JASSS
THE JOURNAL OF ARTIFICIAL SOCIETIES AND SOCIAL SIMULATION

Success-Driven User Activity Contributes to Online Polarization

Sophia Horn[1], Sven Banisch[1], Veronika Batzdorfer[1], Andreas Reitenbach[1], Fabio Sartori[1], Daniel Schwabe[2], Michael Mäs[1]

[1] *Karlsruhe Institute of Technology, Institute of Technology Futures, Department of Sociology, Douglasstrasse 24, Karlsruhe, 76133, Germany*
[2] *Jozef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slowenien*
Correspondence should be addressed to *sophia.horn@kit.edu*

Journal of Artificial Societies and Social Simulation 29(2) 1, 2026
Doi: 10.18564/jasss.5947 Url: http://jasss.soc.surrey.ac.uk/29/2/1.html

Received: 18-03-2025 Accepted: 03-02-2026 Published: 31-03-2026

Abstract: Online social networks are often seen as a breeding ground of political polarization. This study introduces an additional explanation why, attributing polarization to success-driven user activity. Using an agent-based model, we demonstrate that polarization intensifies when users become more active after experiencing rewarding interactions on the platform. We compare a basic version of Axelrod's cultural dissemination model, which lacks success-driven activity, with an extended version that does include it. Our analyses replicate key findings from the literature and show that success-driven activity consistently enhances polarization, even in scenarios like complete networks or minimal network clustering, where Axelrod's model typically predicts uniformity to be unavoidable. Success-driven activity triggers a self-reinforcing "rich-get-richer" dynamic, where success leads to more activity and vice versa, resulting in a highly skewed success distribution. A few highly successful users dominate discussions, causing local convergence and fragmenting the population into distinct groups. This polarization arises in a model without biased media, polarized elites, algorithmic echo chambers, or users intentionally distancing themselves from others. We discuss the implications for designing online social networks that do not exacerbate polarization and for creating digital twins of online platforms for regulatory and analytical purposes.

Keywords: Opinion Polarization, Success-driven User Activity, Online Social Networks, Agent-based Modeling, Reinforcement Learning

Introduction

1.1 The World Wide Web and, in particular, online social networks are considered a key contributor to the growing polarization of political opinions observed in many countries. Scholars debate, however, what mechanisms are responsible for the polarizing effects of online social networks, pointing for instance to the effects of personalized recommender systems (Keijzer & Mäs 2022; Pariser 2011), social bots (Keijzer & Mäs 2021), strong homophily (Sunstein 2001; Bakshy et al. 2015), a lack of digital literacy (Gaultney et al. 2022), and malicious actors (Badawy et al. 2018). Here, we demonstrate that a seemingly innocent aspect is missing in this list of mechanisms explaining why online social networks contribute to opinion polarization: *success-driven user activity*. That is, we show with an agent-based model that polarization is intensified when users become more active after having experienced successful interactions, compared to a world where there is no variation in user activity.

1.2 To this end, we elaborated Axelrod's seminal model of cultural dissemination (Axelrod 1997) and compare his original model, where agents' cultural features are communicating in random order, with a version of his model where agents are activated with an increased probability when they managed to successfully influence network neighbors' features in the past. Simulation experiments demonstrate that success-driven activity generates stronger cultural polarization even under conditions where Axelrod found monoculture (consensus) to

be inevitable. We also provide an explanation for our findings, demonstrating that success-driven user activity generates a highly skewed distribution of user activity, where a few users are highly active and many generate online content at a very low rate. Highly active users pull apart the network into internally homogeneous and mutually distinct clusters.

1.3 The central innovation of our work is the study of success-driven user activity, an assumption supported by decades of empirical research on the motivational framework of social rewards (Homans 1974; Banisch & Olbrich 2019; Lindström et al. 2021). The concept of reward learning follows from classical learning theory, specifically social reinforcement learning. In a nutshell, it implies that humans tend to repeat behavior that they have experienced as rewarding in the past. In online social networks, users generate and emit content that may exert influence on others' opinions. Importantly, users also receive social feedback from their interaction partners who consume their content by receiving likes, retweets, upvotes, and other forms of social approval. Learning theory demonstrates that behavior that receives positive feedback is reinforced and will likely be repeated (Wu et al. 2009). Accordingly, users who have successfully influenced others are assumed to grow increasingly active in the future (Grinberg et al. 2016; Lindström et al. 2021).

1.4 However, while success-driven user activity is a very plausible micro-level process, its emergent consequences on macro-dynamics are not understood. We demonstrate here that success-driven user activity has the potential to generate a polarizing feedback loop. When successful agents are more active, they also will experience additional successful events with an increased frequency and become even more successful. This generates a rich-get-richer dynamics, producing a highly skewed distribution of success where a small number of individuals experience huge success and a large number of individuals are described as having limited success (as can be seen with YouTube feedback loops of attention; Wu et al. 2009). A similar dynamic has been studied in networks generated by so-called "preferential attachment" (Barabási & Albert 1999). While this work studies networks where nodes with a higher number of network connections likely develop further network ties, we study fixed networks, excluding preferential attachment as an explanation for our findings. The central contribution of our work is the finding that success-driven user activity does not only generate heterogeneity in activity. In addition, it contributes to the emergence of polarization.

1.5 Contrary to many earlier contributions to the debate about online social networks and processes of opinion polarization, we advocate the notion that polarization may be an emergent consequence of success-driven user activity. Unlike existing work studying so-called "influencers" in online settings, we do not assume that agents grow successful because of certain characteristics (Vrontis et al. 2021). Likewise, we do not assume psychological processes that change the behavior of users when they grow successful. We also do not assume that users intentionally seek to increase polarization or benefit from growing divisions (Lüders et al. 2022). Instead, we introduce the notion that successful users are more active. The resulting rise in polarization is therefore an unintended consequence of individual user behavior. To be sure, we do not argue that there are no actors who do seek to boost opinion polarization on real online social networks. However, our findings suggest that the widely observed polarization of political opinions may also result from emergent dynamics that act in parallel with psychological mechanisms generating polarization (Liu et al. 2022; Keijzer et al. 2018). We stress that this is an important insight, as it may imply vastly different interventions designed to counter opinion polarization.

1.6 Emergent phenomena tend to conflict with intuition (Liu et al. 2022). As a consequence, we applied a rigorous formal method to demonstrate the validity of our reasoning. In particular, we extended Axelrod's seminal agent-based model of cultural dissemination to include success-driven activity (Axelrod 1997). We demonstrate the effects of success-driven activity on polarization under the same conditions that Axelrod studied since they are very well understood and allow us to identify the mechanism responsible for the emergent effect. Next, we conducted Monte-Carlo simulation experiments to identify conditions under which the effect of success-driven activity on polarization is stronger. In the concluding section, we discuss implications for research on the causes of opinion polarization, approaches to modeling online social networks, and attempts to regulate online social networks.

Literature

2.1 Success-driven activity, the core innovation of our model, follows from classical theories of reinforcement learning, one of the most fundamental paradigms explaining human behavior (Homans 1974; Bandura & Walters 1977; Sutton & Barto 2018; Ruff & Fehr 2014; Lockwood & Klein-Flügge 2021). Whilst this paradigm stems from a line of research to explain non-human behavior (see the Skinner Box and classical conditioning) recent work in computational modeling has shown that it can be leveraged for social media behavior (Das & Lavoie 2014;

Lindström et al. 2021). Lindström et al. (2021), for instance, investigated the optimization of rewards, which are underpinned by dopamine, across various social media platforms, specifically examining the impacts of positive online social feedback. They tested the computational hypotheses of the predictive value of social rewards (such as likes) on social media posting rates and posting latency and found that when the average reward is higher and the posting rate is increased, the average response latency (the time distribution of postings) decreases (Lindström et al. 2021).

2.2 If success leads to higher activity and if increased activity translates into even more success, then a rich-get-richer process can unfold where a few individuals grow very active while many remain comparatively inactive. Empirical research on online social networks actually documented such activity patterns (Riquelme & González-Cantergiani 2016). A seminal study on Twitter users, for instance, found that "20K elite users, comprising less than 0.05% of the user population, attract almost 50% of all attention within Twitter." (Wu et al. 2011).

2.3 Models of social influence constitute one of the main classes of formal models in the social sciences, providing a rigorous method to demonstrate and understand the macro-consequences of micro behavior. These models typically represent individuals, their stances on relevant issues and their social relationships (Friedkin & Johnsen 2011; Flache et al. 2017; Deffuant et al. 2000; Hegselmann & Krause 2002). Next, assumptions are added about how individuals connected by social relationships exert influence on each other's opinions. Analytical methods and simulation are then used to derive often counter-intuitive predictions about collective dynamics (Banisch & Olbrich 2019; Liu et al. 2022; Keijzer & Mäs 2022). In the past decades, the literature on models of opinion dynamics has received much attention and has already made contributions to the debate about polarization on online social networks (Geschke et al. 2019; Chavalarias et al. 2024; Keijzer & Mäs 2022; Baumann et al. 2020).

2.4 One of the most influential contributions to the literature was Axelrod's model of the dissemination of culture (Axelrod 1997). Axelrod's work had lasting impact since it was one of the first to combine social influence with homophily (Carley 1991). In addition, his model was also adopted and elaborated in fields outside the social sciences since it was mathematically very similar to models for instance from statistical physics (Starnini et al. 2025; Castellano et al. 2009). He defined culture as the set of individuals' characteristics that are open to social influence, including beliefs, attitudes, and behaviors. Accordingly, social networks on the Internet fall within the scope of this model. Axelrod was one of the first to combine the process of social influence with homophily, the notion that similarity breeds interaction. When acting in tandem, homophily and social influence can give rise to opinion polarization, because social influence causes local convergence in networks. When local segments of a network converge, however, differences between segments can grow, leading to global divergence or polarization. This process of local convergence and global divergence is a typical example of an emergent process. Homophily and social influence generate polarization even though individuals do not seek to generate this pattern or seek to intensify differences to other individuals. Polarization, in other words, is an unintended consequence of individual behavior.

2.5 Agent-based models of complex systems in general and models of social influence in particular typically assume homogeneous agent activity, implementing so-called "updating schedules" where all agents have the same probability to be activated at a given point in time. However, various authors criticized that the literature as a whole failed to address the sometimes critical effects of this modeling decision (Aracena et al. 2009; Weimer et al. 2019; Caron-Lormier et al. 2008; Bandini et al. 2012; Thaler & Siebers 2019).

2.6 Some models assume a *synchronous updating* schedule. That is, agents are updated in parallel and base their updating decisions on other's past characteristics. This approach is often used because of purely technical reasons, as it can simplify mathematical analyses. However, synchronous updating can also be very plausible. Individuals conditioning their decision to participate in political protests, for instance, update their beliefs about the size of a protest whenever there is a demonstration and, thus, at the same time. *Asynchronous updating*, in contrast, assumes that agents update in some order and that agents base their decisions on the most recent state of their interaction partners. Asynchronous updating has been implemented in various ways that may lead to very different dynamics (Huberman & Glance 1993; Banisch & Araújo 2012). The most standard approach is certainly to update agents in a random order, which also implies homogeneity in agent activity. Axelrod's model also falls into this category. Axtell (2000) compared an activation schedule where all agents are active exactly once every point in time with a schedule where each agent has a random number of activations per period with average 1. Strikingly, he showed that the two schedules can generate different dynamics.

2.7 While we are not aware of updating schedules implementing success-driven activity, there are a few contributions to the literature that dropped the assumption of homogeneous activity. Baumann et al. (2020), for instance, implemented that every agents is described by an activity propensity, an agent characteristic that was drawn from a power-law distribution and that remained unchanged over time. In our model, we do not hard-wire a specific activity distribution, but study their emergence. Page (1997) studied *incentive-based updating*,

implementing that agents are updated with an increased probability when this results in an bigger rise in their personal utility. For instance, when an agent i seeks to coordinate behavior with surrounding agents and happens to be surrounded by agents choosing different behavior, then agent i has an increased desire to update and is, therefore, updated first. Like success-driven activity, incentive-based updating is based on assumptions about individuals' motivation to become active. The central difference between the two approaches, however, is that an agent who had a very high activity based on incentive-based updating, likely adjusts behavior and, thus reduces activity. Success-driven activity, in contrast, can generate a rich-get-richer dynamic, because activity can foster success and more activity.

2.8 An alternative approach is to elaborate the *Poisson process*, where every agent has a personal activation rate (Galante et al. 2023; Alizadeh & Cioffi-Revilla 2015). This rate can be fixed or depend on a variety of agent characteristics. Alizadeh & Cioffi-Revilla (2015), for instance, assumed that agents holding an extreme opinion are updated with an increased rate. Technically speaking, this approach is certainly the most similar one to our work.

2.9 There is a class of models of opinion dynamics on online social networks that included assumptions about learning behavior (Banisch & Olbrich 2019; Gaisbauer et al. 2020; Banisch et al. 2024). Gaisbauer et al. (2020) analyzed a model where agents can choose to express their opinion on a platform or to be silent. This decision is guided by reward-based learning. The model has been extended to multiple platforms in Banisch et al. (2024). In these models, agents were described by a fixed opinion and experienced platforms more rewarding when they encountered users with similar views. They show that a population can experience online segregation in that agents with similar opinions self-select into the same platforms. Our work assumes that there is only a single platform and that opinions are not fixed. We show that polarization can emerge even on a single platform.

Material and Methods

3.1 We demonstrate that success-driven activity contributes to the emergence of polarization, elaborating Axelrod's model of cultural dissemination. We built on Axelrod's model of cultural dissemination and not on an alternative model of social influence (Flache et al. 2017), for three reasons. First, Axelrod's model is certainly one of the most studied and best understood models in the literature (Huckfeldt et al. 2004), which makes it a useful point of comparison. Second, Axelrod's core model ingredients, selection based on homophily, and assimilative social influence are uncontroversial assumptions that have been supported by empirical research (McPherson et al. 2001; Flache et al. 2016; Mäs & Flache 2013; Keijzer et al. 2024). Third, Axelrod's framework allows us to implement success-driven activity in a very straight-forward way. While Axelrod did not implement success-driven activity, his model allows us to distinguish between successful and unsuccessful social influence events. Whenever an agent managed to transmit a cultural trait to another agent, we count this as a success and measure past success as the count of previously transmitted traits. Accordingly, we were able to implement success-driven activity with minimal adjustments to the original model.

3.2 We compare dynamics in Axelrod's classical model assuming homogeneous agent activity with a version of the same model that assumes success-driven activity, keeping all other model aspects the same. We show that polarization is consistently stronger when success-driven activity is added. In fact, even under conditions where the original model virtually always generated monoculture, the model version with success-driven activity still generates strong polarization.

The model

3.3 Axelrod's model of the dissemination of culture is a typical agent-based model, explicitly representing all N members of a population using a cellular automaton. Axelrod described every agent i by a set of F features, representing cultural characteristics that are open to social influence. On each feature f, agents can adopt one of Q discrete traits. Features are measured on a nominal scale. For instance, a feature might represent individuals' stances on the origin of SARS-CoV-2 and traits represent alternative theories.

3.4 In every time step of Axelrod's original model, an agent i and one of his neighbors j is randomly picked for update, implementing asynchronous updating. Whether or not this update is actually executed, however, depends on the cultural overlap between the two agents. That is, Axelrod included the principle of homophily by assuming that the probability of interaction between i and j equals the share of features where the two have

adopted the same trait. When the two agents, for instance, share traits on two out of 10 features, they will interact with a probability of twenty percent. If they do interact, then i will adopt one of j's traits that i had not adopted before.

3.5 We deviated from Axelrod's framework in three ways. First, we moved from a cellular automaton framework to a network setting in order to be able to exploit the flexibility of network-structure manipulation. In our model, agents are integrated in a network, where connected agents can exert social influence on each others' cultural features. In particular, we implemented small-world topologies, assuming symmetric circle networks where agents are arranged in a circle and connected to their closest neighbors to the left and the right (Watts & Strogatz 1998). This results in networks with a very high network clustering — that is, agents who are connected to the same agent are also likely to be connected to each other. Although this setup differs from the originally used cellular automaton, it preserves the principle of local clustering. Previous research has shown that Axelrod's model exhibits similar dynamics on networks characterized by local clustering as it does on cellular grids (Klemm et al. 2003b). Randomly rewiring network links, decrease network clustering and introduces shortcuts decreasing the average path length.

3.6 Second, to be able to study success-driven activity we swapped the role of sender and receiver in Axelrod's model. Axelrod implemented that an agent i and one of his neighbors j are picked for update and, if influence is executed, i *receives* a trait from j. That is, Axelrod chose the receiver of influence first. However, to be able to study that some agents are more active in sending content, we had to swap the sender-receiver selection. When receivers are selected first, the selected sender is not the most successful agent in the population, but merely the most successful among the receiver's neighbors, which is not what we are studying here. Accordingly, we now select the sender i first who then transmits a trait to receiver j. This adjustment ensures that success is evaluated on a population-wide basis rather than within local neighborhoods.

3.7 While we cannot exclude that swapping senders and receivers changes dynamics in Axelrod's model, we are confident that our findings are unaffected by this change compared to the original model. To keep our result comparable to Axelrod's work, we restricted ourselves to social network topologies where all agents have the same degree. When all agents have the same number of network connections, it does not matter whether the sender or the receiver of a content is picked first: all agents always have the same probability of being selected as a sender or a receiver.

3.8 Third, we use Equation 1 to select the sender i from the population of agents, implementing activity heterogeneity and success-driven activity. Equation 1 implements a so-called "roulette wheel selection" where the probability that an agent i is selected for sending a trait to one of his network neighbors is proportional to the relative past success $s_{i,t}$ of the agent. The success motivation m controls the degree to which success $s_{i,t}$ motivates user activity. When $m = 0$, success does not influence activity, which implements Axelrod's model where all agents have a probability of $1/N$ to be selected for update. For values above zero, in contrast, past success $s_{i,t}$ plays an increasing role, allowing us to move from Axelrod's original model to a version of the same model with success-driven activity by tweaking parameter m.

$$P(a_{i,t} = 1) = \frac{{s_{i,t}}^m}{\sum_{j=1}^{N} {s_{j,t}}^m} \tag{1}$$

3.9 Table 1 summarizes how Axelrod implemented interaction and how we deviated from his work.

3.10 At the outset of all simulations, we set success to $s_{i,1} = 1$ for all agents i. After every simulation event where a sender i successfully influenced a network contact j, $s_{i,t}$ is raised by one. That is, $s_{i,t}$ is a count of i's past successful influence events.

3.11 The remainder of the model is adopted from Axelrod. Once the sender i is selected, one of i's neighbors is randomly picked to act as the receiver. The simulation program — in our case NetLogo 6.4.0 — then calculates the overlap between i and j. Agent j is influenced by i with a probability equal to the share of features where the two agents hold the same trait.

3.12 The inclusion of success-driven activity does not affect the equilibria of Axelrod's model. Dynamics and thus simulation runs either end in perfect consensus where all agents have adopted the same traits on all features, which Axelrod referred to as "monoculture". This is stable because further interaction events cannot result in trait changes. Alternatively, dynamics (meaning simulation runs) can end the population has fallen apart into multiple segments where agents within a segment are identical but connected segments differ on all feature dimensions. This pattern is an equilibrium because within-segment interaction cannot lead to feature changes. Furthermore, agents belonging to different segments are maximally different and, according to the homophily principle, have an interaction probability of zero.

Original Axelrod Model	Success-driven Activity
(1) Select an agent i	
Pick an agent with probability $1/N$	Pick an agent with probability proportional to success $s^m_{i,t}$
(2) Select neighbor	
Select a random neighbor j as content sender	Select a random neighbor j as content receiver
(3) Homophily	
With probability equal to similarity between i and j move to Step (4), otherwise move to Step (1)	With probability equal to similarity between i and j move to Steps (4) and (5), otherwise move to Step (1)
(4) Social influence	
Pick a feature on which i and j differ and let i adopt j's trait	Pick a feature on which i and j differ and let j adopt i's trait
(5) Update agent's success count s_i	
—	Increase i's count of successful influence events $s_{i,t}$

Table 1: Core assumptions of Axelrod's original model and its version with success-driven activity.

3.13 While Axelrod's original model already has several stochastic elements (e.g. in Steps 1 and 2 of Table 1), its predictions have been shown to change dramatically when, in addition, random perturbations are introduced (Macy & Tsvetkova 2015; Mäs & Helbing 2020). In particular, Klemm et al. (2003a) included that at the end of every time step, with a probability r a randomly selected agent adopts a random trait on a randomly picked feature. Most of our analyses do follow Axelrod's work and exclude random perturbations ($r = 0$). However, to examine the robustness of our findings on success-driven activity under random perturbations, we implemented perturbations following the same approach as Klemm et al. (2003a).

Computational inefficiencies

3.14 Axelrod's original model is computationally inefficient in that it can select in Step 1 of Table 1 an agent who is connected only to perfectly similar or perfectly dissimilar network neighbors. As a consequence, the resulting interaction will never result in an influence event and will not have any effect on future dynamics. While this inefficiency is usually not a big issue in the original model, it can turn into a serious waste of computational resources under success-driven activity, because here successful agents are selected with an increased probability. What is more, these agents are successful because they have influenced many neighbors and are, thus, likely to have only perfectly similar neighbors. In consequence, the simulation systematically selects agents as senders who will not change model dynamics (receiver j does not adopt a trait from sender i). As a result, the model dynamics remain unchanged for an extended period, requiring significantly more time steps—and consequently greater computational resources—before any trait adoption occurs. This inefficiency becomes particularly problematic when simulations are run until equilibrium and experiments involve numerous replications, making it essential to minimize unnecessary computational overhead.

3.15 To deal with this computational inefficiency, we implemented that only agents who have at least one neighbor that is neither maximally similar or maximally dissimilar can be selected in Step 1. Our findings are unaffected by this decision. As a result, the number of simulation events needed to reach an equilibrium state is hard to interpret across different parameter combinations. Accordingly, we do not address this in the present paper. Note that we did not address other inefficiencies. Also in Step 2, for instance, it is possible that a neighbor j is selected who is already perfectly similar or perfectly dissimilar to i. Like Axelrod, we did not exclude these

events, since it is unclear whether model dynamics are affected, when neighbors who are still open to influence are selected with an increased probability.

3.16 Success-driven activity can, however, generate another highly inefficient livelock dynamic. Imagine, as a very simple example, a population of three agents with a line network where Agent B is connected to Agents A and C. There are no additional links. Next, assume that A and C are highly successful and that B is very unsuccessful. Finally, assume that A and C agree on all but one cultural features. Because of their success, A and C will be selected as senders with a very high probability, always communicating a trait to Agent B. As a consequence, Agent B will always swap between being identical to Agent A or being identical to Agent C. This dynamic can be interrupted only if at some moment Agent B influences either A or C. This, however, becomes increasingly unlikely because in the swapping phase, A and C grow increasingly successful. As a consequence, the dynamics is deeper and deeper trapped in a state out of equilibrium. The same dynamic can emerge in more complex networks, when two successful agents who agree on all but one feature share at least one contact. We refrained from implementing assumptions that avoid such dynamic, because it emerges from success-driven activity, the focus of our work. What is more, we observed these dynamics only for values of success motivation m above 1.5, and decided to focus our study on less extreme values of m.

Outcome measures

3.17 We used four different outcome measures to report our findings. First, inspired by the seminal work of Axelrod, we report the number of different cultures in the population. To be more precise, we counted the number S of distinct feature vectors in the population in equilibrium. This count, obviously, adopts its minimal value of one when the population has reached a state of consensus where all agents have the exact same set of traits. Axelrod referred to this state as "monoculture". Higher values indicate that the population split up into multiple internally homogeneous but mutually distinct subgroups. The number of cultural clusters alone, however, fails to inform about the subgroup structure. When there are two cultural clusters, for instance, the population may have split into two equally large subgroups or into one huge subgroup and a single isolate. Accordingly, we also studied a second outcome measure that is often used in the literature: the size of the biggest cultural subgroup S_{max} (Castellano et al. 2000; Keijzer et al. 2018).

3.18 Third, we quantified the degree of *polarization* P_t in the population (Flache & Mäs 2008), calculating the variance of the pairwise cultural dissimilarities between all $N \times N$ pairs of agents. Dissimilarity was measured as two times the share of features where two agents hold different traits. Multiplication by two linearly transforms the dissimilarities to a scale of width two, which implies that the maximal value of the variance of the dissimilarities is one. When the population is characterized by perfect consensus, then this measure has a value of zero. Higher values indicate that the population consists of agent pairs with either maximal or minimal cultural dissimilarity. The maximal value of $P = 1$ is reached when the population consists of two equally large internally homogeneous but mutually different clusters.

3.19 We had to resort to the polarization measure P_t for our model analyses with random perturbations, since perturbations can distort the validity of the other two outcome measures. The problem is that the other two measures consider an agent part of a cluster only when it shares all traits with the other cluster members. Random perturbations, thus, exclude agents from clusters even when they affect only a single feature. P_t is less affected since it is based on a continuous dissimilarity measure where perturbations increase dissimilarities only slightly. However, also P_t comes with disadvantages. This measure assumes high values also when the population consists of one large homogeneous cluster and a set of isolated agents with random traits on all features. The model with success-driven activity can generate such a pattern when random perturbations are included. When successful and therefore highly active agents influence their neighbors, they can form a large cluster. Furthermore, there can be another section of the network where all agents have a very low activity and, thus, hardly converge. When they are, in addition, exposed to random perturbations, each of them will adopt a different random feature vector. As a result, the population will be described by a large number of high pairwise dissimilarities (from the large homogeneous cluster) and a large number of low pairwise dissimilarities (from the inactive agents). This translates into a high value of P_t, although the population actually consists of only one cluster and a number of isolated agents. To prevent that characteristic of P_t affects our findings, we carefully selected only areas of the parameter space, where we never observed these misleading patterns.

3.20 Finally, we included a measure to describe the distribution of success $s_{i,t}$ in the population. In particular, we used the Gini-coefficient of the $s_{i,t}$ distribution to quantify the degree of success inequality. The Gini-coefficient adopts its minimal value of zero when all N agents have the exact same success $s_{i,t}$. Its maximum of one is reached when a single agent has all success and the remainder has a success of zero.

Results

4.1 We analyzed the described model in various ways. In the first subsection, we describe two typical simulation runs with and without success-driven activity to illustrate their differences in terms of model dynamics. Next, we report the results of a series of simulation experiments replicating core findings from the literature with the original model and our extension. In particular, we replicated the analyses from Axelrod (1997), research on the effects of network structure (Klemm et al. 2003b), and seminal modeling work on the effects of random perturbations (Klemm et al. 2003a). For all analyses, we show stronger polarization under success-driven activity.

Ideal-typical simulation runs

4.2 Figure 1 describes the dynamics observed in a typical simulation run with Axelrod's original model ($m = 0$) and a typical simulation run with our extension ($m = 1$). In both runs, we studied populations of 100 agents, each described by 5 features with 15 possible traits per feature. We assumed a perfectly symmetric circle network with a degree of 24 and no rewiring. To initialize the runs, we assigned random traits on all features to all agents.

4.3 The two orange lines inform about the run with Axelrod's model ($m = 0$) and the blue lines are based on the run with $m = 1$. The solid lines show the trajectories of the populations' levels of polarization. The dotted lines describes the dynamics of success inequality measured with the Gini coefficient. In Appendix A, we provide additional illustrations of the dynamics, using network visualization.

4.4 Despite assuming the exact same parameter values in both runs, the dynamics and the equilibria reached differ vastly. In the run with the classical Axelrod model, polarization remains low during the dynamics and never exceeded a P_t of .33. Eventually, the population reached a state of monoculture. Counter to our expectations, we observed that success inequality was rising in the early stages of run, reaching a maximal value of .29 (see orange dotted line). We argue that this dynamic unfolds because initially agents had random traits, which implies that most pairs of connected agents were very different. Those agents, however, who happened to be similar to their neighbors also interacted with an increased probability and, thus, built up a considerable success $s_{i,t}$. However, under $m = 0$, this success does not increase agent's activity, such that no rich-get-richer dynamic can unfold. Accordingly, also success inequality decreased when the initial differences were overcome. In equilibrium, the Gini coefficient was only .04.

4.5 With success-driven activity ($m = 1$), dynamics looked very different and even ended in a state of polarization with two maximally different subgroups and three isolated agents. Polarization P_t amounted to a nearly maximal value of .99. Initially, the level of success inequality rises very quickly, reaching a maximal Gini-coefficient of .62 and then remains constant at a high level. In equilibrium, we measured a Gini coefficient of .56. The maximal success in the population was $s_{i,27149} = 217$. However, 50 percent of the population had a success count lower than 23. That is, half of the population achieved less than 10% of the success count of the most successful agent.

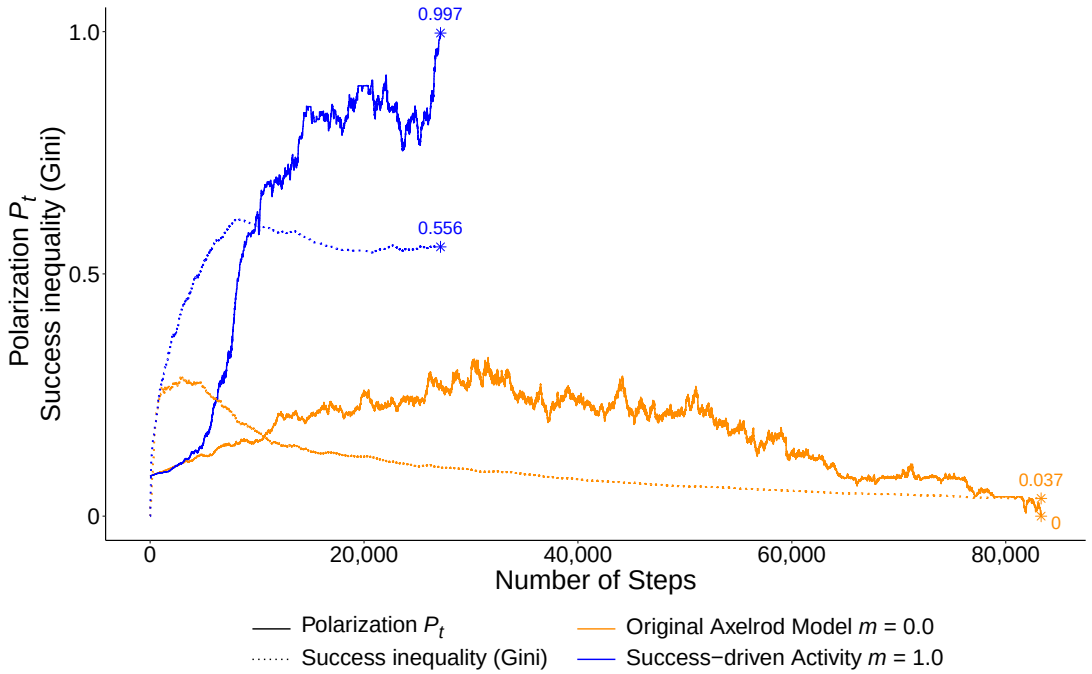

Figure 1: Ideal-typical simulation runs, one under success motivation $m = 0$ (orange lines) and one under $m = 1$ (blue lines). We assumed five features and 15 traits per feature. The population size was $N = 100$ with circle network and a degree of 24. The asterisks mark the respective equilibrium.

Mechanism

4.6 Success-driven activity in the Axelrod model generates a rich-get-richer process: successful individuals are more likely to be activated and subsequently have an increased chance to influence others growing even more successful. The strength of this effect is governed by m.

4.7 Similar dynamics are also known in the social sciences (Broido & Clauset 2019). Most prominently, Merton described what he called the "Matthew effect" in science (Merton 1968). Likewise, network formation mechanisms like preferential attachment (Barabási & Albert 1999) and cumulative advantage feedback processes in economics (Piketty 2014) generate similar inequalities.

4.8 Figure 2 compares the distribution of agent success $s_{i,t}$ for different values of m. Each distributions is based on 10 independent runs of the model with $N = 1000$ agents. Each simulation is run for $T = 3000$ macro steps, where a macro step consists of $N = 1,000$ iterations of Axelrod's model (see Table 1). The remaining parameters and the network topology were adopted from the previous section.

4.9 For $m = 0.0$ and $m = 0.5$, the success distributions resemble a normal distribution with an average value of $\bar{s}_{i,t} = 460$ successes. The variance of $s_{i,t}$ grows as m is increased from $m = 0.0$ to $m = 0.5$. In the case of $m = 0.75$, the distribution becomes broad and is no longer described by a normal distribution. Some agents accumulate more than 600 successful events while others remain unsuccessful. Under $m = 1.0$, we observe a skewed success distribution, in which single individuals accumulate more than 1,000 success events while most of the population has fewer than 10 successful interactions. In Appendix B, we show more detailed analyses that reveal that the distribution of success is highly skewed, but does not resemble a power law.

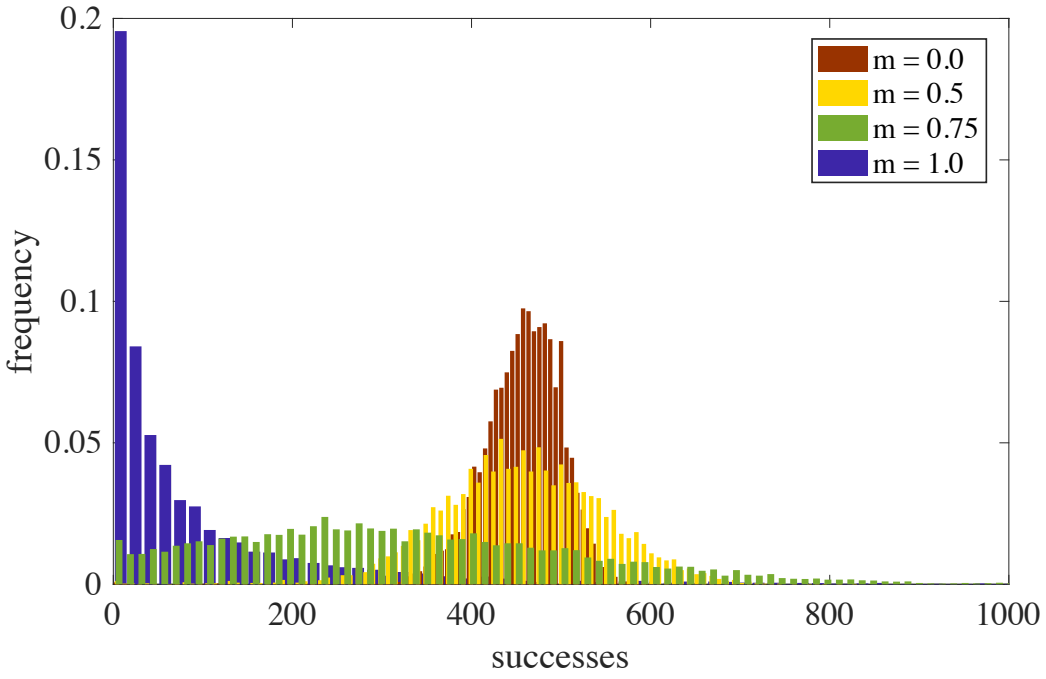

Figure 2: Distributions of individual success $s_{i,t}$ in a setting with $N = 1,000$ agents, $F = 5$ features with $Q = 15$ traits, and a circle network with degree 24. Comparison of the distributions that emerge from different values of m after $T = 3,000$ macro steps ($N \times T = 3,000,000$ interaction events). All data shown in the figure is based on 10 runs per parameter.

Simulation experiments

Number of features and number of traits

4.10 One of the most surprising findings presented by Axelrod is that the number F of features and the number Q of traits per feature have opposite effects on polarization. Axelrod found that increasing the number of features typically generates monoculture, while increasing the number of traits generates more polarization. More features lead to monoculture, because every additional cultural dimension increases the chances that two agents share a trait and, thus, exert influence on each other. Conversely, rising the number of traits per feature generates polarization, because additional traits decrease the probability that two agents happen to be similar, making influence less likely.

4.11 We replicated Axelrod's anaylses with his original model ($m = 0$), a version with a mild form of success-driven activity ($m = .5$), and a version with strong success-driven activity ($m = 1$). Like Axelrod, we varied the number of traits (Q) and features (F) between 5 and 15 in increments of 5 and conducted 100 independent simulation runs for each of the 27 experimental treatments. Across all treatments, we assumed populations of $N = 100$ agents interacting in a circle network with a homogeneous degree of 16 and no rewiring. There were no random perturbations.

4.12 Results are visualized in Figure 3 with scatter plots and box plots of the number S of clusters counted when the respective simulation run reached a state of equilibrium. The black dots linked with solid lines indicate averages. The three subpanels of the left-hand side show the results for Axelrod's original model in orange ($m = 0$). The light blue panels in the center show findings under mild success-driven activity ($m = .5$) and the dark blue panels represent the treatments under ($m = 1$), using the same color coding as in the remainder of the paper.

4.13 We replicted Axelrod's findings that populations virtually always ended up in monoculture under $m = 0$. Only when the number of features is low and the number of traits per feature was high, more than one cluster obtained. In Appendix C, Figure 9 shows that even under this condition most runs ended with a single, huge cluster and a number of very small clusters or isolated agents.

4.14 Under success-driven activity, we observed the same effects of the number of features and traits. However, the Figure 3 also shows a substantially higher degree of polarization even in experimental treatments where the classical Axelrod model always generates monoculture. This supports the main claim of the paper that success-driven activity fosters polarization.

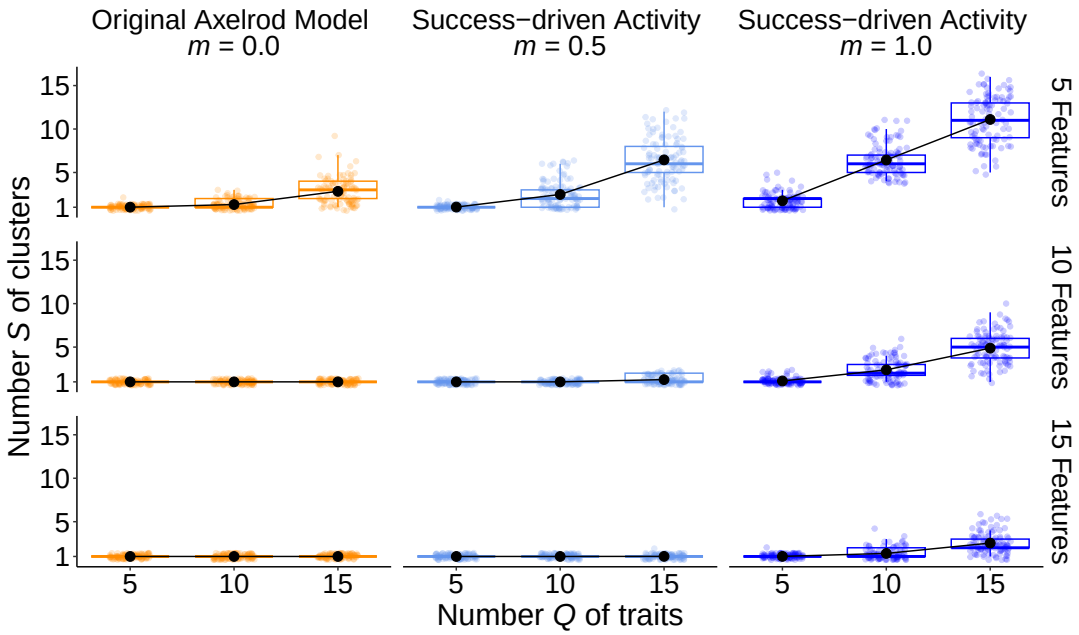

Figure 3: Effect of the number F of features and the number Q of traits on the number S of clusters in equilibrium. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$; a degree of 16 for all agents and no rewiring. All simulations were executed until equilibrium. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 0.5$; $m = 1$) is shown in different shades of blue. Effects on the size of the biggest cluster and success inequality are reported in Appendix C.

Neighborhood size

4.15 Axelrod also studied effects of the size of agents' neighborhoods (degree), showing that bigger neighborhoods result in less polarization because a larger range of interaction fosters convergence. To replicate this finding, we again studied populations of $N = 100$ agents integrated in a circle network without rewiring. We implemented a condition where polarization should be an expected outcome according to Axelrod's model, assuming 5 features and 15 trait per feature. The size of agents' neighborhoods was varied between 8 and 96 in increments of 8. In addition, we included a treatment with 99 neighbors, which implements a complete network where every agent is connected to all $N - 1 = 99$ remaining agents. Again, we studied three degrees of success-driven activity ($m = 0$; $m = .5$; $m = 1$) and conducted 100 independent replications per treatment.

4.16 The left-hand side panel of Figure 4 shows that we could replicate Axelrod's observation that large neighborhoods translate into lower polarization. According to the original model, in fact, the vast majority of the runs ended in monoculture when the degree exceeded 24. The center panel of the figure shows slightly more polarization under mild success-driven activity ($m = .5$). However, under $m = 1$ the differences are pronounced. In fact, even in complete networks, the model generates strong polarization in most cases. This is a remarkable finding, since Axelrod's model seems to be unable to explain polarization under this condition.

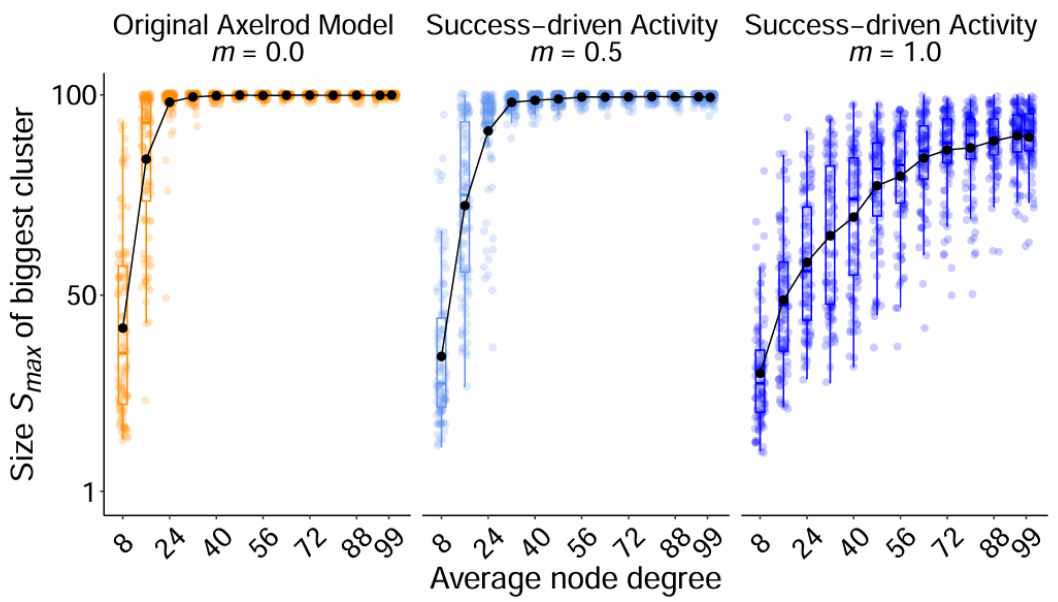

Figure 4: Effect of the size of agents' network neighborhoods, on the size S_{max} of the biggest cultural cluster in equilibrium. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$; zero MS-rewiring Iterations, 5 features, and 15 traits. All simulations were executed until equilibrium. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 0.5$; $m = 1$) is shown in different shades of blue. Effects on the number S of clusters and success inequality are reported in Appendix D.

Network clustering

4.17 Axelrod did not study effects of structural aspects of the interaction network other than the degree. However, previous work has shown that network structure affects model predictions (Keijzer et al. 2018). In particular, network clustering, the degree to which nodes create closed triangles, tends to foster polarization, since it fosters local convergence, which generates polarization.

4.18 To be able to study structural aspects, we applied the Maslov-Sneppen rewiring algorithm (Maslov & Sneppen 2002). In every rewiring iteration, this algorithm drops two randomly picked network links (e.g. $A \rightarrow B$ and $C \rightarrow D$) and reconnects the four affected nodes (e.g. $A \rightarrow C$ and $B \rightarrow D$) in a new way. For the purpose of our work, the strength of this algorithm is that it manipulates network clustering while keeping unchanged agents' degree. A small number of rewiring iterations generates so-called "small worlds", structures with high clustering but short chains of network connections between any two agents. Very high numbers of rewiring iterations lead to network structures with very low clustering.

4.19 In a simulation experiment, we manipulated the degree of network clustering by varying the number of Maslov-Sneppen rewiring iterations (Maslov & Sneppen 2002). Starting from a circle network where all agents have links to the twelve closest neighbors to the left and the twelve closest neighbors to the right, we studied four experimental treatments. First, without any rewiring, clustering was maximal. Second, we added a treatment with 10 and one with 100 rewiring iterations. Finally, we studied a treatment with 100,000 rewiring iterations, which completely breaks up the clustering and generates a random network. As before, we studied populations of 100 agents described by five features and 15 traits per feature.

4.20 Figure 5 reveals that under success-driven activity network clustering promotes polarization, replicating findings derived from Axelrod's original work. Without success-driven activity, most runs ended in monoculture or a state with one very large group and a few isolated agents. Under $m = 1$, in contrast, we observed much more polarization even when 100,000 rewiring iterations were performed. That is, even when the networks displayed minimal network clustering, we observed very strong polarization dynamics. Nevertheless, the panel on the right-hand side depicts a strong effect of network clustering under $m = 1$: network clustering leads to stronger polarization.

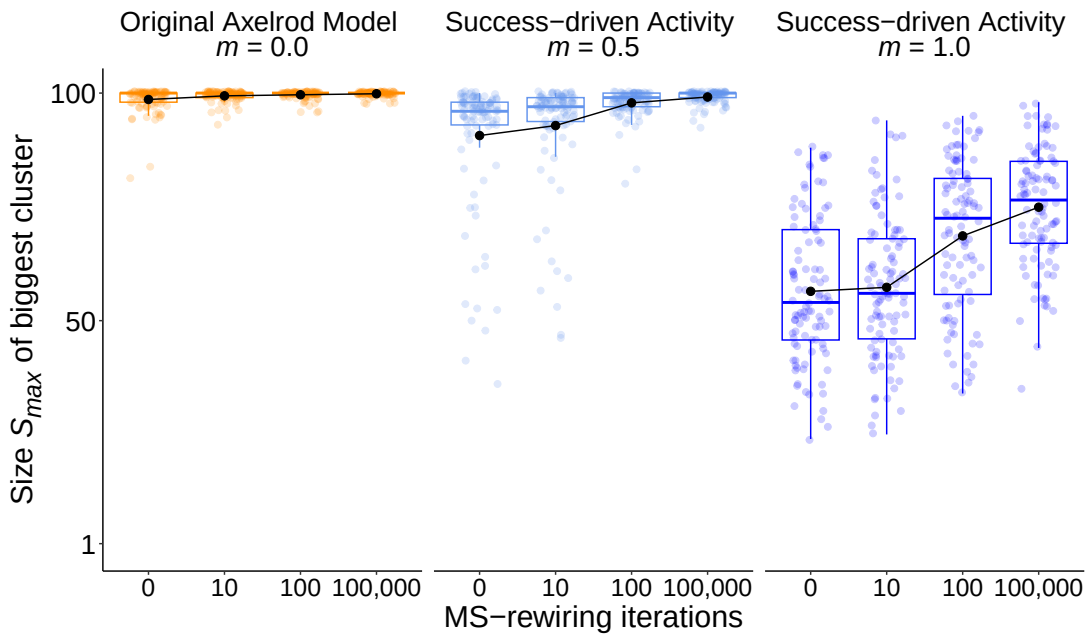

Figure 5: Effect of network clustering on the size of the biggest cluster in equilibrium. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$, number F of features = 5, number Q of traits = 15; and a degree of 24 for all agents. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 0.5$; $m = 1$) is shown in different shades of blue. Effects on the number S of clusters and success inequality are reported in Appendix E.

Random Perturbations

4.21 Klemm et al. (2003a) show, that model predictions change dramatically under random perturbations. When perturbations are included but remain very unlikely, the model always generates cultural consensus. That is, with only a few perturbations, the model loses its ability to explain persistent cultural diversity. Diversity breaks down, because perturbations can induce similarity between otherwise maximally different cultural clusters, which reactivates influence between these clusters and fosters cultural convergence. Once the population has coordinated on a set of shared traits, further perturbations can generate collective shifts. However, it is very unlikely that multiple distinct clusters will form. Second, when there are many perturbations, the system is in an unordered phase, because influence is too weak to create local convergence. In other words, perturbations occur so frequently that locally forming clusters fall apart due to further perturbations. Third, there can be an intermediate phase where perturbations induce new cultural traits in otherwise homogeneous parts of the network. These traits are locally spreading, creating new cultural regions. At the same time, these regions are also merging again due to the perturbations. However, the constant fusion and fission of cultural clusters generates a dynamic equilibrium with many changes at the micro-level but a high collective fragmentation.

4.22 With the findings of Klemm et al. (2003a) in mind, we sought to test whether the predictions of the model with success-driven activity are also dependent on the inclusion of random perturbations. To this end, we added random perturbations in the same way as Klemm et al. (2003a) (see Paragraph 3.14) and varied parameter r between 0.0025 and 0.1 in increments of .0025. In addition, we enforced additional perturbations whenever the system reached a state of equilibrium. To keep the number of time steps comparable, we then advanced the count of time steps by a random number drawn from the geometric distribution with parameter r. We studied two treatments of success-activity: $m = 0$ and $m = 1$ and conducted 100 independent runs per treatment as in the other experiments.

4.23 The initialization of cultural traits was more complicated than in the other experiments. Replicating Klemm et al. (2003a), we did not initialize agents with random traits, but departed from a state of consensus where all agents held the same feature vector. Next, to avoid starting from perfect equilibrium, we initially inserted a first deviation, assigning to one feature of one agent a trait different from the remainder of the population. With the original Axelrod-model ($m = 0$), this is straightforward. Under $m = 1$, however, an initial consensus generates distributions of agent success $s_{i,t}$ that are very different from those observed in runs starting from a random setup. That is, when all agents agree, then there will only be successful interaction events after an

agent happens to have experienced a random perturbation. This agent will then communicate the new trait to its neighbors, becoming more successful and also allowing these neighbors to grow more successful. The result is that a connected cluster of successful agents forms, a distribution of success $s_{i,t}$ in the network that is very unlikely to emerge when dynamics depart from a random feature setup. To avoid that our findings are affected by such a biased success distribution, we started the simulations under $m = 1$ with random feature vectors and allowed the runs to develop a typical success distribution which usually emerges after not more than 50,000 iterations. To be on the safe side, we enforced monoculture at iteration 100,000, assigning to all agents the same feature vector while keeping individual success values $s_{i,t}$ unchanged. Again, we included that one agent differed on one feature from the rest of the population.

4.24 In Figure 6, we show a scatter plot of the polarization measured after 250,000 ticks for each simulation run. The filled dots linked with solid lines show the average polarization for each experimental treatment. In addition, we also stored the average polarization during the last 50,000 iterations of every run and found that the average of these averages was very similar to the average polarization measured at the end of the runs. This shows that we ran simulations long enough to be able to interpret the findings. The dashed horizontal line indicates the expected level of polarization P_t when random traits are assigned to all features, for comparison.

4.25 The orange markers and the respective line shows that we were able to replicate the findings of Klemm et al. (2003a). When a very high perturbation rate is assumed, dynamics lead to an unordered state where trait distributions are very similar to randomness. However, there is an intermediate level of randomness where perturbations generate a continuing fusion and fission of cultural clusters that translates into weak forms of polarization. When success-driven activity is assumed, we find the exact same pattern but polarization was much stronger. In fact, runs are described by long phases of strong polarization, as the blue markers show.

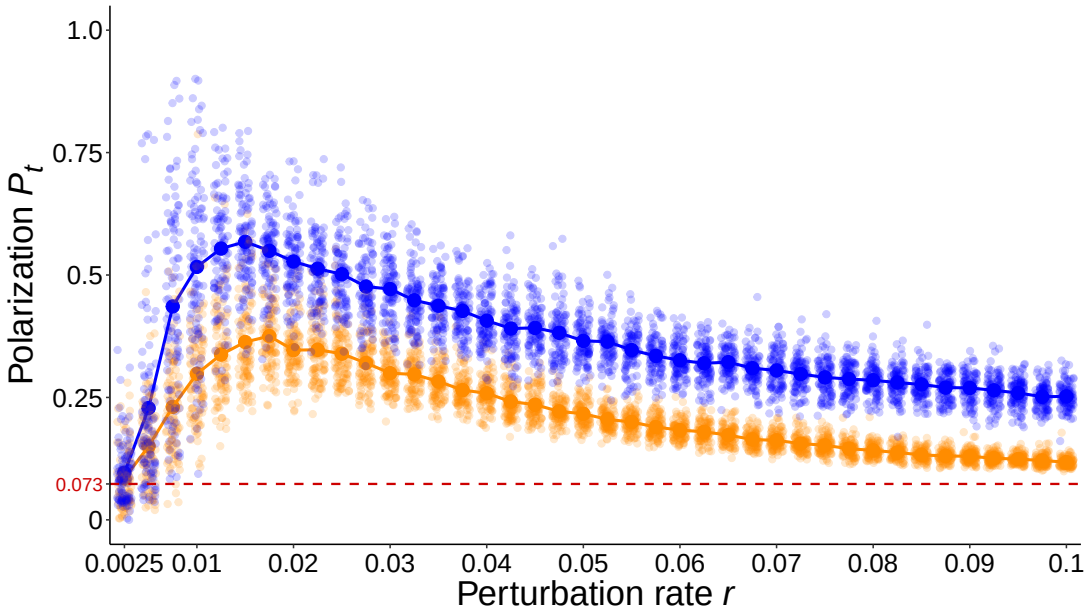

Figure 6: Effect of different perturbation rates on polarization P_t after 250,000 iterations. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$, number F of features = 5, number Q of traits = 20; and a degree of 8 for all agents. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 1$) is shown in blue. Dots linked with lines show treatment averages.

Discussion

5.1 Extreme forms of fragmentation and polarization of political opinions have been identified as a major threat to public discourse and democratic decision making, which has triggered public and academic debate about factors contributing to polarization. The central conclusion of our work is that polarization may be an unintended consequence of individual behavior, rather than polarized elites, similarity-based ranking algorithms creating filter bubbles, and foreign influences or other malicious actors seeking to interfere with public discourse. In particular, we demonstrated that adding the assumption of success-driven activity to a model of social-influence

dynamics results is much stronger polarization than the same model without this aspect would generate. In fact, we observed strong polarization even under conditions where Axelrod's model of cultural polarization typically leads to monoculture (e.g. random networks, complete networks).

5.2 Success-driven activity causes polarization because it activates a self-reinforcing rich-get-richer dynamic where users who happen to experience success early in the process are growing more active and, as a consequence, even more successful. The highly unequal distribution of success paves the ground for polarization, since a few highly successful agents dominate the debate and tear apart the remainder of the population into distinct camps.

5.3 At the outset of the simulations, all agents are characterized by identical levels of success and, consequently, identical activity. Differences in activity due to personality traits, opportunity costs, or other motivational factors are not considered. Therefore, the pronounced inequality in success observed in the model arises solely from the self-reinforcing, rich-get-richer dynamics described above.

5.4 The rich-get-richer dynamic in our model leads to highly skewed distributions of success. However, our results show that success does not follow a power-law distribution. We suggest that a key factor limiting the emergence of a power-law lies in an assumption from the original Axelrod model that we have not yet modified: namely, that senders only transmit traits that are new to the receiver. As a result, agents who have already influenced many others eventually reach a point where they can no longer increase their success, effectively capping the rich-get-richer effect. Relaxing this assumption, and redefining success beyond actual influence, presents a promising direction for future modeling. Such changes could amplify inequalities in both success and activity, potentially intensifying polarization within the population.

5.5 We observed this polarization dynamic in a model that does not assume any form of distancing behavior, opposing media, political elites, or social-media algorithms that foster polarization. In Axelrod's model, polarization emerges because social influence generates local convergence, which leads to global divergence. Success-driven activity fosters local convergence, because some agents grow successful and exert very strong influence in their local neighborhoods. This translates into increased differences between neighborhoods.

5.6 We focused our analysis on the Axelrod model of cultural dissemination due to its widespread recognition and because it provides a straight-forward way to measure success. As Flache et al. (2017) argued, the Axelrod model shares core assumptions (homophily and influence) and models predictions with the Bounded-Confidence model (Deffuant et al. 2000; Hegselmann & Krause 2002), which assumes continuous rather than nominal cultural features. Despite these similarities, future modeling work should replicate our analyses with the various models of continuous opinion dynamics as reviewed by Flache et al. (2017).

5.7 The insight that polarization can be an emergent phenomenon, however, does not imply that it is inevitable or unmanageable. Algorithms sorting content for users can contribute to the polarizing effect of success-driven activity. When content emitted by users who have experienced success is ranked higher in other users' feeds, the same rich-get-richer dynamic can unfold as observed in our analyses. In other words, ranking algorithms that rank content based on the popularity of its source have the potential to foster polarization in the same way success-driven activity does in our model. Conversely, polarizing effects of success-based ranking can be avoided by adjusting these algorithms.

5.8 Our findings also have important implications for developers of digital Twins of Online Social Networks, so-called TWONs (`https://www.twon-project.eu/`). These models are highly realistic representations of online social networks, developed to study and regulate online platforms. Being highly realistic, these models allow one to study dynamics on the real complex system without facing the practical, technical, and ethical restrictions of such empirical research. For this purpose, however, a digital twin needs to be empirically validated with great care. What our analyses revealed is that according to one of the most seminal models from the literature, sufficiently strong success-driven activity contributes to polarization. Accordingly, digital twins need to be calibrated on this aspect, in order to not misrepresent this potentially important determinant of polarization.

5.9 The present paper illustrates that the analysis of toy models can play an important role in the development of digital twins of complex systems, as they allow one to study the effects of a given aspect and to demonstrate its potential to also affect predictions of the digital twin, a model that is much harder to analyze. Hence, toy models point developers of digital twins to aspects that need to be explored with the digital twin and that require empirical validation.

5.10 Future modeling work should explore alternative implementations of success-driven activity. A key strength of our approach is the integration of success-driven activity with a single parameter into the well-established paradigm of Axelrod's model of cultural dissemination. This integration allows for direct comparison of our findings with the extensive literature building on Axelrod's seminal work, as analyses on the effects of noise

illustrate. However, online social networks quantify success not by actual successful interaction events but by success indicators such as the number of likes or shares that content has accumulated. While these indicators may be highly correlated with actual success, it remains an open research question whether different measures of success generate the same polarizing dynamics. These analyses may also allow for the formulation of recommendations for developing ranking algorithms that do not foster polarization.

5.11 While empirical research has provided evidence for success-driven activity already, there remain important empirical research questions. First, it is unclear how the design of online social networks affects users activity. Some platforms, such as YouTube, allow users to monetize their accounts. Once users have reached a certain number of subscribers and watch hours, they can earn money from ads displayed on their videos. Hence, YouTube provides successful users financial incentives to publish more content, which suggest that on YouTube activity may be more success-driven than on other platforms, a hypothesis that deserves to be tested empirically. Another open question is whether or not success decays over time. We assumed that success can only grow, but it appears plausible that recently experienced success has much stronger motivational effect than past experiences.

5.12 Online social networks are a fundamental technological achievement but there is growing concern that foster highly undesired external effects on societies. Our work extends the list of publications demonstrating that the dynamics on online social networks are highly complex and generate dynamics that are hard to anticipate without rigorous analysis (Liu et al. 2022; Keijzer et al. 2018; Keijzer & Mäs 2022; Geschke et al. 2019; Waldrop 2021; Macy et al. 2021). Regulating online social networks likely remains ineffective or may even backfire, if it is not informed by formal methods and empirical validation.

Acknowledgments

The work of Horn, Mäs, Reitenbach, Sartori, and Schwabe is funded by TWON (project number 101095095). The contributions by Banisch and Batzdorfer is funded by SoMe4Dem (project number 101094752). Both projects are funded by the European Union under the Horizon framework (HORIZON-CL2-2022-DEMOCRACY-01-07). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union.

Appendix A: Ideal-Typical Simulation Runs

In the main text, we compared two ideal-typical simulations runs with the original Axelrod model ($m = 0$) and our extension ($m = 1$). Here, we add network graphs describing the dynamics of the two runs. To generate these graphs, we indicate the degree of similarity between the two respective nodes by the width and the shade of grey of the network links. That is, black and thick lines indicate maximal similarity. Furthermore, we used the Fruchterman-Reingold algorithm to place similar and connected nodes closer to each other (Fruchterman & Reingold 1991). The size of the nodes corresponds to their success $s_{i,t}$.

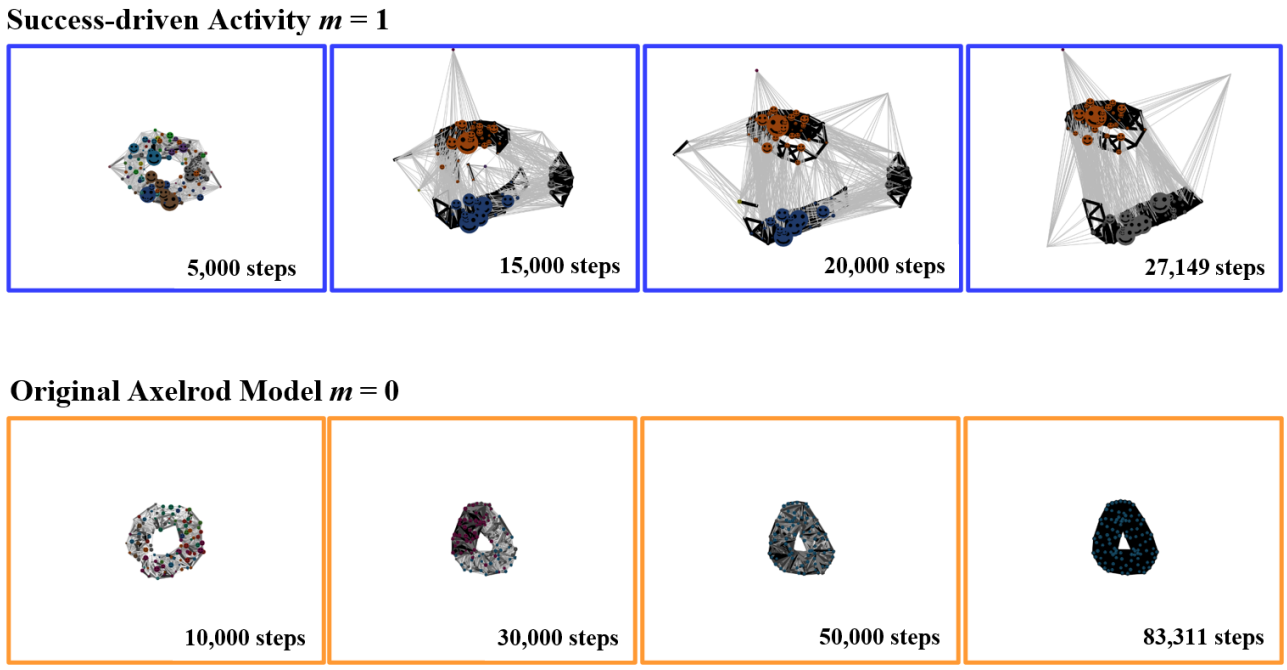

Figure 7: Sequence of network snapshots for the two ideal-typical simulation runs reported in Figure 1.

Appendix B: Mechanism

In Figure 8, we take a dynamical perspective on the model with $m = 1.0$, and show how a skewed success distribution emerges over time. For this purpose, we show the distributions for $T = 25, 50, 100, 200$ and 500 macro steps. The figure shows that a large inequality of success is emerging from the onset of the simulation, and self-sustained by the success-driven activity mechanism.

However, we note that the model with success-driven activity does not generate a power-law distribution of success (Lux & Alfarano 2016; Newman 2005; Clauset et al. 2009; Broido & Clauset 2019). Power laws are interesting because they have been observed in complex systems as diverse as financial markets, the universe, and biological systems and emerge from rich-get-richer dynamics. While the model does generate strong rich-get-richer dynamics and strong success inequality, Figure 3 does not suggest that the success distribution resembles a power-law. We argue that this is because the rich-get-richer dynamics is restricted by the limited size of agents' neighborhoods. Whenever an agent has convinced all neighbors, this agent may be very successful but cannot grow more successful since there are no further agents to influence. This limits the strength of success inequality. In the discussion section, we point to future modeling work studying this phenomenon.

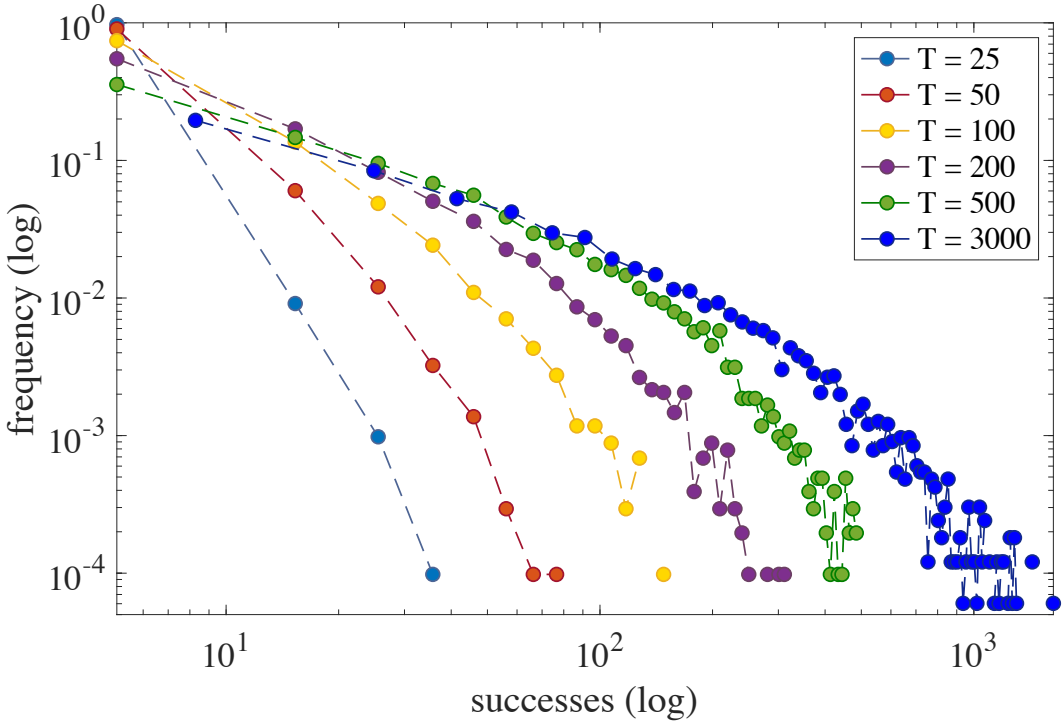

Figure 8: The model does not generate power-law distributed success: Distribution of individual success $s_{i,t}$ in a setting with $N = 1,000$ agents, $F = 5$ features with $Q = 15$ traits, and a circle network with degree 24. Temporal evolution of the success distribution for success motivation $m = 1$ for $T = 25, 50, 100, 200, 500$ macro steps. Data is shown on a logarithmic scale. The final distribution ($T = 3,000$) is also shown. All data shown in the figure is based on 10 runs per parameter.

Appendix C: Number of Features and Number of Traits

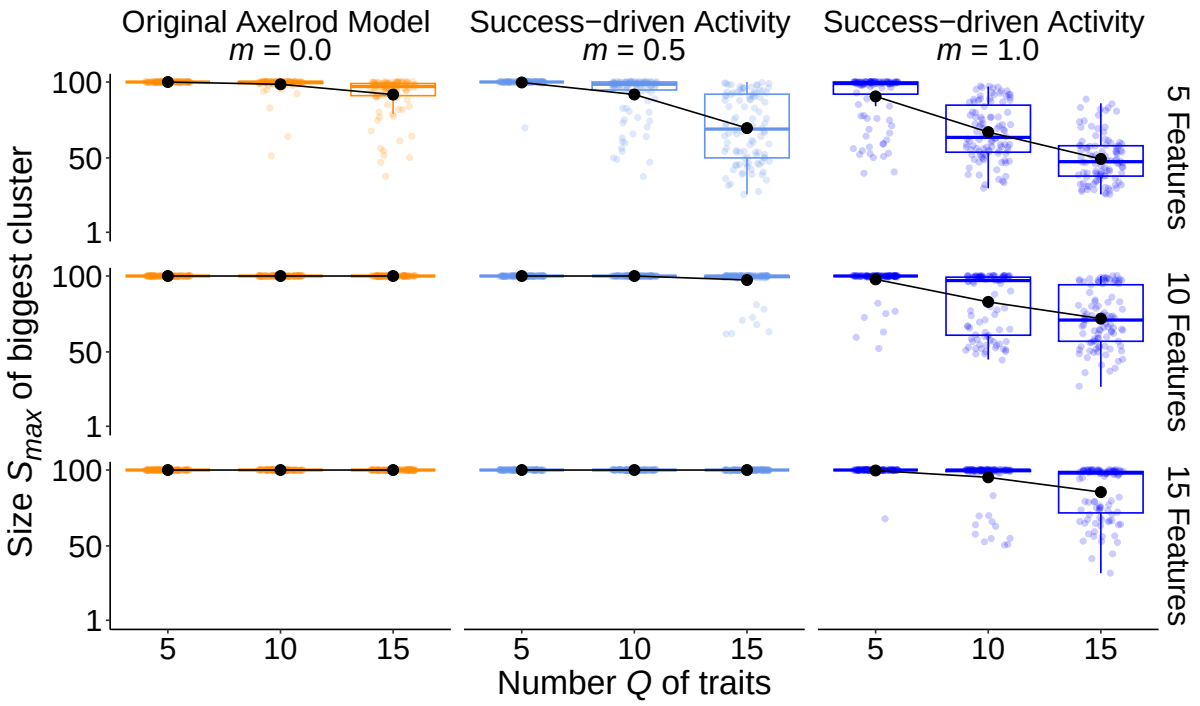

Figure 9: Effect of the number F of features and the number Q of traits on the size S_{max} of biggest cluster in equilibrium. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$; a degree of 16 for all agents and no rewiring. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 0.5$; $m = 1$) is shown in different shades of blue.

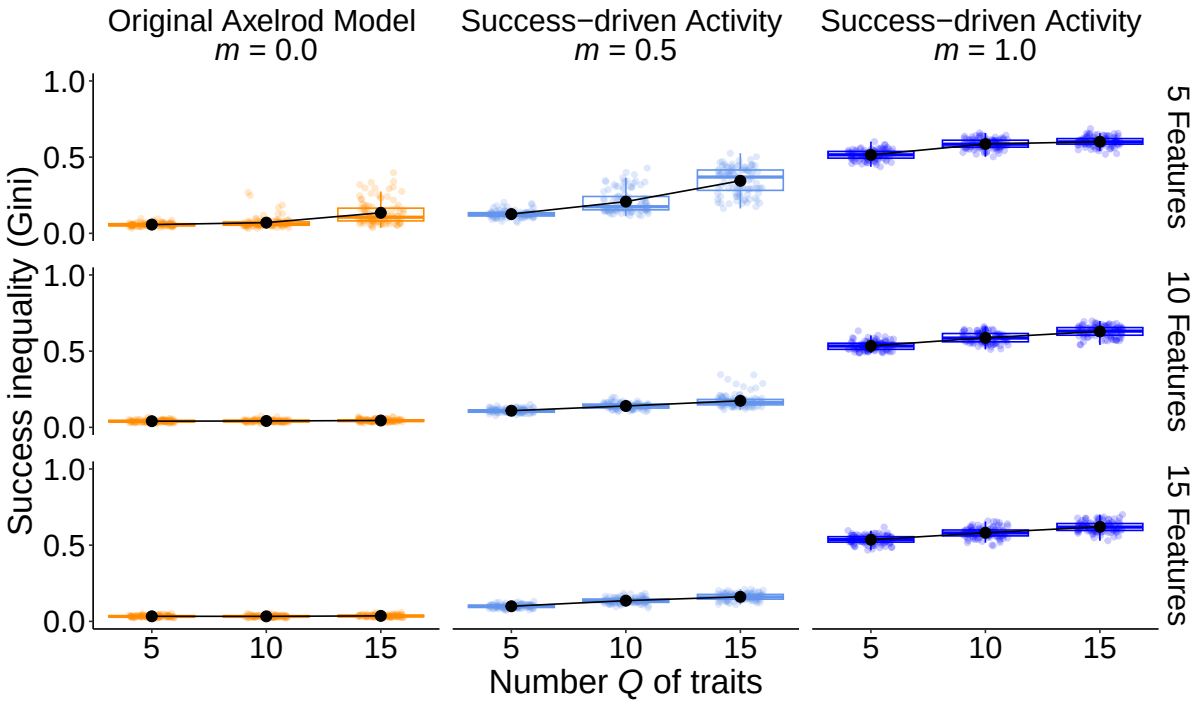

Figure 10: Effect of the number F of features and the number Q of traits on success inequality (Gini) in equilibrium. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$; a degree of 16 for all agents and no rewiring. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 0.5$; $m = 1$) is shown in different shades of blue.

Appendix D: Neighborhood Size

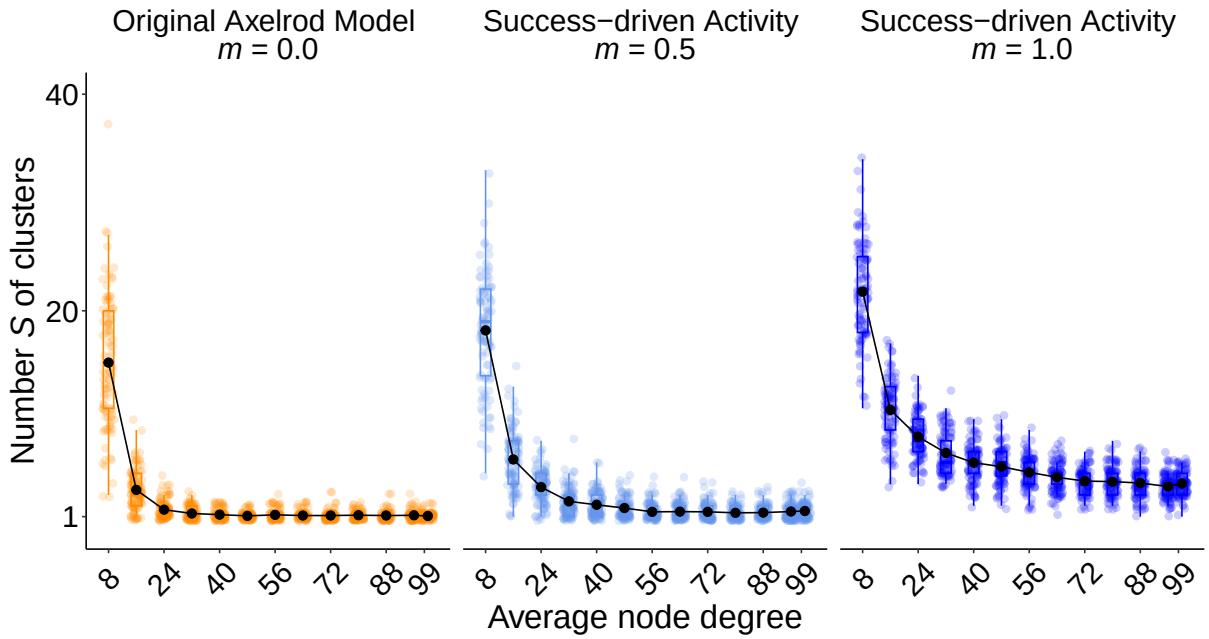

Figure 11: Effect of size of the neighborhood on number S of clusters in equilibrium. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$; number F of features = 5; number Q of traits = 15 and no rewiring iterations. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 0.5$; $m = 1$) is shown in different shades of blue.

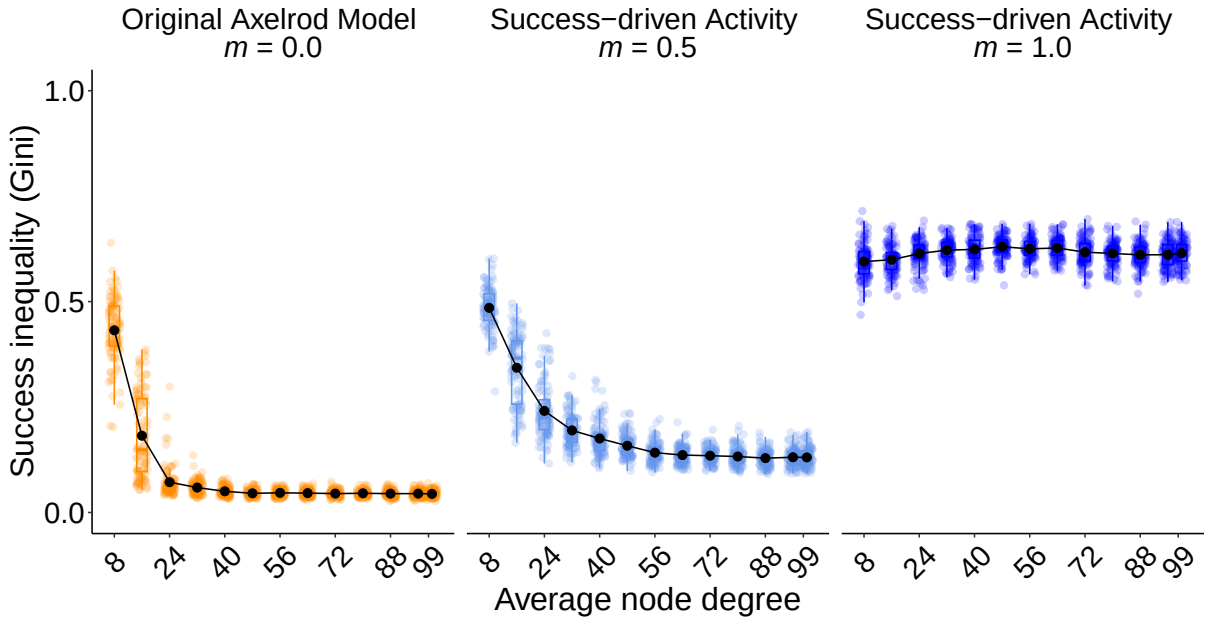

Figure 12: Effect of size of the neighborhood on succes inquality (Gini) in equilibrium. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$; number F of features = 5; number Q of traits = 15 and no rewiring iterations. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 0.5$; $m = 1$) is shown in different shades of blue.

Appendix E: Network Clustering

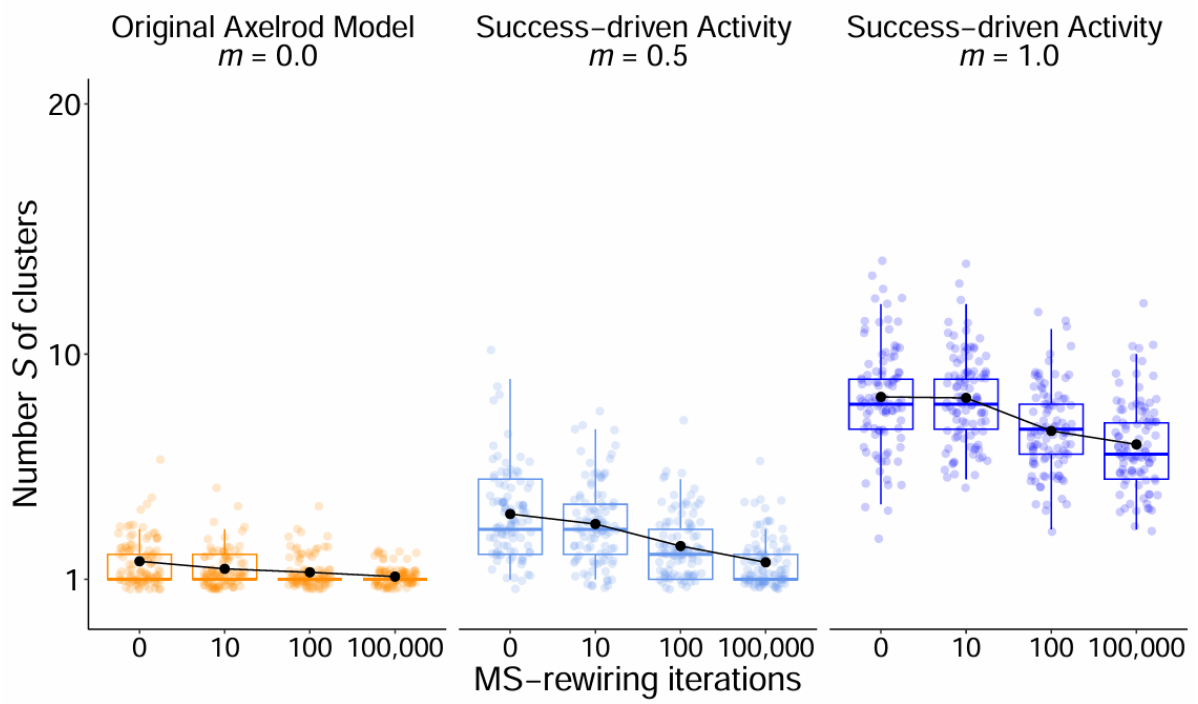

Figure 13: Effect of number of rewiring iterations on number S of clusters in equilibrium. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$; number F of features = 5; number Q of traits = 15 with a degree of 24. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 0.5$; $m = 1$) is shown in different shades of blue.

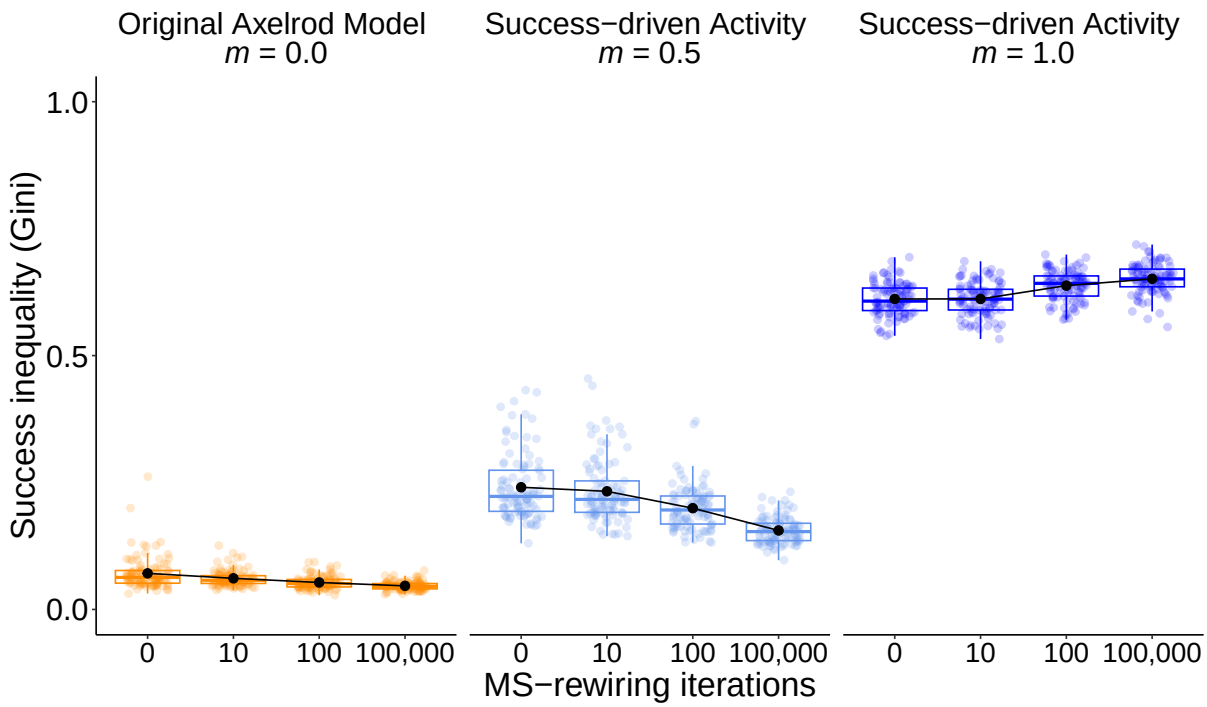

Figure 14: Effect of number of rewiring iterations on success inequality (Gini) in equilibrium. We conducted 100 independent simulation runs per experimental treatment, always assuming a population size of $N = 100$; number F of features = 5; number Q of traits = 15 with a degree of 24. The original Axelrod model ($m = 0$) is shown in orange. The success-driven activity ($m = 0.5$; $m = 1$) is shown in different shades of blue.

References

Alizadeh, M. & Cioffi-Revilla, C. (2015). Activation regimes in opinion dynamics: Comparing asynchronous updating schemes. *Journal of Artificial Societies and Social Simulation*, *18(3)*

Aracena, J., Goles, E., Moreira, A. & Salinas, L. (2009). On the robustness of update schedules in boolean networks. *Biosystems*, *97*(1), 1–8

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, *41*(2), 203–226

Axtell, R. (2000). Effects of interaction topology and activation regime in several multi-agent systems. International Workshop on Multi-Agent Systems and Agent-Based Simulation

Badawy, A., Ferrara, E. & Lerman, K. (2018). Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Bakshy, E., Messing, S. & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132

Bandini, S., Bonomi, A. & Vizzari, G. (2012). An analysis of different types and effects of asynchronicity in cellular automata update schemes. *Natural Computing*, *11*, 277–287

Bandura, A. & Walters, R. H. (1977). *Social Learning Theory*. Englewood Cliffs, NJ: Prentice Hall

Banisch, S. & Araújo, T. (2012). Who replaces whom? Local versus non-local replacement in social and evolutionary dynamics. *Discontinuity, Nonlinearity, and Complexity*, *2*(1), 57–73

Banisch, S., Jacob, D., Willaert, T. & Olbrich, E. (2024). The social dilemma of online segregation: A dynamical model of platform choice. arXiv preprint. arXiv:2411.04681

Banisch, S. & Olbrich, E. (2019). Opinion polarization by learning from social feedback. *The Journal of Mathematical Sociology*, *43*(2), 76–103

Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512

Baumann, F., Lorenz-Spreen, P., Sokolov, I. M. & Starnini, M. (2020). Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, *124*(4), 048301

Broido, A. D. & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, *10*(1), 1017

Carley, K. (1991). A theory of group stability. *American Sociological Review*, (pp. 331–354)

Caron-Lormier, G., Humphry, R. W., Bohan, D. A., Hawes, C. & Thorbek, P. (2008). Asynchronous and synchronous updating in individual-based models. *Ecological Modelling*, *212*(3-4), 522–527

Castellano, C., Fortunato, S. & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, *81*(2), 591–646

Castellano, C., Marsili, M. & Vespignani, A. (2000). Nonequilibrium phase transition in a model for social influence. *Physical Review Letters*, *85*(16), 3536

Chavalarias, D., Bouchaud, P. & Panahi, M. (2024). Can a single line of code change society? Optimizing engagement in recommender systems necessarily entails systemic risks for global information flows, opinion dynamics and social structures. *Journal of Artificial Societies and Social Simulation*, *27*(1). 9

Clauset, A., Shalizi, C. R. & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703

Das, S. & Lavoie, A. (2014). The effects of feedback on human behavior in social media: An inverse reinforcement learning model. Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems

Deffuant, G., Neau, D., Amblard, F. & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, *3*(01n04), 87–98

Flache, A. & Mäs, M. (2008). How to get the timing right. a computational model of the effects of the timing of contacts on team cohesion in demographically diverse teams. *Computational and Mathematical Organization Theory*, *14*(1), 23–51

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), 2

Flache, A., Takács, K. & Mäs, M. (2016). Discrepancy and disliking do not induce negative opinion shifts. *PLoS One*, *11*(6)

Friedkin, N. E. & Johnsen, E. C. (2011). *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Cambridge: Cambridge University Press

Fruchterman, T. M. & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164

Gaisbauer, F., Olbrich, E. & Banisch, S. (2020). Dynamics of opinion expression. *Physical Review E*, *102*(4), 042303

Galante, F., Vassio, L., Garetto, M. & Leonardi, E. (2023). Modeling communication asymmetry and content personalization in online social networks. *Online Social Networks and Media*, *37*, 100269

Gaultney, I. B., Sherron, T. & Boden, C. (2022). Political polarization, misinformation, and media literacy. *Journal of Media Literacy Education*, *14*(1), 59–81

Geschke, D., Lorenz, J. & Holtz, P. (2019). The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology*, *58*(1), 129–149

Grinberg, N., Dow, P. A., Adamic, L. A. & Naaman, M. (2016). Changes in engagement before and after posting to Facebook. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems

Hegselmann, R. & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, *5*(3), 2

Homans, G. C. (1974). *Social Behavior: Its Elementary Forms*. San Diego, CA: Harcourt Brace Jovanovich

Huberman, B. A. & Glance, N. S. (1993). Evolutionary games and computer simulations. *Proceedings of the National Academy of Sciences*, *90*(16), 7716–7718

Huckfeldt, R., Johnson, P. E. & Sprague, J. (2004). *Political Disagreement: The Survival of Diverse Opinions Within Communication Networks*. Cambridge: Cambridge University Press

Keijzer, M., Mäs, M. & Flache, A. (2024). Polarization on social media: Micro-level evidence and macro-level implications. *Journal of Artificial Societies and Social Simulation*, *27*(1)

Keijzer, M. A. & Mäs, M. (2021). The strength of weak bots. *Online Social Networks and Media*, *21*, 100106

Keijzer, M. A. & Mäs, M. (2022). The complex link between filter bubbles and opinion polarization. *Data Science*, *5*(2), 139–166

Keijzer, M. A., Mäs, M. & Flache, A. (2018). Communication in online social networks fosters cultural isolation. *Complexity*, *2018*, 1–18

Klemm, K., Eguíluz, V. M., Toral, R. & San Miguel, M. (2003a). Global culture: A noise-induced transition in finite systems. *Physical Review E*, *67*(4), 045101

Klemm, K., Eguíluz, V. M., Toral, R. & San Miguel, M. (2003b). Nonequilibrium transitions in complex networks: A model of social interaction. *Physical Review E*, *67*(2), 026120

Lindström, B., Bellander, M., Schultner, D. T., Chang, A., Tobler, P. N. & Amodio, D. M. (2021). A computational reward learning account of social media engagement. *Nature Communications*, *12*(1), 1311

Liu, S., Maes, M., Xia, H. & Flache, A. (2022). When intuition fails: The complex effects of assimilative and repulsive influence on opinion polarization. *Advances in Complex Systems*, *25*(08), 2250011

Lockwood, P. L. & Klein-Flügge, M. C. (2021). Computational modelling of social cognition and behaviour – A reinforcement learning primer. *Social Cognitive and Affective Neuroscience*, *16*(8), 761–771

Lüders, A., Dinkelberg, A. & Quayle, M. (2022). Becoming "us" in digital spaces: How online users creatively and strategically exploit social media affordances to build up social identity. *Acta Psychologica*, *228*, 103643

Lux, T. & Alfarano, S. (2016). Financial power laws: Empirical evidence, models, and mechanisms. *Chaos, Solitons & Fractals*, *88*, 3–18

Macy, M. & Tsvetkova, M. (2015). The signal importance of noise. *Sociological Methods & Research*, *44*(2), 306–328

Macy, M. W., Ma, M., Tabin, D. R., Gao, J. & Szymanski, B. K. (2021). Polarization and tipping points. *Proceedings of the National Academy of Sciences*, *118*(50), e2102144118

Mäs, M. & Flache, A. (2013). Differentiation without distancing. explaining bi-polarization of opinions without negative influence. *PloS One*, *8*(11), e74516

Mäs, M. & Helbing, D. (2020). Random deviations improve micro–macro predictions: An empirical test. *Sociological Methods & Research*, *49*(2), 387–417

Maslov, S. & Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, *296*(5569), 910–913

McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444

Merton, R. K. (1968). The matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63

Newman, M. E. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics*, *46*(5), 323–351

Page, S. E. (1997). On incentives and updating in agent based models. *Computational Economics*, *10*, 67–87

Pariser, E. (2011). *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think*. New York, NY: Penguin

Piketty, T. (2014). *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press

Riquelme, F. & González-Cantergiani, P. (2016). Measuring user influence on twitter: A survey. *Information Processing & Management*, *52*(5), 949–975

Ruff, C. C. & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience*, *15*(8), 549–562

Starnini, M., Baumann, F., Galla, T., Garcia, D., Iñiguez, G., Karsai, M., Lorenz, J. & Sznajd-Weron, K. (2025). Opinion dynamics: Statistical physics and beyond. arXiv preprint. arXiv:2507.11521

Sunstein, C. R. (2001). *Republic.com*. Princeton, NJ: Princeton University Press

Sutton, R. S. & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press

Thaler, J. & Siebers, P.-O. (2019). The art of iterating: Update-strategies in agent-based simulation. *Social Simulation for a Digital Society: Applications and Innovations in Computational Social Science*, (pp. 21–36)

Vrontis, D., Makrides, A., Christofi, M. & Thrassou, A. (2021). Social media influencer marketing: A systematic review, integrative framework and future research agenda. *International Journal of Consumer Studies*, *45*(4), 617–644

Waldrop, M. M. (2021). Modeling the power of polarization. *Proceedings of the National Academy of Sciences*, *118*(37), e2114484118

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*, *393*(6684), 440–442

Weimer, C., Miller, J. O., Hill, R. & Hodson, D. (2019). Agent scheduling in opinion dynamics: A taxonomy and comparison using generalized models. *Journal of Artificial Societies and Social Simulation*, *22*(4)

Wu, F., Wilkinson, D. M. & Huberman, B. A. (2009). Feedback loops of attention in peer production. 2009 International Conference on Computational Science and Engineering

Wu, S., Hofman, J. M., Mason, W. A. & Watts, D. J. (2011). Who says what to whom on Twitter. Proceedings of the 20th International Conference on World Wide Web