Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Decoupling Belief and Belonging: Latent Polarization as a Timescale Effect in Small Friendship Groups

Antonis Chatzipanagiotou

chatzipanagiotou.antonios@outlook.com

Hellenic American University

Article

Keywords: agent-based modeling, opinion dynamics, social fragmentation, co-evolutionary networks, belief and belonging, latent polarization, small groups, network assortativity

Posted Date: September 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10408176/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

In small face-to-face groups a puzzling pattern recurs: members converge on a shared view while the ties that bind them fray or turn openly hostile. Belief converges while belonging fractures. Classical models of opinion dynamics reproduce one side of this pattern at a time. Mean-field averaging and the Voter model drive global consensus, while Axelrod's cultural model and bounded-confidence models sustain stable polarization, but they rarely produce agreement and division together. This paper names that joint condition latent polarization: attitudinal agreement coexisting with structural antagonism.

The condition is studied with an agent-based model, implemented in NetLogo, that couples a global mean-field influence signal (the Net-Attitude) to a co-evolving weighted social network in which agreement strengthens ties (bridge-building) and disagreement erodes them (bridge-burning). Agents are heterogeneous in two traits: Openness, which sets how fast an agent moves, and Assertiveness, which sets how loudly it enters the aggregate. Attitudes occupy a bipolar dispositional-structural scale with an explicit neutral pole. Because every agent holds a weighted tie to every other, the model addresses groups small enough for mutual acquaintance, and populations of three to twelve are swept. A full-factorial analysis of 1,458,000 runs (162,000 parameter combinations under nine fixed random seeds) maps the phase structure of the model, recording the complete network state at every tick.

Three findings emerge. First, whenever agents are at all open, opinions collapse onto a Neutral Attractor irrespective of initial composition; because global coupling makes this near-inevitable, it is treated as a baseline rather than a discovery. Second, and centrally, structure lags opinion: when a group reaches agreement, between 21.91 and 44.69 per cent of its ties remain severed, and attitude assortativity is undefined because opinions no longer vary. Third, that division is a transient rather than a scar. Groups repair between 65 and 81 per cent of it, settling at a residual of 7.70 per cent identical across every positive rate of repair, so the rate governs how long a group stays divided rather than how divided it ends; where no repair exists, fracture instead deepens after agreement. Structural damage falls monotonically with Openness, making stubbornness rather than prolonged negotiation the most corrosive regime.

Gauging the health of a group by opinion agreement alone can therefore mask a fractured social fabric, since consensus does not imply cohesion. But the fracture is a window rather than a permanent wound, and its width is set by the relative speed of two processes, so when a group is observed is part of what is measured.

Introduction

Agreement is supposed to bring people together. Yet across many modern societies the opposite pairing is increasingly visible: publics that have not moved especially far apart on the substance of many issues nonetheless report rising distrust of, and unwillingness to associate with, those on the other side (Iyengar et al., 2019). Online, the pattern is starker still: a "trending" consensus can crystallize on a platform at the very moment its underlying interaction network splinters into mutually walled-off

clusters. This is the puzzle of belief without belonging: how can a population grow more agreed and less connected at the same time?

The theoretical tension behind the puzzle is old. Abelson (1964) proved that in any connected network in which individuals repeatedly average their opinions toward their neighbors, the system must collapse into a single global consensus. Empirical societies, however, exhibit durable disagreement and division. The standard way to reconcile the theorem with reality is to cut interaction off: bounded-confidence models let agents ignore anyone too different from themselves (Deffuant et al., 2002), and fixed-geometry models prevent information from flowing across cultural boundaries (Axelrod, 1997). In both, polarization becomes a matter of who cannot reach whom.

In a hyper-connected, algorithmically mediated environment that assumption of local isolation is increasingly untenable. Individuals are simultaneously embedded in local relationships and continuously exposed to global aggregates: trending topics, viral content, opinion polls, a running sense of "what everyone is saying." A global mean-field runs alongside local structure and cannot be escaped by simply ignoring a disagreeable neighbor. This reopens Abelson's problem in a new form, and it motivates separating two processes that most models entangle: how opinions change, and how relationships are maintained.

This paper therefore presents a hybrid agent-based model in which every agent responds to a global Net-Attitude signal while its pairwise ties co-evolve by homophily: agreement strengthens a tie and disagreement erodes it. Agents differ in Openness (how quickly they move toward the signal) and Assertiveness (how strongly they contribute to it). Attitudes occupy a bipolar dispositional-structural scale with a substantive neutral midpoint. By letting belief dynamics and tie dynamics run on separate but coupled tracks, the model can express latent polarization: agreement laid over a substrate of antagonism.

This paper makes three substantive contributions and one methodological one.

- It formalizes *latent polarization* as a measurable decoupling of attitudinal variance from network structure, and identifies the conditions under which a population reaches opinion consensus while its ties remain fractured.
- It shows that *intermediate* Openness, not the extremes, maximizes structural damage (a non-linear "sweet spot" in which the transient of disagreement lasts longest), which cautions against interventions that prize open-mindedness without accounting for its relational cost.
- It situates global-field averaging as a limiting idealization of algorithmic aggregation, clarifying that the resulting drift toward neutrality is a property of the coupling rather than of human psychology, and connecting it to the consensus bias recently reported in societies of language-model agents.
- Methodologically, it reports an exhaustive sensitivity analysis of 1,458,000 runs (162,000 parameter combinations under nine fixed random seeds) that maps the phase structure of the model rather

than illustrating a single trajectory, and it records the full state of the network at every tick so that the trajectory of social structure can be separated from its endpoint.

A word on scope. What follows is not an empirical account of any particular population but the computational scaffold for one. Its role is twofold: to specify a co-evolutionary architecture in which opinion updating and tie formation are driven by measurable personality traits, and, through the sensitivity analysis, to chart that architecture's full behavioral repertoire. That map is the reference frame for the empirical work to come: once Openness and Assertiveness are measured in real groups, each group can be located at a coordinate in the parameter space swept here, and its observed trajectory tested against the regime the model predicts there. The regimes reported below are therefore the model's predictions, not claims about an observed society.

The remainder of the paper is organized as follows. The next section reviews the modeling traditions the model draws on, identifying for each what the model borrows and what it changes. The conceptual-framework section then sets out the attitude scale, the two traits, and the tie dynamics, keeping only the constructs that map onto model parameters. The model is then specified in full, with equations, a parameter table, and pseudo-code. The Results and Discussion report and discuss the findings.

Related Work

The model borrows ingredients from several traditions in computational social science. They are reviewed not for completeness but to locate each ingredient the model adopts or revises, organizing by the locus of complexity each tradition emphasizes: social geometry, cultural homophily, the statistical mechanics of consensus, and cognitive architecture. Each subsection closes by stating what the present model takes and what it changes.

Interacting-Particle Systems and the Mean Field

The earliest sociophysics models adapted tools from magnetism and particle systems, representing individuals as spins or discrete states that update by local alignment, majority rule, or noisy imitation (Castellano et al., 2009). These models delivered sharp results about phase transitions and the conditions for consensus, but by design they carried little psychological content: an agent had a state and a rule, and nothing resembling motivation or a relationship.

What the model takes: the mean field as a tractable form of global coupling, a single aggregate signal to which every agent responds. **What the model changes:** it attaches trait heterogeneity (Openness, Assertiveness) to the agents and lets the ties among them evolve, so that the same global pull produces very different structural outcomes.

Dynamic Social Impact Theory

Latané's Dynamic Social Impact Theory (DSIT) treats influence as a spatial force field whose magnitude is a multiplicative function of the strength, immediacy, and number of surrounding sources (Latané, 1981, 1996). In simulation, agents on a grid update toward the weighted state of their neighbors, and even simple local rules generate robust emergent regularities: clustering, consolidation, and a continuing diversity in which minority clusters shelter their own members (Latané & L'Herrou, 1996; Harton & Bullock, 2007).

What the model takes: the "strength" term, reinterpreted here as Assertiveness, an agent's expressive weight in the aggregate. **What the model changes:** it replaces spatial immediacy with a global field and represents relationships as explicit, continuously valued ties rather than fixed grid adjacency, so that closeness is a state variable that the dynamics can rewrite.

Axelrod's Dissemination of Culture

Axelrod's (1997) model asks why diversity survives if interaction makes people more similar. Culture is a vector of features; neighbors interact with probability proportional to their similarity and, when they do, become more similar still. The system freezes into regions that are either identical or share nothing, so that local convergence produces global polarization behind "cultural walls." Subsequent work showed these walls to be fragile: even infinitesimal noise reopens interaction channels and pushes the system toward monoculture (Castellano et al., 2009; Klemm et al., 2003).

What the model takes: the homophily principle that drives tie change: similarity builds relationships, dissimilarity dissolves them (the model's bridge-building and bridge-burning). **What the model changes:** ties here are continuous weights that can record antagonism, not merely gate interaction on or off; and a global field continues to average opinions even across ties that have been severed, which is precisely what lets belief and structure part ways.

Voter and Sznajd Models

The Voter model is the paradigmatic null model of opinion dynamics: agents with no self-confidence simply imitate a neighbor, and finite systems reach consensus on a timescale that grows with size (Holley & Liggett, 1975; Sood & Redner, 2005). The Sznajd model adds outflow dynamics ("united we stand") in which an agreeing pair convinces its neighbors, producing a sharp transition in which an opinion above a critical initial density sweeps the population (Sznajd-Weron & Sznajd, 2000).

What the model takes: consensus as a reference outcome, and initial composition as a factor worth testing explicitly. **What the model changes and finds:** under global coupling the initial composition becomes irrelevant to the final opinion state, a direct contrast with the critical-mass behavior of Sznajd-type models, and a useful diagnostic that the mean field, not the starting mix, governs the outcome.

Cognitive Architectures: Dissonance and Motivated Reasoning

A second lineage moves the complexity from the network into the agent. Constraint-satisfaction networks model cognitions as nodes linked by consonant and dissonant weights and treat attitude change as energy minimization, reproducing counter-intuitive dissonance effects (Shultz & Lepper, 1996). Motivated-reasoning models such as John Q. Public attach affective tags to concepts and update non-neutrally, yielding the sophistication paradox in which more knowledgeable agents polarize more, because they can better counter-argue what they dislike (Taber & Lodge, 2006; Kim et al., 2010).

What the model takes: the principle that resistance to influence is active and heterogeneous rather than mere inertia, captured here by Openness as a per-agent plasticity and by a local bias term that can amplify or damp it. **What the model changes:** it deliberately keeps the agent light (a scalar plasticity plus a bias term rather than a full cognitive network), trading representational fidelity for a model whose entire phase space can be swept.

Generative Agents and Language-Based Simulation

The newest frontier replaces update rules with language. Generative agents endowed with memory, reflection, and planning produce believable coordinated behavior (Park et al., 2023). Applied to opinion dynamics, however, such agents display a consensus bias: fine-tuned to be agreeable, societies of language-model agents converge toward neutral or fact-based agreement far faster than human groups, and must be explicitly prompted into stubbornness to polarize at all (Chuang et al., 2024).

What the model connects: the Net-Attitude plays a role structurally analogous to the normative anchor that alignment training imposes on language models: both pull agents away from the extremes. That two very different modeling traditions converge on the same drift-to-neutral suggests the effect is a property of centralized averaging, not of any particular representation of an agent. On the contested question of *negative* influence (whether distant or disliked others actively push an agent away), the literature is far from settled, and reviews of social-influence models treat repulsive mechanisms with caution (Flache et al., 2017); this point is revisited below, where the distant-tie term is justified.

Conceptual Framework

This section defines the constructs the model represents. Only constructs that map onto a model parameter or output are included; supporting psychological detail is kept to the minimum needed to justify each choice.

The Dispositional-Structural Attitude Scale

Attitudes are commonly understood in social psychology as evaluative orientations toward objects, people, or ideas, combining cognitive, affective, and behavioral components (Petty et al., 1997). Here an attitude is represented as a position on a single bipolar content dimension, not as a valence (for or against an object) but as an orientation. One pole, labelled dispositional, foregrounds individual-level, short-term, need-based concerns; the other, labelled structural, foregrounds system-level rules, institutions, and long-term stability. An agent's attitude is simply its position along this one axis, exactly as a left-right ideology score encodes position by magnitude and side.

Two clarifications follow, and they answer a natural objection: that the numeric value of the attitude should not by itself fix its orientation. First, because the axis is bipolar and its poles carry the dispositional and structural meanings, a value necessarily implies a pole: this is a coordinate convention on one dimension, not a claim that low values are dispositional as a type that could vary independently of content. Orientation and position are two descriptions of the same quantity. Second, the midpoint (a Neutral attitude, value 4) is not the mere absence of opinion. It is a substantive, cross-pressured stance that exerts no directional pull on others and responds weakly to polarized cues, and it plays a central dynamical role: as an attractor for agents under cross-pressure and, as later sections show, as a universal bridge in the network.

A note on terminology. Because "structural" is easily overloaded, it is reserved throughout for this attitudinal pole. Where the graph of ties is meant, the terms *network structure* or *topology* are used; where the topology-focused sociophysics models discussed above are meant, the terms *structure-based* or *topology-focused* models are used. The term *orientation* denotes the dispositional-structural axis, and the word *disposition* is avoided in its generic "evaluative disposition" sense to prevent collision with the dispositional pole.

Openness: Plasticity

Openness enters the model as susceptibility to influence: the rate at which an agent moves toward the prevailing signal. It is treated as a plasticity coefficient rather than a personality inventory: at one extreme an agent is effectively frozen, at the other it adopts the signal almost immediately. Framing openness-to-influence as a functional, adaptive property rather than a simple weakness is consistent with work showing that receptivity to peers tracks the quality of an individual's social relationships (Allen et al., 2022). In the model, this trait is the single scalar that separates a rigid population from a fluid one, and, as the results show, its intermediate values, not its extremes, generate the most interesting dynamics.

Assertiveness: Expressive Weight

Assertiveness enters as expressive weight, or social loudness: how much an agent's position contributes to the aggregate signal that others perceive. This corresponds directly to the "strength" term in social-

impact theory and to influence weight in network-diffusion models (Latané, 1996). Separating what an agent believes from how loudly that belief enters the shared field lets the model represent vocal minorities (high assertiveness, extreme attitude) and silent majorities (low assertiveness, moderate attitude), a distinction with obvious purchase in online settings, where amplification magnifies a subset of voices out of proportion to their number.

Ties, Closeness, and Homophily

Relationships are represented as weighted links whose value stands for closeness, following Granovetter's (1973) treatment of tie strength as graded rather than binary: strong ties are high-weight, weak ties low-weight, and intermediate weights represent ambivalent or acquaintance relationships. The weights are not fixed. They evolve by homophily, the well-documented tendency for similar individuals to associate (McPherson et al., 2001), implemented as two opposing rules: agreement strengthens a tie (bridge-building) and disagreement weakens it (bridge-burning). This couples the relational fabric to the flow of opinion, so the network can reorganize as attitudes move, and it is the mechanism through which structure can diverge from belief.

Manifest and Latent Polarization

It is useful to distinguish two forms of polarization. Polarization is manifest when attitudes themselves split into distinct, divergent clusters. It is latent when attitudes converge (the population agrees) yet the pattern of ties and trust encodes antagonism, so that agreement sits atop a fractured or hostile relational structure. Traditional measures that survey attitudes alone detect only the manifest form. The central aim of this model is to make the latent form visible: to show when and why a society can reach consensus in belief while remaining divided in structure.

The Model

The model is described following the logic of the ODD protocol (purpose, then entities and state variables, then the process schedule, then each sub-process in turn), so that the specification is complete enough to reimplement. The model is implemented in NetLogo 7.0.0, with the parameter sweep run under BehaviorSpace.

Purpose

The model exists to examine whether, and under what conditions, a population subject to a global influence signal reaches opinion consensus while its social network fractures (the conditions for latent polarization), and how the two agent traits, Openness and Assertiveness, shape both the opinion outcome and the structural one.

Entities, State Variables, and Scales

The model contains agents (nodes) embedded in a complete weighted graph whose link weights evolve over time. Each agent carries a current opinion and two stable traits. Table 1 lists the state variables and parameters; *N* denotes the number of agents.

Table 1

State Variables and Parameters of the Model

Meaning	Type / range	Role in the model
Attitude of agent i	Discrete, {1,…7}	1-3 dispositional, 4 neutral, 5-7 structural
Polarity (continuous image of the attitude)	Continuous, [-1, 1]	Linear map: 1 to -1, 4 to 0, 7 to +1
Openness	Static, [0, 10]	Plasticity / step size (0 frozen, 10 hyper-liquid)
Assertiveness	Static, [0, 10]	Weight of agent i in the global signal
Tie weight between i and j	[0, 100]	close at least 56, distant at most 44, neutral in between
Net-Attitude and Net Signal	Derived each tick	Global mean-field every agent perceives
same-growth	Swept, {0, 2.5, 5, 7.5, 10}	Bridge-building increment
opposite-growth	Swept, {0, 2.5, 5, 7.5, 10}	Bridge-burning decrement
Initial composition s	Swept, {0, 3, 6}	Equal mixture, Neutral rising, Structural hegemony
Population size	Swept, {3, 6, 9, 12}	Number of agents, on a complete graph
Tick ceiling	Fixed, 200	Safety cap; a run halts earlier on structural stability

opposite-growth | Swept, [0, 10] | Bridge-burning decrement |
Initial composition | Swept, [0, 6], step 0.25 | Path through Dispositional (s = 2), Neutral (s = 4), Structural (s = 6) hegemony |
Population size | Swept, {3, 6, 9, 12} ‡ | Number of agents |

† *Both bounds are enforced by explicit clamping in the update rules (Equations 5-6); weights are real-valued and are never rounded. Initial weights* ***Wij*** *are drawn independently and uniformly from [weight-mean − weight-sd, weight-mean + weight-sd], clamped to [0, 100], with weight-mean ∈{0, 25, 50, 75, 100} and weight-sd ∈{0, 25, 50}.* ‡ *N is swept over {3, 6, 9, 12} on a complete graph.* ***Oi*** *and* ***Si*** *are drawn*

independently and uniformly from [mean − sd, mean + sd], clamped to [0, 10], where sd is the half-width of the uniform jitter around the mean (not a Gaussian standard deviation); sd = 0 yields a homogeneous population fixed at the mean.

Scale and Scope: The Model Is a Model of Friendship Groups

The population sizes swept here, N in {3, 6, 9, 12}, are small by the standards of opinion dynamics, and the choice is deliberate rather than a concession to computation. The model places agents on a complete graph, in which every agent holds a weighted tie to every other agent. That assumption is defensible only at the scale of a face-to-face group in which everyone genuinely does know everyone: a household, a project team, a circle of friends. At populations of a hundred or a thousand, a complete graph is not a simplification of a social network, it is a different object, and results obtained on it would not transfer.

The scope of the model follows from the topology, and the two are stated together here so that neither is read as an oversight. This is a model of small groups bound by mutual acquaintance, and its claims are claims about such groups. Mesoscale community structure, the level of organization at which sparse large networks divide into identifiable and internally dense subgroups, is not a property a complete graph possesses at any size, and this work makes no claim about it.

The attitudes at issue are likewise not political attitudes. The dispositional-structural axis introduced above is a general orientation toward how problems are framed, whether in terms of immediate individual need or of system-level rule and institution. Such orientations surface in ordinary group disagreement about anything from how to divide a shared expense to how to handle a friend who has behaved badly. Political attitudes carry an additional apparatus, in particular identity sorting and group-based affect, which the model does not represent, and to which its mechanisms should not be assumed to extend. Where the results below are set beside the literature on affective polarization, the comparison is offered as an analogy between mechanisms at different scales, not as an identity claim.

Process Overview and Scheduling

Each tick proceeds in a fixed order. The global signal is computed first from the attitudes held at the start of the tick. Every agent then updates its opinion in response to that signal and to its local neighborhood, reading tie weights as they stood at the start of the tick. Every tie then updates, reading the attitudes as they stood at the start of the tick rather than the values just committed. Because both update phases draw on a single snapshot taken before either phase begins, no agent and no tie can respond to a change made within the same tick, which is what gives the scheme its quasi-synchronous character. The stopping condition is evaluated last, after the tick counter has advanced, so that the final state of every run is recorded rather than discarded.

A run terminates when the network has stopped changing, defined as three successive ticks in which no tie weight moves, or when a ceiling of 200 ticks is reached. This is a departure from a stopping rule based on opinion. Because the neutral clause in Equation (5) reinforces every tie once all agents have become Neutral, a run halted shortly after opinions settle is halted in the middle of a structural process that has not finished. Stopping on structural stability instead allows the network to reach its own equilibrium, and the tick at which opinions first settled is recorded separately so that the state of the network at consensus and at structural equilibrium can both be examined. The distinction turns out to matter a great deal, and it is the subject of the section on latent polarization below. Figure 1 shows the full cycle.

Figure 1.

Note. Setup initializes agents, traits, attitudes, and tie weights, either at random or by the composition parameter s. The Go loop computes the global Net-Attitude and Net Signal, takes a snapshot of the start-of-tick state, runs the update-attitudes sub-routine for every agent, updates the network by bridge-building and bridge-burning, records the tick of opinion consensus, advances the clock, and then halts on three successive ticks without a change in tie weight. The right panel lists the global state and entities. ## The Global Mean-Field: Net-Attitude The Net-Attitude is a single global variable standing for the aggregate public opinion (the “zeitgeist”) that every agent perceives, regardless of its local ties. It is the assertiveness-weighted sum of individual polarities:

$$\Phi t = \Sigma i=1N \ *P*i,t \cdot *S*i \quad (1)$$

This sum is normalized to a Net Signal bounded on the interval from −1 to 1 by dividing by the theoretical maximum, in which every agent holds an extreme polarity at maximum assertiveness (a value of 10):

$$\Phi t = \Phi t / (10 \cdot N) \quad (2)$$

Normalizing by the theoretical maximum rather than by the realized sum of weights is a deliberate choice: it means a more assertive population generates a stronger signal for the same distribution of opinions, which is the lever behind the effect of Assertiveness reported in the results. Unlike bounded-confidence models, in which an agent perceives only neighbors within an opinion radius, here the signal is genuinely global.

Attitude Update

Attitude change occurs in two phases. In the first, each agent computes a desired change to its continuous polarity; in the second, that continuous value is committed back to the discrete 1-7 scale.

The agent first forms a *Social Bias*, *B*i, by surveying its close and distant neighbors. The bias records the net local support *for the global signal*, not agreement with the agent itself, and is the mean over the agent’s neighbors of the contributions in Table 2.

Table 2. *Neighbor Contributions to the Social Bias, by Tie Type and Stance Toward the Net Signal*

Neighbor type	Neighbor's stance toward the Net Signal	Contribution to bias	Interpretation
Close tie (weight at least 56)	agrees with the signal	+1	Validation from a trusted ally reinforces the move (an Asch-type ally effect).
Close tie (weight at least 56)	disagrees with the signal	-1	Dissent from a trusted ally creates local resistance: echo-chamber insulation.
Distant tie (weight at most 44)	agrees with the signal	-1	Reactance: alignment with an out-group cue provokes differentiation.
Distant tie (weight at most 44)	disagrees with the signal	+1	An opponent's rejection of the signal reassures the agent.
Ambivalent tie (weight 45 to 55)	either	excluded	Ties in the neutral band carry no directional information and are omitted from the average.
Any tie to a Neutral agent	treated as agreeing	as above	A Neutral neighbor exerts no counter-pull and is coded as aligned with the signal.

The bias then modulates the agent's innate Openness to give an effective step size, α_i:

$$\alpha_i = ({}^*O^*_i / 10)(1 + 0.5 \cdot {}^*B^*_i) \quad (3)$$

Strong local support can thus amplify an agent's openness by up to fifty per cent, and strong opposition can halve it, but the step size affects only the *speed* of movement, never its *direction*: every agent moves toward the Net Signal. The new continuous polarity is

$${}^*P^{*\prime}_{i,t+1} = {}^*P^*_{i,t} + \alpha_i (\Phi_t - {}^*P^*_{i,t}) \quad (4)$$

In the second phase the updated polarity is mapped back to the nearest integer on the 1-7 scale. This discretization introduces a hysteresis: small changes in internal polarity do not immediately register as a change of attitude category, which damps oscillation and stabilizes the system. The neutral category (4) is what makes both the opinion attractor and the bridge rule below operate, so the discretization is load-bearing rather than cosmetic.

Network Co-Evolution

In parallel with the opinion updates, ties evolve by two opposing rules that together implement homophily and a structural-balance flavor. When two agents fall on the same side of the scale (both

dispositional or both structural), or when at least one of them is Neutral, their tie is reinforced:

*W*ij,t+1 = min(100, *W*ij,t + *g*+) (5)

The neutral clause makes Neutral agents universal bridges, able to hold strong ties with partisans of either side. When two agents fall on opposite sides and neither is Neutral, their tie decays:

*W*ij,t+1 = max(0, *W*ij,t − *g*−) (6)

Weights are real-valued and are never rounded. The only adjustment either rule makes is the hard clamp to [0, 100] written explicitly in Equations (5) and (6). Initial weights *W*ij,0 are drawn independently and uniformly from weight-mean plus or minus weight-sd and clamped to [0, 100], so the sweep starts networks anywhere between uniformly severed (weight-mean = 0) and uniformly saturated (weight-mean = 100).

One consequence of the neutral clause deserves stating in advance, because it governs the structural results. Once every agent has drifted to Neutral, every pair satisfies the reinforcement condition, and so every surviving tie grows by same-growth on every subsequent tick. Network fracture is therefore not a fixed point of the model whenever same-growth is positive. It is a state the network passes through, and how much of it remains at the end depends on when the observation stops. The sweep is designed to measure this directly rather than to inherit an answer from the stopping rule.

Pseudo-Code

The full tick, in compact form:

```
SETUP:
create N agents; place them on a complete graph
for each agent i:
draw O_i, S_i ~ U(mean - sd, mean + sd), clamp to [0, 10]
set A_i (random, or by composition s)
for each tie (i,j):
draw W_ij ~ U(weight-mean - weight-sd, weight-mean + weight-sd),
clamp to [0, 100]

GO (repeat each tick):

# 0 - snapshot the start-of-tick state
A_snapshot <- A # used by phase 3
W_snapshot <- W # used by phase 2

# 1 - global signal
```

```
Phi <- sum over i of P_i * S_i
Phi_hat <- Phi / (10 * N) # Net Signal, in [-1, 1]

# 2 - update attitudes (reads W_snapshot)
for each agent i:
B_i <- mean over neighbors of (+/- 1) # see Table 2
alpha_i <- min(1, (O_i / 10) * (1 + 0.5 * B_i))
P_i <- P_i + alpha_i * (Phi_hat - P_i)
A_i <- round(3 * P_i + 4), clamp to [1, 7]

# 3 - update network (reads A_snapshot)
for each tie (i,j):
if same side, or either agent is Neutral:
W_ij <- min(100, W_ij + same_growth) # bridge-building
else if opposite sides (and neither Neutral):
W_ij <- max(0, W_ij - opposite_growth) # bridge-burning

# 4 - record the tick at which opinions first settled
if attitudes unchanged for 3 successive ticks and consensus_tick unset:
consensus_tick <- current tick

# 5 - advance the clock, then test
tick <- tick + 1
if tie weights unchanged for 3 successive ticks, or tick >= 200: STOP
```

Two details are worth making explicit. The step size alpha is capped at 1, since a value above 1 would carry an agent past the signal it is moving toward rather than to it. And the tick counter advances before the stopping test, so that the state recorded for a terminating run is the state the run actually ended in.

Simulation Design and Sensitivity Analysis

To characterize the model rather than illustrate it, a full-factorial sweep was run over ten parameters. Population size N took the values {3, 6, 9, 12}, for the reasons set out above. The composition parameter *s* took the values {0, 3, 6}, spanning the path from the Equal mixture through Neutral Hegemony to Structural Hegemony. Traits were heterogeneous: openness-mean and assertiveness-mean each took the values {0, 5, 10}, crossed with a spread (sd) of {0, 5}, a uniform jitter of that half-width around the mean, clamped to [0, 10], with sd = 0 collapsing to a homogeneous population. Initial tie weights were likewise drawn uniformly, with weight-mean in {0, 25, 50, 75, 100} and weight-sd in {0, 25, 50}, clamped to [0, 100]. The two tie increments were swept at full resolution as same-growth, opposite-growth in {0, 2.5,

5, 7.5, 10}, since these are the parameters that drive network structure and the resolution of the structural results depends on them.

This grid yields 4 x 3 x 3 x 2 x 3 x 2 x 5 x 3 x 5 x 5 = 162,000 unique parameter combinations. Each combination was run under nine fixed random seeds, giving 1,458,000 independent runs in total. Fixing the seed list, rather than drawing fresh seeds per run, makes every reported run exactly reproducible. Ten seeds were planned; the output for one of them was lost to a data handling error during archiving and could not be recovered, and the analyses reported here therefore rest on the nine intact seeds.

Nineteen measures were recorded at every tick of every run. Thirteen are run-level outcomes carried over from the original design: the tick count, the mean and variance of attitudes, the polarization share (agents at levels 1 to 2 or 6 to 7), the Net-Attitude, the counts of dispositional, neutral, and structural agents, the mean tie weight, the percentages of close (at least 56) and distant (at most 44) ties, the population size, and the seed used. Two record the timing of opinion convergence: the tick at which attitudes first settled, and whether they had settled at all. The remaining four record the state of the system itself rather than a summary of it: the full vector of tie weights, and the per-agent vectors of attitude, Openness, and Assertiveness. Recording the network state rather than aggregate statistics of it is what allows the community-aware measures reported below to be computed, and recording it at every tick rather than only at termination is what allows the trajectory of network structure to be separated from its endpoint.

Runs lasted a mean of 18.38 ticks and a maximum of 86, and 92.9 per cent of runs reached opinion consensus before the network settled. The scale of the sweep allows transient artifacts to be separated from robust emergent structure, in particular the conditions for consensus and for network fracture.

The NetLogo model, analysis code and flowchart **[find them in the supplementary material]**

Raw Data [link Removed Due to Peer Review]

Results

The results below are derived from the full-factorial sensitivity analysis specified above: 1,458,000 independent runs spanning 162,000 parameter combinations, each evaluated under the same nine fixed random seeds. The regimes reported here are properties of the model, mapped across its parameter space; whether real groups occupy them is the question the forthcoming empirical study is designed to answer. The scale of the sweep allows the attractors and phase transitions of the model to be identified robustly, and transient artifacts to be separated from stable emergent structure.

Every measure reported below is read from the state of a run at the tick it ended, or, where stated, at the tick its opinions first settled. Because the network continues to evolve after opinions converge, the

distinction between these two moments is load-bearing rather than procedural, and it is treated explicitly in the section on latent polarization.

The Neutral Attractor: Convergence of Opinion Dynamics

Openness governs whether the group converges at all, and the effect is stark. Averaged over the whole sweep, terminal attitude variance falls from 3.0268 at openness-mean = 0, through 0.4105 at openness-mean = 5, to 0.0117 at openness-mean = 10. The mean terminal attitude sits at 3.96, 3.99, and 3.99 respectively, in every case within a twentieth of a scale point of exact Neutrality. Groups with any capacity to absorb the global signal do not settle on a compromise between their starting positions; they settle on Neutral.

The population composition tells the same story from the other direction. The mean number of Neutral agents at termination rises from 2.78 to 6.20 to 7.30 as openness-mean increases, against a mean population of 7.5, while the polarization share, the proportion of agents holding positions at the extremes of the scale, collapses from 35.63 per cent to 5.23 per cent to 0.02 per cent. At openness-mean = 10 the extremes are essentially emptied. This is not the survival of a moderate majority over a radical minority. It is the active abandonment of extreme positions by the agents who held them, since the dispositional count falls in step, from a mean of 2.38 to 0.60 to 0.11.

The initial composition of the group is irrelevant to any of this. Across the three values of the composition parameter *s*, each covering 486,000 runs, the terminal mean attitude is 3.9797, the terminal variance is 1.1496, and the polarization share is 13.6264 per cent. The three values agree to four decimal places. Whether a group begins evenly mixed, entirely Neutral, or entirely Structural makes no measurable difference to where it ends up. Under global coupling the mean field, and not the starting mix, governs the outcome, in direct contrast to the critical-mass behavior of Sznajd-type models.

Assertiveness, which scales each agent's contribution to the global signal, proves similarly inert with respect to the final state. Terminal variance is 1.1479, 1.1549, and 1.1461 at assertiveness-mean of 0, 5, and 10, and the share of runs reaching consensus is 92.95, 92.85, and 92.83 per cent. The reason is visible in Equation (2): the Net-Attitude is normalized by the maximum possible assertiveness-weighted sum, so scaling every agent's expressive weight by a common factor leaves the normalized signal, and therefore the attractor, where it was. Assertiveness redistributes influence when it varies across agents, but its overall level does not move the destination.

Table 3. *Terminal Opinion State by Openness*

openness-mean	Mean attitude	Attitude variance	Neutral agents	Dispositional agents	Polarization share (%)
0	3.9607	3.0268	2.78	2.38	35.63
5	3.9899	0.4105	6.20	0.60	5.23
10	3.9885	0.0117	7.30	0.11	0.02

Note. Values are means over the whole sweep; mean population size is 7.5 agents.

Figure 2.

Note. The uniform color across the horizontal axis shows that initial group composition has no effect on the outcome. All groups with non-zero Openness converge to a low-variance consensus.

Table 4. *Irrelevance of Initial Composition*

s	Mean attitude	Attitude variance	Polarization share (%)	Runs
0 (Equal mixture)	3.9797	1.1496	13.6264	486,000
3 (Neutral rising)	3.9797	1.1496	13.6264	486,000
6 (Structural hegemony)	3.9797	1.1496	13.6264	486,000

Note. Each value of s covers 486,000 runs. The three rows agree to four decimal places.

Figure 3.

Note. As Openness rises, agents holding extreme positions abandon them rather than being outnumbered. ## Network Topology: The Structural Tug-of-War The two tie rules pull against each other, and the sweep separates their contributions cleanly. Increasing opposite-growth erodes cohesion: the terminal percentage of close ties falls from 84.95 at opposite-growth = 0 to 81.11, 81.04, 80.95, and 80.86 across the remaining levels, while distant ties rise from 11.61 to 16.97 per cent. Almost the whole of that effect is delivered by the move from zero to 2.5. What matters structurally is whether antagonism erodes ties at all, far more than how fast it does so.

The same pattern, in a sharper form, governs bridge-building. Terminal cohesion at same-growth = 0 stands at 40.58 per cent close ties. At every positive value of same-growth it stands at 92.08 per cent, and the four values agree to three decimal places: 92.0794, 92.0802, 92.0821, and 92.0815. What the rate of repair changes is not the destination but the time taken to reach it, which falls from 33.98 ticks at same-growth = 2.5 to 20.19, 16.02, and 13.72 ticks as repair accelerates. A group that repairs its ties slowly arrives at the same network as a group that repairs them quickly; it simply takes longer.

This result also resolves a question about the initial conditions. Because tie weights are drawn from weight-mean, a network observed shortly after initialization will largely reflect that draw rather than

anything the dynamics did. Measured at structural equilibrium, the dependence disappears wherever repair is available. With same-growth = 0, terminal fracture varies from 97.23 per cent to 7.07 per cent as weight-mean rises from 0 to 100, a spread of 90 percentage points that is essentially the initial draw persisting untouched. With same-growth positive, the same sweep yields 8.82, 8.61, 7.48, 6.90, and 6.71 per cent, a spread of 2 percentage points. Given any capacity to rebuild ties, the network forgets where it started.

Table 5. *Terminal Network State by Tie Increment*

same-growth	Close ties (%)	Distant ties (%)	Onnela clustering	Modularity	Ticks to equilibrium
0	40.58	47.74	0.4847	0.1045	7.97
2.5	92.0794	7.7057	0.8136	0.0181	33.98
5	92.0802	7.7046	0.8137	0.0181	20.19
7.5	92.0821	7.7031	0.8137	0.0181	16.02
10	92.0815	7.7039	0.8137	0.0181	13.72

opposite-growth	Close ties (%)	Distant ties (%)	Onnela clustering	Modularity	Ticks to equilibrium
0	84.95	11.61	0.8626	0.0192	15.11
2.5	81.11	16.49	0.7224	0.0312	22.36
5	81.04	16.61	0.7205	0.0316	18.71
7.5	80.95	16.87	0.7180	0.0322	17.99
10	80.86	16.97	0.7160	0.0325	17.72

Note. The upper panel varies bridge-building, the lower bridge-burning. Note that every positive value of same-growth yields the same endpoint and differs only in the time taken to reach it.

Figure 4.

Note. Positive reinforcement builds cohesion; negative reinforcement erodes it. **Table 6**

Dependence of Terminal Fracture on the Initial Tie-Weight Draw, Split by Whether Repair Is Available

weight-mean	Distant ties (%), same-growth = 0	Distant ties (%), same-growth > 0
0	97.23	8.82
25	86.38	8.61
50	35.42	7.48
75	12.57	6.90
100	7.07	6.71
Spread	**90.16**	**2.11**

Note. Given any capacity to rebuild ties, the network forgets where it started.

Figure 5.

Community Structure of the Close-Tie Network

Three families of community-aware measure were computed on the recorded network states: weighted clustering, modularity, and attribute assortativity. Before reporting them, a property of the topology has to be stated, because it determines which forms of these measures carry information at all.

On a complete graph, the standard unweighted clustering coefficient is identically 1 for every run, since every triple of agents closes a triangle by construction. The Barrat weighted generalization is likewise constant, because the edge weights in its numerator cancel against node strength in its denominator. Modularity on the complete weighted graph is maximized at Q = 0 by the trivial single-community partition, since all nodes are structurally equivalent. And unweighted attribute assortativity reduces to minus one over (N minus one), a constant determined by population size alone and independent of the attribute values supplied to it. Four otherwise standard measures are therefore degenerate here, not because the model lacks structure but because these measures read structure off the presence and absence of edges, and in a complete graph that information is constant.

Two routes recover the information. The first is the Onnela weighted clustering coefficient, which is informative on a complete weighted graph because it takes the geometric mean of all three edge weights in each triangle rather than normalizing by node strength. The second is to threshold the graph at the close and distant cut points already defined in this model, at weights of at least 56 and at most 44, and to compute the conventional measures on the resulting close-tie graph. Both routes are used below, and they agree.

Onnela clustering tracks cohesion closely. It stands at 0.4847 when same-growth is zero and at 0.8137 at every positive value, mirroring the close-tie percentages exactly, and it falls from 0.8626 to 0.7160 as opposite-growth rises from 0 to 10. Strong ties in this model do not scatter across the group; they consolidate into coherent triads, which is what the friendship-group framing would predict.

Modularity of the close-tie graph behaves as the mirror image. It stands at 0.1045 when repair is unavailable and at 0.0181 whenever repair is present, and it rises from 0.0192 to 0.0325 as antagonism strengthens. These are low values in absolute terms, as they should be for groups of at most twelve mutually acquainted agents, but their direction is unambiguous: the close-tie network divides into identifiable groupings only when bridge-burning is allowed to run without repair.

Attribute assortativity is the measure that speaks most directly to the central claim of this paper, and it does so twice. Assortativity by Openness, computed on the weighted graph, is indistinguishable from the complete-graph baseline of minus one over (N minus one): the observed values of -0.4877, -0.2432, -0.1270, and -0.1141 at N of 3, 6, 9, and 12 sit against baselines of -0.5000, -0.2000, -0.1250, and -0.0909. Assertiveness behaves the same way, at -0.5257, -0.2098, -0.1198, and -0.0947 against the same baselines. Figure 8 plots both against it. Tie strength in this model does not sort agents by either of their traits. Tie strength in this model does not sort agents by their plasticity. Assortativity by attitude is undefined in 49.6 per cent of terminal observations, and the reason is that the attribute has no variance left to correlate with anything: once every agent is Neutral, the denominator of the assortativity coefficient is zero. Among the observations where it is defined, it averages 0.0258, which is to say approximately nothing.

That an attitude assortativity coefficient cannot be computed is not a failure of measurement. It is latent polarization stated in a second vocabulary. The opinions have converged so completely that they no longer distinguish one agent from another, while the ties among those agents remain unequal. Belief has stopped varying; belonging has not.

Figure 6.

Note. Strong ties consolidate into coherent triads rather than scattering across the group. >

Figure 7. *Modularity of the Close-Tie Graph (weights of at Least 56) at Structural Equilibrium, by Same-Growth and Opposite-Growth, Over Nine Seeds*

Note. The close-tie network divides into identifiable groupings only where bridge-burning runs without repair.

Table 7.

N	Baseline, -1/(N-1)	Observed, by Openness	Excess over baseline
3	-0.5000	-0.4877	-0.5257
6	-0.2000	-0.2432	-0.2098
9	-0.1250	-0.1270	-0.1198
12	-0.0909	-0.1141	-0.0947

Latent Polarization Is a Matter of Timescale

Recording the network at the tick of opinion consensus as well as at structural equilibrium makes the central phenomenon of this paper measurable in a way the original design could not.

At the moment opinions converge, the network carries substantial fracture. The percentage of distant ties at consensus runs from 44.69 per cent at same-growth = 0 down through 40.51, 34.20, and 27.52 to 21.91 per cent at same-growth = 10. A group reaching agreement is therefore not a group that has repaired itself. Between a fifth and nearly a half of its relationships remain severed at the instant its members stop disagreeing.

Allowed to continue, that fracture largely closes. Terminal fracture settles at 7.70 per cent for every positive value of same-growth, identical to four significant figures, so that between 14.21 and 32.80 percentage points of fracture heal after consensus is reached. Expressed as a proportion, between 65 and 81 per cent of the fracture present at consensus eventually closes. The exception is the case where repair is unavailable. At same-growth = 0, fracture does not heal but deepens, rising from 44.69 per cent at consensus to 47.74 per cent at equilibrium, because ties in that regime can only burn.

The claim this supports is therefore narrower and more falsifiable than a claim that structural scars persist indefinitely. Latent polarization is real, it is large, and it is a transient. Whether an observer finds a group cohesive or fractured depends on when the group is observed relative to the two processes running inside it, one convergence of opinion and one repair of ties. Where opinion converges quickly relative to the rate of repair, a group will present as unanimous while remaining structurally divided, and that window is wide: for the slowest repair rate examined here, consensus arrives at a mean of 4.61 ticks while structural equilibrium is not reached for 33.98. Where repair is absent altogether, the division is permanent and continues to deepen after the disagreement that caused it has been forgotten.

Table 8. *Network Fracture at Opinion Consensus and at Structural Equilibrium*

same-growth	Distant ties at consensus (%)	Distant ties at equilibrium (%)	Healed (pp)	Proportion healed
0	44.69	47.74	-3.04	none; fracture deepens
2.5	40.51	7.71	+32.80	81.0%
5	34.20	7.70	+26.50	77.5%
7.5	27.52	7.70	+19.81	72.0%
10	21.91	7.70	+14.21	64.9%

Note. Positive values in the final column are fracture that healed after agreement was reached; the negative value is fracture that deepened.

Figure 8.

Note. Unweighted assortativity on a complete graph reduces to −1/(N−1) regardless of the attribute supplied. Observed weighted values for both Openness and Assertiveness track that baseline at every group size, indicating that tie strength does not sort agents by either trait.

Figure 9.

Note. The gap between the two series is the fracture that heals after agreement is reached. ## The Sweet Spot of Dynamics: Non-Linearity of Openness Openness does not accelerate convergence monotonically. Measured as the tick at which attitudes first settle, the relationship is single-peaked: consensus arrives at a mean of 2.68 ticks at openness-mean = 0, is delayed to 4.61 ticks at openness-mean = 5, and quickens again to 3.55 ticks at openness-mean = 10.

The two ends of the range are fast for opposite reasons, and neither is interesting. At openness-mean = 0, agents cannot absorb the signal at all, so attitudes never move and the stagnation criterion is satisfied almost immediately. These groups do not converge; they are frozen at their initial configuration, which is why their terminal variance remains at 3.0268. At openness-mean = 10, agents adopt the global signal in a single step, and the population collapses to Neutral before any structure of interest can develop. The dynamics are over before they begin.

The intermediate regime is where the model is actually dynamic. At openness-mean = 5, agents move toward the signal but do not arrive in one step, and the resulting gradual drift is the only regime in which the local bias term of Equation (4) has time to operate on the trajectory. It is also the regime in which consensus and structural equilibrium are most widely separated, and therefore the regime in which latent polarization is most pronounced.

Time to structural equilibrium follows a different pattern from time to consensus, which is itself informative: 20.81, 20.71, and 16.93 ticks at the three levels of Openness. The network takes roughly as long to settle whether its agents are frozen or moderately open, because tie repair proceeds on its own schedule once opinions stop changing. Group size lengthens both processes modestly, with equilibrium reached at 15.59, 18.45, 19.19, and 20.28 ticks for populations of 3, 6, 9, and 12.

Table 9. *Timing of the Two Convergence Processes, by Openness and by Group Size*

Grouping	Level	Ticks to consensus	Ticks to structural equilibrium
Openness	0	2.68	20.81
Openness	5	4.61	20.71
Openness	10	3.55	16.93
Group size	3	2.81	15.59
Group size	6	3.23	18.45
Group size	9	3.44	19.19
Group size	12	3.57	20.28

Figure 10.

Note. The single-peaked relationship places the dynamic regime at moderate Openness. # Discussion ## The Inevitability of Consensus in Mean-Field Models The findings of this study provide a robust computational validation of Abelson's (1964) diversity dilemma. Abelson mathematically demonstrated that in a fully connected network where agents reduce differences with their neighbors, universal agreement is the only stable equilibrium. By implementing a Net-Attitude, which is functionally a weighted average of the entire group, this model enforces a mean-field topology.

In this architecture, every agent is effectively connected to every other agent through the medium of the global signal. Even if agent A and agent B have severed their direct tie, so that the weight between them is zero, agent A is still influenced by the contribution of agent B to the global Net-Attitude. This influence of unlimited range bypasses local network fractures. Manifest polarization, the formation of stable opposing opinion clusters, is consequently impossible in this model whenever Openness is greater than zero, and the sweep confirms it: terminal attitude variance falls to 0.0117 at maximal Openness, and the composition parameter has no measurable effect on where a group arrives.

This contrasts sharply with bounded-confidence models (Deffuant et al., 2002) and with the culture model of Axelrod (1997), in which interactions are strictly local and conditional. In those models, if the distance between agents exceeds a threshold, influence cuts off completely, allowing divergent clusters to survive. The Neutral Attractor observed here is therefore not a reflection of human psychology as such, but a consequence of global coupling. Whether any real communication structure approximates a mean field closely enough for the result to transfer is a separate question, and one this model cannot answer. A small group in sustained conversation, in which every member hears what every other member says, is the case the present specification is meant to describe. Extension of the argument to larger and more mediated settings would require a model in which the field is local rather than global, and is left to the future work outlined below.

Latent Polarization as a Lag Rather Than a Scar

The central theoretical contribution of this work is the identification of latent polarization, a state in which attitudinal distance is zero while relational distance remains large. The sweep establishes that the state is real and that it is common. At the tick on which a group reaches agreement, between 21.91 and 44.69 per cent of its ties are severed, and the proportion depends on how readily that group rebuilds what disagreement has cost it. The coefficient of attitude assortativity cannot be computed at all in 49.6 per cent of terminal observations, and the reason is instructive: once every agent holds the same view, the attribute has no variance left to correlate with tie strength. Agreement has become total and structure has not followed it. That is latent polarization expressed as a statistic rather than as a description.

The results also correct an earlier reading of the same mechanism. It is tempting to describe the severed ties as scars, and to conclude that convergence of opinion is insufficient to repair them. Measured at structural equilibrium rather than shortly after agreement, that conclusion does not hold. Groups repair between 65 and 81 per cent of the fracture present at consensus, and they settle at a residual of 7.70 per cent that is identical to four significant figures across every positive rate of repair examined. The rate of repair determines how long a group remains divided, not how divided it finally is. What looked like a permanent structural injury is better understood as a lag between two processes running at different speeds, one that resolves opinion and one that rebuilds relationships.

The lag is nonetheless substantial, and the practical consequence survives the correction. For the slowest rate of repair examined, agreement arrives at a mean of 4.61 ticks while structural equilibrium is not reached until 33.98. A group observed anywhere in that interval will present as unanimous while remaining materially divided. The condition that does not resolve is the one in which repair is unavailable altogether: at a bridge-building increment of zero, fracture rises from 44.69 per cent at consensus to 47.74 per cent at equilibrium, deepening after the disagreement that caused it has been settled, because ties in that regime can only burn.

This pattern bears an instructive resemblance to work on affective polarization (Iyengar et al., 2019), which finds that partisan animosity can persist in the absence of current policy disagreement, sustained instead by identity sorting and by the residue of past conflict. The resemblance is an analogy between mechanisms operating at very different scales, and it should not be read as a claim that the present model explains mass political behavior. This model contains no representation of social identity, no group membership, and no mechanism by which affect toward a category of person could form. What it does show is that a much simpler ingredient, namely a difference in the speed at which opinions converge and relationships mend, is by itself sufficient to produce agreement coexisting with antagonism. If so minimal a mechanism suffices in a group of twelve, it is worth asking what proportion of the analogous pattern at larger scales requires the heavier apparatus normally invoked to explain it.

For the small groups this model addresses, the implication is that measuring cohesion by surveying attitudes alone can yield false positives. A group may be uniform in what it thinks while lacking the strong ties that collective action and resilience to future shocks require. But the diagnosis must include

timing. Because fracture is transient wherever repair is possible, a single observation cannot distinguish a group that was never divided from one that has already recovered, nor a group in temporary difficulty from one that is permanently broken.

Temporal Nonlinearity and the Dynamic Range

The relationship between Openness and the duration of the opinion process is not monotonic. Consensus arrives at a mean of 2.68 ticks when agents are frozen, is delayed to 4.61 ticks at moderate Openness, and quickens again to 3.55 ticks when agents are maximally receptive. The two extremes are fast for opposite and equally uninteresting reasons. Frozen groups satisfy the stagnation criterion because nothing ever moves, and hyper-liquid groups because everything moves at once. Complexity arises only when agents are receptive enough to change but resistant enough to avoid immediate conversion, and this intermediate regime is the only one in which the local bias term has time to shape a trajectory.

It does not follow, as an earlier reading of these dynamics suggested, that intermediate plasticity is the most dangerous state for structural cohesion. The sweep shows the reverse. Terminal fracture falls monotonically as Openness rises, from 28.07 per cent through 10.26 to 8.80 per cent, and weighted clustering rises correspondingly from 0.4462 to 0.8776 to 0.9199. Modularity of the close-tie graph, which measures the extent to which the network divides into identifiable groupings, falls from 0.0640 to 0.0128 to 0.0112 across the same range. By every structural measure computed here, receptiveness protects the network rather than endangering it.

The reason is visible in the mechanism. Bridge-burning requires two agents on opposite sides of the scale, and neither of them Neutral. Openness is what removes agents from that condition. A group of stubborn members never leaves it, so opposed pairs continue to erode their ties for as long as the simulation runs, which is why the frozen regime produces both the highest fracture and the widest gap between agreement and structural settlement. Prolonged negotiation is therefore not the costly option that a superficial reading of the sweet spot implies. Refusal to move is the costly option, and moderate plasticity merely extends the interesting part of the process without extending the damage.

What the Complete Graph Permits and Forecloses

Because every agent in this model holds a weighted tie to every other, several standard instruments of network analysis return constants here rather than measurements. The unweighted clustering coefficient is identically one, since every triple closes a triangle by construction. The Barrat weighted coefficient is likewise constant, because its edge weights cancel against node strength. Modularity computed on the full weighted graph is maximized at zero by the trivial single-community partition, since all nodes are structurally equivalent. Unweighted attribute assortativity reduces to minus one over the quantity N minus one, a value fixed by group size alone and wholly insensitive to the attribute supplied to it. These

are not failures of the model. They are properties of the topology, and any analysis of a complete graph will encounter them.

The information is recoverable by two routes, and this study uses both. The Onnela coefficient remains sensitive to weights because it takes the geometric mean of all three edges in each triangle rather than normalizing by node strength, and it tracks cohesion closely across the sweep. Thresholding the graph at the close and distant cut points the model already defines yields a close-tie subgraph on which conventional modularity, transitivity and component counts are meaningful. The two routes agree, which is some assurance that neither is an artifact of its own construction.

What this forecloses is any claim about mesoscale organization. Community structure at that level is a property of sparse large networks, and a complete graph does not possess it at any size. The close-tie subgraph does divide when bridge-burning runs unopposed, and the modularity values reported above document that division, but these are groupings within a dozen mutually acquainted people, not communities in a social network of thousands. The distinction matters for how the results should be read and for what a sparse-topology successor to this model would be able to add.

Comparison With Generative Agents and LLMs

The drift to Neutral observed in this mathematical model parallels recent findings in the domain of generative agents (Park et al., 2023). Chuang et al. (2024) found that societies of agents powered by large language models exhibit a strong consensus bias. Standard language models, trained with reinforcement learning from human feedback, are fine-tuned to be agreeable, non-confrontational, and moderate.

In this model, the Net-Attitude performs the same function as alignment training performs in a language model: it is a normative anchor that pulls agents away from the extremes. Just as Chuang et al. found that they had to prompt models explicitly to be stubborn or irrational in order to generate polarization, this model shows that unless Openness is set to zero the system cannot sustain diversity. That two very different modeling traditions, mathematical sociophysics and neural language processing, converge on the same result suggests that centralized alignment mechanisms, whether a mathematical average or a safety objective, are in tension with the preservation of diversity.

The correspondence has a limit worth stating. In the present model the anchor is exactly specified and its strength is a swept parameter, which is what makes the phase structure mappable. In a language model the anchor is implicit in the training procedure and its strength is neither specified nor directly adjustable. The parallel is therefore useful for generating hypotheses about how such systems behave in aggregate, and not for drawing quantitative conclusions about either from the other.

Future Directions: From Behavioral Map to Empirical Test

The regimes charted above are predictions awaiting data. An approved prospective study will supply that test by instantiating the model on real friendship groups: Openness and Assertiveness will be measured directly, agents will be initialized from those measured traits rather than from swept parameters, and the simulated trajectories will be compared against the observed evolution of the groups over time. This procedure locates each real group at a coordinate in the parameter space mapped here and asks whether its predicted regime, neutral drift, structural tug-of-war, or the moderate-Openness dynamic range, actually obtains.

The timescale result reported above sharpens what such a study needs to measure. Because network fracture in this model is a transient rather than a fixed point, a single observation of a real group cannot distinguish a cohesive group from a fractured one that has had time to repair itself. The empirical design therefore requires at least two measurements separated in time, and ideally a measurement taken close to the moment a disagreement resolves. What the model predicts is not a level of fracture but a relationship between two rates, the speed at which a group reaches agreement and the speed at which it rebuilds the ties that the disagreement cost it. Estimating both rates in real groups is the substantive empirical task, and it is more demanding than estimating either alone.

A second direction follows from the finding that the rate of tie repair sets the time to structural equilibrium without affecting its location. If this holds outside the model, it implies that interventions aimed at repairing damaged relationships change how long a group spends in a vulnerable state rather than how the group eventually settles. That is a testable and practically consequential claim, and it would be worth examining in settings where the repair process can be manipulated directly.

Limitations

The primary limitation of this model remains the assumption of the global mean field. Individuals are rarely exposed to the true average opinion of the population around them. They reside in filter bubbles and echo chambers in which the perceived Net-Attitude is heavily skewed by the local network. Future iterations should localize the Net-Attitude calculation, allowing agents to perceive a local field derived only from their connected neighbors. It is hypothesized that this modification would break the Neutral Attractor and permit the emergence of manifest polarization, as local clusters would converge on different means. The present results should be read as describing what global coupling does, not as a claim that real groups experience global coupling.

The complete graph carries a second set of limitations. It is the right topology for the small mutually acquainted groups this model addresses, but it forecloses questions about sparse structure. It also renders several standard network measures uninformative, as set out above: unweighted clustering, the Barrat weighted coefficient, modularity on the full weighted graph, and unweighted attribute assortativity are each constant on a complete graph regardless of the weights or attributes supplied to them. The measures reported here work around that constraint by using the Onnela coefficient, which remains sensitive to weights, and by thresholding the graph at the close and distant cut points the model already

defines. Those are defensible choices, but they are choices, and a sparse-topology variant of the model would permit a wider and more conventional set of structural diagnostics.

Three features of the sweep itself constrain what can be claimed from it. Ten random seeds were planned and nine were used, because the output for one seed was lost to a data handling error and could not be recovered; the nine remaining seeds are intact and independent, but the robustness set is smaller than intended. The grid resolution was reduced relative to an earlier design in order to make per-tick recording of the full network state tractable, and while the parameters governing network structure were retained at full resolution, the composition parameter, the assertiveness mean, and the two trait dispersions were sampled more coarsely. And the stopping rule, three successive ticks without a change in any tie weight, subject to a ceiling of 200 ticks, is itself a modeling decision rather than a property of the system. It was chosen because a rule based on opinion halts runs in the middle of an unfinished structural process, but a different rule would place the measured endpoint elsewhere.

The model has no calibration to real time. Ticks are not days, weeks, or conversations, and the timescale result should be read as a statement about the ordering and relative duration of two processes rather than about their absolute speed. Whether the window between consensus and structural repair is hours or years in a real friendship group is exactly what the empirical study is designed to establish, and nothing in the present results bears on it.

Finally, two mechanisms are retained with acknowledged uncertainty. The distant-tie contributions to the social bias encode reactance, the idea that alignment by a disliked other pushes an agent the other way, and repulsive influence of this kind is empirically contested (Flache et al., 2017). Traits are static: Openness and Assertiveness are drawn at setup and never change, which excludes the possibility that a group’s experience of conflict alters the plasticity of its members.

Conclusion

This study presents a hybrid agent-based model that integrates the micro-foundations of personality, Openness and Assertiveness, with the macro-dynamics of network evolution in small friendship groups. Two results follow, and the second qualifies the first.

Global influence mechanisms lead to opinion homogenization. Across the sweep, any group whose members possess a non-zero capacity to absorb the global signal converges on Neutrality, with terminal attitude variance falling from 3.03 to 0.01 as Openness rises, and it does so irrespective of how the group was composed to begin with. This confirms Abelson’s theoretical prediction that universal agreement is the only stable equilibrium under global coupling, and it locates the effect in the architecture rather than in the psychology of the agents.

Homogenization does not, however, guarantee social integration. At the moment a group reaches agreement, between 21.91 and 44.69 per cent of its relationships remain severed, depending on how readily it rebuilds them. Belief has converged while belonging has not, and the assortativity coefficient by

attitude cannot even be computed at that point, because the attribute has no variance left to correlate with anything. That condition is what this paper terms latent polarization, and it is distinct both from traditional polarization, which is disagreement, and from cohesion, which is agreement together with connection.

The central refinement this study offers is that latent polarization is a matter of timescale rather than a permanent state. Allowed to continue past the point of agreement, a group repairs between 65 and 81 per cent of that fracture, settling at a residual of 7.70 per cent that is identical across every positive rate of repair examined. What the rate of repair changes is how long the group spends divided, not how divided it ends up. The exception is instructive: where no capacity for repair exists at all, fracture does not heal but deepens, rising after agreement has been reached, because ties in that regime can only burn. Whether a group appears cohesive or fractured therefore depends on when it is observed relative to two processes running inside it at different speeds, and the practical question is not whether fracture occurs but how long the window between agreement and repair stays open.

This also revises a cautionary note that an earlier reading of these dynamics appeared to support. Moderate Openness does prolong the opinion process, with consensus arriving at 4.61 ticks against 2.68 for frozen groups and 3.55 for maximally receptive ones, and it is the only regime in which the local bias term has time to shape a trajectory. But it does not maximize structural damage. Terminal fracture falls monotonically with Openness, from 28.07 per cent through 10.26 to 8.80 per cent, and weighted clustering rises correspondingly from 0.45 to 0.88 to 0.92. The most structurally damaging regime in this model is not prolonged negotiation but stubbornness, in which agents never move, opposition never resolves, and ties burn without any countervailing process. Receptiveness to others protects the social fabric in this model rather than endangering it.

These findings advocate for a multi-dimensional approach to the health of small groups, one that monitors not only the variance of opinions but the density and condition of the ties that connect their members, and that treats the timing of observation as part of the measurement rather than incidental to it. A group that agrees is not thereby a group that has healed, and the interval between those two states is where the interesting dynamics live.

Declarations

Ethical Approval

This article does not contain any studies with human participants performed by any of the authors

Informed Consent

This article does not contain any studies with human participants performed by any of the authors

Conflict of Interest

The author has no conflict of interest to disclose

Ai Usage

No Ai was used at any step of the manuscript

Consent to Publish

Consent to Participate declaration: not applicable

Author Contribution

A.Ch. Conceptualization, Analysis, Write-up, Editing

Data Availability

NetLogo, analysis code and Flowchart [find them in the supplementary material]Raw Data [Repository link removed due to peer review]

References

1. Abelson, R. P. (1964). Mathematical models of the distribution of attitudes under controversy. In N. Frederiksen & H. Gulliksen (Eds.), Contributions to mathematical psychology (pp. 142–160). Holt, Rinehart & Winston.
2. Allen, J. P., Loeb, E. L., Kansky, J., & Davis, A. A. (2022). Beyond susceptibility: Openness to peer influence is predicted by adaptive social relationships. *International Journal of Behavioral Development, 46*(3), 180–189. https://doi.org/10.1177/0165025420922616
3. Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193*(5), 31–35. https://doi.org/10.1038/scientificamerican1155-31
4. Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution, 41*(2), 203–226. https://doi.org/10.1177/0022002797041002001
5. Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics, 81*(2), 591–646. https://doi.org/10.1103/RevModPhys.81.591
6. Chuang, Y.-S., Goyal, A., Harlalka, N., Suresh, S., Hawkins, R., Yang, S., Shah, D., Hu, J., & Rogers, T. (2024). Simulating opinion dynamics with networks of LLM-based agents. In K. Duh, H. Gomez, & S. Bethard (Eds.), *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 3326–

3346). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-naacl.211

7. Deffuant, G., Amblard, F., Weisbuch, G., & Faure, T. (2002). How can extremism prevail? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation, 5*(4), Article 1. https://www.jasss.org/5/4/1.html
8. Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation, 20*(4), Article 2. https://doi.org/10.18564/jasss.3521
9. Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380. https://doi.org/10.1086/225469
10. Harton, H. C., & Bullock, M. (2007). Dynamic social impact: A theory of the origins and evolution of culture. *Social and Personality Psychology Compass, 1*(1), 521–540. https://doi.org/10.1111/j.1751-9004.2007.00022.x
11. Holley, R. A., & Liggett, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability, 3*(4), 643–663. https://doi.org/10.1214/aop/1176996306
12. Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science, 22*, 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034
13. Kim, S. Y., Taber, C. S., & Lodge, M. (2010). A computational model of the citizen as motivated reasoner: Modeling the dynamics of the 2000 presidential election. *Political Behavior, 32*(1), 1–28. https://doi.org/10.1007/s11109-009-9099-8
14. Klemm, K., Eguíluz, V. M., Toral, R., & San Miguel, M. (2003). Global culture: A noise-induced transition in finite systems. *Physical Review E, 67*(4), Article 045101. https://doi.org/10.1103/PhysRevE.67.045101
15. Latané, B. (1981). The psychology of social impact. *American Psychologist, 36*(4), 343–356. https://doi.org/10.1037/0003-066X.36.4.343
16. Latané, B. (1996). Dynamic social impact: The creation of culture by communication. *Journal of Communication, 46*(4), 13–25. https://doi.org/10.1111/j.1460-2466.1996.tb01501.x
17. Latané, B., & L'Herrou, T. (1996). Spatial clustering in the conformity game: Dynamic social impact in electronic groups. *Journal of Personality and Social Psychology, 70*(6), 1218–1230. https://doi.org/10.1037/0022-3514.70.6.1218
18. McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415–444. https://doi.org/10.1146/annurev.soc.27.1.415
19. Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3586183.3606763

20. Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology, 48*, 609–647. https://doi.org/10.1146/annurev.psych.48.1.609
21. Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review, 103*(2), 219–240. https://doi.org/10.1037/0033-295X.103.2.219
22. Sood, V., & Redner, S. (2005). Voter model on heterogeneous graphs. *Physical Review Letters, 94*(17), Article 178701. https://doi.org/10.1103/PhysRevLett.94.178701
23. Sznajd-Weron, K., & Sznajd, J. (2000). Opinion evolution in closed community. *International Journal of Modern Physics C, 11*(6), 1157–1165. https://doi.org/10.1142/S0129183100000936
24. Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Figures

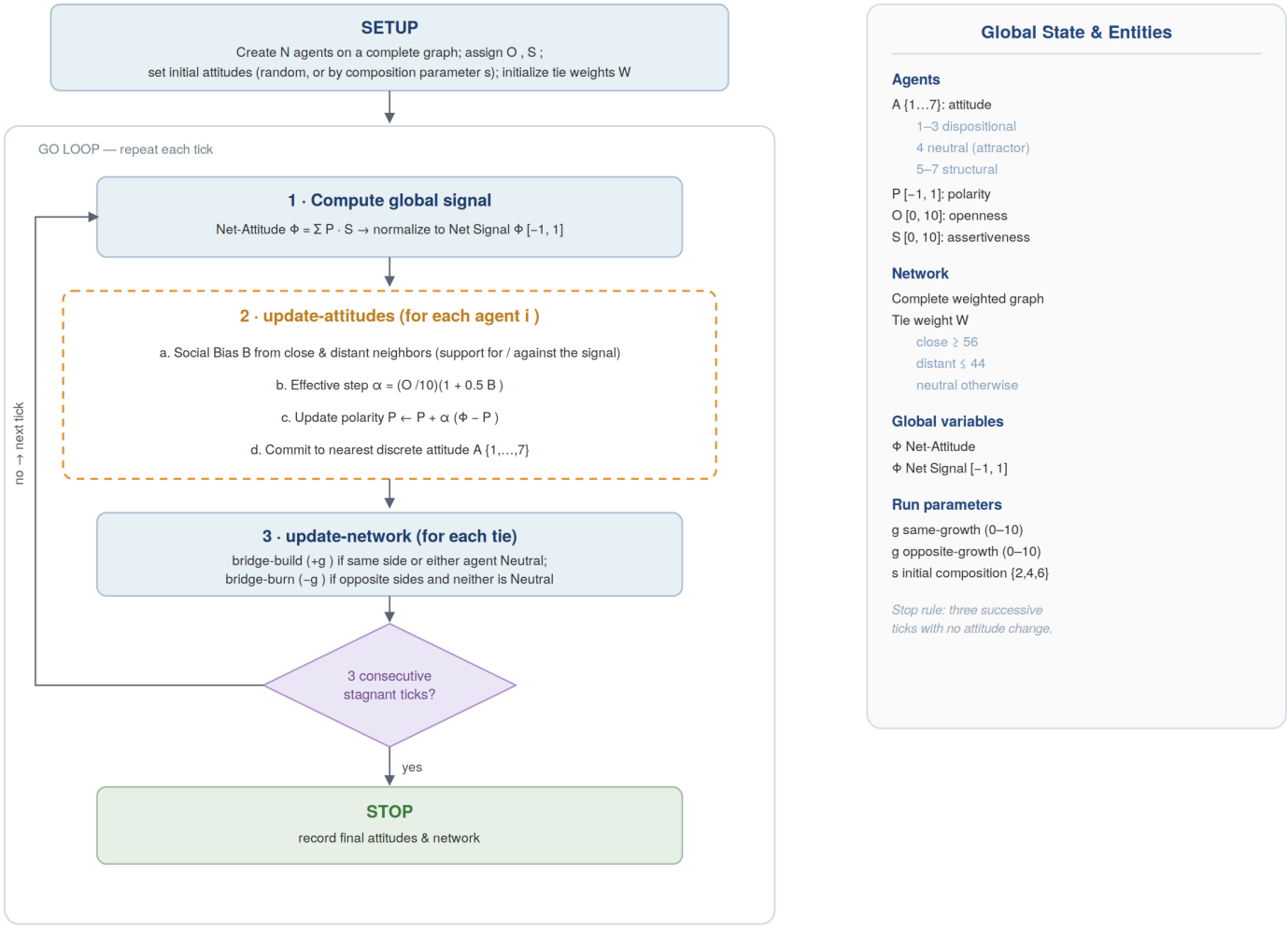


Figure 1

Model Lifecycle

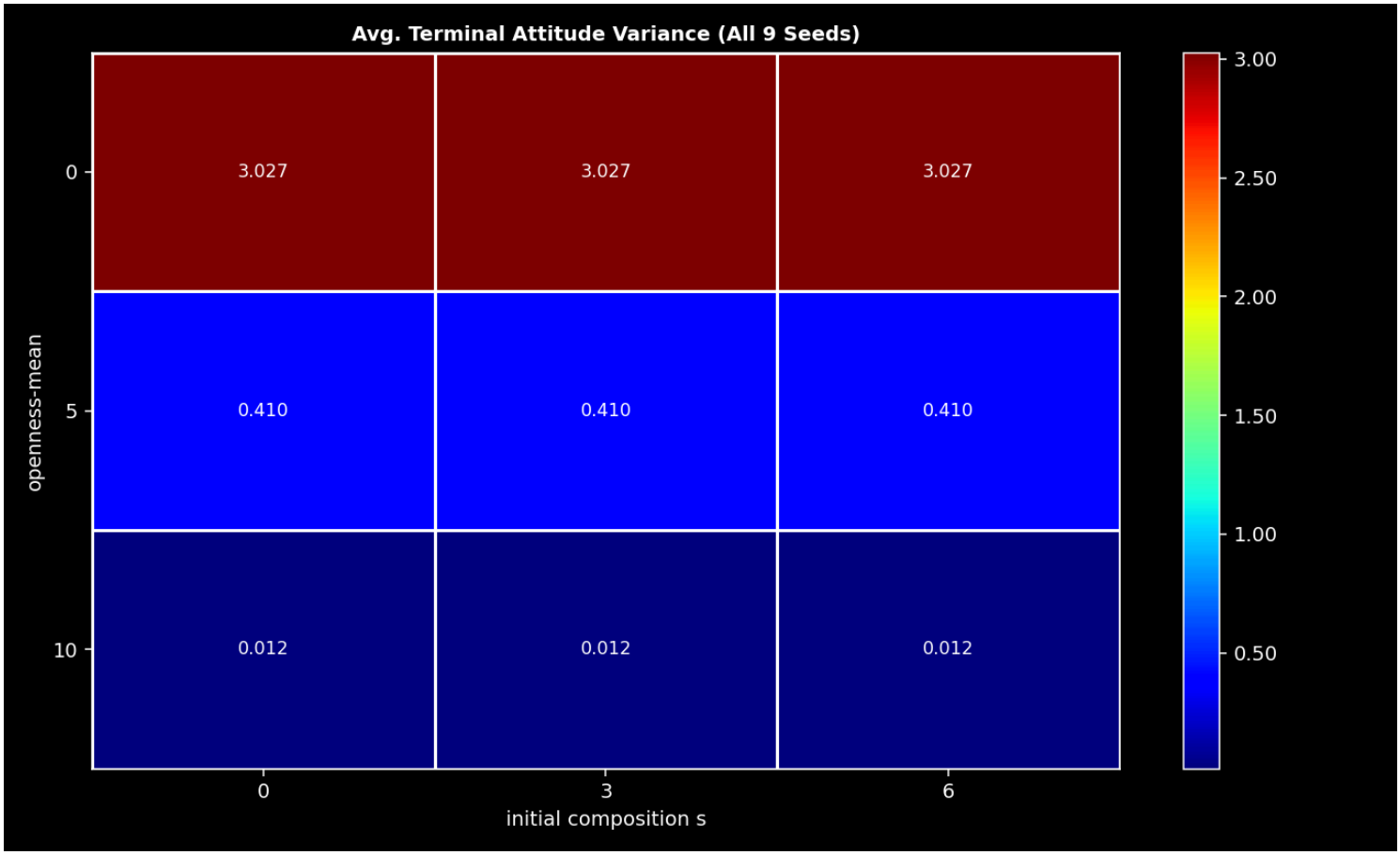


Figure 2

Average Terminal Attitude Variance, by Openness and Initial Composition s

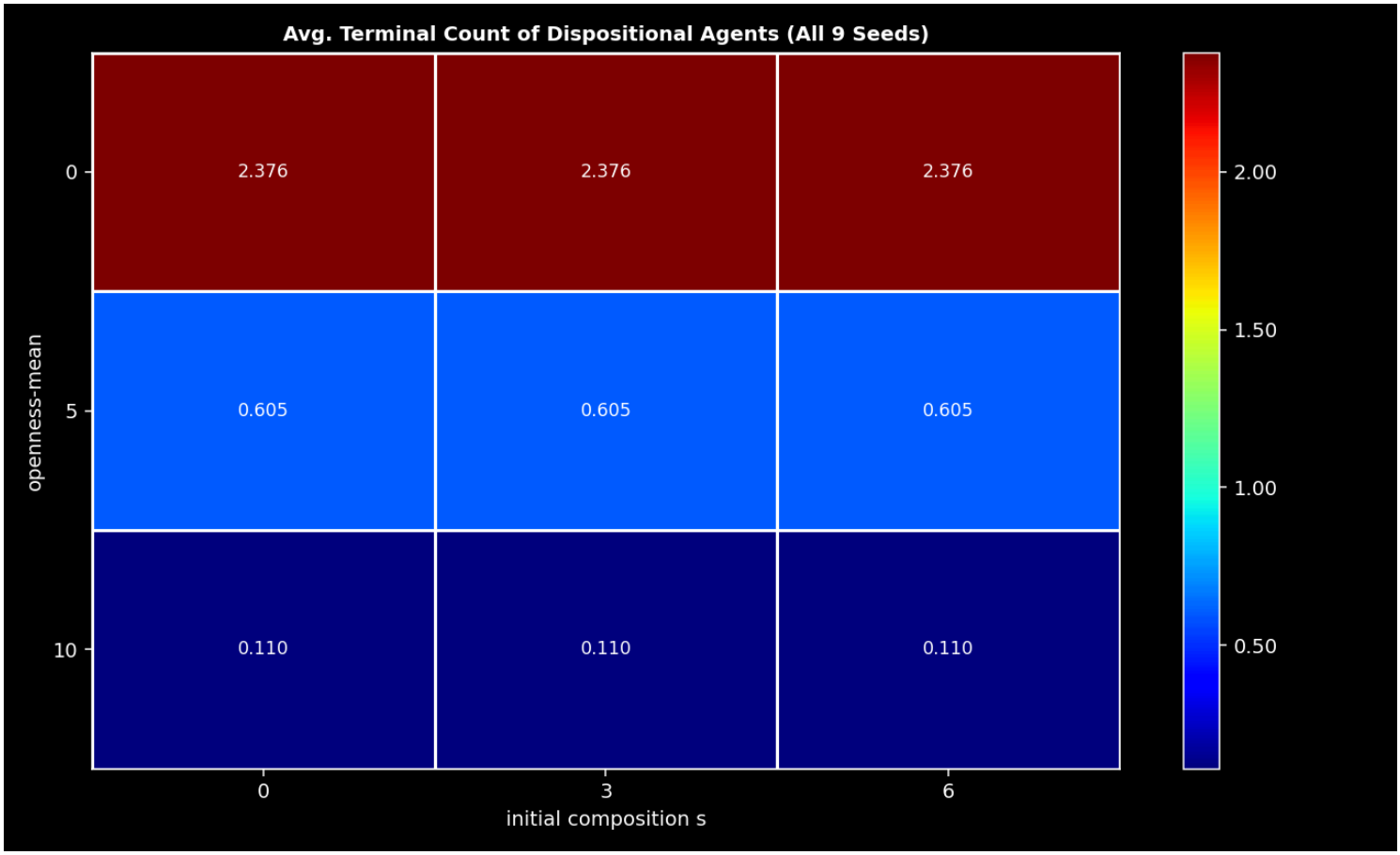


Figure 3

Average Terminal Count of Dispositional Agents, by Openness and Initial Composition s

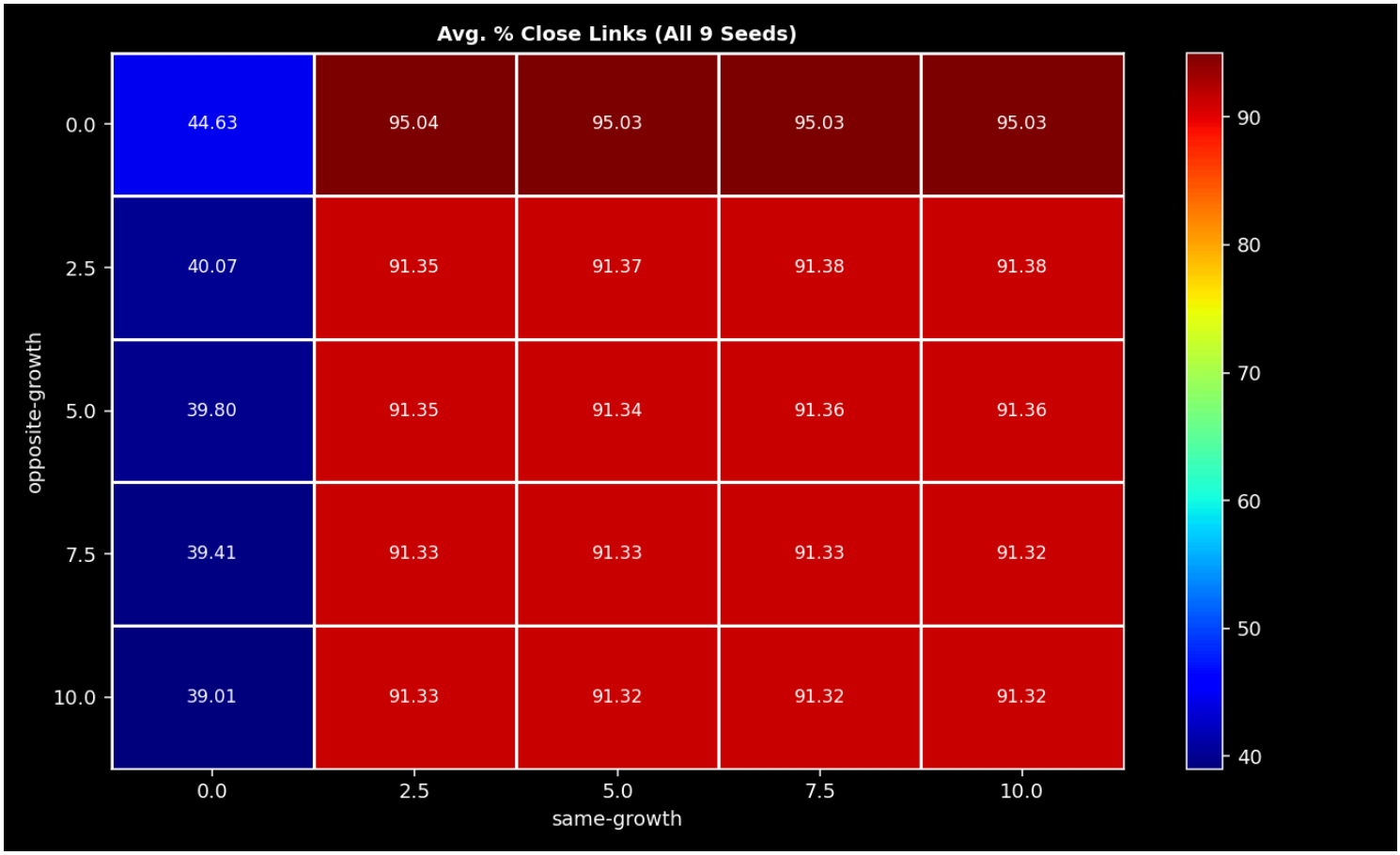


Figure 4

Average Percentage of Close Ties at Structural Equilibrium, by Same-Growth and Opposite-Growth, Over Nine Seeds

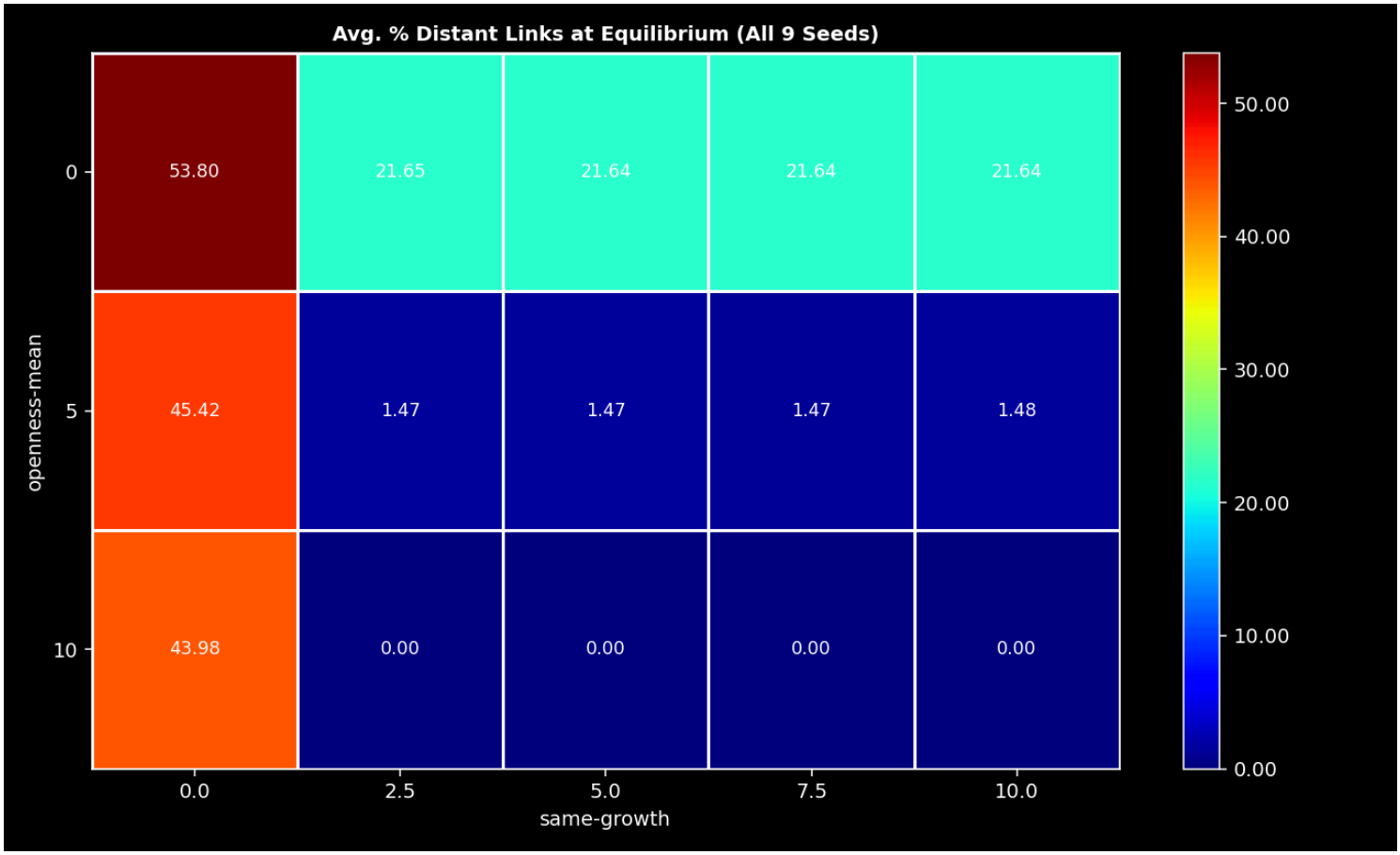


Figure 5

Average Percentage of Distant Ties at Structural Equilibrium, by Openness and Same-Growth, Over Nine Seeds

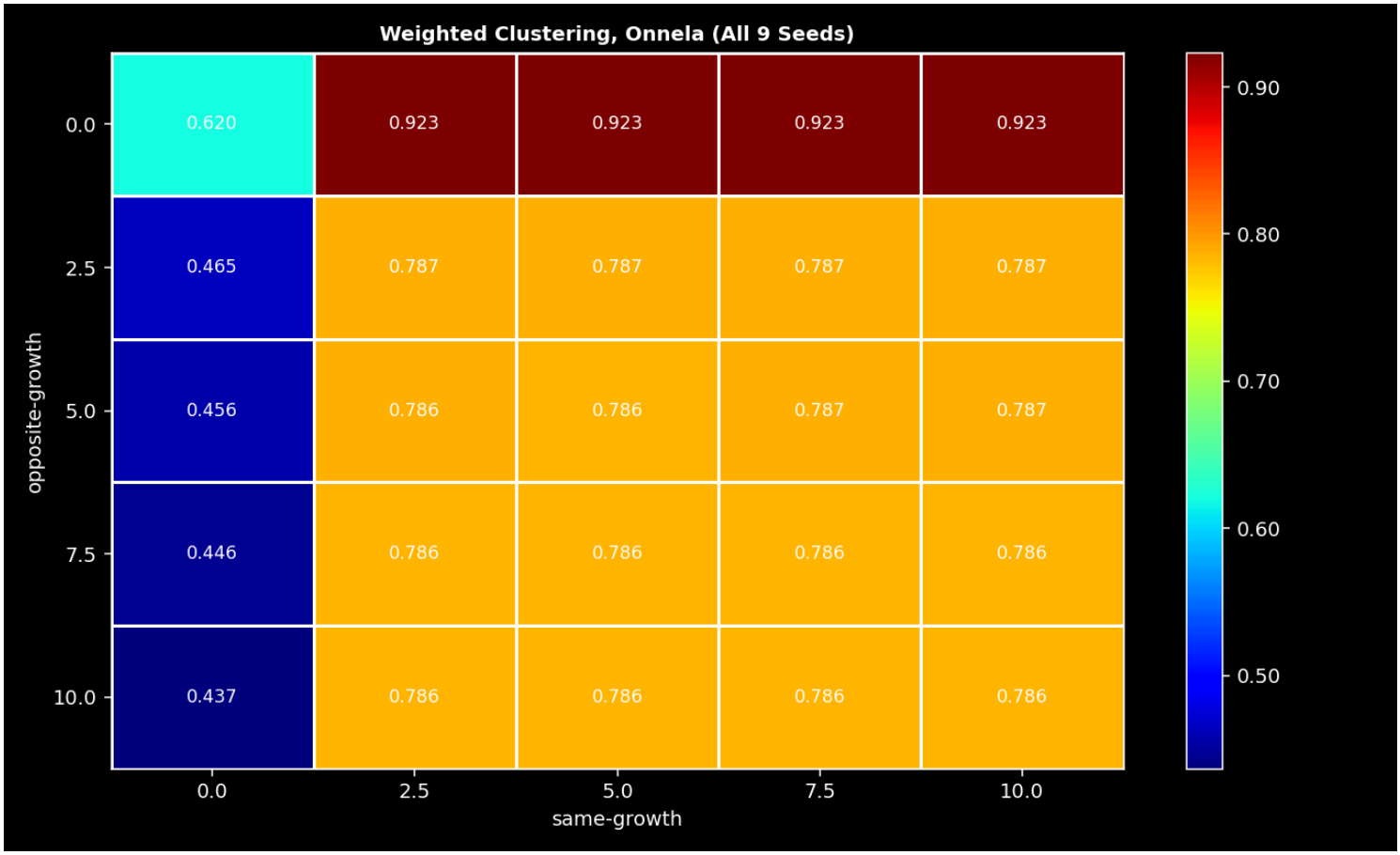


Figure 6

Onnela Weighted Clustering Coefficient at Structural Equilibrium, by Same-Growth and Opposite-Growth, Over Nine Seeds

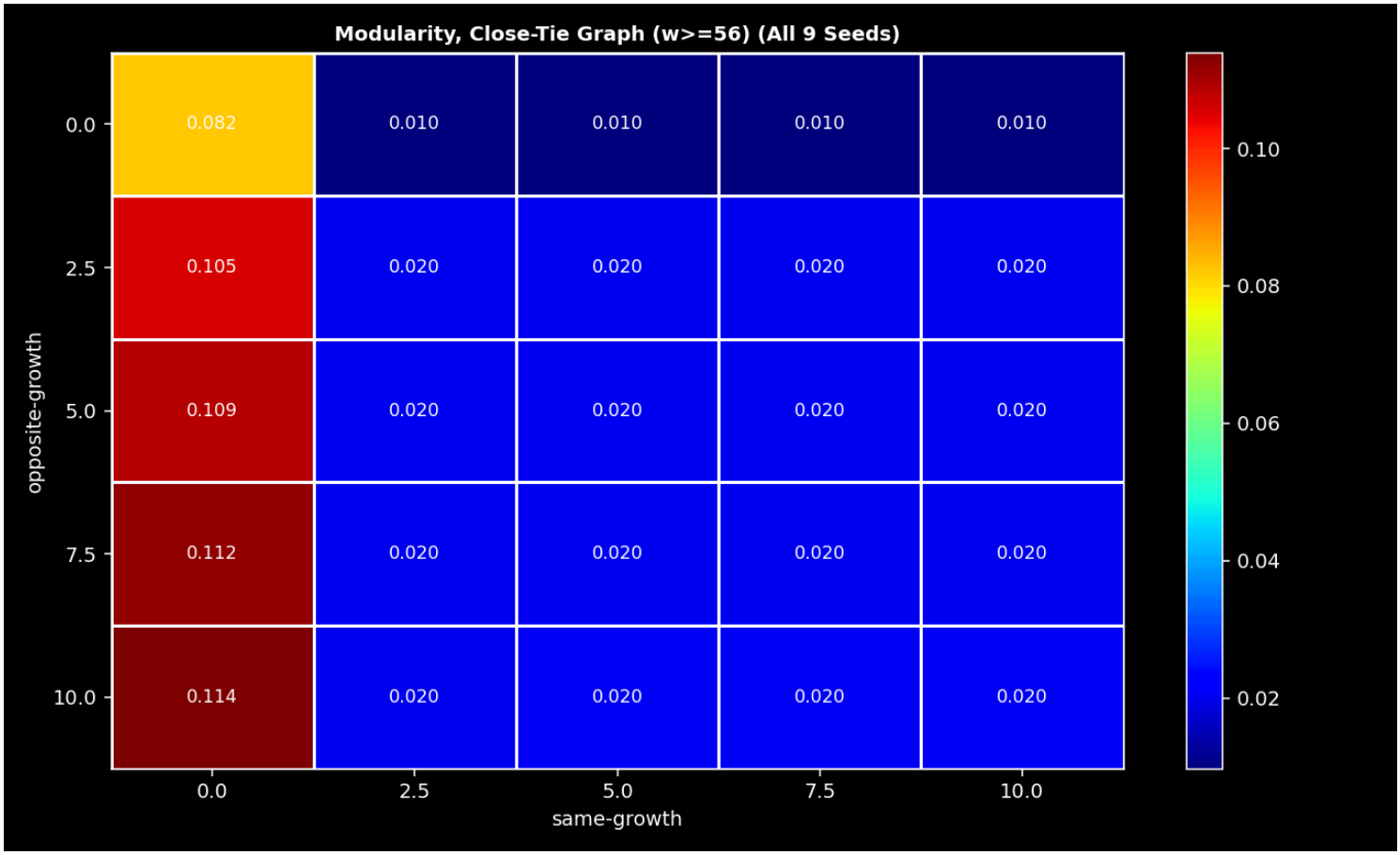


Figure 7

Modularity of the Close-Tie Graph (Weights of at Least 56) at Structural Equilibrium, by Same-Growth and Opposite-Growth, Over Nine Seeds

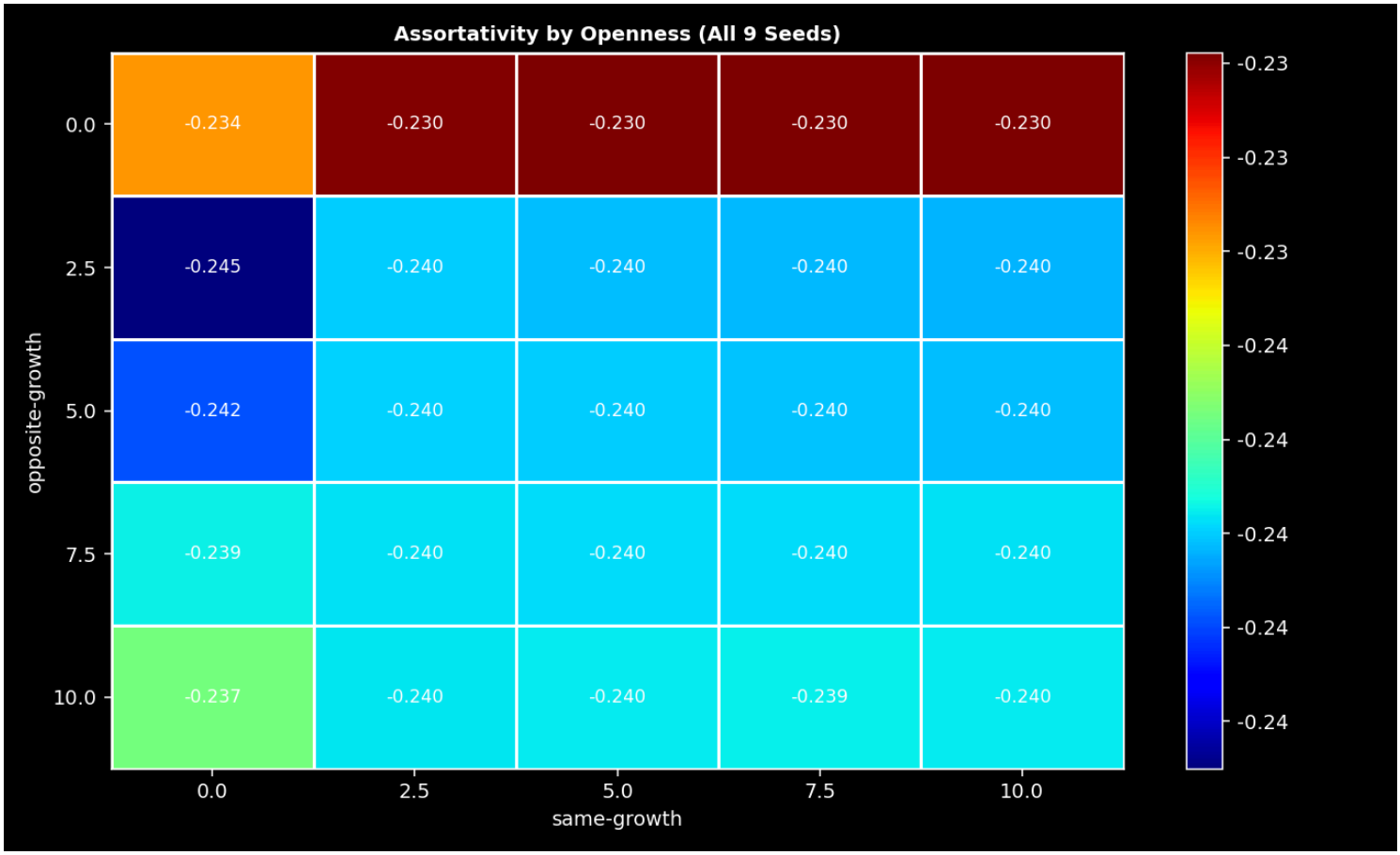


Figure 8

Attribute Assortativity Against the Complete-Graph Baseline, by Group Size

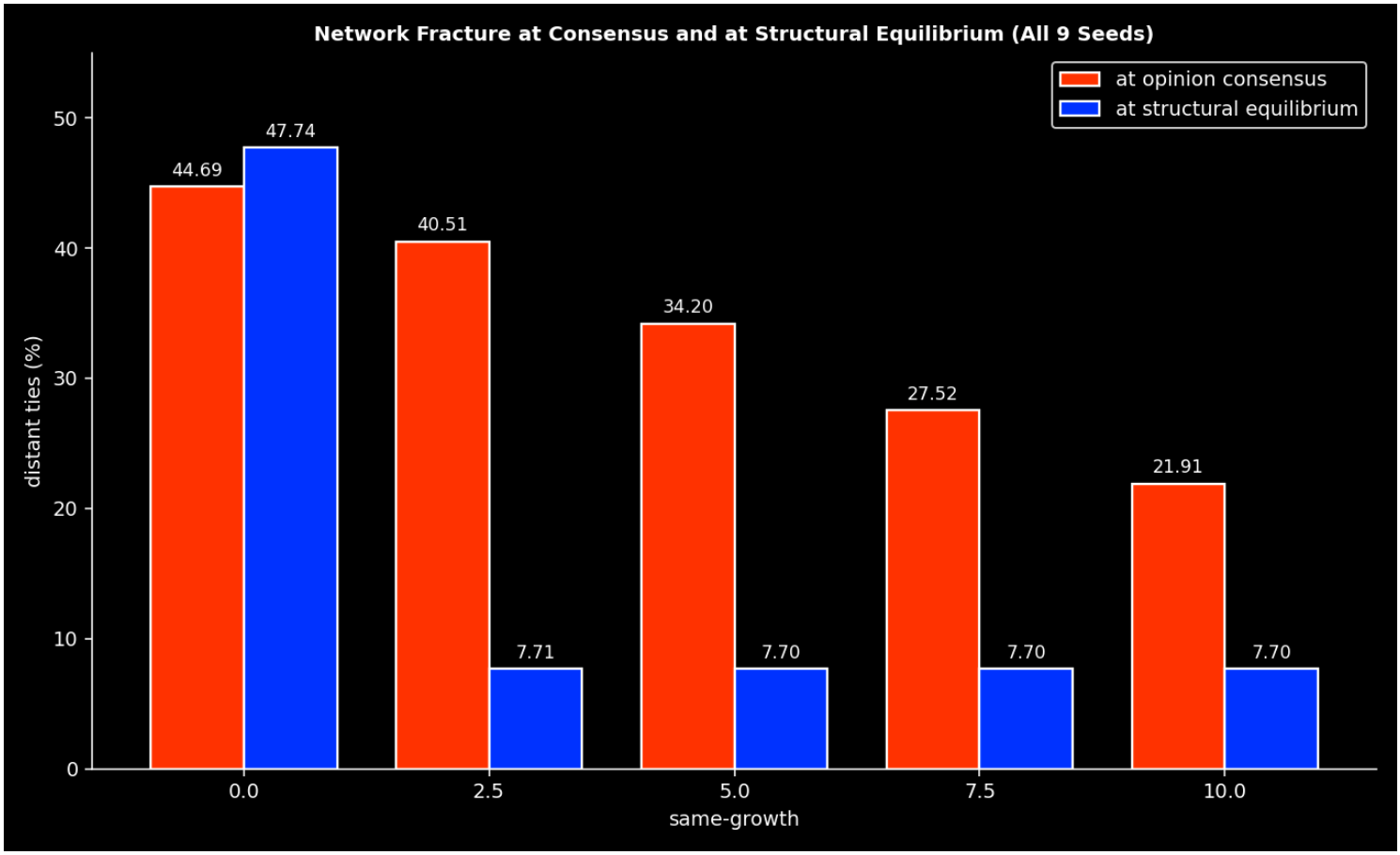


Figure 9

Network Fracture at the Tick of Opinion Consensus and at Structural Equilibrium, by Same-Growth, Over Nine Seeds

Image not available with this version

Figure 10

Average Ticks to Opinion Consensus, by Openness and Same-Growth, Over Nine Seeds

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryCode.zip