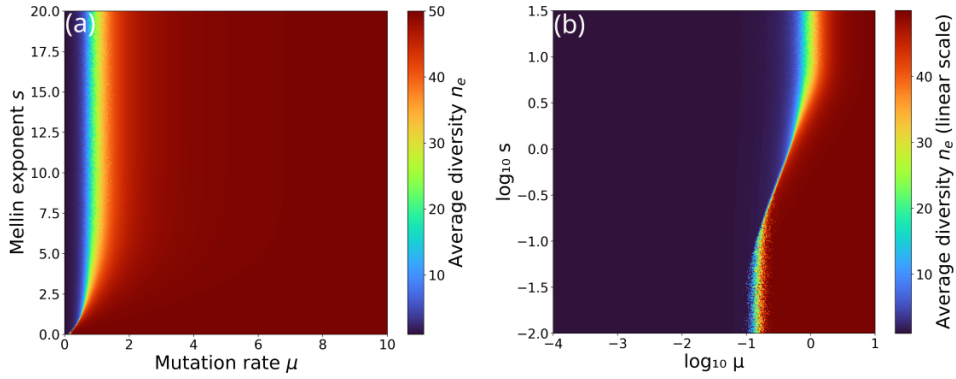


FIGURE S2.18. Contour plots for the Barabási-Albert network obtained from a 20-realization ensemble average, showing the effective diversity n_e as a function of the mutation rate μ and the Mellin exponent s . Subfigure (a) is shown on a linear scale, whereas subfigure (b) is shown on a logarithmic scale. The common simulation parameters are $n = 50$, $a = 0.3$, and $m = 3$.

combines responses with different transition locations and basin selections, thereby smoothing the resulting ensemble-averaged response.

S2.5. Conclusion. The phrase “first-order phase transition” traditionally refers to a discontinuous transition between phases in the large-system limit. For the finite networks considered here, our numerical analysis instead provides evidence for a first-order dynamical transition at the level of individual network realizations.

Specifically, strong spatial nonlocality can induce a discontinuous collapse of diversity in the generalized replicator-mutator dynamics. For each of the 20 network

realizations examined, the transition is associated with three complementary signatures: (a) progressive sharpening of the transition with increasing system size, (b) a finite hysteresis interval under forward and backward parameter continuation, and (c) bistability, manifested by the coexistence of distinct low- and high-diversity attracting states at identical parameter values. Taken together, these observations provide strong numerical evidence for a first-order dynamical transition in individual network realizations.

The ensemble-averaged response exhibits a different appearance. When the results are averaged over the 20 network realizations, the sharp transition observed for individual realizations becomes broadened over a finite range of the mutation rate, particularly at strong spatial nonlocality. This smoothing does not imply that the individual realizations undergo a continuous transition. Rather, it results from the averaging of realization-dependent responses: the transition location and the extent of the bistable region can vary between network realizations, while different initial conditions can select different coexisting attractors within the bistable regime.

Therefore, the discontinuous character of the transition is most clearly resolved at the level of individual network realizations, where system-size sharpening, hysteresis, and bistability are observed simultaneously. Ensemble averaging subsequently rounds this discontinuity because it combines responses with different transition locations and basin selections. Thus, within the finite-size numerical systems studied here, strong spatial nonlocality robustly induces a first-order dynamical transition at the level of individual network realizations, while network-to-network variability and initial-condition dependence produce a broadened crossover in the ensemble-averaged response.

S3. NUMERICAL METHODS AND SIMULATION DETAILS FOR DYNAMICAL RESPONSE DISTANCE (DRD)

Here we discuss the numerical methods and experiments focusing on the DRD and average normalized dynamical response χ_G . Initially, we follow the same computational framework discussed in section S1. However, the present DRD calculations use the normalized diversity surface

$$(S3.1) \quad r_G(\mu, s) = \frac{n_e(\mu, s)}{n},$$

where n_e is the effective number of behaviors. The rows of each stored response matrix correspond to s , and the columns correspond to μ .

For the numerical calculation of χ_G , the integrals are evaluated using a two-dimensional trapezoidal quadrature. If $w_i^{(s)}$ and $w_j^{(\mu)}$ denote the one-dimensional trapezoidal weights, the two-dimensional weight is

$$(S3.2) \quad W_{ij} = w_i^{(s)} w_j^{(\mu)},$$

and hence

$$(S3.3) \quad \chi_G = \frac{\sum_{i,j} W_{ij} r_G(s_i, \mu_j)}{\sum_{i,j} W_{ij}}.$$

Moreover, the numerical realization of the DRD between two graphs G and H can be written as

$$(S3.4) \quad D(G, H) = \left[\frac{\sum_{i,j} W_{ij} (r_G(s_i, \mu_j) - r_H(s_i, \mu_j))^2}{\sum_{i,j} W_{ij}} \right]^{1/2}.$$

S3.1. Parameter grids. For the following networks, including artificial and real-world networks, we consider the same parameter domain and the same scaling of (μ, s) . Here, we work with and collect the data in linear scale. For instance, the linear response surfaces use

$$(S3.5) \quad \mu \in [0, 100], \quad s \in [0, 20],$$

with a 200×200 grid.

Note that different realizations and initial conditions may introduce variability in the values of χ_G and $D(G, H)$. However, the overall trends of the dynamical response distance matrix are similar. Therefore, considering the computational cost and simplicity of the analysis, we work with a single realization (when applicable) and random initial conditions for all the following artificial and real-world networks.

S3.2. Artificial networks. Here we consider $n = 50$ nodes and $a = 0.3$ for all the networks, including the path graph, cycle graph, star graph, complete graph, Watts–Strogatz network (initial nearest-neighbour parameter $k = 10$ and rewiring probability $p = 0.1$), Erdős–Rényi network (edge probability $p_{\text{edge}} = 0.1$), and Barabási–Albert network (attachment parameter $m = 5$).

Note that previously $p_{\text{edge}} = 0.118$ and $m = 3$ were chosen to produce approximately comparable mean degrees in the Erdős–Rényi and Barabási–Albert networks used in the phase-diagram comparison, whereas the dynamical response distance analysis used its own common parameter specification.

We perform the simulations, and then the network labels $(G_1, G_2, \dots)$ are arranged according to the ascending values of χ_G , shown in table S3.1. After the arrangement, the dynamical response distances are calculated and shown in table S3.2. The heatmap in the main manuscript is based on these matrix values.

TABLE S3.1. Artificial networks with average response values

Graph	value (χ_G)	Label
Path graph P_{50}	0.40737976	G_1
Cycle graph C_{50}	0.56940109	G_2
Watts–Strogatz graph	0.98469906	G_3
Star graph S_{50}	0.98592810	G_4
Erdős–Rényi graph (ER)	0.98756040	G_5
Barabási–Albert graph (BA)	0.98957422	G_6
Complete graph K_{50}	0.99753769	G_7

TABLE S3.2. Dynamical response distance matrix for artificial network classes. The matrix is symmetric with zero diagonal.

	G_1	G_2	G_3	G_4	G_5	G_6	G_7
G_1	0	0.21876351	0.62761461	0.63106577	0.63137093	0.63269126	0.64057226
G_2	0.21876351	0	0.47587262	0.48037688	0.48078790	0.48251778	0.49274332
G_3	0.62761461	0.47587262	0	0.04724648	0.03405533	0.04109385	0.08585411
G_4	0.63106577	0.48037688	0.04724648	0	0.04963422	0.05703456	0.09793783
G_5	0.63137093	0.48078790	0.03405533	0.04963422	0	0.02674359	0.08356863
G_6	0.63269126	0.48251778	0.04109385	0.05703456	0.02674359	0	0.06992247
G_7	0.64057226	0.49274332	0.08585411	0.09793783	0.08356863	0.06992247	0

S3.3. Social networks. In a similar fashion, we perform the simulations for social networks. The network labels ($G_1, G_2, \dots$) are arranged according to the ascending values of χ_G , shown in table S3.3. After the arrangement, the dynamical response distances are calculated and shown in table S3.4.

TABLE S3.3. Social networks with average response values

Graph	value (χ_G)	Label
Drugs	0.36425865	G_1
Colorado Springs	0.38770681	G_2
Heterosexual	0.61700833	G_3
Twitter Russian Troll	0.75519917	G_4
Jazz	0.86240200	G_5
PolBlogs	0.95305557	G_6
Prison	0.95308206	G_7
Sawmill	0.96374978	G_8
Corporate People	0.97333248	G_9

TABLE S3.4. Dynamical response distance matrix for social network classes.

	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9
G_1	0	0.30873500	0.43687028	0.53934319	0.63836907	0.71191365	0.72261853	0.72619931	0.73281986
G_2	0.30873500	0	0.37646670	0.48325190	0.58292864	0.67634018	0.68485418	0.69100577	0.70091285
G_3	0.43687028	0.37646670	0	0.26569404	0.33277437	0.43807016	0.44524426	0.45295304	0.46668987
G_4	0.53934319	0.48325190	0.26569404	0	0.17248325	0.29313625	0.30260341	0.31096572	0.32690077
G_5	0.63836907	0.58292864	0.33277437	0.17248325	0	0.17170272	0.17622043	0.18766214	0.21191416
G_6	0.71191365	0.67634018	0.43807016	0.29313625	0.17170272	0	0.08318601	0.04426474	0.05041607
G_7	0.72261853	0.68485418	0.44524426	0.30260341	0.17622043	0.08318601	0	0.07190690	0.11071831
G_8	0.72619931	0.69100577	0.45295304	0.31096572	0.18766214	0.04426474	0.07190690	0	0.06557063
G_9	0.73281986	0.70091285	0.46668987	0.32690077	0.21191416	0.05041607	0.11071831	0.06557063	0

S3.4. Biological networks. Along the same line, we perform the simulations for biological networks. The network labels ($G_1, G_2, \dots$) are arranged according to the ascending values of χ_G , shown in table S3.5. After the arrangement, the dynamical response distances are calculated and shown in table S3.6.

TABLE S3.5. Biological networks with average response values

Graph	value (χ_G)	Label
PIN E. coli validate-edges	0.46379291	G_1
Trans E. coli	0.79549093	G_2
Transc Yeast	0.86115082	G_3
PIN B. subtilis	0.88823386	G_4
PIN E. coli validated-main	0.91176194	G_5
Trans Urchin	0.92212798	G_6
PIN A. fulgidus	0.92976567	G_7
PIN Human	0.96606599	G_8
PIN Malaria	0.96974037	G_9
KSHV	0.97498106	G_{10}

TABLE S3.6. Dynamical response distance matrix for biological network classes.

	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9	G_{10}
G_1	0	0.58155931	0.61024086	0.61319126	0.59898247	0.60760556	0.62539036	0.63915361	0.64290975	0.64288529
G_2	0.58155931	0	0.28013149	0.25474562	0.27378249	0.30134336	0.29372761	0.34114842	0.34574414	0.35172668
G_3	0.61024086	0.28013149	0	0.17787486	0.19870629	0.23370231	0.21303633	0.26824230	0.27173249	0.28015857
G_4	0.61319126	0.25474562	0.17787486	0	0.12532062	0.17276081	0.13748845	0.20908220	0.21359876	0.22216607
G_5	0.59898247	0.27378249	0.19870629	0.12532062	0	0.08994389	0.08965296	0.10928823	0.11372168	0.11798834
G_6	0.60760556	0.30134336	0.23370231	0.17276081	0.08994389	0	0.13256012	0.12042096	0.12492996	0.12082729
G_7	0.62539036	0.29372761	0.21303633	0.13748845	0.08965296	0.13256012	0	0.13605674	0.13525711	0.14786403
G_8	0.63915361	0.34114842	0.26824230	0.20908220	0.10928823	0.12042096	0.13605674	0	0.05742764	0.06746188
G_9	0.64290975	0.34574414	0.27173249	0.21359876	0.11372168	0.12492996	0.13525711	0.05742764	0	0.06779828
G_{10}	0.64288529	0.35172668	0.28015857	0.22216607	0.11798834	0.12082729	0.14786403	0.06746188	0.06779828	0

S4. VERIFICATION OF THE LARGE- μ IMPLEMENTATION

Several of the numerical experiments presented in this work consider values of the mutation parameter $\mu > 1$. In this regime,

$$(S4.1) \quad Q = I - \mu L_{\text{rw}} = (1 - \mu)I + \mu P$$

generally has negative diagonal entries and therefore cannot, in general, be interpreted as a mutation-probability matrix. More precisely, since P is row-stochastic, Q remains row-stochastic for all μ , but it is nonnegative only when

$$(S4.2) \quad 0 \leq \mu \leq \min_i (1 - p_{ii})^{-1}.$$

For larger values of μ , we interpret Q as a mass-preserving mixing operator rather than as a matrix of transition probabilities. The numerical validation below is therefore designed to verify that the continuous-time implementation remains numerically well controlled in the large- μ regime.

For each trajectory, we monitor both conservation of the total population and the non-negativity of the state variables throughout the integration. In addition, because the simulations are interpreted in terms of terminal dynamical states, we monitor the norm of the vector field at the final integration time. The three diagnostics are

$$(S4.3) \quad E_{\text{mass}} = \max_t \left| \sum_{i=1}^N x_i(t) - 1 \right|,$$

$$(S4.4) \quad x_{\min} = \min_{i,t} x_i(t),$$

and

$$(S4.5) \quad R_{\text{term}} = \|\dot{\mathbf{x}}(t_f)\|_2,$$

where t_f denotes the final integration time. The first quantity measures the maximum deviation from the conserved total population, the second measures the largest negative excursion of any state component, and the third measures the distance of the terminal state from a dynamical equilibrium.

For every point (s, μ) , the initial condition is generated according to

$$(S4.6) \quad \mathbf{x}(0) \sim \text{Dirichlet}(1, \dots, 1).$$

Consequently,

$$(S4.7) \quad x_i(0) \geq 0, \quad \sum_{i=1}^N x_i(0) = 1,$$

so that every numerical trajectory starts on the standard $(N - 1)$ -simplex.

The generalized replicator–mutator equation is integrated using the adaptive RK45 method implemented in `solve_ivp`. To provide a more stringent numerical validation than that used in the main production simulations, the relative and absolute solver tolerances are set to

$$(S4.8) \quad \text{rtol} = 10^{-10}, \quad \text{atol} = 10^{-10}.$$

The integration interval is

$$(S4.9) \quad 0 \leq t \leq 5000.$$

The longer integration time is used to reduce the possibility that a non-negligible terminal dynamical residual is caused simply by insufficient relaxation time.

The validation is performed on a 400×400 parameter grid containing 400 values of μ and 400 values of s . Thus, each network realization generates

$$(S4.10) \quad 400 \times 400 = 160000$$

numerical trajectories.

To avoid adjusting the acceptance criterion after inspecting the numerical results, fixed tolerances are specified in advance for all three diagnostics:

$$(S4.11) \quad \varepsilon_{\text{mass}} = 10^{-8}, \quad \varepsilon_{\text{pos}} = 10^{-8}, \quad \varepsilon_{\text{res}} = 10^{-8}.$$

A trajectory is considered to satisfy the complete numerical validation if

$$(S4.12) \quad E_{\text{mass}} \leq \varepsilon_{\text{mass}}, \quad x_{\min} \geq -\varepsilon_{\text{pos}}, \quad R_{\text{term}} \leq \varepsilon_{\text{res}}.$$

The diagnostics are evaluated on the raw numerical solution before any subsequent clipping or normalization used for the calculation of the diversity measure.

The validation was carried out for three representative network structures: a Barabási–Albert network, an Erdős–Rényi network, and a Watts–Strogatz network. The corresponding numerical results are summarized below.

For the Barabási–Albert network with

$$(S4.13) \quad n = 50, \quad m = 3, \quad a = 0.3,$$

the maximum mass error was

$$(S4.14) \quad E_{\text{mass}}^{\max} = 6.814682 \times 10^{-9},$$

while the global minimum state component was

$$(S4.15) \quad x_{\min}^{\text{global}} = -1.581010 \times 10^{-322}.$$

No solver failures occurred among the 160000 trajectories. All trajectories satisfied the mass-conservation and positivity criteria, while 159996 trajectories satisfied the terminal-residual criterion. Thus, 159996 trajectories, corresponding to 100.00% when rounded to two decimal places, satisfied all three criteria.

For the Erdős–Rényi network with

$$(S4.16) \quad n = 50, \quad p_{\text{edge}} = 0.118, \quad a = 0.3,$$

the maximum mass error was

$$(S4.17) \quad E_{\text{mass}}^{\text{max}} = 6.842939 \times 10^{-9},$$

and the global minimum state component was

$$(S4.18) \quad x_{\min}^{\text{global}} = -1.581010 \times 10^{-322}.$$

Again, no solver failures were observed. All 160000 trajectories satisfied the mass-conservation and positivity criteria, while 159952 trajectories satisfied the terminal-residual criterion. Hence, 159952 trajectories, corresponding to 99.97% of the full set, satisfied all three fixed criteria.

For the Watts–Strogatz network with

$$(S4.19) \quad n = 50, \quad k = 10, \quad p = 0.1, \quad a = 0.3,$$

the maximum mass error was

$$(S4.20) \quad E_{\text{mass}}^{\text{max}} = 6.386569 \times 10^{-9},$$

and the global minimum state component was

$$(S4.21) \quad x_{\min}^{\text{global}} = -1.581010 \times 10^{-322}.$$

No solver failures occurred. All 160000 trajectories satisfied both the mass-conservation and positivity criteria, while 158880 satisfied the terminal-residual criterion. Therefore, 158880 trajectories, corresponding to 99.30%, satisfied all three fixed criteria.

The results are summarized in Table S4.1.

TABLE S4.1. Numerical validation of the large- μ implementation using `rtol` = `atol` = 10^{-10} and an integration time $t_f = 5000$. The acceptance thresholds were fixed before evaluating the results at 10^{-8} for mass conservation, positivity, and terminal dynamical residual.

Network	$E_{\text{mass}}^{\text{max}}$	$x_{\min}^{\text{global}}$	$R_{\text{term}}^{\text{max}}$	Solver failures	All criteria
Barabási–Albert	6.81×10^{-9}	-1.58×10^{-322}	1.10×10^{-5}	0	100.00%
Erdős–Rényi	6.84×10^{-9}	-1.58×10^{-322}	4.79×10^{-4}	0	99.97%
Watts–Strogatz	6.39×10^{-9}	-1.58×10^{-322}	2.42×10^{-4}	0	99.30%

The validation shows that the total population is conserved to better than 7×10^{-9} in all three network classes. Moreover, the smallest state components observed over all trajectories are of order 10^{-322} below zero, which is at the level of numerical underflow and is therefore effectively zero. Thus, all 480000 tested trajectories satisfy the prescribed mass-conservation and positivity criteria.

The terminal dynamical residual exhibits some variation across the different network structures. The largest observed values are 1.10×10^{-5} for the Barabási–Albert network, 4.79×10^{-4} for the Erdős–Rényi network, and 2.42×10^{-4} for the Watts–Strogatz network. Nevertheless, the fraction of trajectories satisfying the fixed terminal-residual criterion is 100.00% for the Barabási–Albert network, 99.97% for the Erdős–Rényi network, and 99.30% for the Watts–Strogatz network. No solver failures were observed in any of the 480000 trajectories considered. The small fraction of trajectories that do not satisfy the strict terminal-residual criterion therefore does not correspond to integration failure, but rather indicates that some trajectories have not reached the prescribed residual threshold by the final integration time $t_f = 5000$.

Importantly, the acceptance criteria were fixed at

$$(S4.22) \quad \varepsilon_{\text{mass}} = \varepsilon_{\text{pos}} = \varepsilon_{\text{res}} = 10^{-8}$$

before evaluating the numerical results. No post-hoc relaxation of these criteria was introduced. The calculations were performed using the stricter solver tolerances

$$(S4.23) \quad \text{rtol} = \text{atol} = 10^{-10},$$

together with the extended integration time $t_f = 5000$. Under these conditions, the maximum mass-conservation error remains below 10^{-8} for all three network classes, while the global minimum state component is at the level of numerical underflow, indicating no meaningful negative excursion.

These calculations provide numerical evidence that the large- μ implementation is well controlled over the tested parameter ranges and network structures. In particular, the total population is preserved to within

$$(S4.24) \quad E_{\text{mass}}^{\text{max}} < 7 \times 10^{-9}$$

for all three network classes, and all 480000 trajectories satisfy the prescribed positivity criterion. Moreover, more than 99% of the trajectories satisfy all three fixed validation criteria in each network class. The remaining trajectories should not be interpreted as solver failures; rather, they retain a terminal dynamical residual larger than 10^{-8} at the prescribed final time.

Finally, the distinction between the mutation-probability interpretation and the generalized dynamical interpretation of Q is important. For sufficiently large μ , Q need not be nonnegative and therefore need not represent a matrix of transition probabilities. Nevertheless, the generalized continuous-time equation can be evaluated with Q acting as a mass-preserving mixing operator. The present numerical validation shows that the resulting trajectories remain numerically controlled in the tested large- μ regime. Therefore, the use of $\mu > 1$ in the generalized dynamics is not excluded merely because Q ceases to be a mutation-probability matrix; rather, its interpretation changes from a probability transition matrix to a mass-preserving mixing operator.