Social Networks 84 (2026) 164–179

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Use of aggregated relational data in agent-based modeling

Yunsub Lee [a,*], Xinwei Xu [b]

[a] COALESCE Lab, Department of Social and Cultural Anthropology, Autonomous University of Barcelona, Barcelona, Spain
[b] Social Networks Lab, ETH Zürich, Zürich, Switzerland

ARTICLE INFO

Keywords:
Aggregated relational data
Agent-based modeling
Small-world network
Large scale networks
Acquaintance relationships

ABSTRACT

Aggregated relational data (ARD) provides valuable information for inferring structural features of personal social networks at scale. Following recent ARD studies, we suggest a formal parameter for agent-based modeling (ABM) that helps reflect multiple structural features of extended social networks (e.g., size; variation; distribution) and apply it to a widely known classic ABM—Axelrod's cultural dynamic model. Results show that when incorporating realistic network features estimated from ARD, the model generates outcomes substantially different from its original results. Our study highlights ARD's potential to enrich ABM in reflecting more realistic networks that better connect micro-processes with macro-phenomena.

1. Introduction

Agent-based models (hereafter ABMs) are powerful tools to study complex systems and emergent phenomena. They often reveal how simple micro-level processes under specific network conditions (e.g., connected actors share similar cultural/opinion states after interactions) can lead to the unexpected emergence of macro-level social phenomenon, such as political polarization, residential segregation, and collective action (Bianchi and Squazzoni, 2015; Macy and Willer, 2002; Wilensky and Rand, 2015). While these models are often distinguished by their assumptions of micro-processes, network conditions also play a critical role, as the way in which actors are connected may largely shape the overall emergent patterns (e.g., a virus spreads slower in clustered networks) (Manzo, 2014; Raub et al., 2011; Rolfe, 2014; Squazzoni, 2008).

While a large body of ABMs delves into the emergence of macro social phenomena under various network conditions, most ABMs often rely on stylized network models that do not fully resemble real-world network features. For instance, many ABMs often default to small-world network models, originating from Watts and Strogatz (1998) (DellaPosta et al., 2015; Fagiolo et al., 2007; Flache and Macy, 2011b; Goldberg and Stein, 2018; Rahmandad and Sterman, 2008). Although small-world network models share theoretical similarities with real-world networks in terms of connectivity structures—with dense clusters sparsely connected by long-range ties that result in a shorter average length of shortest paths, they fall short in capturing other important real-world network features, such as the wide variation in node degrees and the skewness of the degree distribution.

Filling this gap, we draw on recent findings based on aggregated relational data (hereafter ARD) to propose a realistic and customizable network parameter for ABMs and apply it to a widely known classic agent-based model—Axelrod's cultural dynamic model (1997). ARD provides rich information that can be used to infer structural features of social networks, including proportions of hard-to-reach populations, network segregation level, and size and distribution of degree. This information is especially useful for understanding real-world network features at the societal level, which is almost impossible to estimate unless every individual is sampled.

Our proposed parameter extends the classic small-world network to reflect three network structural features: average size, variation, and skewness in the distribution of non-core social relationships, analogous to long-range ties in small-world networks.[1] Stylized network models used in most ABMs assume that nodes tend to have similar and smaller degrees (< 50) and may assign a small number of long-range ties

* Corresponding author.
E-mail addresses: Yunsub.lee@uab.cat (Y. Lee), xinwei.xu@gess.ethz.ch (X. Xu).

[1] Here we should point out that the estimate of extended network size in ARD studies is not a strict one-to-one mapping onto "long-range" ties in small-world models. As we explain later, acquaintances defined in ARD studies—based on "knowing" someone—reflect overall social connectedness rather than a specific type of tie (see also DiPrete et al., 2011). This makes ARD estimates closer to *total* network degree in stylized network models. However, recent ARD studies consistently find that variation in acquaintances is primarily due to variation in non-core ties (see summary in Table 1). Thus we approximate long-range ties in small-world models using the residual variation in "acquaintances" (i.e., excluding core ties).

https://doi.org/10.1016/j.socnet.2025.09.004
Received 30 October 2024; Received in revised form 31 July 2025; Accepted 26 September 2025
Available online 4 October 2025

randomly with uniform probability. However, many ARD studies support that when extended networks—such as acquaintance relationships—are considered, people may have far more social connections (~600). Additionally, while individuals tend to have similar numbers of close ties, the size of acquaintance relationships tends to have a wide variation and exhibit a skewed distribution (DiPrete et al., 2011; Hofstra et al., 2021; Lubbers et al., 2019). Following these insights, we draw an empirical parallel between core and extended social networks in ARD research and clustered and long-range ties in small-world network models. However, as in many stylized networks used in ABMs, the distinction is not based on tie strength, rather on differences in topological and distributional features such as clustering and degree variability. Thus, by allowing the customization of the average size of each actor's long-range ties, their variation, and the distribution pattern (e.g., normal; left- or right-skewed), our parameter helps to consider more realistic structural features of social networks in ABMs and enable researchers to consider various structural features of long-range ties that are revealed in recent studies on population-level social network studies, especially those using ARD. A NetLogo model that can generate small-world networks with our proposed parameter is accessible through this link.

To demonstrate the value of our approach, we apply the parameter to Axelrod's cultural dynamic model (1997) with the same assumptions about micro-process and initial assignment of each actor's cultural states. Axelrod assumes that actors are in a 10×10 grid network and have a 5-dimensional cultural vector. Each dimension's state (0−9) becomes identical with the state of one of the connected neighbors after an interaction. When this process is iterated, actors are likely to share homogeneous cultural states with their neighbors (i.e., local convergence), but at the same time, multiple culturally heterogeneous "regions" emerge due to the limited connections between actors in the grid network (i.e., global diversity). Influenced by this pioneering study, many ABM studies extend the model to explain not only cultural changes but also opinion dynamics (e.g., polarization or convergence of political opinions) with various micro-process assumptions and network conditions (Baldassarri and Bearman, 2007; Banisch and Olbrich, 2019; DellaPosta, 2020; DellaPosta et al., 2015; Flache and Macy, 2011a, b; Goldberg and Stein, 2018; Mark, 2003; Mäs and Flache, 2013). Yet, the classic model and follow-up studies tend to underestimate the large size of acquaintance relationships in social networks and overlook the skewed variation in their distribution, a discrepancy contradicted by recent ARD findings.

Based on a recent landmark ARD study (Lubbers et al., 2019), we configure the small-world network in Axelrod's model to include a large size of long-range ties (mean = 658) with wide variation (standard deviation = 458) and right-skewed distribution. To better align with the scale of the empirical estimates, we set a larger network size of 10,000, exceeding typical network sizes used in ABMs. This reconfigured small-world setup, where each node starts with 20 locally clustered ties on a ring lattice, is compared with the outcomes of the original small-world model, where long-range ties are a small fraction of the rewired clustered ties. The results show that even under the same micro-process, the small-world network with the ARD-informed parameter generates substantially different macro-level outcomes, compared to the original model. More specifically, convergences of both local and global levels rapidly emerge after a very long time of no change, and the homogeneities of neighbors (i.e., node-level local convergence) become polarized. In contrast, in the small-world networks without the parameter (i.e., 0.05 of locally clustered ties are rewired as long-range ties), the main findings are similar to the original results. We further explore how the three structural features of long-range ties characterized by L (e.g., larger size than locally clustered ties; wide variation; skewness) contribute to the diverging results.

Our study underscores the importance of considering realistic network features in agent-based modeling, which resonates with previous studies embedding ABMs into realistic empirical contexts (Bruch and Atwell, 2015; Hamill and Gilbert, 2009; Rolfe, 2014) and responds to recent calls for integrating ABMs with large-scale empirical data (Flache et al., 2022). We discuss the potential of using other ARD-derived information about network structural features—such as segregation level in social networks and proportions of social groups—to better represent realistic social network conditions in ABMs. More broadly, we reflect upon some general challenges ABM research faces when integrating insights from empirical network research, including the challenge of aligning stylized network structures and micro-interaction rules with the ambiguity and complexity of real-world social networks.

2. Background – ARD for the estimation of real-world social network structures at scale

ARD have proven useful in several aspects where researchers intend to draw inferences about specific characteristics of personal networks at the population level, such as the size and the distributional features of ego-centric networks (Hofstra et al., 2021; Lubbers et al., 2019) and the degree of segregation between subpopulations (DiPrete et al., 2011; Park, 2021). Such inferences are often made based on survey data from representative population samples asking about the number of persons respondents know with a specific attribute X (such as whose name is Maria, or whose occupation is lawyer) using the Network Scale-up Method (NSUM, see a recent review in McCormick 2020). NSUM was originally designed to estimate the size of a hard-to-reach population (Bernard et al., 2010; Killworth et al., 1998), and its basic idea is that in a general population under a random-mixing assumption, the size of the subpopulation with attribute X should be in proportion to the share of persons with attribute X in one's personal network, i.e., population-level group composition can be estimated by scaling up the group composition in the average personal network of a random sample drawn from the population.

Recently NSUM has been leveraged to estimate structural features of personal social networks, such as the size of acquaintance networks and its variation among subsegments of the population (Hofstra et al., 2021; Lubbers et al., 2019) and levels of homogeneity and segregation in personal networks according to a certain attribute (DiPrete et al., 2011; Park, 2021). Instead of estimating the unknown size of a target subpopulation, these studies leverage the known subpopulation size and elicit the number of people that the respondents know from various subpopulations. By averaging responses of many subpopulations, the size of one's personal network (i.e., degree or the total number of connections) can be estimated using the scale-up estimator (McCormick et al., 2010):

$$\widehat{d}_i = \frac{\sum_{k=1}^{K} y_{ik}}{\sum_{k=1}^{K} N_k} \cdot N \tag{1}$$

Where y_{ik} is the number of people that person i knows in subpopulation k, N_k is the known size of subpopulation k, and N is the size of the population. The corresponding standard errors can be estimated by the following (McCormick et al., 2010).

$$SE(\widehat{d}_i) \approx \sqrt{\widehat{d}_i} \cdot \sqrt{\frac{1 - \sum_{k=1}^{K} \frac{N_k}{N}}{\sum_{k=1}^{K} \frac{N_k}{N}}} \tag{2}$$

Several extensions of the original scale-up estimator have been proposed to address three potential sources of errors: transmission error (when the known number of persons from subpopulation k is underreported by respondents due to a lack of awareness or visibility of the attribute), barrier effects (when individuals systematically know more or fewer persons from subpopulation k than under random-mixing), and recall error (when respondents may under- or over-recall their known number of persons from subpopulation k). For instance, Zheng et al. (2006) explicitly estimated the overdispersion parameter to account for

Table 1
Summary of degree-related findings from recent NSUM studies.

Authors (year)	Estimated core network size /SD	Estimated extended network size /SD	Study context
DiPrete et al. (2011)	Trust network: Median = 17/IQR 10–26; right-skewed	Median = 550/ IQR 400–800; right skewed	National probabilistic sample of adults in the U.S. (N = 1371)
Lubbers et al. (2019)	Total: Mean = 23.2; Median= 19.5; Relatives: Mean = 17.4/ SD = 13.4; Friends: Mean = 5.8/ SD= 5.6	Mean = 658/SD = 489 (barrier model); Median = 536/IQR 337–830; right skewed	National probabilistic sample of adults with Spanish nationality in Spain (N = 2276)
Hofstra et al. (2021)		*NSUM estimates:* Mean = 980.39/SD = 529.24; Median = 892/IQR: 631–1295; skewness = 1.778 *Facebook estimates:* Mean = 381.71/SD = 206.71; Median = 351/IQR = 236–491; skewness = 1.172	Dutch adolescents (18–19 years old, N = 2546)
Jeroense et al. (2024)		Mean = 943.46/SD = 639.82; Median = 778 /IQR = 478–1271	Probabilistic sample of the Dutch population (16–45 years old, N = 1096) with an oversample of ethnic minority inhabitants

the heterogeneity in individuals' propensity to form ties with different subpopulations. McCormick et al. (2010) extended this work and proposed a latent non-random mixing model to account for varying mixing rates between subgroups. Most recently, Breza et al. (2020) developed a parametric approach to infer node- and network-level characteristics based on ARD and latent space models for network formation.

These methodological developments have enabled more accurate estimates of personal social network features on a population scale with cost-efficient data collection strategies. Recent empirical studies using NSUM provide novel insights into structural features of extended social networks, and more importantly, how these features diverge from well-established theoretical properties of social networks. For example, in an early study, McCormick et al. (2010) noted that the commonly established power-law distribution poorly approximates the ARD-based estimates of network degree distribution, which rather aligns better with a log-normal distribution.

Additionally, in contrast to traditional social network research, which often relies on statistical inferences using complete network data that are typically in bounded settings such as schools or organizations (e.g., see an early review in Snijders 2011), ARD-based studies open up new analytical avenues by offering valuable insights into the organizational dynamics of extended social networks that are not confined to specific settings.[2] Despite the theoretical prominence of "weak" ties (Granovetter, 1973), social relationships outside core networks—such as acquaintances—remain largely overlooked in existing social network theory and research (Rivera et al., 2010), which predominantly focus on close-relationship networks such as friendship, advice, mentorship, and core discussion networks.

Indeed, as surfaced from recent ARD studies across different contexts, the average size of personal networks—when accounting for non-core ties—is rather large, with huge variation across subgroups of the population. For instance, using a probabilistic adult sample in Spain, Lubbers et al. (2019) estimated that the average size of acquaintance networks is 658, with a standard deviation of 489. In particular, individuals with higher educational levels and active participation in voluntary associations tend to have a larger volume of acquaintances. Linking survey data with social media data on a sample of Dutch adolescents, Hofstra et al. (2021) provided two estimates for extended network size—one based on ARD (median = 892) and the other based on the number of links on Facebook (median = 351)—and found that girls, ethnic majority members, and individuals participating in multiple social foci tend to have larger network size. Similarly, Jeroense et al. (2024) reported estimates of extended network size within the same range for the broader Dutch population. All studies observed a right-skewed degree distribution, with very few individuals having an extremely large network size.

In Table 1 we summarize the most recent findings from empirical ARD studies using the NSUM method to estimate degree-related features of personal social networks on a population scale. Although the contexts vary, all studies support that 1) when acquaintance ties are considered, individuals tend to have a large network size on average; 2) sizes of core networks are similar, but sizes of extended networks exhibit very large variation; 3) the degree distribution is typically right-skewed with a long tail, and in certain cases, a fat-tailed distribution. These patterns are consistent with other studies using different measurements and data sources (Bokányi et al., 2023; Hill and Dunbar, 2003; Ishiguro, 2016; Wang and Wellman, 2010).

In fact, recent network studies aiming to infer population-level network patterns using alternative data sources — such as administrative records (Bokányi et al., 2023) and social media platforms (Chetty et al., 2022) — align with the degree-related features of personal networks observed in NSUM research. For instance, Chetty et al. (2022) conducted a study of Facebook ties (as a proxy for offline social relationships) among 72.2 million U.S. users aged 25–44, representing 82 % of the population, and found that, on average, individuals have 568.2 connections (SD = 601.0), with a median of 382 connections (IQR = 182–1251). Comparable estimates of median degree (351) were also found by Hofstra et al. (2021) using Facebook links among a sample of Dutch adolescents, albeit with a narrower range of variation (IQR = 236–491). In a population-scale network study of 17.2 million registered residents in the Netherlands, Bokányi et al. (2023) examined co-affiliation-based social networks (e.g., attending the same school, residing in the same household, employed by the same company) using administrative records and found a right-skewed fat-tailed degree distribution, with 80 % of nodes having 28–177 connections. This range can be seen as an approximate representation of the "opportunity set" for core networks—family, friends, neighborhood and workplace contacts—typically measured in ego network surveys (Marsden et al.,

[2] Note that the definition of "acquaintances" (e.g., whom you know) is much looser than traditional social network definitions such as friendships. In fact, acquaintances—defined on the basis of "knowing someone"—transcends the traditional distinction between strong versus weak ties, as the acquaintance network tends to contain a handful of strong contacts and hundreds of weaker ones. It reflects overall social connectedness rather than ties with specific content (also see DiPrete et al. 2011).

2021).

Altogether, these studies reveal a few blind spots in our existing knowledge of personal social networks. While much of social network research–so far–has focused on close relationships, such as friendships and core discussion networks, NSUM studies using aggregated relational data reveal the vast and highly variable nature of extended social networks, such as acquaintance relationships — social fabrics that have hitherto remained largely hidden and overlooked. Extended social networks are not only larger in volume than close relationship networks (as traditionally measured in social network research) but also exhibit substantial heterogeneity across different subgroups. Right-skewed and long-tailed degree distributions suggest that a small subset of individuals maintains exceptionally large networks, which may have critical implications for access to resources, social influence, and information diffusion. To explore the theoretical implications of these structural features, we turn to the Agent-based Modeling approach, as it allows explicit manipulation of network properties and thus an examination of how changing network topology may shape collective outcomes (Manzo, 2014; Raub et al., 2011; Squazzoni, 2008).

3. Background – stylized networks in ABMs

Many ABMs aim to explain how macro-level phenomena (e.g., residential segregation; political polarization; collective action) can emerge through iterations of certain micro-level interdependent actions (e.g., moving to another neighborhood; influenced by others' political opinions; interpersonal punishment) (Bianchi and Squazzoni, 2015; Macy and Willer, 2002; Schelling, 2006; Wilensky and Rand, 2015). For example, in Schelling's segregation model (1971), there are 138 actors split equally into two groups, and each group of actors is of the same race. They are randomly positioned in a 13×16 grid network, the structure of which is identical to the second network of Fig. 1. This means that 70 spots in the network are assumed vacant. In each iteration, one randomly chosen actor moves to a vacant spot when more than 50 % of their neighbors are of a different race from the actor. These actions are then interdependent because one's move will determine the racial proportion of another's neighbors, consequently affecting another's move or stay. Due to the interdependence, when the process is iterated, extreme racial segregation rapidly emerges even if each actor has a fairly low preference for similar neighbors (≥ 0.5).

This micro-to-macro explanation is the basic framework of ABMs. Yet, this does not mean that macro-level structural features are unimportant in ABMs. Especially, social networks are considered important because a network structure can largely shape the overall pattern of micro-processes (Manzo, 2014; Rolfe, 2014). For example, when Schelling's segregation model is applied in scale-free and small-world networks with lower average degrees (~10), the two models similarly generate a higher level of segregation. However, when their average degrees increase (~30), while the segregation decreases in the small world network, there is no change in the scale-free network. This is because when the average degree increases in a small-world network, the connectivity between nodes increases due to the increased number of long-range ties. Then, one's moving is highly likely to affect the racial proportions of many others' neighbors due to the connectivity, thereby segregation decreases (i.e., many actors keep moving). On the contrary, the increase in average degree does not change the connectivity of a scale-free network because only a few nodes have large degrees, which implies that one's moving is likely to affect only those few nodes (Fagiolo et al., 2007, 2009). Thus, in general, how actors are connected determines each actor's local action, and thereby the interdependence of micro-processes is largely shaped by a network structure, eventually affecting the emergence of a certain macro-phenomenon (i.e., macro-micro-macro; Coleman Boat) (Raub et al., 2011; Squazzoni, 2008).

In many cases, ABMs assume one of three network models—grid (i.e., Von Neumann neighborhood), scale-free, and small-world—which are defined and stylized theoretically (Barabási and Albert, 1999; Toffoli and Margolus, 1987; Watts, 1999; Watts and Strogatz, 1998) (See also Hamill and Gilbert 2009; Rolfe 2014). Fig. 1 shows three example networks generated under the actual simulation conditions of existing ABM studies (Axelrod, 1997; Bravo et al., 2012; Flache and Macy, 2011b). Grid network is often used in many early ABMs (Axelrod, 1997; Centola et al., 2007; Mark, 2003; Schelling, 1971). Scale-free and small-world networks are used in relatively recent ABMs. While small-world networks tend to be more preferred in ABMs that focus on actions under social relationship contexts (e.g., discussion of political opinions) (Banisch and Olbrich, 2019; DellaPosta et al., 2015; Fagiolo et al., 2007, 2009; Flache and Macy, 2011b; Giardini and Vilone, 2021), scale-free networks are more used in ABMs for actions that do not necessarily happen in existing social relationships (e.g., spread of viruses or rumors; selecting exchange partners) (Bravo et al., 2012; Hosseini et al., 2016; Hui et al., 2010; Liu and Chen, 2011; Lombardo et al., 2022).

Since many ABMs assume networks following the theoretical models, some structural features of those generated networks tend to substantially deviate from what is documented by recent ARD studies. For example, actors in ABMs tend to have average lower degrees in small-scale networks (~100), whereas ARD studies show that people have a substantially larger volume of social relationships (~600). In the classic small-world network model, the number of long-range ties maintained by each node must be smaller than the node's clustered ties in the initial lattice structure. This contradicts observations from ARD studies, which consistently show that, on average, individuals tend to maintain far more non-core ties—such as acquaintance relationships (~500) —than locally clustered core ties (~20). Moreover, many ARD studies support that individuals tend to have a wide variation of degrees and the distribution is right-skewed. However, traditional grid and classic small-world models inherently produce networks with narrower degree variation more uniform and non-skewed distributions of degrees. Although

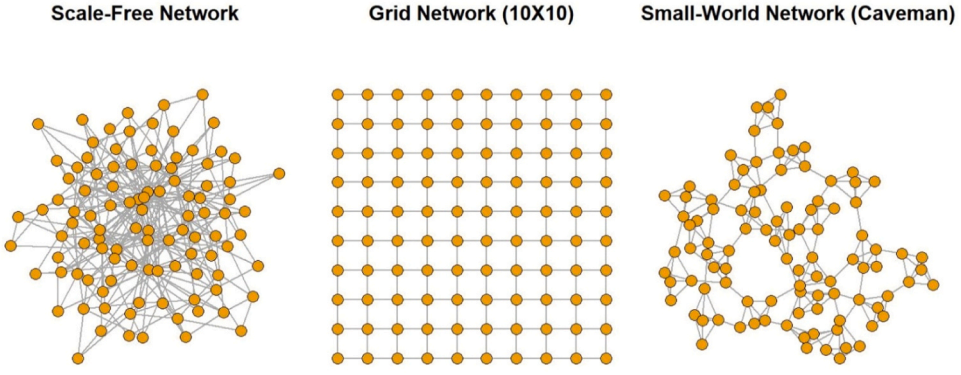

Fig. 1. Three example networks of agent-based models: Scale-Free, Grid (10 ×10), and Small-World (Caveman). Every network consists equally of 100 actors.

scale-free networks generate degree distributions with a large variation and extremely right-skewness (i.e., power law), they do not adequately capture distinct properties of core ties—which tend to be locally clustered with low variation—and extended non-core ties, which are more dispersed with wide variation.

These discrepancies suggest that integrating the findings of ARD studies into existing network models in ABMs can open up new analytical possibilities for exploring how structural network conditions may shape emergent macro-level outcomes, thus refining theoretical models of social networks — particularly, the role of weak or non-core ties in shaping network structural connectivity and emergent collective dynamics, a task well-suited for ABM studies.

4. ARD-informed network parameter for agent-based modeling

We suggest an ARD-informed network parameter for agent-based modeling as an extension to the original small-world model of Watts and Strogatz (1998). This parameter shapes the mean, standard deviation, and skewness of the distribution of long-range ties. As noted, many NSUM studies find that acquaintance relationships—which can be considered long-range ties in small-world networks—have much larger sizes (mean = 500–600), wide variation (standard-deviation = 400–500), and strong skewness (= right) (DiPrete et al., 2011; Hofstra et al., 2021; Lubbers et al., 2019). With this parameter, agent-based modelers can customize the structure of long-range ties in a small-word network based on the findings of ARD studies.

$$SW = (v,\ n,\ p) \tag{3}$$

Eq. (3) represents the small-world network model of Watts and Strogatz. It consists of three parameters: v, n, and p. The network size (i.e., number of nodes) is determined by v. The size of each node's close neighbors in the initial ring-lattice network is determined by n. The rewiring probability for long-range ties is determined by p. For instance, let $SW = (100, 20, 0.1)$. Then, 100 nodes are positioned in a lattice space following a ring shape, and each node is connected to 20 close neighbors in the lattice space. Among the generated ties in the initial ring-lattice network, 0.1 of randomly selected ties are removed and rewired as long-range ties also randomly (i.e., each pair of unconnected nodes has an equal probability of being connected by a long-range tie). Thus, each node's expected degree and size of long-range ties are 20 and 2 (=20*0.1), respectively. As p increases and approaches 1, the network structure gradually approximates a random graph.

We propose a new parameter L replacing p, which determines the rewiring probability of clustered ties and thus a proxy for the share of long-range ties. Since every pair of unconnected nodes has the same probability of being rewired, each node (i.e., actor of an ABM) in a generated small-world network is expected to have the same number of long-range ties, which is smaller than their number of clustered ties (i.e., close neighbors). Also, its distribution is expected to follow a binomial distribution with non-skewness and fixed variation. In contrast, when the parameter L suggested here is applied to a small-world network model, the structure of long-range ties can be empirically calibrated to consider recent findings from ARD studies, such as people tend to have many more extended social ties (~ 500) than close relationships (~20), and the distribution tends to exhibit wide variation and right-skewness. The extended small-world network $SW(L)$ with the parameter L can be formulated as follows with the assumption that $p = 0$ (i.e., initially assigned clustered ties are not rewired):

$$SW(L) = (v,\ n,\ p,\ L),\ s.t\ p = 0\ and\ L = (\mu,\ \delta,\ S) \tag{4}$$

In Eq. (4), the parameter $L = (\mu, \delta, S)$ is added to the small-world model, such that μ is the mean, δ is the standard deviation, and S is the skewness of the distribution of long-range ties, mapping onto the distributional features of extended social networks in ARD studies mentioned above. Note that $0 < S < 1$, and when the value is 0.5, the distribution becomes non-skewed (i.e., close to normal and polarized when μ is small and large enough, respectively). On the contrary, when S is close to 0 and 1, the distribution is right-skewed and left-skewed, respectively (i.e., actors tend to have fewer and many long-range ties). For simplicity, we assume that every actor has the same number of clustered ties. This is in line with the original small-world model SW (i.e., every node is connected to n close neighbors in a ring-lattice space) as well as the findings of ARD studies that people tend to have a similar number of close relationships. Thus, with the newly added parameter L, ABMs relying on the extended small-world network model $SW(L)$ can preserve not only its desiring theoretical properties—such as local clustering and the short average of shortest paths—but also flexibly specify the structure of long-range ties.

The three elements of the parameter L in $SW(L)$— μ, δ, S—are based on the statistical characteristics of a beta distribution. Essentially, each actor's expected number of long-range ties is randomly assigned from a beta distribution characterized by the three elements. Since beta distribution can take many forms in terms of mean, variation, and skewness, the parameter L can be flexible in generating different distributions of long-range ties. This is made possible by mapping the three elements onto the parameters of a beta distribution, α and β. Given a beta distribution $Beta\ (\alpha,\ \beta)$, its mean and standard deviation are $\frac{\alpha}{\alpha+\beta}$ and $\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$, respectively. Since the range of a beta distribution is [0,1], we add a weight value w to adjust for the mean and the standard deviation of long-range ties, which exceeds the range of [0,1]. Thus, we have the following definitions:

$$\mu = w\frac{\alpha}{\alpha+\beta},\ \delta = w\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}},\ and\ 0 < S < 1 \tag{5}$$
$$s.t\ \alpha = Sk\ and\ \beta = (1-S)k$$

S is defined as the ratio of α to β, under the assumption that $Sk : (1-S)k = \alpha : \beta$. Another weight value k is added to the definition of S in order to determine the actual values of α and β under the assigned ratio S. For example, when $S = 0.01$, $\alpha : \beta = 0.01k : (1 - 0.01)k = 0.01 : 0.99$. Since a beta distribution becomes right-skewed and left-skewed when $\alpha < \beta$ and $\alpha > \beta$, respectively, the customization of S can determine the skewness of long-range tie distribution. w and k can be calculated based on Eq. (5). When $L = (\mu, \delta, S)$ is given, we replace α and β in μ and δ using Eq. (5), yielding $\frac{\alpha}{\alpha+\beta} = \frac{Sk}{Sk+(1-S)k} = S$ and $\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} = \sqrt{\frac{Sk(1-S)k}{(Sk+(1-S)k)^2(Sk+(1-S)k+1)}} = \sqrt{\frac{S(1-S)}{k+1}}$. Thus, w and k can be calculated via the following:

$$w = \frac{\mu}{S}\ and\ k = \frac{w^2}{\delta^2}S(1-S) - 1 \tag{6}$$

For example, let $L = (5, 3, 0.05)$ in $SW(L)$. Since the skewness index S is defined under the assumption $\alpha = Sk$ and $\beta = (1 - S)k$, we immediately have $\alpha = 0.05k$ and $\alpha + \beta = 1k$. Then, based on $\mu = w\frac{\alpha}{\alpha+\beta}$ from Eq. (6), we get $w = \frac{5}{0.05} = 100$. With the same approach, using $\delta = w\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$, we get the equation $3 = 100\sqrt{\frac{0.05*(1-0.05)(k)^2}{(k)^2(k+1)}}$. Thus, $k = \frac{100^2}{3^2}0.05(1-0.05) = 52.78$. With the two weighted values w= 100 and $k = 52.78$, a weighted beta distribution w*$Beta\ (\alpha,\ \beta)$ can be

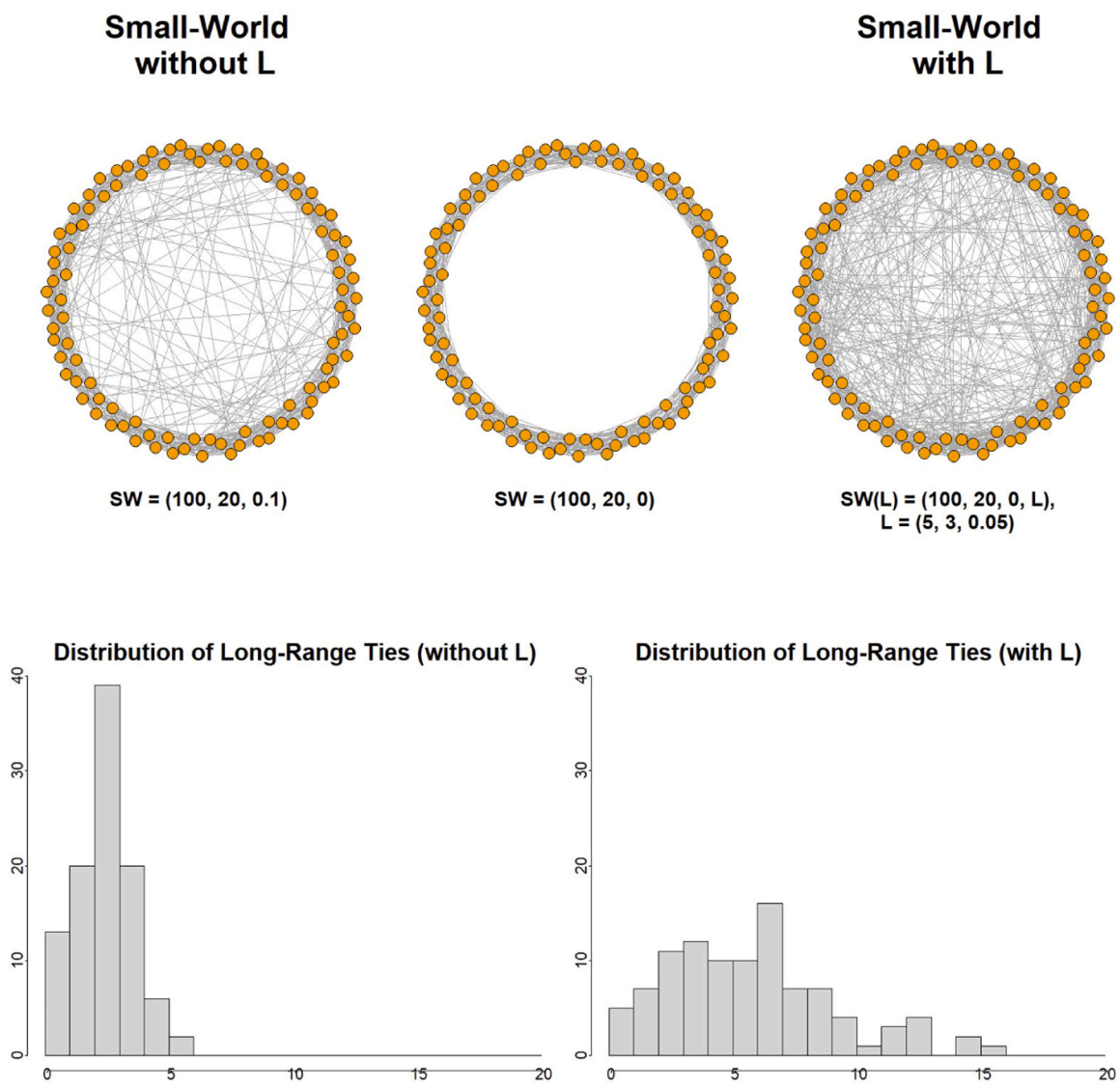

Fig. 2. Small-world networks without and with L and their distributions of long-range ties (frequency). Bin-width = 1. The first network (left) has long-range ties with the rewiring probability p = 0.1. The second network (middle) has no long-range ties with p = 0. The third network (right) has long-range ties randomly drawn from a pre-specified distribution L = (5, 3, 0.05).

constructed with $\alpha = 0.05*52.78 = 2.64$ and $\beta = (1–0.05)*2475 = 50.14$ that have the mean = 5, standard deviation = 3, and right-skewness (2.64 < 50.14).

To illustrate how parameter L shapes the structure of long-range ties in a small-world network, Fig. 2 (above graphs) compares two small-world networks SW = (100, 20, 0.1) without L and $SW(L)$ = (100, 20, 0, L) with L = (5, 3, 0.05). The center network in Fig. 2 visualizes the same starting condition, a ring-lattice network G (100, 20), consisting of 100 nodes with 20 closely clustered neighbors per node, all positioned on the initial ring-lattice. The network on the left illustrates a classic small-world network after rewiring ($p = 0.1$), with long-range ties consisting of, on average, 10 % of the existing clustered ties. The network on the right illustrates the extended small world $SW(L)$: instead of randomly rewiring existing clustered ties (i.e., $p = 0$), long-range ties are additionally drawn from a pre-specified distribution L = (5, 3, 0.05). Again, this difference reflects a key insight emerging from recent ARD studies that people's acquaintance network is much larger than their core network, meaning that the average ratio of long-range to clustered ties should, in many cases, be much larger than 1.

Fig. 2 (below histograms) also shows that the realized distributions of long-range ties in the two small-world networks are very different. Nodes in the classic small-world network SW (left) have, on average, 2 long-range ties with little variation and no skewness, due to the binomial process of long-range tie generation. In contrast, nodes in the extended small-world network $SW(L)$ (right) have much more long-range ties on average (= 5) with a wide variation (= 3) and a right-skewed distribution, where many nodes tend to have fewer long-range ties, but some nodes have many. This difference arises as each node's number of long-range ties is assigned by the randomly generated number that follows 100*$Beta$(2.64, 50.14) distribution, which is shaped by the L = (5, 3,

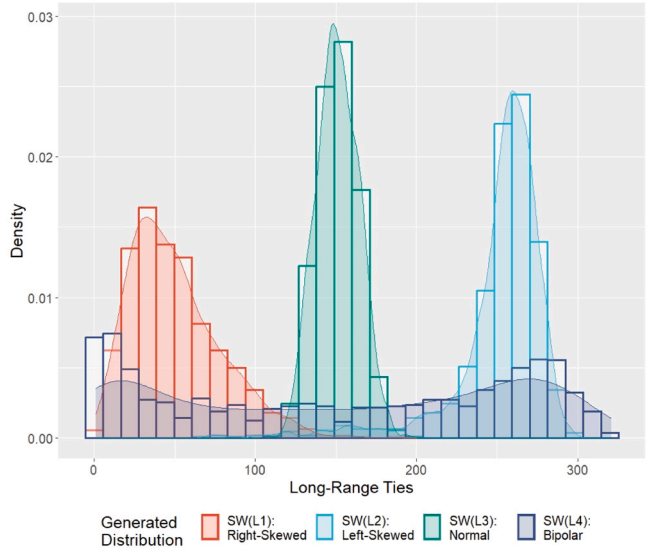

Fig. 3. Distributions of long-range ties (density) in four small-world networks *SW(L)* = (1000, 20, 0, *L*) with different *L* parameters. Bin-width = 10. *SW(L1)* and *SW(L2)* show the right- and left-skewed distributions, respectively, due to their different skewness values in *L1* = (50, 30, 0.01) and *L2* = (250, 30, 0.95). *SW(L3)* and *SW(L4)* show the normal and bipolar distributions, respectively, due to their different standard-deviation values in *L3* = (150, 10, 0.5) and *L4* = (150, 110, 0.5).

0.05) as explained in the example.[3]

Fig. 3 shows distributions of long-range ties in four small-world networks with different parameters of L. Note that every network shares the same condition of *SW(L)* = (1000, 20, 0, *L*). In turn, only the structures of long-range ties are differentiated by each network's unique parameter *L*, shown in the label of Fig. 3. Following the features of beta distribution, the generated long-range ties in four networks show different skewness of distributions (e.g., left, right, normal, and bi-polarized), different variations (e.g., narrow and wide), and different means (e.g., the blue distribution shows much larger mean value than the red distribution). This implies that by the flexible customization of L, various types of long-range tie structures based on ARD studies can be reflected in ABMs.

5. Application to Axelrod's cultural dynamic model

We apply our extended small-world model, *SW(L)*, to a widely known classic ABM—Axelrod's cultural dynamic model (1997). Axelrod's model is arguably one of the most influential ABMs and serves as the foundation for numerous cultural and opinion dynamic studies (Baldassarri and Bearman, 2007; Banisch and Olbrich, 2019; DellaPosta, 2020; DellaPosta et al., 2015; Flache and Macy, 2011a, b; Goldberg and Stein, 2018; Mark, 2003; Mäs and Flache, 2013). Although subsequent ABMs adopting Axelrod's model vary in their assumptions about social networks, micro-interaction rules, and measures of macro-outcomes (e.g., cultural diversity; opinion polarization), they similarly rely on the central idea in Axelrod's model that actors' cultural and opinion states are influenced by those of their connected neighbors in social networks (Flache et al., 2017; Squazzoni, 2012). Thus, the implications of our findings here will not only inform the original Axelrod's study but also have broader applicability to a wider range of ABM studies of cultural and opinion dynamics. In other words, our goal here is to explore the theoretical implications of having a large and highly variable number of long-range ties on local and global outcomes of cultural change, while holding the micro-interaction rules constant. Next we describe the original Axelrod model in detail and how these micro-interaction rules are implemented in our computational experiments. The program code of our replicated model (NetLogo) is available at this link.

In Axelrod's original model, 100 actors are connected in a 10 × 10 grid network, and each actor has a 5-dimensional cultural vector that can change over time. At the initial assignment (iteration = 0), each cultural dimension has a random state between 0 and 9. Formally, let $S_{i,t}$ be node i's 5-dimensional cultural vector and $a_{ik,t}$ be a kth element of the vector at iteration t. Then, $S_{i,t}$ is defined as the Eq. (7) and node i's initial cultural states (= $S_{i,0}$) are based on the random assignment of $a_{ik,t}$ at t = 0.

$$S_{i,t} = \left(a_{i1,t},\ a_{i2,t},\ a_{i3,t},\ a_{i4,t},\ a_{i5,t}\right)\ \ s.t\ \left\{a^{\forall}_{ik,t}\ \middle|\ 0 \le a^{\forall}_{ik,t} \le 9\ and\ a^{\forall}_{ik,t} \in \mathbb{N}\right\} \tag{7}$$

At each iteration, one actor is randomly chosen and interacts with one of the connected neighbors whose cultural states are relatively similar to the actor's (i.e., actors are more likely to interact with similar neighbors). The similarity is measured by the number of cultural states (= $a_{ik,t}$) that are identical with each other. For instance, when actor i's $S_{i,t} = (1, 4, 6, 7, 8)$ and neighbor j's $S_{j,t} = (1, 4, 6, 0, 0)$, then their similarity index at iteration t becomes 3/5 = 0.6, and when other neighbors' indices are lower than 0.6, actor i is likely to interact with neighbor j due to their relatively higher similarity index. Then, one of their cultural states is randomly chosen and becomes identical after the interaction. For example, when their 4th cultural state is chosen, since the actor and neighbor's state is 7 and 0, respectively, the actor i's 4rd cultural state becomes 0 after the interaction, following the neighbor j's state. In a formal way, $S_{i,t} = (1, 4, 6, 7, 8)$ changes to $S_{i,t+1} = (1, 4, 6, 0, 8)$ after interaction, influenced by $S_{i,t} = (1, 4, 6, 0, 0)$. Note that when the randomly chosen cultural state is the 2nd dimension one, there is no change between t and t + 1 because actor i and neighbor j already share the identical cultural state (= 4).

When this homophily process is iterated, similar actors become more similar, and thereby each actor is likely to have homogeneous neighbors in terms of cultural states (i.e., local convergence). Yet, since these interactions are only constrained to the connected neighbors in the grid network, actors who have a long graph-theoretic distance are likely to be heterogeneous (i.e., global diversity).

Assuming all modeling conditions (e.g., homophily process; random assignment of actors' cultural tastes) except the network structure, we apply two different types of networks—small-world networks with and without *L (SW* and *SW(L))*—to the classic Axelrod model and compare the results. The first network—following the classic small-world formulation—is *SW* = (1000, 20, 0.05). There are 1000 actors in the network, and each actor has 20 clustered ties in the initial ring-lattice network. Then, 0.05 of clustered ties are randomly chosen and rewired as long-range ties between clusters. This means that each actor's expected number of long-range ties is 1 (= 20*0.05). The second network—based on our extension to the classic small-word—is *SW(L)* = (1000, 20, 0, *L*), where *L* = (658, 489, 0.01). This network has the same cluster size and network size as the first network. Yet, unlike the first network— where clustered ties are randomly rewired based on a fixed probability (*p = 0.05*), the second network's long-range ties are randomly drawn from a pre-specified distribution that has mean = 658, standard deviation = 489, and right-skewness (= 0.01 in *L*) without a rewiring process (i.e., *p = 0*)

We calibrate the distribution of long-range ties—as represented by

[3] In this process, any number generated by the beta distribution is transferred to an integer number. For instance, when 37.25 is generated from the assumed beta distribution for an actor, the actor is expected to have 37 long-range ties. They consist of in-degree and out-degree, which are undirected. Thus, the actor is expected to have 18.5 out-degree from the actor and 18.5 in-degree from others. Since those assigned numbers are different by actors, the generation process of long-range ties is based on the weighted random choice function in Netlogo. In turn, actors with higher long-range tie numbers generate more out-degrees and are more likely to be selected as in-degrees alters.

the L parameter—based on estimates from Lubbers et al. (2019), a recent high-quality NSUM study with a nationally representative sample of adults in Spain. We choose this landmark study as our basis for empirical calibration because of its methodological robustness, comprehensive coverage of sociodemographic groups (e.g., across age, gender, education, and occupation), and its consistency with prior empirical findings from earlier NSUM studies (e.g., DiPrete et al., 2011). Additionally, Lubbers et al. (2019)' study is particularly useful for our purposes as it provides estimates on both the core network and total acquaintance network size, making it easier to calibrate the clustered to long-range ties ratio in our model. Their findings support that people tend to have similar sizes of core networks (mean = 23), but they have many (mean = 658) and diverse (standard deviation = 489) numbers of extended social relationships, the distribution of which is strongly right-skewed.

Here we outline a few scope conditions for our model. Since our proposed parameter builds on the classic small-world network, our model inherently assumes that both clustered ties and long-range ties can provide the infrastructure for social influence and thus the potential to shape individual attitudes and opinions (Flache and Macy, 2011b). This implies that our model is only applicable to scenarios where long-range ties have proven relevant for social influence processes. In the small-world model with L, we assume that, beyond a fixed number of clustered ties, nodes can have a large and highly variable number of long-range ties. These ties can span a wide spectrum ranging from superficial encounters to regular interactions beyond one's core network, and they may be inherently less clustered compared to one's core networks (often consisting of close, trusted relationships). This assumption is informed by findings from recent NSUM studies and, in general, reflects increasing social connectedness due to online and digital platforms (Lewis, 2024) and increasing intergroup and intercultural contacts driven by migrations on a local or global scale (Vacca et al., 2025).

Although the homophily process of Axelrod's model is equally applied to the two networks following the original interaction rule, some details and assumptions are modified. First, the number of actors interacting at each iteration is defined as 0.1 of the actors. In the original model, only one actor interacts at each iteration among 100 actors. Yet, since the network size in this study is large (= 1000), this one-actor rule requires too many iterations. Moreover, comparing the results of networks with different sizes—which is necessary for additional analysis—becomes difficult under this rule. To solve these problems, we apply the 0.1 proportion rule to every model. Note that the interactions of 0.1 actors in each iteration proceed independently rather than simultaneously. This means that one actor is randomly selected for interaction with one of neighbors (i.e., one-to-one rule), and the process is repeated with replacement (i.e., each choice of an actor is independent) until the number of chosen actors is equal to the number of 0.1 actors. In other words, 1000 iterations of Axelrod's original model are exactly the same as 1 iteration of our model in the network of 10,000 actors.[4]

Secondly, global homogeneity is measured by the proportion of the largest number of actors (or size of a region) whose 5-dimensional cultural states are identical to each other's — i.e., the share of actors forming the global dominant culture (Eq. 9), and local homogeneity is measured by the average proportion of neighbors whose 5-dimensional cultural states are identical to ego's (Eq. 8). Thus, the more actors (and neighbors) share identical cultural states (with ego), closer the measures will be to 1. Global homogeneity is based on the measurement suggested by follow-up studies of Axelrod's model since there are no clear global homogeneity measures defined in the original model (Flache and Macy, 2011a; Klemm et al., 2003). The two measures are summarized in the equations below. Let N be the number of actors in the whole network, n_i be the number of neighbors in node i's ego network, $S_{\max}$ be the size of the largest group (or region) of actors who share identical cultural states, and S_i be the number of ego i's neighbors whose cultural states are identical to ego i's. Then, local and global homogeneities are measured by:

$$Local\ Homogeneity = \frac{1}{N}\sum_{i=1}^{N}\frac{S_i}{n_i} \tag{8}$$

$$Global\ Homogeneity = \frac{S_{\max}}{N} \tag{9}$$

While the two measures represent the cultural homogeneity in network-level, we also add node-level measures to explore an actor's proportion of homogeneous neighbors (i.e., node-level local homogeneity) and whether an actor shares the same cultural states with others of the largest cultural region ($Smax,i = 1$) or not ($Smax,i = 0$) (i.e., node-level global homogeneity). Note that the averages of node-level local and global homogeneities are the same as the measures of local and global homogeneities in network-level, respectively.

$$Local\ Homogeneity\ in\ Node\ Level = \frac{S_i}{n_i} \tag{10}$$

$$Global\ Homogeneity\ in\ Node\ Level = S_{\max,i} \tag{11}$$

It should be also noted how the models are run. Since the network size is much larger than the typical network size used in existing ABMs and each node has many long-range ties, the network construction time and simulation iterations may take much longer with a personal computer's specs. To solve this issue effectively, we use a cloud computing service publicly available.[5]

6. Results

We present findings comparing the results of Axelrod's cultural dynamic model in two different small-world networks *SW* and *SW(L)*—without and with L—considering the ARD-informed structure of long-range ties (or non-core ties). For the extended small-world networks with L, we sequentially introduce the structural features derived from Lubbers et al. (2019) in two steps: first, we adjust the average number of long-range ties to match with mean estimates of extended network = 658 with almost no variation (SD = 0.1) and no skewness (= 0.5 in L) of the distribution (i.e., every actor has almost the same number of long-range ties); second, we introduce variation (SD = 458) in the number of long-range ties and account for the distribution's right-skewness (= 0.01 in L) (i.e., actors have heterogeneous numbers of long-range ties). This approach allows us to isolate the effect of the average long-range tie size on macro-level outcomes and how this effect can change when the distribution becomes more heterogenous and

[4] Note that we choose 0.1, instead of 0.01, which is actually more similar to the original rule of Axelrod's model (0.01 = 1/100). This is because in our programmed model, we find that more iterations with a small number of interacting actors take much more time to simulate than fewer iterations with a large number of interacting actors. But again, due to the non-simultaneous (or loop) process of interaction, any assignment of the proportion does not change Axelrod's original one-to-one interaction rule.

[5] More specifically, we use the Amazon Elastic Compute Cloud (Amazon EC2) service with 512 G RAM and 32 cores of 3.2 GHz CPU (Intel Xeon Scalable). The use of cloud computing that provides high-performance computing (HPC) is known to remarkably improve the speed of iterations in a large-scale ABM (Antelmi et al., 2024; Wittek and Rubio-Campillo, 2012). With these specs, the generation of SW(L) = (10000, 20, 0, L) with L = (658, 489, 0.01) takes around 2 min, and 100 repetitions of Axelrod's model (3000 iterations for each) in the same network condition take around 3 h.

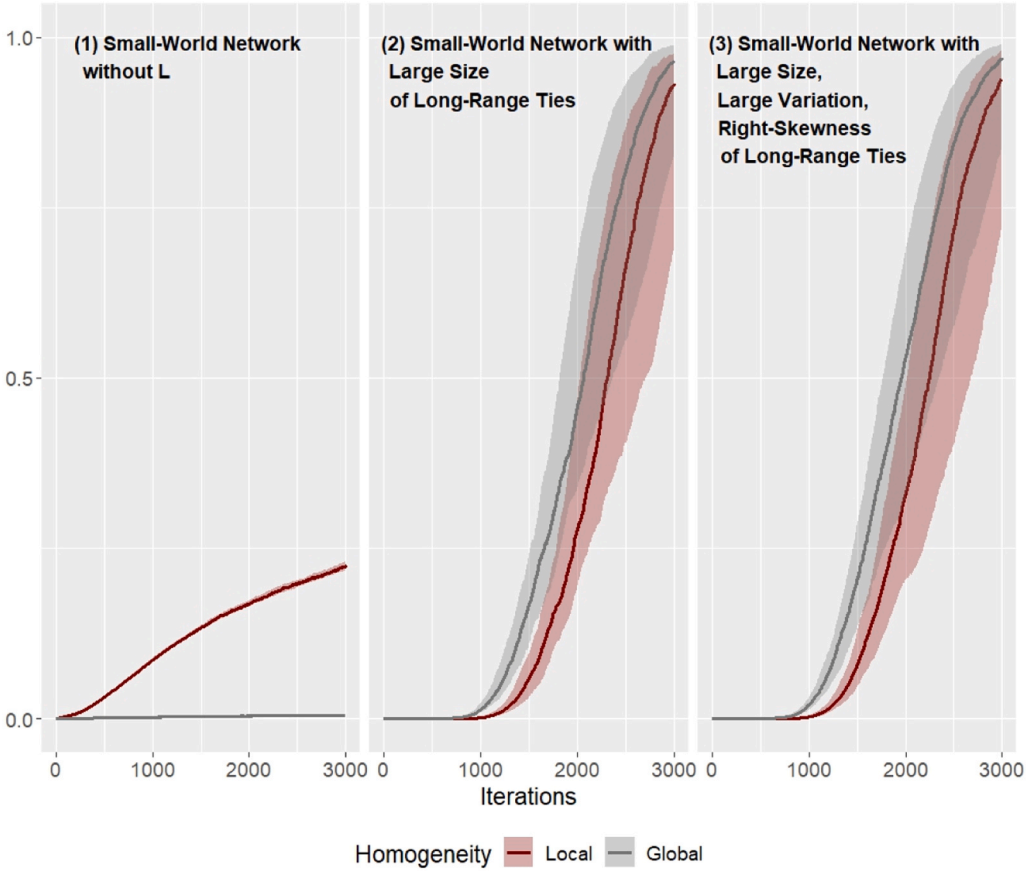

Fig. 4. Results of Axelrod's cultural dynamic model in small-world networks without and with L. The line plots represent the median of local and global homogeneities (during 3000 iterations), and the areas represent the interquartile range of the values, which are computed from 100 simulations per network model. Note that each set of 100 simulations per network model is conducted on 100 graph instances generated by the same network model (*SW* or *SW(L)*), differences among instances arise due to random variations in the degree and cultural vector assignments. While the classic *SW* (1) replicates outcome from the original Axelrod model (local homogeneity while globally diverse), both extended *SW(L)* (2) and (3) result in the emergence of local and global homogeneity.

skewed.[6] Note again that in both *SW(L)* models, the original rewiring probability in the classic small-world model is set to be zero ($p = 0$), since the long-range ties are randomly drawn from a pre-specified distribution and no additional rewiring is needed.

Fig. 4 compares outcomes of local and global homogeneity of culture for the two versions of small-world networks, *SW* and *SW(L)*. Note that the line plots and areas represent the median and interquartile range of the two values observed from 100 simulation results in each network, respectively. The simulations in each network are operated in 100 graph instances that are generated by the same network model (e.g., for (3) network, *SW(L)* = (10000, 20, 0, *L*) with $L = (658, 489, 0.01)$), but are different due to the randomness of the degree and cultural vector assignments. Thus, the results show the change of local and global homogeneities under the given network conditions—not in a particular network graph—clarifying the relationship between the parameter *L* and cultural change more rigorously.

The first network, *SW* = (10000, 20, 0.05), small-world network without *L,* yields a steady increase in local homogeneity, meaning that actors increasingly share similar cultural states as their neighbors—an indicator of local convergence. However, no change in global homogeneity is observed, meaning that global diversity remains stable across time. Both are in line with the original findings of Axelrod's model. In contrast, the two variants of the extended small-world network (*SW(L)* = (10000, 20, 0, *L*) with $L = (658, 0.1, 0.5)$ and $L = (658, 489, 0.01)$) produce a non-linear increase in both local and global homogeneity, and the pace of change becomes very rapid after an initial extended period of stability—the trend of local convergence and global convergence start to pick up after around 1000 iterations. In turn, since the two networks both have the large size of long-range ties (= 658), this result supports that the increased size of long-range ties contributes to the emergence of cultural convergence in both local and global level.

We further explore whether wide variation and right-skewness of the long-range tie distribution affect the outcomes differently. Fig. 5 shows the trends of local and global convergence for the two variants of the extended small-world network starting from iterations = 1000. This comparison demonstrates that increased variation and the right-skewness of the distribution of long-range ties have led to an earlier onset of converging trend and a steeper incline in both local and global homogeneities. More specifically, during 1000–1500 iterations, the medians and interquartile ranges of both local and global homogeneities observed from the simulation results in the network (3) are higher than those observed in the network (2). In other words, the two network features—wide variation and right-skewness—boost the average size effect of long-range ties on cultural convergence.

Beyond differences in macro-level trends, Fig. 6 reveals another crucial difference in the node-level outcomes of the two small-world networks. In Fig. 6, the distributions are estimated based on 100 simulation runs per network model. Each bar's height represents the average number of actors whose node-level local homogeneity falls within the bin. Each bar's color represents the average proportion of actors in the

[6] Here, we didn't separately apply the variation and skewness. This is because when mean = 659 and standard-deviation = 489 is applied in L with no skewness (= 0.5), the generated Beta distribution is bi-polarized—not normal—due to the too large standard-deviation compared to the mean.

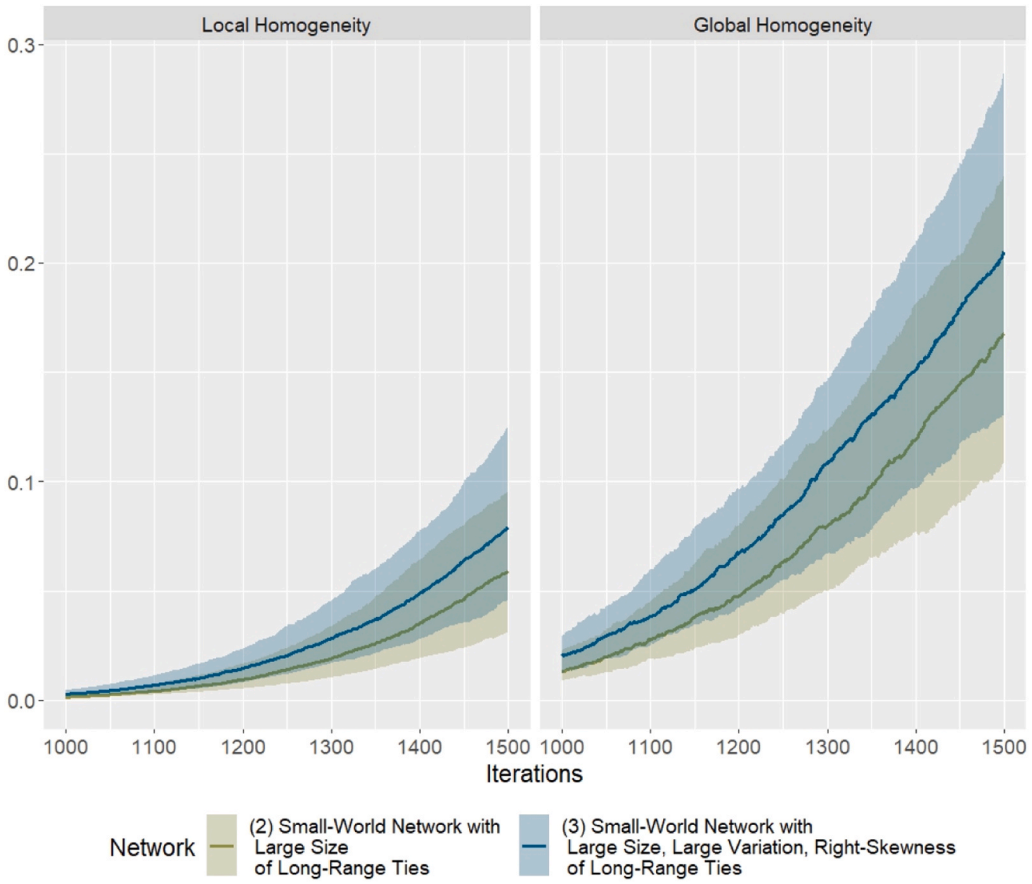

Fig. 5. Effects of the variation and skewness of the long-range tie distribution on local (left) and global (right) homogeneities. The line plots and areas represent the median and interquartile ranges of the two values during 1000–1500 iterations, computed from 100 simulations per network model (2) and (3). When long-range ties are drawn from a distribution with greater variation and right-skewness (model (3), trends of local and global homogeneity exhibit an earlier onset and steeper incline.

bar whose node-level global homogeneity = 1 (gradient from white to black), i.e., share of actors whose cultural states are identical to the global dominant cultural state.

Based on Fig. 6, it is expected that after 3000 iterations, the distribution of local and global homogeneities at the node-level (see Eqs. (10) and (11)) differs markedly between the two networks. In the classic small-world network *SW*, the distribution shows a unimodal pattern—most actors tend to have a high proportion of dissimilar neighbors. This suggests that although local convergence is gradually emerging in a collective sense—as shown by the steady increase of local homogeneity in Fig. 4, individual actors still tend to have a small proportion of neighbors who share their cultural traits even after 3000 iterations. In addition, they tend to be embedded in diverse cultural regions—that is, most actors' node-level global homogeneity is 0, meaning that they do not share the same cultural state as the dominant cultural group in their surroundings. In contrast, in the extended small-world networks *SW(L)*, the proportion of homogenous neighbors exhibits a bi-modal, polarized distribution: one group of actors has a very high proportion of homogeneous neighbors (>0.95), while another group has very few (<0.05). Moreover, the first group of actors tends to align their cultural traits with the mainstream culture (i.e., most of their node-level global homogeneity is equal to 1), whereas the second group of actors don't. This means that a group of culturally "isolated" actors—i.e., those with a high share of dissimilar neighbors locally and no alignment with any dominant culture globally—is likely to emerge in the extended small-world networks.

Next, we explore which structural features of the network may lead to these contrasting outcomes. More specifically, we examine the extent to which nodal degrees and the ratio of long-range ties to clustered ties are related to the divergent pattern of cultural homogeneity shown in Figs. 4 and 5. A key difference between the two versions of small-world networks is network density. This is a mechanical implication of the two different procedures of generating long-range ties. In the classic small-world network, long-range ties are considered as a small portion of existing clustered ties, a small fraction of "randomness" injected into the regular ring lattice. In our extended small-world model, we start from the empirical insight that the volume of long-range ties, in many cases, can be much larger than clustered ties and can exhibit a right-skewed distribution. Accordingly, we fix the number of clustered ties while randomly adding long-range ties following a pre-specified distribution. In the current implementation, the density level for small-world without *L*, $SW = (1000, 20, 0.05)$, is much lower than in the extended small-world network $SW(L) = (1000, 20, 0, L)$ with $L = (658, 489, 0.01)$ due to the latter having a much larger volume of long-range ties, resulting in a higher average node degree and a higher ratio of long-range to clustered ties. In this sense, although it could be assumed that the differing results are due to network density, further testing indicates this is not the case.

Fig. 7 presents results from nine different network conditions with the same network density (~0.068) but varying in the ratio of long-range ties to clustered ties (hereafter *L/C* ratio) and average node degree. The L/C ratio shapes the average shortest-path length, a key feature of the small-world model: as more long-range ties are added, the average shortest-path length decreases. With density held constant, varying the average node degree leads to changes in network size. In other words, each *column* shows three networks with a similar average

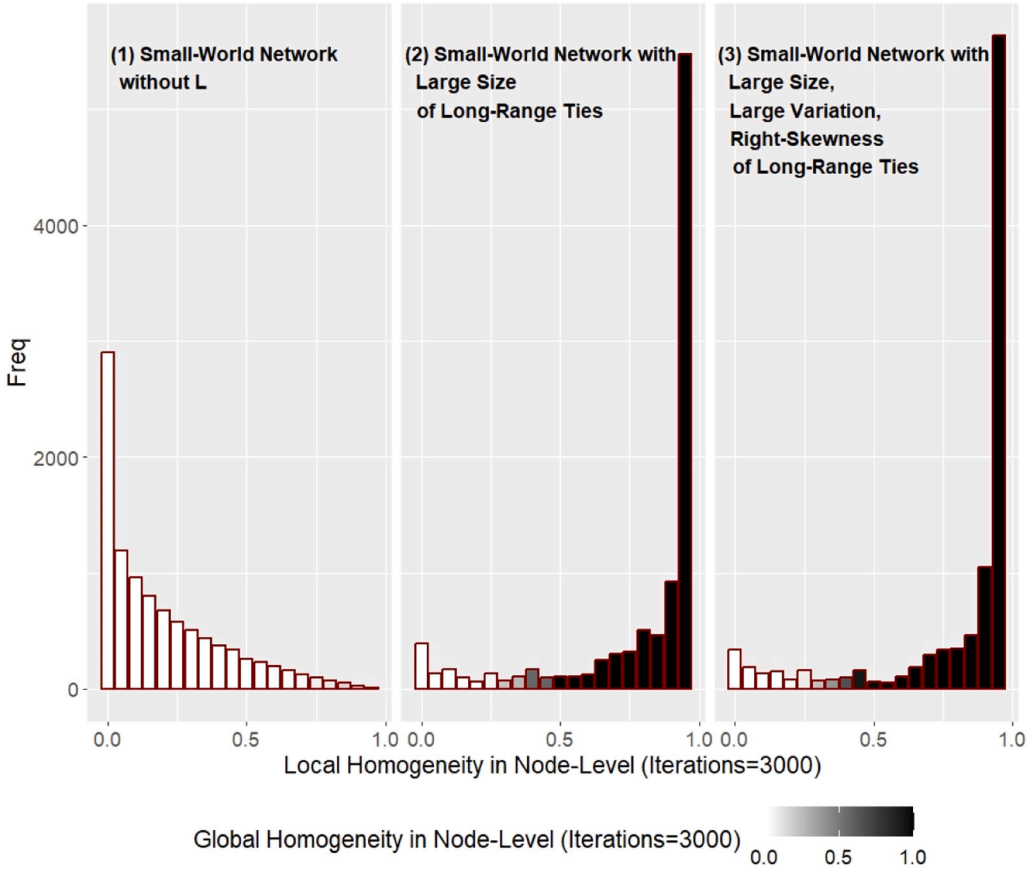

Fig. 6. Estimated distributions of node-level local and global homogeneities in small-world networks without and with L after 3000 iterations, based on 100 simulations per network model. Bar height indicates the average number of actors whose node-level local homogeneity is within the bin. Bar color represents the average proportion of those actors whose node-level global homogeneity = 1, i.e., whose cultural states align with the global dominant cultural states. Bin-width = 0.05. Both extended *SW* models result in a polarized distribution of node-level local homogeneity, i.e., share of neighbors having the same cultural states, between nodes that align with the global dominant cultural states and those don't.

shortest-path length and the same L/C ratio, while each *row* shows three networks with a similar average node degree and the same network size. For instance, the network in Plot (1) has a longer average shortest path than the network in Plot (7) (approximately $7.5 > 2$). This is because there is no long-range tie in the network in Plot (1) ($L/C = 0/(0.068*1000)$), whereas the network in Plot (7) has a few ($L/C = 6.58/(0.002*1000)$). Similarly, the network in Plot (9)—which also appears as the fully calibrated network (3) in Fig. 4—has a larger average node degree than the network in Plot (7) ($0.002*10000 + 658 > 0.002*1000 + 6.58$) because Plot (9)'s network has more nodes (10000 vs. 1000), even though both networks share the same density.

The results show that the higher L/C ratio contributes to the emergence of local and global homogeneities—not the density. In the first column (Plots (1), (2), and (3)), local convergence emerges, but global convergence remains minimal, a pattern similar to what we observed from the classic small-world of $SW = (1000, 20, 0.05)$ in Fig. 4 (left panel). In contrast, in the second and third columns (Plots (4), (5), (6), and Plots (7), (8), (9)), we observe a consistent pattern of both local and global homogeneities, following an initial period of no change. The convergence patterns are notably stronger in Plots (7), (8), (9) than Plots (4), (5), (6). This suggests that increasing the L/C ratio—and hence shortening the average shortest path length—promotes both local and global cultural homogeneity even under the same network density level.

In addition, looking across the rows, we see that average node degree may also influence the emergence pattern of Axelrod's model. In both the second and the third columns, for instance, we see that increasing the average node degree (and thus network size) prolongs the time before the emergence of local and global homogeneity. In Plot (7), local homogeneity starts to emerge after 900 iterations, whereas in plot (9), it only appears after 1100 iterations. In other words, not only the L/C ratio, but also the actual numbers of those two types of ties may lead to different patterns of macro-level emergence.

The key puzzle, then, is how a higher L/C ratio contributes to global and local homogeneities—and how this effect is further amplified by the variation and skewness of the long-range tie distribution. Fig. 8 presents a single simulation run as a case study, and in the left panel, illustrates how the distributions of node-level local and global homogeneities evolve over 3000 iterations in the small-world network with $L = (658, 489, 0.01)$—the same network condition as network (9) in Fig. 7. At the beginning (iteration = 0), all actors have a very low node-level local homogeneity (< 0.05), and there is no dominant cultural region (i.e., global homogeneity is close to 0). By iteration 1300, some actors begin to have culturally homogeneous neighbors (near 0.1). Among them, a subgroup of 68 actors (Group B) with local homogeneity above 0.1 actually leads the change, becoming the members of the largest cultural group (i.e., node-level global homogeneity = 1). By iteration 2000, all actors are clearly separated into two groups. Group B actors have more than half of culturally homogeneous neighbors, sharing the mainstream culture, whereas other actors in Group A have heterogeneous neighbors (local homogeneity ~ 0.1) and cultural states that are different from the majority. Finally, by iteration 3000, the two groups become polarized: a large majority of actors align with Group B, while a small number of marginal actors remain in Group B. Most actors in Group B have almost entirely homogeneous neighbors, driving the global homogeneity. In contrast, the remaining "isolated" actors in Group A have cultural states that are different from their immediate neighbors' as well as the

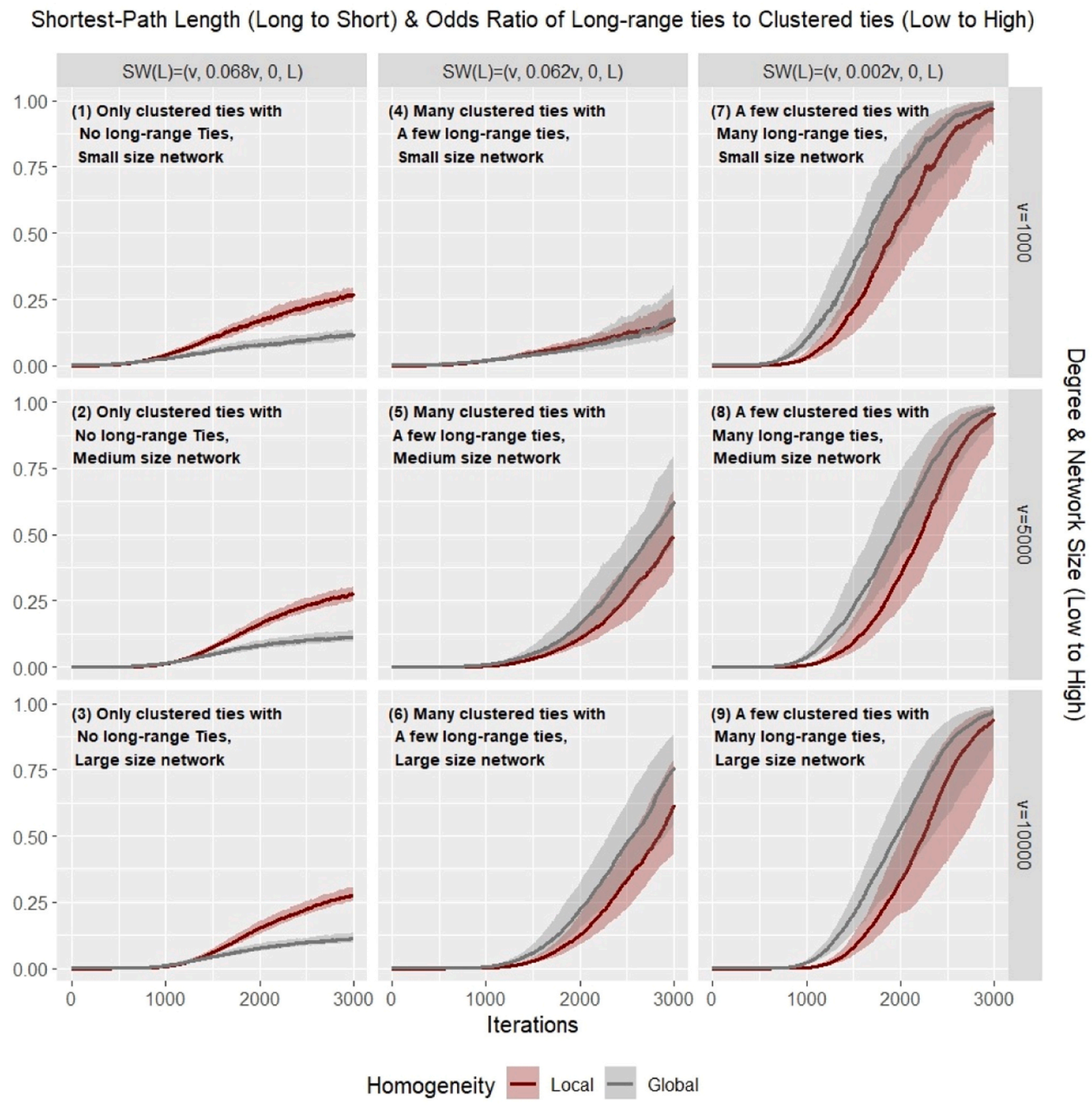

Fig. 7. Results of Axelrod's cultural dynamic model in 9 different small-world network models, with 100 simulations run per model. All networks share the same level of network density (~ 0.068). Network (9) uses $L = (658, 489, 0.01)$, and networks (8) and (7) use $L = (329, 244.5, 0.01)$ and $L = (65.8, 48.9, 0.01)$, respectively, scaled to their network sizes. Similarly, network (6) uses $L = (65.8, 48.9, 0.01)$, and networks (5) and (4) use $L = (32.9, 24.45, 0.01)$ and $L = (6.58, 4.89, 0.01)$, respectively. Networks (1), (2), and (3) represent the baseline condition of classic SW with $L = (0, 0, 0)$. Increasing the L/C ratio promotes the emergence of local and global homogeneities (within-row comparison).

mainstream culture (i.e., node-level local and global homogeneities are close to 0).

We find that this pattern can be explained by comparing the evolving network characteristics of Groups A and B. First of all, the variation and skewness of the long-range tie distribution boosts cultural homogeneity by creating conditions where many actors are connected to neighbors with much higher degrees. As shown in the box plots at iteration = 0, actors' average degrees tend to be lower than the average degrees of their neighbors, which is in line with the well-known "friendship paradox" found in empirical networks (Feld, 1991). This is because, the skewed distribution creates a small set of actors with extremely high degrees (> 3000), making them disproportionately likely to appear as neighbors for many actors—including those with degrees much lower than the mean 658 + 20.

This network feature plays an important role in facilitating the diffusion of cultural traits of the actors in Group B when they first begin to form homogeneous neighborhoods at around iteration 1300. As shown in the boxplots on the right panel, Group B actors initially have a much smaller mean degree than Group A actors at iteration 1300. Intuitively, actors with fewer connections are more likely to achieve local homogeneity earlier than actors connected with larger sets of neighbors — especially when they have the same chances of one-to-one

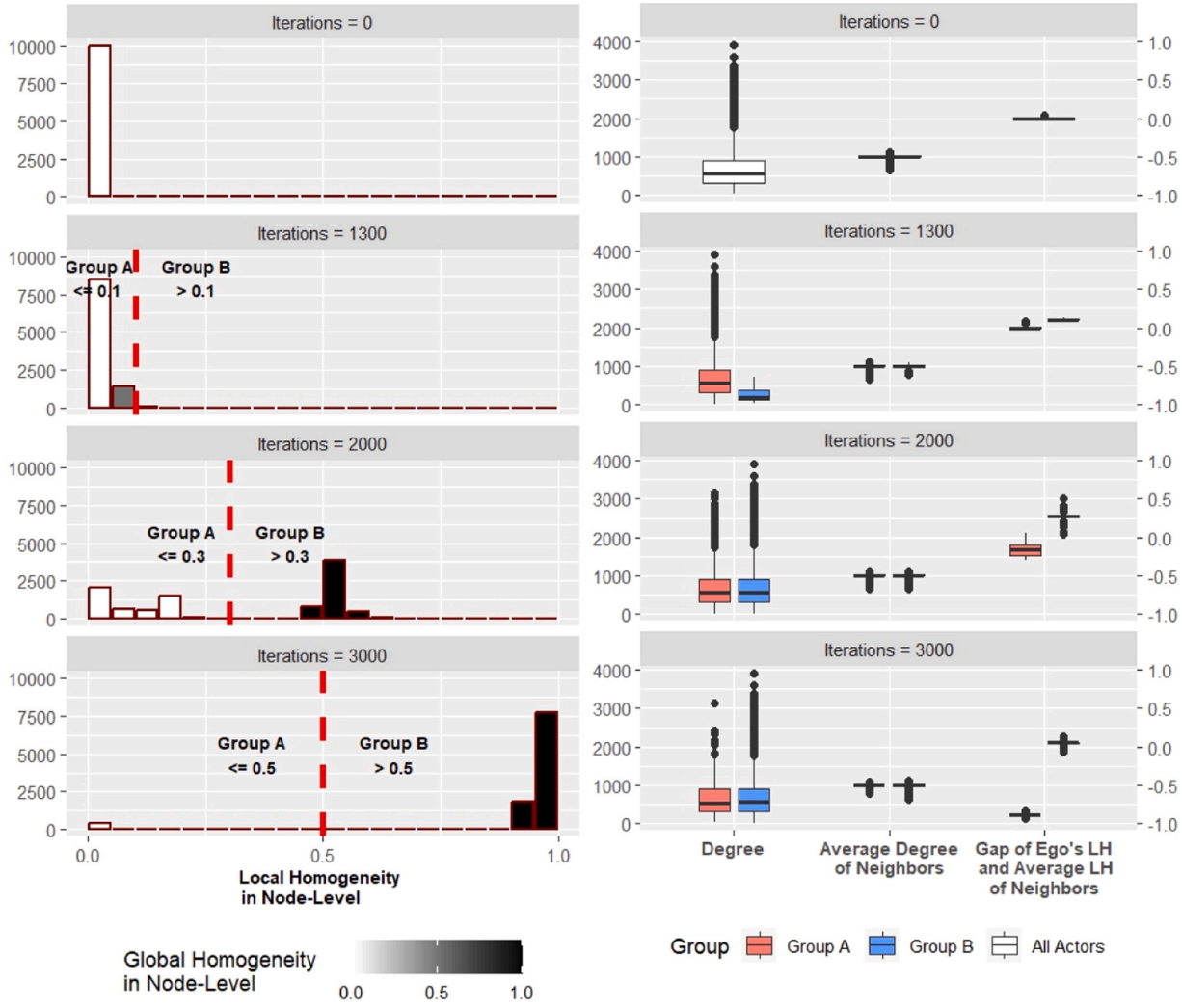

Fig. 8. Changing distributions of local and global homogeneities in node-level during 3000 iterations observed from a single simulation result. LH represents Local Homogeneity in node-level. In each iteration, for illustrative purposes, Groups A and B are defined by applying a threshold to the current distribution of local homogeneity.

interactions. The point is that although Group B actors have lower degrees, they are connected with higher-degree neighbors, which enables Group B's cultural states to easily diffuse, thereby attracting more actors to align with their dominant cultural states. This pattern is supported by the box plot at iterations 2000. The average degree of Group B's actors is on par with that of Group A, as Group B has absorbed many high-degree nodes and their cultural traits have diffused to a larger share of the overall population. In short, the variation and skewness of long-range tie distribution provide opportunities for some low-degree actors to achieve local homogeneity early, and through their higher-degree neighbors, their cultural states can quickly diffuse to a larger share of the population, contributing to the formation of a dominant global cultural mainstream (as seen in the histogram at iteration 3000).

Secondly, the higher L/C ratio contributes to global and local homogeneities—eventually producing the cultural polarization—because it creates differences in actors' and their neighbors' levels of local homogeneity (node-level, Eq. (10)). As shown in the box plots, actors in Group B begin to have slightly more homogeneous neighbors than their neighbors at iteration 1300 (i.e., positive gap) — meaning that Group B actors have a higher share of neighbors with the same cultural states than the average share for all their neighbors. The average gap increases at iteration 2000, and it becomes close to 0 at iteration 3000 since most actors—and their neighbors as well—have entirely homogeneous neighbors. On the contrary, Group A actors have negative gaps (i.e., their local homogeneity is lower than the average local homogeneity of their neighbors)—meaning that they are connected with more heterogeneous neighbors, while those neighbors themselves tend to be embedded in more homogenous networks—and this trend persists over time.

These differences in neighborhood homogeneity gaps lead to cultural homogeneity as well as polarization. When an actor has homogeneous first-order neighbors, and these neighbors have more diverse second-order neighbors (i.e., a positive gap between the ego's and the average of alter's neighborhood homogeneity), the first-order neighbors are more likely to interact with the actor than with their own neighbors, due to homophily. In this positive gap situation, the actor and their neighbors are likely to maintain their cultural states over time, reinforcing each other and allowing their cultural traits to diffuse, eventually producing both local and global homogeneity. In contrast, when an actor has heterogeneous first-order neighbors whose second-order neighbors are more homogenous (i.e., negative gap), the first-order neighbors are more likely to interact with the second-order neighbors, rather than with the actor. Thus, these neighbors are less likely to be influenced by the actor, making the actor increasingly "isolated" since their neighbors adopt others' cultural states over time.

These gaps and the subsequent diffusion process emerge because, on

average, the actor's network consists of many more long-range ties than clustered ties (i.e., higher *L/C* ratio). A high *L/C* ratio means little overlap between an actor's first-order and second-order neighbors, making the network less clustered locally. This creates opportunities for both positive and negative gaps between ego's and the average of alter's neighborhood homogeneity.

Consider a scenario with low *L/C* ratio (as in the classic small-world network) where most ties are clustered and only a few long-range ties exist. An actor and her neighbors share the same cultural states, and their second-order neighbors largely overlap due to the local clustering, interactions are likely to remain within this locally tight-knit group. This reinforces locally shared cultural traits but hinders diffusion through long-range ties, thus preserving local cultural homogeneity and global diversity— a pattern seen in the classic small-world model in Fig. 4.

In contrast, in a network with high *L/C* ratio, where actors have a much larger size of long-range ties (= 658) than clustered ties (= 20), their first-order neighbors have less overlap with their second-order neighbors. A positive gap can emerge when an actor shares the same cultural states with some of their neighbors (e.g., Group B actors at iteration 1300 in Fig. 8). Then, by the probabilistic partner choice, second-order neighbors may interact with dissimilar first-order neighbors, and thereby their cultural states are able to diffuse through long-range ties — eventually leading to global homogeneity due to the shorter average shortest-paths.[7] Meanwhile, a negative gap may arise when an actor's culturally diverse first-order neighbors are largely influenced by their more homogenous second-order neighbors eventually making the actor[8]culturally isolated (e.g., Group A at iteration 2000 in Fig. 8). In short, a higher L/C ratio shapes both types of neighborhood homogeneity gaps,[9] producing wider diffusion in some parts of the network and isolation in others, and thereby contributing to the emergence of cultural homogeneity as well as polarization.

7. Discussion and conclusion

Our study suggests an ARD-informed parameter for network modeling in ABMs. Extending the small-world network model of Watts and Strogatz (1998), the parameter *L* helps to specify various structures of long-range ties, allowing the customization of the mean, standard deviation, and skewness in the distribution. We empirically calibrate the parameter based on a recent ARD study (Lubbers et al., 2019) and apply it to a widely known classic ABM (Axelrod, 1997). Results show that, once we incorporate a realistic distribution of long-range ties, the emergent patterns of local and global cultural homogeneity are very different from the original model. Both local and global cultural homogeneities start to emerge rapidly after a long time of stasis, and node-level homogeneities (i.e., proportion of neighbors whose cultural states are identical to ego's) become polarized. Additional exploratory analyses show that these patterns are strongly related to the distributional features of long-range ties specified by the ARD study (e.g., large number of long-range ties compared to clustered ties; wide variation and skewness) that are under-explored in existing stylized network models.

A major contribution of our study is that it shows that the usefulness of ARD studies is not only limited to empirical understanding of social networks at the population scale, but also can be extended to theory building. Agent-based modeling is widely known to be useful for clarifying how simple micro-level mechanism under certain macro-level constraints can lead to the emergence of (unexpected) macro-phenomenon (i.e., macro-micro-macro) (Bianchi and Squazzoni, 2015; Flache et al., 2022; Schelling, 2006; Smaldino et al., 2015; Smith and Conrey, 2007). Although many ARD and NSUM-based studies are effective in revealing hidden social network features and their implications for empirical questions (Bernard et al., 2010; DiPrete et al., 2011; Hofstra et al., 2021; McCormick, 2020), there has been a lack of theoretical exploration into how these network features may be related to broader social phenomena (e.g., political polarization, cultural convergence, social cohesion and conflict). The modeling parameter *L* suggested here allows these explorations with a customizable long-range tie structure in a small-world network. Our study shows how ARD-based information about extended social networks can be formally integrated and modeled in ABMs under the assumption that non-core social ties—such as acquaintance relationships—can be used as an empirical analogue to long-range ties. In addition, when ARD data include multiple types of acquaintance relationships (e.g., whose name is Maria or occupation/race/migrant status is *X*) — including through the use of positional generators (Baum and Marsden, 2023; Marsden et al., 2021), this approach can support building richer network theories that account for varying types of non-core relationships—an under-theorized aspect in existing network theories which tend to focus on close relationships (Rivera et al., 2010).

From an agent-based modeling perspective, our study underscores the importance of considering realistic social networks in ABMs. Our findings show that the large number of long-range ties compared to clustered ties as well as their distributional characteristics (i.e., wide variation and skewness), which are found in ARD studies but under-explored in stylized network models, crucially shape the outcomes of Axelrod's cultural dynamic model. This means applying stylized network models in ABMs may be inherently limited in revealing the full possible range of macro-outcomes, as the scopes of these stylized models do not fully reflect realistic social networks but rather aim to solve some graph-theoretic puzzles embedded in social networks. For instance, the scope of the small-world network model of Watts and Strogatz (1998) is to answer the puzzle of how people know each other within six degrees in the sparsely connected world. In this sense, the condition that nodes have more long-range ties than clustered ties is unnecessary because it obviously shortens the average shortest-path length between nodes. Yet, this can be an important condition to reflect realistic social networks in ABMs. Our solution to address the gap is to suggest an ARD-informed (or empirically driven) network modeling parameter, which is in line with previous studies embedding ABMs into realistic empirical contexts (e.g., modeling social networks; using population-level data) (Bruch and Atwell, 2015; Hamill and Gilbert, 2009; Rolfe, 2014) and recent calls for integrating ABMs with large-scale empirical data (Flache et al., 2022).

It is also noteworthy that ARD provides valuable insights into various aspects of personal social networks at the population level, although we focused only on degree-related features, specifically the distributional properties of acquaintance relationships. ARD can also be used to estimate proportions of social groups with certain attributes (McCormick, 2020), or levels of segregation between groups (DiPrete et al., 2011; Park, 2021). Integrating this information into ABMs could further enhance their *realism* — that is, the extent to which the underlying network models reflect actual patterns of social structure and interactions observed in real populations, including non-random patterns of connections (e.g., network segregation). This, in turn, can help reveal a wider range of potential emergent outcomes than what is possible with simple, stylized network models based on overly abstract or unrealistic

[7] In other words, global diversity cannot be sustained in the network due to the noise or randomness of homophily process (i.e., actors sharing identical cultural states are likely to interact, but not always), which is in line with the finding of modified Axelrod's model in Flache and Macy (2006).

[8] This is in line with the finding of Keijzer et al. (2018) that culturally isolated actors are emerging when the influence process is one-to-many. Although our study assumes the one-to-one influence process, following Axelrod's original assumption, the same cultural isolation emerges because the negative gaps and the extended network structure make some actors have neighbors, many of which are largely influenced by their neighbors (2nd order neighbors), which can be also similarly driven by the one-to-many rule.

[9] We note that a similar phenomenon of node-level homogeneity gaps was also documented by (Evtushenko and Kleinberg, 2021) in static networks (with fixed node attributes). Our study illustrates the dynamic process through which such gaps emerge and evolve (where node attributes can change), and their implications for global outcomes.

structural assumptions. Although our study takes a small step in this direction by calibrating degree distribution based on empirical ARD findings, the network generation process remains limited in its ability to incorporate empirically realistic generative mechanisms. Future research could further explore how additional ARD-based estimates could help constrain the sampling space of random graphs, support inferences of full network structures—particular in the case of egocentric network samples (Krivitsky and Morris, 2017)—or be integrated into generative models like exponential random graph models (ERGM) to yield more realistic conditional distributions of random networks (Felmlee et al., 2021).

While our study represents an initial step in bridging ABM studies and ARD research, it is important to acknowledge a few limitations. First, although the proposed parameter allows researchers to flexibly specify the distributional properties of long-range ties, the choice of this parameter should be guided by empirical insights. This implies that researchers should have well-informed estimates about the distribution of long-range ties, carefully choosing and calibrating the model parameters either based on existing studies or through post hoc analysis. In other words, in our implementation of the extended small-world network, consideration of more realistic social network features in ABMs is built into the starting condition of the model by drawing findings from recent ARD studies. However, this empirical anchoring may not always be desirable, as it can introduce constraints depending on the purpose of specific ABMs. For instance, for ABM studies focusing exclusively on theoretical exploration rather than empirical plausibility, calibrating the network structure based on realistic features may limit the generality and the tractability of the model. A closely related practical constraint may simply be the computational cost of simulating on very large networks. This trade-off between empirical plausibility and computational feasibility should be considered when designing ABMs to simulate social processes at scale.

Additionally, the use of the parameter L must be theoretically grounded within the context of assumed micro-processes involving long-range ties. Many follow-up ABM studies influenced by Axelrod's model extend beyond cultural dynamics to study changes in political opinions, such as polarization or convergence of political opinions (Baldassarri and Bearman, 2007; Banisch and Olbrich, 2019; DellaPosta et al., 2015; Flache and Macy, 2011b; Giardini and Vilone, 2021). Yet, this does not mean that the parameter L is universally applicable across different models, as each model assumes a specific social influence process—some of which may not necessarily be mediated by long-range or non-core ties. For instance, discussions about politically sensitive topics (e.g., abortion, immigration, or climate change) may only occur within close, trusted relationships. This means that in order to claim the theoretical implications of an ABM with the parameter L, the assumed micro-process needs to be plausible within the scope of long-range ties.

A more general challenge underlying this issue is how we define what constitutes more "realistic" networks. In our case, we draw on NSUM estimates of extended social networks—such as acquaintance relationships—as an empirical analogue for "long-range" ties. This approach prioritizes having more realistic structural, topological properties of networks while abstracting away from various distinctions in tie strength or the specific content of the relationship. For instance, the definition of "acquaintances" used in NSUM studies often encompasses a mixture of strong and weak ties (DiPrete et al., 2011). It is closer to a measure of a "global" or active network (Marsden and Hollstein, 2023) rather than ties related to a focused domain or associated with a specific content (Perry and Roth, 2021). Can we use "acquaintances" as a meaningful and realistic proxy when constructing stylized networks in ABMs? This issue reflects a deeper, well-known problem in empirical social network research: the problem of network boundary specification (Laumann et al., 1989; Marsden, 1990; Perry and Roth, 2021) — namely, determining which set of nodes and relationships should be included when defining and measuring a network. This problem has largely been sidestepped in ABM research through theoretical abstraction of network structures. In this light, the theoretical exercises in our study also reveal a critical gap between ABM studies and empirical social network research — how stylized network structures map onto specific types of social ties in real populations, and how assumed micro-interaction rules align with empirically plausible mechanisms in real-world contexts (see Keijzer et al. 2024 for a recent example using behavioral experiments to ground micro-interaction rules empirically). In other words, ABM research faces its own set of boundary specification problems—often not explicitly acknowledged—that merit more scrutiny. This issue becomes especially critical when we attempt to integrate empirical social network research into ABMs. Despite these limitations, we believe that our study provides a fruitful starting point for future research in leveraging large-scale empirical data and insights to inform and enrich ABM research.

CRediT authorship contribution statement

Xinwei Xu: Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Yunsub Lee:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization.

Acknowledgements

We would like to thank Miranda J. Lubbers, Michal Bojanowski, Alejandro Ciordia, and members of COALESCE Lab for their helpful comments and advice. We are also grateful to two anonymous reviewers for their valuable feedback. This article is part of the PATCHWORK project. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No.101020038). The second author is also supported by funding from the Swiss National Science Foundation (SNSF, grant No.10001512).

References

Antelmi, A., Caramante, P., Cordasco, G., D'Ambrosio, G., De Vinco, D., Foglia, F., Postiglione, L., Spagnuolo, C., 2024. Reliable and efficient Agent-Based modeling and simulation. J. Artif. Soc. Soc. Simul. 27.
Axelrod, R., 1997. The dissemination of culture: a model with local convergence and global polarization. J. Confl. Resolut. 41, 203–226.
Baldassarri, D., Bearman, P., 2007. Dynamics of political polarization. Am. Sociol. Rev. 72, 784–811.
Banisch, S., Olbrich, E., 2019. Opinion polarization by learning from social feedback. J. Math. Sociol. 43, 76–103.
Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.
Baum, D.S., Marsden, P.V., 2023. Uses and limitations of dichotomous aggregate relational data. Soc. Netw. 74, 42–61.
Bernard, H.R., Hallett, T., Iovita, A., Johnsen, E.C., Lyerla, R., McCarty, C., Mahy, M., Salganik, M.J., Saliuk, T., Scutelniciuc, O., 2010. Counting hard-to-count populations: the network scale-up method for public health. Sex. Transm. Infect. 86, ii11–ii15.
Bianchi, F., Squazzoni, F., 2015. Agent-based models in sociology. Wiley Interdiscip. Rev. Comput. Stat. 7, 284–306.
Bokányi, E., Heemskerk, E.M., Takes, F.W., 2023. The anatomy of a population-scale social network. Sci. Rep. 13, 9209.
Bravo, G., Squazzoni, F., Boero, R., 2012. Trust and partner selection in social networks: an experimentally grounded model. Soc. Netw. 34, 481–492.
Breza, E., Chandrasekhar, A.G., McCormick, T.H., Pan, M., 2020. Using aggregated relational data to feasibly identify network structure without network data. Am. Econ. Rev. 110, 2454–2484.
Bruch, E., Atwell, J., 2015. Agent-Based models in empirical social research. Sociol. Methods Res. 44, 186–221.
Centola, D., Gonzalez-Avella, J.C., Eguiluz, V.M., San Miguel, M., 2007. Homophily, cultural drift, and the co-evolution of cultural groups. J. Confl. Resolut. 51, 905–929.
Chetty, R., Jackson, M.O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R.B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., 2022. Social capital I: measurement and associations with economic mobility. nature 608, 108–121.
DellaPosta, D., 2020. Pluralistic collapse: the "oil spill" model of mass opinion polarization. Am. Sociol. Rev. 85, 507–536.
DellaPosta, D., Shi, Y., Macy, M., 2015. Why do liberals drink lattes? Am. J. Sociol. 120, 1473–1511.
DiPrete, T.A., Gelman, A., McCormick, T., Teitler, J., Zheng, T., 2011. Segregation in social networks based on acquaintanceship and trust. Am. J. Sociol. 116, 1234–1283.

Evtushenko, A., Kleinberg, J., 2021. The paradox of second-order homophily in networks. Sci. Rep. 11, 13360.
Fagiolo, G., Valente, M., Vriend, N.J., 2007. Segregation in networks. J. Econ. Behav. Organ. 64, 316–336.
Fagiolo, G., Valente, M., Vriend, N.J., 2009. A dynamic model of segregation in small-world networks. Springer.
Feld, S.L., 1991. Why your friends have more friends than you do. Am. J. Sociol. 96, 1464–1477.
Felmlee, D., McMillan, C., Whitaker, R., 2021. Dyads, triads, and tetrads: a multivariate simulation approach to uncovering network motifs in social graphs. Appl. Netw. Sci. 6, 63.
Flache, A., Macy, M.W., 2006. What sustains cultural diversity and what undermines it? Axelrod and beyond. arXiv preprint physics/0604201.
Flache, A., Macy, M.W., 2011a. Local convergence and global diversity: from interpersonal to social influence. J. Confl. Resolut. 55, 970–995.
Flache, A., Macy, M.W., 2011b. Small worlds and cultural polarization. J. Math. Sociol. 35, 146–176.
Flache, A., Mäs, M., Keijzer, M.A., 2022. Computational approaches in rigorous sociology: agent-based computational modeling and computational social science. Handbook of Sociological Science. Edward Elgar Publishing, pp. 57–72.
Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., Lorenz, J., 2017. Models of social influence: towards the next frontiers. J. Artif. Soc. Soc. Simul. 20.
Giardini, F., Vilone, D., 2021. Opinion dynamics and collective risk perception: an agent-based model of institutional and media communication about disasters. JASSS J. Artif. Soc. Soc. Simul. 24, 4.
Goldberg, A., Stein, S.K., 2018. Beyond social contagion: associative diffusion and the emergence of cultural variation. Am. Sociol. Rev. 83, 897–932.
Granovetter, M.S., 1973. The strength of weak ties. Am. J. Sociol. 78, 1360–1380.
Hamill, L., Gilbert, G., 2009. Social circles: a simple structure for agent-based social network models. J. Artif. Soc. Soc. Simul. 12.
Hill, R.A., Dunbar, R.I., 2003. Social network size in humans. Hum. Nat. 14, 53–72.
Hofstra, B., Corten, R., van Tubergen, F., 2021. Beyond the core: who has larger social networks? Soc. Forces 99, 1274–1305.
Hosseini, S., Abdollahi Azgomi, M., Rahmani Torkaman, A., 2016. Agent-based simulation of the dynamics of malware propagation in scale-free networks. Simulation 92, 709–722.
Hui, C., Goldberg, M., Magdon-Ismail, M., Wallace, W.A., 2010. Simulating the diffusion of information: an agent-based modeling approach. Int. J. Agent Technol. Syst. (IJATS) 2, 31–46.
Ishiguro, I., 2016. Extroversion and neuroticism affect the right side of the distribution of network size. Soc. Netw. 44, 219–225.
Jeroense, T., Hofstra, B., Spierings, N., Tolsma, J., 2024. Size and ethnic homogeneity of extended social networks in the Netherlands: differences between migrant groups and migrant generations. Int. Migr.
Keijzer, M.A., Mäs, M., Flache, A., 2018. Communication in online social networks fosters cultural isolation. Complexity 2018, 9502872.
Keijzer, M.A., Mäs, M., Flache, A., 2024. Polarization on social media: Micro-level evidence and macro-level implications. JASSS 27, 7.
Killworth, P.D., Johnsen, E.C., McCarty, C., Shelley, G.A., Bernard, H.R., 1998. A social network approach to estimating seroprevalence in the United States. Soc. Netw. 20, 23–50.
Klemm, K., Eguíluz, V.M., Toral, R., San Miguel, M., 2003. Nonequilibrium transitions in complex networks: a model of social interaction. Phys. Rev. E 67, 026120.
Krivitsky, P.N., Morris, M., 2017. Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. Ann. Appl. Stat. 11, 427.
Laumann, E.O., Marsden, P.V., Prensky, D., 1989. The boundary specification problem in network analysis. Res. Methods Soc. Netw. Anal. 61, 176–179.
Lewis, K., 2024. Digital networks: elements of a theoretical framework. Soc. Netw. 77, 31–42.
Liu, D., Chen, X., 2011. Rumor propagation in online social networks like twitter–a simulation study. Proceedings of the 2011 Third International Conference on Multimedia Information Networking and Security. IEEE, pp. 278–282.
Lombardo, G., Pellegrino, M., Tomaiuolo, M., Cagnoni, S., Mordonini, M., Giacobini, M., Poggi, A., 2022. Fine-grained agent-based modeling to predict COVID-19 spreading and effect of policies in large-scale scenarios. IEEE J. Biomed. Health Inform. 26, 2052–2062.
Lubbers, M.J., Molina, J.L., Valenzuela-García, H., 2019. When networks speak volumes: variation in the size of broader acquaintanceship networks. Soc. Netw. 56, 55–69.
Macy, M.W., Willer, R., 2002. From factors to factors: computational sociology and agent-based modeling. Annu. Rev. Sociol. 28, 143–166.
Manzo, G., 2014. Data, generative models, and mechanisms: more on the principles of analytical sociology. Anal. Sociol. 1–52.
Mark, N.P., 2003. Culture and competition: homophily and distancing explanations for cultural niches. Am. Sociol. Rev. 319–345.
Marsden, P.V., 1990. Network data and measurement. Annu. Rev. Sociol. 16, 435–463.
Marsden, P., Fekete, M., Baum, D., 2021. On the general social survey. Egocentric Network Studies Within The General Social Survey: Measurement Methods, Substantive Findings, And Methodological Research. Structural Analysis in the Social Sciences. Cambridge Univ. Press, pp. 519–552.
Marsden, P.V., Hollstein, B., 2023. Advances and innovations in methods for collecting egocentric network data. Soc. Sci. Res. 109, 102816.
Mäs, M., Flache, A., 2013. Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. PloS One 8.
McCormick, T.H., 2020. The network scale-up method. The Oxford Handbook of Social Networks. Oxford University Press New York, pp. 153–169.
McCormick, T.H., Salganik, M.J., Zheng, T., 2010. How many people do you know?: efficiently estimating personal network size. J. Am. Stat. Assoc. 105, 59–70.
Park, B., 2021. Segregated in social space: the spatial structure of acquaintanceship networks. Sociol. Sci. 8, 397–428.
Perry, B.L., Roth, A.R., 2021. On the boundary specification problem in network analysis: an update and extension to personal social networks. Pers. Netw. 431–443.
Rahmandad, H., Sterman, J., 2008. Heterogeneity and network structure in the dynamics of diffusion: comparing agent-based and differential equation models. Manag. Sci. 54, 998–1014.
Raub, W., Buskens, V., Van Assen, M.A., 2011. Micro-macro links and microfoundations in sociology. J. Math. Sociol. 35, 1–25.
Rivera, M.T., Soderstrom, S.B., Uzzi, B., 2010. Dynamics of dyads in social networks: assortative, relational, and proximity mechanisms. Annu. Rev. Sociol. 36, 91–115.
Rolfe, M., 2014. Social networks and agent-based modelling. Anal. Sociol. 233–260.
Schelling, T.C., 1971. Dynamic models of segregation. J. Math. Sociol. 1, 143–186.
Schelling, T.C., 2006. Micromotives and Macrobehavior. WW Norton & Company.
Smaldino, P.E., Calanchini, J., Pickett, C.L., 2015. Theory development with agent-based models. Organ. Psychol. Rev. 5, 300–317.
Smith, E.R., Conrey, F.R., 2007. Agent-based modeling: a new approach for theory building in social psychology. Personal. Soc. Psychol. Rev. 11, 87–104.
Snijders, T.A., 2011. Statistical models for social networks. Annu. Rev. Sociol. 37, 131–153.
Squazzoni, F., 2008. The micro-macro link in social simulation. Sociologica 2, 0-0.
Squazzoni, F., 2012. Agent-Based Computational Sociology. John Wiley & Sons.
Toffoli, T., Margolus, N., 1987. Cellular Automata Machines: A New Environment for Modeling. MIT press.
Vacca, R., Bilecen, B., Lubbers, M.J., 2025. Social networks in migration and migrant incorporation: new developments and challenges. Int. Migr. 63, e13373.
Wang, H., Wellman, B., 2010. Social connectivity in america: changes in adult friendship network size from 2002 to 2007. Am. Behav. Sci. 53, 1148–1169.
Watts, D.J., 1999. Networks, dynamics, and the small-world phenomenon. Am. J. Sociol. 105, 493–527.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world'networks. nature 393, 440.
Wilensky, U., Rand, W., 2015. An Introduction to Agent-based Modeling: Modeling Natural, Social, and Engineered Complex Systems With NetLogo. Mit Press.
Wittek, P., Rubio-Campillo, X., 2012. Scalable agent-based modelling with cloud hpc resources for social simulations. 4th IEEE International Conference on Cloud Computing Technology and Science Proceedings. IEEE, pp. 355–362.
Zheng, T., Salganik, M.J., Gelman, A., 2006. How many people do you know in prison? Using overdispersion in count data to estimate social structure in networks. J. Am. Stat. Assoc. 101, 409–423.