

## NETWORK SCIENCE

# Cumulative effects of triadic closure and homophily in social networks

Aili Asikainen<sup>1</sup>, Gerardo Iñiguez<sup>2,1,3,4</sup>, Javier Ureña-Carrión<sup>1</sup>, Kimmo Kaski<sup>1,5</sup>, Mikko Kivela<sup>1\*</sup>

Social network structure has often been attributed to two network evolution mechanisms—triadic closure and choice homophily—which are commonly considered independently or with static models. However, empirical studies suggest that their dynamic interplay generates the observed homophily of real-world social networks. By combining these mechanisms in a dynamic model, we confirm the longheld hypothesis that choice homophily and triadic closure cause induced homophily. We estimate how much observed homophily in friendship and communication networks is amplified due to triadic closure. We find that cumulative effects of homophily amplification can also lead to the widely documented core-periphery structure of networks, and to memory of homophilic constraints (equivalent to hysteresis in physics). The model shows that even small individual bias may prompt network-level changes such as segregation or core group dominance. Our results highlight that individual-level mechanisms should not be analyzed separately without considering the dynamics of society as a whole.

## INTRODUCTION

One of the most important traits of human sociality is homophily (1), the tendency of similar people to be connected to each other due to their shared biological and cultural attributes such as gender, occupation, or political affiliation. Homophily has been observed across various social networks (1–5), and it is a major force behind several pressing social issues including inequality, segregation, and online echo chambers (6–8). Thus, a thorough quantitative understanding of the network mechanisms leading to homophily (9–12) is essential for promoting a sufficient flow of information (1, 13) and equal opportunity in social networks of individuals with diverse personal preferences.

The homophily observed in social networks is often attributed either to choice homophily, defined by people's preference when choosing whom to connect with, or to induced homophily, rising from constraints in the opportunities of individuals to make connections (2). These two mechanisms and their relative importance have long been a subject of study in social network research (14, 15). However, as suggested by longitudinal empirical results (16, 17), the two mechanisms of homophily generation cannot be separated without considering the cumulative advantage-like dynamics (10) driving the evolution of social networks: Choice homophily creates circumstances for induced homophily, such as groups of similar people interacting, which are then further reinforced in cycles of choice and induced homophily. While the dynamics of homophily is well understood in the case of tipping point models of residential segregation (18, 19), a similar understanding of the dynamics in social networks is still needed to validate and measure homophily amplification (17).

Here, we introduce a minimal model of social network evolution to analyze to what extent the structural constraints caused by triadic closure and choice homophily interact. The triadic closure mechanism

uses the existing social network to create new connections between people who share common friends, acquaintances, or other connections. This mechanism has been reported as the most common structural constraint (16) and can explain many salient features of empirical social networks. These include a high number of closed triangles between acquaintances and fat-tailed degree distributions (20–23). Thus, triadic closure should be considered as the main mechanism in most minimal dynamic social network models (20, 24, 25). In our approach, individuals are considered to belong to either of two groups (*a* and *b*) representing the values of a static attribute of interest (gender, class, party, etc.), and they rewire their connections by two mechanisms: triadic closure (modeling the creation of edges via current contacts) and random rewiring [emulating any unknown mechanisms beyond triadic closure, such as focal closure in large foci (16, 26)]. Choice homophily/heterophily is implemented by accepting new links with a bias probability dependent on the similarity of attributes between individuals.

In this study, we characterize the rich tapestry of emergent behavior captured by the model with a mean-field bifurcation analysis for varying relative group sizes, triadic closure probabilities, and choice homophily rates. We measure the amount of observed homophily in the network, which we interpret as the sum of choice homophily (a parameter in the model) plus the induced homophily caused by triadic closure acting on a homophilic social network. By tuning the parameters of the system with empirical data on friendship and communication networks, we find that, under the right circumstances, even a small amount of choice homophily may be greatly amplified by triadic closure to produce large amounts of observed homophily. Further, we find that the interplay of triadic closure and homophily can explain the emergence of core-periphery structures. These findings suggest that the observations of homophilous patterns of association in society should not be explained solely on the basis of a human preference for similarity, but as a constantly evolving interplay between structural constraints and homophily, one that requires computational simulation as a central part of the analysis.

## Model definition and parameters

We introduce a model of social network evolution with a simultaneous interplay of triadic closure and choice homophily (see Fig. 1, A to E

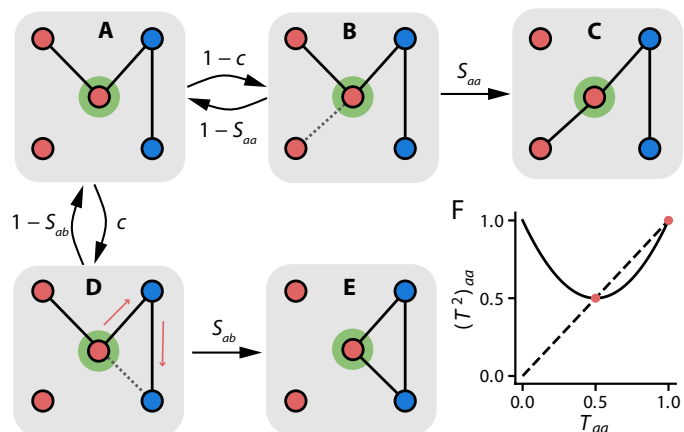
Copyright © 2020  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>Department of Computer Science, School of Science, Aalto University, FI-00076, Finland. <sup>2</sup>Department of Network and Data Science, Central European University, H-1051 Budapest, Hungary. <sup>3</sup>Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, CDMX-01000, Mexico. <sup>4</sup>Next Games, FI-00100 Helsinki, Finland. <sup>5</sup>The Alan Turing Institute, British Library, London NW1 2DB, UK.

\*Corresponding author. Email: mikko.kivela@aalto.fi

for an illustration and further model details). The model is stylized such that it contains a minimal set of simple mechanisms on how social relationships are made and forgotten; details beyond these core mechanisms are modeled by uniform randomness to assume the least amount of information on them. The initial social structure is a random network with static attribute groups  $a$  and  $b$  (of relative sizes  $n_a$  and  $n_b$ , with  $n_a + n_b = 1$ ) distributed among nodes uniformly at random and independently of the initial network structure, such that there is a fraction  $P_{ab} = P_{ba}$  of edges between groups, and fractions  $P_{aa}, P_{bb}$  within each group ( $P_{ab} + P_{aa} + P_{bb} = 1$ ).

From its initial state, the network evolves with nodes updating their connections. At each time step, we select a focal node uniformly at random and a candidate neighbor, representing a social encounter that might lead to a new social relationship. The candidate neighbor is chosen by triadic closure with probability  $c$  or uniformly at random otherwise [emulating any other mechanisms for edge creation beyond the triadic closure; (23, 27–30)]. The triadic closure mechanism selects a candidate neighbor by randomly sampling a neighbor of the focal node and then a neighbor of the sampled neighbor. If the edge between the focal node and the candidate node cannot be created (because it would create a multi-edge or a self-loop, or because the degree of the focal node is zero), then no updates are made for the focal node at this step.



**Fig. 1. Mechanisms of triadic closure and choice homophily.** (A) A focal node selected uniformly at random (node at the center with green boundary) finds a candidate neighbor by either (B) selecting a node uniformly at random with probability  $1 - c$  or by (D) closing a triangle with probability  $c$ . (C and E) For a focal node in group  $i$ , the candidate neighbor in group  $j$  is accepted with bias probability  $S_{ij}$  (where  $i, j \in \{a, b\}$ ). We parametrize  $S_{ij}$  with tunable parameters  $s_a, s_b$  such that  $S_{aa} = s_a, S_{bb} = s_b, S_{ab} = 1 - s_a$ , and  $S_{ba} = 1 - s_b$ . If the potential edge [dashed line in (B) and (D)] is accepted, an edge of the focal node (selected uniformly at random) is replaced by one between the focal and candidate neighbor nodes. Otherwise no edges are rewired. (F) Probability  $(T^2)_{aa}$  of choosing a candidate neighbor with triadic closure from the same group as the focal node ( $a$ ) as a function of observed homophily  $o_a = T_{aa}$  for equally sized ( $n_a = n_b$ ) and equally connected groups ( $T_{aa} = T_{bb}$  and  $T_{ab} = T_{ba} = 1 - T_{aa}$ ). If the network is not randomly mixed ( $T_{aa} \neq 1/2$ ), the probability of triadic closure choosing two nodes of the same group is always larger than the same probability if the selection is done uniformly at random ( $1/2$ ), implying that triadic closure amplifies existing observed homophily in the network (or suppresses heterophily). However, triadic closure without choice homophily is not enough to maintain the observed homophily in the network, which would make  $(T^2)_{aa}$  equal to observed homophily (dotted and solid lines cross). An exception is the case  $T_{aa} = 1$ , where two completely separate components exist, and triadic closure cannot create edges between them.

The edge between the focal node and candidate neighbor is created with probability  $S_{ab}$  if the focal node is in group  $a$  and the candidate neighbor is in group  $b$ . The elements  $S_{ab}$  form a  $2 \times 2$  bias matrix specifying the amount of choice homophily/heterophily in the social network. For simplicity, we parameterize the bias matrix as  $S_{aa} = s_a, S_{ab} = 1 - s_a, S_{bb} = s_b$ , and  $S_{ba} = 1 - s_b$ , where  $s_a$  ( $s_b$ ) is the choice homophily for group  $a$  ( $b$ ). In this way, when  $s_a = s_b = 1/2$ , all of the elements of the bias matrix are also  $1/2$ , i.e., there is no choice homophily bias. Multiplying the bias matrix by a constant changes the speed of network evolution, but not the fixed points of the dynamics. Note that the bias depends only on the groups the two individuals belong to, meaning that individuals have homogeneous choice homophily preferences (31, 32).

Last, as maintaining social connections requires mental capacity and time investment, creating new connections implies forgetting some of the old ones (33). We model this process by randomly removing an edge of the focal node after a successful edge creation. This keeps the degree of the focal node unchanged (the individual making the time investment) but alters the degrees of the other two affected nodes, thus changing the degree distribution of the network. Note that random link removal may open triangles, while in reality, links that are surrounded by triangles are more likely to be strong links (24) and thus less likely to be removed. However, here we opt for random link removal (29) which does not involve these additional assumptions of social behavior and is a typical choice in this type of social network models [along with random node deletion; (26, 27)].

## RESULTS

In our approach, the interplay between homophily and triadic closure (in a social network with two attribute groups) forms a dynamical system in which the evolution from an arbitrary initial network depends on the parameters that regulate choice homophily ( $s_a, s_b$ ) and triadic closure ( $c$ ). As with any other network model, the associated stochastic process exhibits random fluctuations, but the average dynamics of key behavioral quantities such as the degree distribution, clustering coefficient, and, centrally, the observed homophily depend deterministically on the model parameters. We characterize the behavior of the network following our model dynamics using a mean-field approximation and confirm our results with numerical simulations (see Materials and Methods and the Supplementary Materials).

The amount of homophily that can be directly observed in the model network is not necessarily the same as the choice homophily parameter in the model. In a network with two groups of equal size ( $n_a = n_b$ ), the observed homophily within groups  $a$  and  $b$  (denoted by  $o_a$  and  $o_b$ ) is equal to the transition probability that following a link from a group leads to the same group ( $T_{aa}$  and  $T_{bb}$ ). Thus, when there is no triadic closure ( $c = 0$ ), the choice homophily equals the expected fraction of neighboring nodes in the same group ( $s_a = o_a = T_{aa}$  and  $s_b = o_b = T_{bb}$ ). If one of the groups is larger ( $n_a > n_b$ ) and there is no triadic closure ( $c = 0$ ), the fraction of neighbors  $T_{aa}$  in group  $a$  is larger than the choice homophily bias ( $s_a$ ). This is why we define a group-size correction in the observed homophily as

$$o_a = \frac{n_b T_{aa}}{n_a(1 - T_{aa}) + n_b T_{aa}} \quad (1)$$

and equivalently for group  $b$ . Note that this relation simplifies to  $o_a = T_{aa}$  and  $o_b = T_{bb}$  when the attribute groups are of equal size

(see Fig. 1F). The size-corrected observed homophily equals choice homophily ( $o_a = s_a$  and  $o_b = s_b$ ) when there is no triadic closure ( $c = 0$ ; see Materials and Methods). In the case of arbitrary triadic closure ( $c \geq 0$ ), we further define induced homophily  $i_a$  such that the observed homophily is the sum of induced homophily and choice homophily.

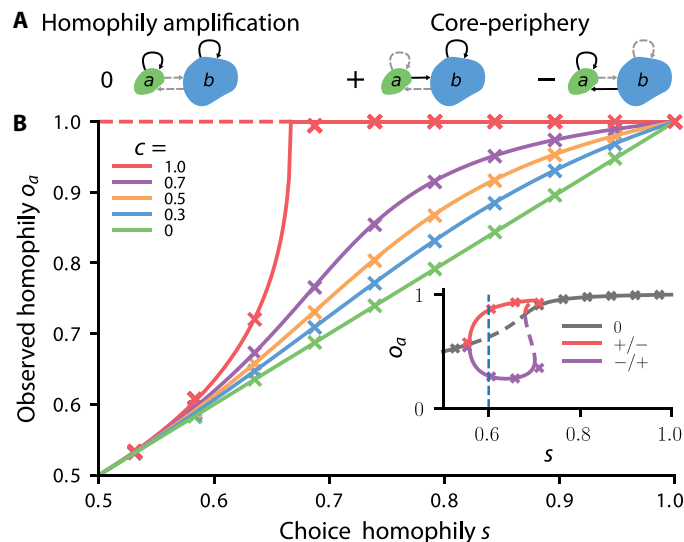
$$o_a = s_a + i_a \quad (2)$$

Consistently, for  $c = 0$ , we have  $i_a = 0$ , and all observed homophily in the social network is choice homophily.

When individuals form new edges using the triadic closure mechanism in a network with existing homophilous patterns of connectivity, they link to their own group even without having an explicit choice homophily bias (Fig. 1F). This process increases the observed homophily ( $o_a$ ) in the network beyond that due to choice homophily ( $s_a$ ), which in turn increases the likelihood for homophilic connections in upcoming triadic closure events. That is, the existing observed homophily originally due to choice homophily creates induced homophily (more opportunities for similar people to meet), which in turn creates even more induced homophily. We call the result of this cumulative advantage-like cycle homophily amplification, since the amount of observed homophily ( $o_a$ ) is larger than the amount of choice homophily ( $s_a$ ) (see Fig. 2A for an illustration). The results of this cumulative homophily amplification are shown for equally sized groups and symmetric choice homophily ( $s = s_a = s_b$ ) in Fig. 2B. In the extreme case of no random rewiring (lack of other mechanisms of edge creation), even a moderate choice homophily bias ( $s \geq 2/3$ ) will segregate the social network into fully disconnected groups (see Fig. 2B for the mean-field solution and the Supplementary Materials for a derivation of this result).

In addition to homophily amplification, the triadic closure mechanism and choice homophily may also lead to a core-periphery (34–36) social structure where the core group mostly connects with itself, while the periphery group almost exclusively connects with the core group even in the presence of choice homophily (see Fig. 2A for an illustration of the core-periphery structure and Fig. 2B inset for the analytic and simulation results). This effect, seemingly opposed to the drive of individuals to find homophilous connections in the periphery group, is due to the large likelihood of finding a candidate neighbor in the core group while attempting to close triangles (Fig. 1D). This likelihood is larger when the connectivity between the two groups is weak, which explains the role of choice homophily in the formation of the core-periphery structure (see the Supplementary Materials for a schematic similar to Fig. 1F). Homophily amplification and core-periphery are two competing results of the interaction between triadic closure and choice homophily. A core-periphery social structure is possible within otherwise symmetric networks with high triadic closure when there is enough choice homophily to boost the core-periphery structure, but not enough to tip the balance toward two tightly interconnected groups. Alternatively, a core-periphery structure can appear when there is enough asymmetry in the social network due to unequal group sizes or choice homophily biases (Fig. 3).

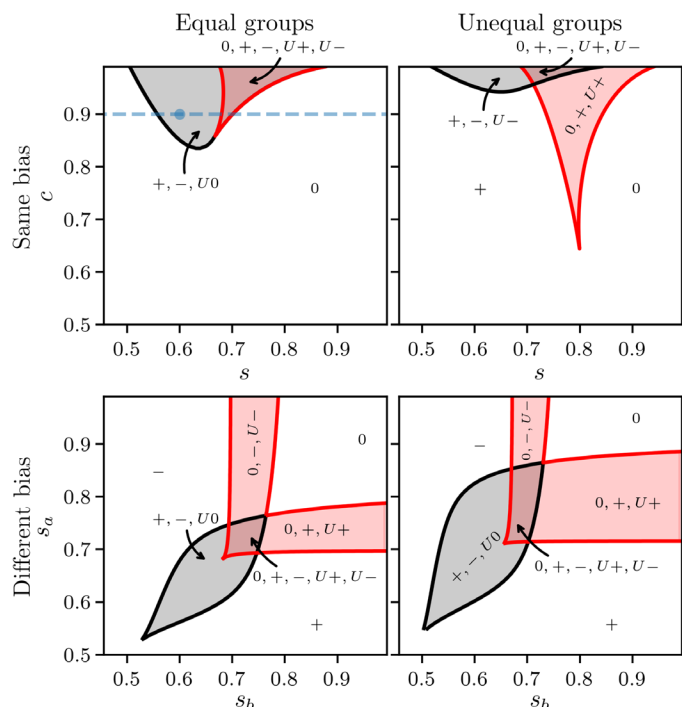
The rise of homophily amplification and core-periphery social structures depends not only on the parameters regulating triadic closure and choice homophily but also on the initial conditions and random fluctuations of network evolution, meaning that the system exhibits memory of previous structural constraints, or homophily hysteresis (see Fig. 2B for an example where either of the two groups can become the core, and Fig. 3 for a more systematic analysis). In



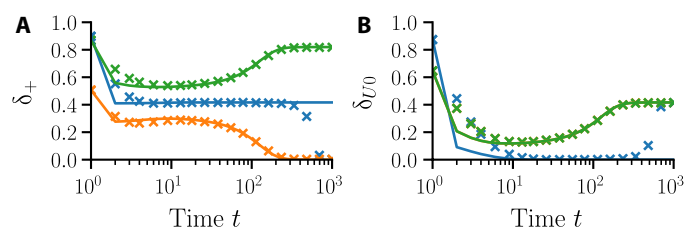
**Fig. 2. Interplay between triadic closure and choice homophily.** (A) Schematics of the stationary states the network can converge to during the rewiring process (stable fixed points of the model dynamics based on a mean-field analysis; see the Supplementary Materials). Continuous (dashed) lines represent large (low) intra- and intergroup connectivity. The homophily amplification fixed point 0 has many edges within groups and a few between (high  $T_{aa}$  and  $T_{bb}$  values). In the core-periphery fixed point + (–), the large (small) group forms the core and attracts edges from the small (large) periphery group:  $T_{ab}$  and  $T_{bb}$  ( $T_{ba}$  and  $T_a$ ) are high, while  $T_{aa}$  and  $T_{ba}$  ( $T_{bb}$  and  $T_{ab}$ ) are low (see the Supplementary Materials for details on fixed point classification). (B) Observed homophily  $o_a = T_{aa}$  (fraction of neighbors in the same group of a focal node in group  $a$ ) at the end of the dynamics as a function of choice homophily  $s = s_a = s_b$ . Group sizes are equal ( $n_a = n_b$ ), and triadic closure probability  $c$  is varied. Mean-field calculations (solid lines correspond to stable fixed points and dashed lines to unstable ones) and numerical simulations (crosses) agree very well (see the Supplementary Materials for systematic analysis). Inset: Case  $c = 0.9$  where both stable/unstable points (continuous/dashed lines) exist. Since  $n_a = n_b$ , fixed points +/– are equivalent, but for suitable choice homophily ( $0.56 < s < 0.70$ ), one group becomes the core and the other periphery depending on the initial network and/or random chance.

other words, if the system parameters change (i.e., choice homophily and probability of triadic closure), the social network can experience dramatic, nonreversible changes such that returning to the previous parameters does not return the system to the same final state (i.e., the same fixed point). This suggests that social networks may have persistent memory of homophily, with a structure dependent both on current choice homophily biases and their history. Therefore, we speculate that the timing of interventions aiming to reduce observed homophily or the formation of core groups in, say, an online social network can be critical. Once the network has reached a stable point of its dynamics or is close to one, it can be much more difficult to drive the system to another stable point by attempting to change the choice homophily of individuals or other parameters.

The time scales at which the social network is driven toward homophily amplification or a core-periphery structure vary greatly (Fig. 4A). Homophily amplification is generally a fast process, requiring only a few rewiring events per edge for the social network to reach a stable point. On the other hand, a core-periphery social structure evolves slowly toward equilibrium, and even if a network would eventually stabilize to a core-periphery structure, it may exhibit fast homophily amplification first (Fig. 4B). This result suggests that



**Fig. 3. Phase diagrams of available fixed points.** Areas of different shading correspond to parameter ranges in which the various fixed points exist (stable points 0, +, and – or the corresponding unstable points U0, U+, and U–) (see Fig. 2A). Note that unstable fixed points are saddle points with qualitatively similar relative connectivities than their stable variants. **(Top)** Choice homophily is the same for both groups ( $s = s_a = s_b$ ), and fixed points are shown as a function of  $s$  and  $c$ . **(Bottom)** Choice homophily for the groups is varied, and fixed points are shown as a function of  $s_a$  and  $s_b$ . Here, the triadic closure probability is fixed to  $c = 0.95$ . **(Left)** Groups are of equal size ( $n_a = n_b$ ). **(Right)** Groups are of unequal size ( $n_a = 0.1$ ). The dashed line corresponds to parameter values in Fig. 2B inset and the dot to Fig. 4A and 4B. Results in the main panel of Fig. 2B are in the white region with a single fixed point (0) (upper left panel).



**Fig. 4. Model temporal evolution.** Euclidean distance  $\delta$  in  $(T_{aa}, T_{bb})$ -space between current state of the dynamics and fixed points **(A)** + and **(B)** U0 as a function of time  $t$  (in units of the average number of times an edge is selected for rewiring) for  $n_a = n_b$ ,  $c = 0.9$ , and  $s = 0.6$  (dashed vertical line in Fig. 2B inset). Markers are simulation results, and lines are mean-field solutions. Colors correspond to three different initial network configurations. **(B)** The dynamics amplifies homophily first by quickly approaching the unstable point U0 and diverting away from it after  $\sim 100$  time steps. **(A)** The blue and orange curves slowly approach the core-periphery stable point +. The green curve approaches the other core-periphery stable point (–).

even if a real social network (in society or in online platforms) would follow our stylistic model accurately, it might not show a stable, fully realized core-periphery structure but a transient state slowly drifting toward the structural dominance of one group over the other.

If the network is first driven toward homophily amplification, the group that eventually becomes the core can depend purely on random chance (Fig. 4A).

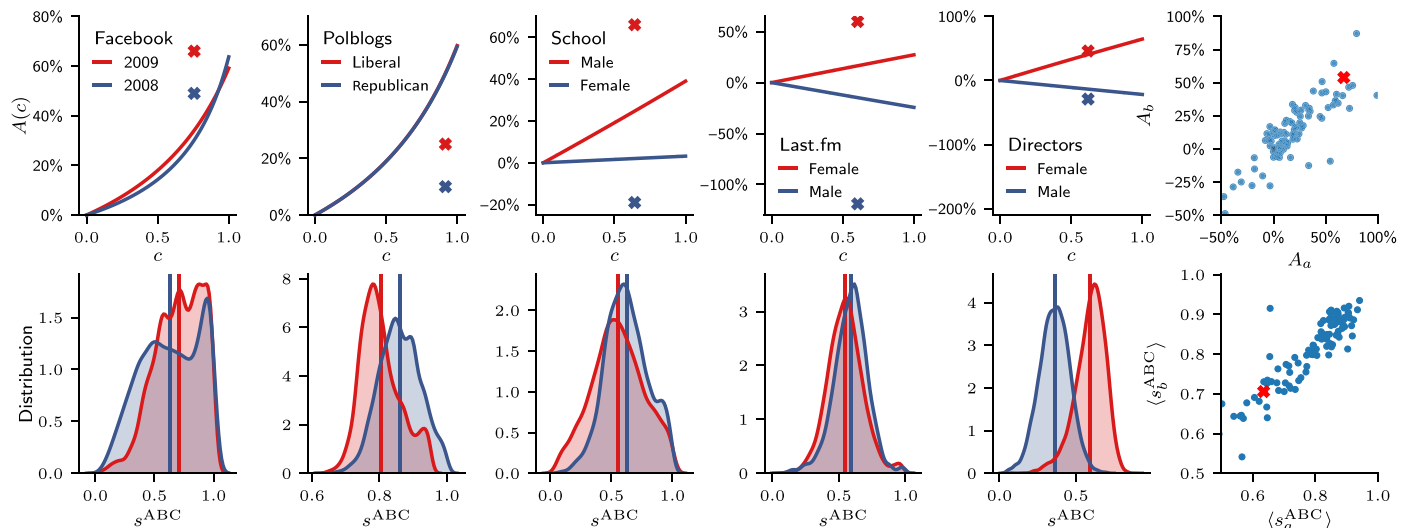
To estimate how much observed homophily differs from choice homophily in real-world social networks, as well as to find the stable point that best corresponds to their structure, we fit several empirical datasets of off- and online social interactions to our model of triadic closure and choice homophily (Fig. 5). We use two approaches for fitting the data: (i) solving for the choice homophily parameters  $s_a$  and  $s_b$  from the fixed points of the mean-field equations, given the  $T$  matrix in data and a triadic closure probability  $c$  [we denote these solutions  $s_a(c)$  and  $s_b(c)$ ], and (ii) using an approximate Bayesian computation (ABC) method (37) to find estimates for the model parameters (denoted by  $s_a^{ABC}$  and  $s_b^{ABC}$ ). To easily compare the effect of triadic closure across datasets, we first normalize choice and observed homophily such that 0 implies no homophily/heterophily in the network, and 1 implies maximum homophily ( $\hat{s} = 2s - 1$  and  $\hat{\delta} = 2\delta - 1$ ). We then define homophily amplification  $A$  as the relative difference between the normalized choice and observed homophilies

$$A = \frac{\hat{\delta} - \hat{s}}{|\hat{\delta}|} = \frac{i}{|o - 1/2|} \quad (3)$$

where induced homophily  $i$  follows Eq. 2. Note that the observed homophily  $\hat{\delta} = \hat{s}(0)$  is the same as the mean-field estimate for  $c = 0$  (see Materials and Methods for details).

In terms of fitting the mean-field behavior of the model to data, three networks show homophily amplification in both groups: a Facebook friendship network consisting of two classes in a U.S. university (38), a 1-day contact network of primary school students divided by gender (39), and a network of political blogs divided by party affiliation (40). The rest of them—a friendship network in a website for sharing music listening habits (Last.fm) and a network of company directors (41), both divided by gender—display a pattern where part of the observed homophily within one of the groups could be explained by homophily amplification, but the choice homophily in the other groups could be underestimated because of the triadic closure mechanism. In the case of the board of directors, we observe choice heterophily within males and choice homophily within females, which together with the triadic closure mechanism explain the core of female directors observed in the study where this network was introduced (41). In the mean-field fitting procedure, the maximum homophily amplification possible [ $A(c = 1)$ ] goes higher than 50% for the social network of political blogs, the largest Facebook network, and the board of directors. The exact estimate of choice homophily depends on the latent tendency for triadic closure in the network ( $c$ ). However, the parity of the amplification [ $A(c)$ ] is independent of this estimate, and the growth of amplification is monotonous as a function of  $c$ .

The qualitative picture we get for the amplification estimates using the ABC method is mostly similar to the mean-field fitting process. The changes are mostly in the scales of the effects for some datasets, and in the primary school data, the almost nonexistent amplification is now estimated as negative amplification. In addition, the ABC method gives us an estimate for the triadic closure probability ( $\langle c^{ABC} \rangle$ ), which ranges from very high (for the Facebook and political blogs networks) to medium (for the other datasets). Full posterior distributions for the choice homophily variables have a fair amount of variance, which means that the point estimates we give are indicative of expected



**Fig. 5. Homophily amplification in real-world social networks. (Top)** Estimated homophily amplification  $A$  (see Eq. 3) for empirical networks. Lines are mean-field fits as a function of triadic closure  $c$ . Points correspond to the estimated mean homophily amplification values using an ABC method, depicted at the estimated mean triadic closure ( $c$ ). Edges, nodes, and attributes in each dataset are as follows: Facebook friendships for users by graduation year in the largest of the 100 networks (Facebook), links between political blogs by party (Polblogs), friendships between students by gender (School), Last.fm friendships for users by gender (Last.fm), and links between people in boards of directors in Norway in 2010 (Directors). The smaller group  $a$  is denoted in red. (Scatter plots) Estimated homophily amplification  $A_a$  and  $A_b$ , and mean choice homophily estimates  $\langle s_a^{ABC} \rangle$  and  $\langle s_b^{ABC} \rangle$  (both from ABC method) in all 100 universities of the Facebook dataset, where the two groups  $i = a, b$  ( $n_a < n_b$ ) are graduation years. The largest network is marked with a red cross (see Materials and Methods and Table 1 for dataset details). **(Bottom)** Estimated distributions for choice homophily  $s_a^{ABC}$  and  $s_b^{ABC}$  using the ABC method for the corresponding top panel dataset (colors denote group), where vertical lines depict mean estimates ( $\langle s_a^{ABC} \rangle$  and  $\langle s_b^{ABC} \rangle$ ).

**Table 1. Properties of empirical datasets.** List of real-world social networks used in this study, their main properties, and estimated model parameters.  $N$  is the number of nodes in the network,  $\langle k \rangle$  is the average degree,  $n_a$  is the fraction of nodes in the smaller attribute group,  $HI_a = (T_{aa} - n_a)/(1 - n_a)$  and  $HI_b = (T_{bb} - n_b)/(1 - n_b)$  are the Coleman homophily indices (60) of the groups,  $s_a(0)$  and  $s_b(0)$  are the estimates of the bias parameters when  $c = 0$  [equal to observed homophily,  $s_a(0) = o_a$  and  $s_b(0) = o_b$ ], and  $s_a(1)$  and  $s_b(1)$  are the estimates when  $c = 1$ .  $\langle s_a^{ABC} \rangle$  and  $\langle s_b^{ABC} \rangle$  are the mean ABC estimates for  $s_a$  and  $s_b$ .

	$a$	$b$	$N$	$\langle k \rangle$	$n_a$	$HI_a$	$HI_b$	$s_a(0)/s_b(0)$	$s_a(1)/s_b(1)$	$\langle s_a^{ABC} \rangle$	$\langle s_b^{ABC} \rangle$
Facebook	2009	2008	14,251	42.49	47.2%	0.768	0.794	0.889/0.893	0.666/0.660	0.63	0.70
Polblogs	Liberal	Conservative	1,224	27.31	48.0%	0.813	0.810	0.909/0.902	0.665/0.664	0.804	0.861
School	Female	Male	212	4.12	49.1%	0.308	0.213	0.656/0.605	0.595/0.602	0.553	0.625
Last.fm	Male	Female	188,672	5.49	31.7%	0.168	0.106	0.621/0.540	0.588/0.550	0.55	0.588
Directors	Female	Male	2,412	6.12	35.8%	0.192	-0.382	0.658/0.321	0.554/0.372	0.594	0.36

behavior only (see Fig. 5 and the Supplementary Materials for posterior distributions for other parameter values). The ABC method also gives us a distribution of possible initial  $T$  matrices or, equivalently,  $P$  matrices (see the Supplementary Materials for related distributions). Although for most networks the initial condition is largely irrelevant, for the political blogs network, the observed homophily needs to be high in the initial condition. This is because of the political blogs network being located in a part of a parameter space with multiple fixed points (i.e., homophily hysteresis). Our model suggests that the network could have been in a core-periphery fixed point if the network had not evolved from a structurally polarized situation historically (see the Supplementary Materials for details).

A systematic analysis of 100 Facebook networks (38) reveals that the bulk of ABC estimated amplification values are positive, ranging up to the 60% value observed for the largest of these networks

(and above in a few extreme cases). Note that there are large differences between the choice homophily estimates ( $\langle s_a^{ABC} \rangle$  and  $\langle s_b^{ABC} \rangle$ ), and they are highly correlated (Pearson's  $r = 0.88$ ), which may indicate that the amount of choice homophily we observe in each university is a feature of the university and not of the student class. Results on both of our fitting procedures suggest that using observed homophily as a naive estimator for choice homophily can lead to a serious overestimation or underestimation of the intensity of homophily (even for a moderate amount of triadic closure) in several real-world social networks, both in society and online platforms.

**DISCUSSION**

Our findings show that the homophilous patterns of association typically seen in empirical social networks not only arise because of

an individual preference for similarity but are also the result of a cumulative advantage-like process that has the tendency to amplify this intrinsic bias for choice homophily due to triadic closure. By means of a minimal model of social network evolution, we find bounds on the amounts of triadic closure and choice homophily necessary for such amplification of homophily to arise. This corroborates theoretically previous observations in organizational (42) and communication (16, 17) networks. In the generic case of a moderate amount of triadic closure events and similarly sized attribute groups, choice homophily is amplified by triadic closure through a tipping point mechanism analogous to the one responsible for residential segregation in the Schelling model (18), in which segregation takes place in the social network topology rather than in physical space.

In addition to homophily amplification, our results suggest that the interplay between triadic closure and choice homophily is a plausible explanation for the emergence of the core-periphery structure found in social, communication, academic, trade, and financial networks (34–36). In such structures, the core group of individuals is so well connected that following edges via triadic closure almost always leads to the same group, making the core even more connected. While triadic closure and homophily are already considered as contributing factors in the formation of communities (cohesive and assortative groups densely connected within), the impact of node attributes on the core-periphery structure is less studied. Our model implies that the dynamic transition to core-periphery networks is slow and often preceded by fast but temporary homophily amplification. This may partly explain why the social network literature has focused on clustered networks rather than other, rarer types of intermediate-scale structures.

The coupled effects of triadic closure and choice homophily also include the memory of homophilic constraints, i.e., systems with multiple, coexisting stable points for a wide range of relative group sizes and amounts of triadic closure and choice homophily. In other words, even if choice homophily or triadic closure tendencies are changed, a social network may preserve memory of their current structural configuration. This makes it difficult to alter the shape of a stable network, for example, by varying the typical choice homophily of its individuals. On the basis of these findings, we expect that, when planning external interventions to reduce homophily-induced social segregation, measures of action should be taken sooner rather than later, since the scale of interventions with meaningful effect on the structure of the social network increases with time. Still, more research is needed before we can deliver concrete suggestions on intervention strategies.

Because choice homophily is not directly observable from static social network data, one needs to infer it from the available information. Such inference is always subject to assuming a model for data creation, and the exact estimates for choice homophily should always be interpreted with this in mind. Fitting the mean-field solution of the model to data assumes that the real-world network is in a stable state, since it is simply based on matching the linking probabilities ( $T$ ) of the data with the stable states of the model. The more elaborate ABC method involves more observables and a finite number of evolution steps. One could also include the number of steps  $t$  as a fitting parameter to investigate its effect in the convergence of the method. The results from the mean-field and ABC fitting procedures match each other qualitatively in the sense that the overall conclusions drawn from them are the same. However, the point estimate values are different in many cases, and there is a notable amount of variance

in the posterior distributions of the parameter estimates for the ABC method. This variance could potentially be reduced by allowing for larger network sizes (which are limited by computational constraints), by including different summary statistics in the discrepancy function, by further tuning the ABC method parameters, or by including additional mechanisms in the model. Applying these methods to mechanistic network models is a relatively new approach (43), and new developments in this area can be expected as these methods mature.

In contrast to our approach, stationary, nonmechanistic models, such as exponential random graphs (ERGMs), can also be used to study the interplay between triadic closure and homophily (44–46). The key conceptual difference to our approach is that ERGMs are static network models, which can be used to balance between the tendency toward triangles, homophilic edges, homophilic triangles, and many other network features as factors explaining network structure. Our approach is rather a model of cumulative interplay of two explicit and microscopic network evolution mechanisms [note, however, that some carefully crafted microscopic network evolution models can converge, under certain assumptions, to ERGMs as stable states (46, 47), and ERGMs are often sampled with Markov chain Monte Carlo methods in which networks are rewired (48)]. This means that ERGMs do not explicitly model cumulative processes or tell anything about multiple time scales or metastable states, which we find as a consequence of combining triadic closure with a choice homophily bias.

Our approach assumes that the biological and cultural attributes underlying homophily are constant in time. While this assumption is mostly true for long-term individual characteristics such as gender or religion, it is less so for traits like political affiliation, occupation, and opinions. Networks where both edges and attributes change adaptively to each other (i.e., following adaptive coevolutionary dynamics) have been studied extensively for biological, economic, and social phenomena (49–51). When edges between nodes with similar attributes are favored, the adaptive dynamics self-organizes into heterogeneous networks where groups of individuals sharing attributes are structurally distinguishable from each other (52–56). Such a generic feature of adaptive networks makes it likely that our observations of the cumulative effects of triadic closure and homophily will hold even in the case of time-dependent individual attributes (57, 58). The study of an adaptive interplay between triadic closure and homophily is a worthy line of future research that may reveal additional, complex feedback loops between social structure and attribute evolution.

The simplicity of our framework suggests that the presence of triadic closure and choice homophily for a given attribute value is enough to explain some salient features of empirical social networks like homophily amplification and core-periphery structures. Yet, the effects of more realistic features of society, such as the existence of more than two values for a single attribute, structural constraints beyond triadic closure, and the coexistence of several attributes in a population [in the spirit of the Axelrod model of cultural dissemination (59)], remain to be studied. We anticipate that our results promote even more interest in the data-driven computational simulation of social interactions and shed further light on the relationship between triadic closure and homophily. This insight will help researchers and policy-makers in devising intervention strategies to decrease the most adverse effects of homophilic decision making, including segregated social structures such as gender-specific workplaces and partisan political systems.

## MATERIALS AND METHODS

### Mean-field bifurcation analysis

We derive approximate analytical expressions for the temporal evolution of the amount of observed homophily in a social network based on a mean-field bifurcation analysis of our model. The key assumption in this approximation is that all nodes within both groups are statistically equivalent, in the sense that we only track the relative number of connections within and between groups to determine the state of the system. Otherwise, the network is considered fully random (i.e., following a stochastic block model) and of infinite size. The validity of this approximation, which omits all higher-order structure such as local clustering or degree distributions, is confirmed by comparing it to extensive numerical simulations (see section S2).

In the case of two attribute groups, the state of the system at time  $t$  can be tracked by a  $2 \times 2$  matrix  $P$ , where the element  $P_{ab}$  is the probability that an edge chosen uniformly at random lies between groups  $a$  and  $b$ . Equivalently, we may follow the dynamics of a  $2 \times 2$  transition matrix  $T$ , where element  $T_{ab}$  is the probability that following a random edge from a node in group  $a$  leads to a node in group  $b$ . We can switch from the  $P$  to the  $T$  matrix via

$$T_{aa} = \frac{2P_{aa}}{2P_{aa} + P_{ab}} \quad (4)$$

using a symmetric equation for  $T_{bb}$ , and noting that  $T_{ab} = 1 - T_{aa}$  and  $T_{ba} = 1 - T_{bb}$ . A similar bijective transformation exists from  $T$  to  $P$ .

The evolution rules of choice homophily and triadic closure in the mean field are captured in the model matrix  $M_{ab}$ , defined as the probability that, in a single time step of the dynamics, we create an edge between nodes in groups  $a$  and  $b$ , respectively, when a node from group  $a$  has been selected as the focal node. It can be written as

$$M_{ab} = [c(T^2)_{ab} + (1 - c)n_b]S_{ab} \quad (5)$$

The first term in the parenthesis on the right side of the equation is the probability that triadic closure chooses a candidate neighbor between groups  $a$  and  $b$ , and the second term gives this probability for random choice of the neighbor. This sum is then multiplied by the choice homophily probability that the link is accepted ( $S_{ab}$ ).

Using  $M_{ab}$ , we then write a rate equation describing the change in the fraction of edges within group  $a$

$$\frac{dP_{aa}}{dt} = n_a M_{aa} - n_a T_{aa}(M_{aa} + M_{ab}) \quad (6)$$

and a symmetric equation for  $P_{bb}$  (or a similar equation for  $P_{ab}$ ). Here, the first term on the right side of the equation is the rate at which the edges are created, and the second term is the rate at which they are deleted because of successful rewiring. We determine the fixed points of the rate equations and their stability, and therefore those of the mean-field dynamics, through linear stability analysis (see the Supplementary Materials for details of the analytical solution of the model and Figs. 2 to 4 for a summary of the analytical results).

### Numerical simulations

We use numerical simulations to verify the accuracy of the mean-field approximation of Eq. 6 (Fig. 2B). We first construct a random network with  $N = 10^5$  nodes and average degree  $\langle k \rangle = 50$ . To create networks with different initial conditions in terms of in- and out-group edges, we choose values for the fractions of same-group neighbors

$T_{aa}$  and  $T_{bb}$ . For simulations in Fig. 2B the initial networks have  $(T_{aa}, T_{bb}) = (0.5, 0.5)$ . For the inset, we use two initial conditions, i.e.,  $(T_{aa}, T_{bb}) = (0.1, 0.9)$  and  $(T_{aa}, T_{bb}) = (0.9, 0.1)$ . We then create two random networks so that the number of edges in each network corresponds to the desired number of in-group edges. Last, we place the remaining edges randomly between the two groups, so that the final network has  $L = N\langle k \rangle/2$  edges.

Simulations follow the model definition described above. Between times  $t$  and  $t + 1$ , we attempt to rewire  $L$  edges, so that on average, each edge in the network is chosen once. For the parameters in Fig. 2B,  $t = 10^2$  is enough for getting convergence to a fixed point, while for the parameters in the inset, we need  $t = 10^3$ . Each point in Fig. 2B is averaged over  $10^2$  realizations, with the SD smaller than the marker size (see the Supplementary Materials for a more detailed analysis of model parameters).

### Social network data

We use several large-scale social network datasets to determine empirically the possible effects of triadic closure on the observed homophily. The first one is Facebook, a friendship network of two classes at the University of Pennsylvania in the United States. The dataset includes friendships and metadata for 100 universities during 2005 (38). For each university, we use the subnetwork of the two largest classes. The second one is Polblogs, a network of political blogs collected in 2005 (40), with edges between two nodes if at least one of the blogs links to the other. Blogs are split into two groups using the classification of liberal and conservative blogs provided by the original study. The third one is School, a network between students collected by automatically sensing proximity between individuals. The original data have a 20-s time resolution for 2 days, which we aggregate into edges by considering two nodes connected if they have been in each other's proximity for at least 20 min during the observation period. Nodes are split into two groups according to gender (39). The fourth one is Last.fm, a snapshot of a self-reported friendship network in a music-listening website. The network is split into two groups according to gender, and it includes only users for which this information is available (see table S1 for a summary of dataset features and the Supplementary Materials for more details). The fifth one is Directors (41), a network of board directors for publicly listed companies in Norway, with groups determined by gender. This dataset was collected to analyze the effect of an affirmative action law enforced in 2008, which required for each gender to have at least 40% representation in any position, with a link established if two people belong to the same board of directors.

### Homophily measures and model fitting

We have three ways of estimating choice homophily in the data: naive choice homophily estimation, choice homophily estimation with the mean-field approximation, and choice homophily estimation with ABC.

#### Estimation with the mean-field approximation

To estimate the amount of choice homophily in both groups of an empirical network, we solve the following inverse problem: Given a certain value  $c$  of triadic closure, we find the choice homophily parameters  $s_a$  and  $s_b$  in our mean-field solution of the model that lead to the observed edge fractions between and within groups,  $P_{aa}$ ,  $P_{bb}$ , and  $P_{ab}$ . We solve this inverse problem by setting  $dP_{aa}/dt = 0$  in Eq. 6 and solving for  $s_a(c)$  and  $s_b(c)$  given the matrix  $P$  or equivalently the transition matrix  $T$  (see the Supplementary Materials for a closed-form formula).

**Estimation with ABC**

ABC methods allow us to infer model parameters without knowing the explicit functional form of the likelihood function of the model. We use a recently developed method called Bayesian Optimization for Likelihood-Free Inference (37). The parameters we fit are the two choice homophily parameters  $s_a$  and  $s_b$ , the triadic closure probability  $c$ , and the transition matrix at the initial condition ( $T_0$ ). We use uniform priors for all of these variables. The relative group size  $n_a$  and average degree  $\langle k \rangle$  are set to the same value as in the data. The number of time evolution steps is set to  $t = 200$ . The expected model results are not sensitive to network size, and for computational reasons, network size is limited to  $N = 1000$  (this limitation might affect the variance of the posterior estimates). The method also requires defining statistics we want to fit, and for this, we use 14 different statistics related to connectivity of the groups, clustering coefficients, and core-periphery structure. For consistency, we find that creating synthetic networks with our model, the ABC fitting method is able to recover all parameter values (with some issues mostly focused around the unstable fixed points) (see the Supplementary Materials for details).

**Naive estimation**

In the naive estimation of choice homophily, we assume that all of the observed homophily is due to choice homophily. For groups of equal size, these estimates are given by  $T_{aa}$  and  $T_{bb}$  (i.e., elements of the transition matrix). Since group sizes will have an effect on the  $T$  matrix, we normalize the estimates separately for each group size (to keep choice homophily estimates comparable to each other). If we set  $c = 0$  in the mean-field estimation, we get a naive estimate of the biases that does not consider any triadic closure but, for example, corrects for a disproportionate amount of links observed within large groups as compared with small groups even if there is no intrinsic bias

$$\begin{aligned} s_a(0) = o_a &= \frac{n_b T_{aa}}{n_a(1 - T_{aa}) + n_b T_{aa}} \\ s_b(0) = o_b &= \frac{n_a T_{bb}}{n_b(1 - T_{bb}) + n_a T_{bb}} \end{aligned} \quad (7)$$

Note, however, that this feature of our estimation process leads to a different size correction than the Coleman homophily index (see Table 1) (60).

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/19/eaax7310/DC1>

**REFERENCES AND NOTES**

1. M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
2. J. M. McPherson, L. Smith-Lovin, Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *Am. Sociol. Rev.* **52**, 370–379 (1987).
3. W. Shrum, N. H. Cheek Jr., S. MacD, Friendship in school: Gender and racial homophily. *Sociol. Educ.* **61**, 227–239 (1988).
4. K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, N. Christakis, Tastes, ties, and time: A new social network dataset using facebook.com. *Soc. Net.* **30**, 330–342 (2008).
5. Y. Volkovich, D. Laniado, K. E. Kappler, A. Kaltenbrunner, Gender patterns in a large online social network, in *Proceedings of the Social Informatics: 6th International Conference (SocInfo'14), Barcelona, Spain, 11 to 13 November 2014*, L. M. Aiello, D. McFarland, Eds. (Springer, 2014), Lecture Notes in Computer Science, vol. 8851, pp.139–150.
6. D. Zeltzer, Gender homophily in referral networks: Consequences for the medicare physician earnings gap (21 February, 2017), available at SSRN; <https://ssrn.com/abstract=2921482> or <http://dx.doi.org/10.2139/ssrn.2921482>.

7. M. D. Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, W. Quattrociocchi, Echo chambers: Emotional contagion and group polarization on facebook. *Sci. Rep.* **6**, 37825 (2016).
8. A. L. Schmidt, F. Zollo, M. D. Vicario, A. Bessi, A. Scala, G. Caldarelli, H. E. Stanley, W. Quattrociocchi, Anatomy of news consumption on facebook. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3035–3039 (2017).
9. P. DiMaggio, F. Garip, Network effects and social inequality. *Annu. Rev. Sociol.* **38**, 93–118 (2012).
10. T. A. DiPrete, G. M. Eirich, Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annu. Rev. Sociol.* **32**, 271–297 (2006).
11. S. Currarini, M. O. Jackson, P. Pin, An economic model of friendship: Homophily, minorities, and segregation. *Econometrica* **77**, 1003–1045 (2009).
12. S. Currarini, J. Matheson, F. Vega-Redondo, A simple model of homophily in social networks. *Eur. Econ. Rev.* **90**, 18–39 (2016).
13. F. Karimi, M. Géniois, C. Wagner, P. Singer, M. Strohmaier, Homophily influences ranking of minorities in social networks. *Sci. Rep.* **8**, 11077 (2018).
14. P. M. Blau, A macrosociological theory of social structure. *Am. J. Sociol.* **83**, 26–54 (1977).
15. H. Louch, Personal network integration: Transitivity and homophily in strong-tie relations. *Soc. Net.* **22**, 45–64 (2000).
16. G. Kossinets, D. J. Watts, Empirical analysis of an evolving social network. *Science* **311**, 88–90 (2006).
17. G. Kossinets, D. J. Watts, Origins of homophily in an evolving social network. *Am. J. Sociol.* **115**, 405–450 (2009).
18. T. C. Schelling, Dynamic models of segregation. *J. Math. Sociol.* **1**, 143–186 (1971).
19. D. Vinković, A. Kirman, A physical analogue of the schelling model. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19261–19265 (2006).
20. R. Toivonen, L. Kovanen, M. Kivela, J.-P. Onnela, J. Saramäki, K. Kaski, A comparative study of social network models: Network evolution models and nodal attribute models. *Soc. Net.* **31**, 240–254 (2009).
21. N. Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, D. Song, Evolution of social-attribute networks: Measurements, modeling, and implications using google+, in *Proceedings of the 2012 Internet Measurement Conference* (Organization ACM, year 2012) pp. 131–144.
22. P. Klimek, S. Thurner, Triadic closure dynamics drives scaling laws in social multiplex networks. *New J. Phys.* **15**, 063008 (2013).
23. G. Bianconi, R. K. Darst, J. Iacovacci, S. Fortunato, Triadic closure as a basic generating mechanism of communities in complex networks. *Phys. Rev. E* **90**, 042806 (2014).
24. M. S. Granovetter, The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
25. T. A. B. Snijders, Statistical models for social networks. *Annu. Rev. Sociol.* **37**, 131–153 (2011).
26. J. M. Kumpula, J.-P. Onnela, J. Saramäki, K. Kaski, J. Kertész, Emergence of communities in weighted networks. *Phys. Rev. Lett.* **99**, 228701 (2007).
27. J. Davidsen, H. Ebel, S. Bornholdt, Emergence of a small world from local interactions: Modeling acquaintance networks. *Phys. Rev. Lett.* **88**, 128701 (2002).
28. A. Vázquez, Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **67**, 056104 (2003).
29. M. Marsili, F. Vega-Redondo, F. Slanina, The rise and fall of a networked society: A formal model. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1439–1442 (2004).
30. R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen, K. Kaski, A model for social networks. *Phys. A* **371**, 851–860 (2006).
31. L. Peel, J.-C. Delvenne, R. Lambiotte, Multiscale mixing patterns in networks. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4057–4062 (2018).
32. G. T. Cantwell, M. E. J. Newman, Mixing patterns and individual differences in networks. *Phys. Rev. E* **99**, 042306 (2019).
33. J. Saramäki, E. A. Leicht, E. López, S. G. B. Roberts, F. Reed-Tsochas, R. I. M. Dunbar, Persistence of social signatures in human communication. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 942–947 (2014).
34. S. P. Borgatti, M. G. Everett, Models of core/periphery structures. *Soc. Net.* **21**, 375–395 (2000).
35. P. Holme, Core-periphery organization of complex networks. *Phys. Rev. E* **72**, 046111 (2005).
36. P. Rombach, M. A. Porter, J. H. Fowler, P. J. Mucha, Core-periphery structure in networks (revisited). *SIAM Rev.* **59**, 619–646 (2017).
37. M. U. Gutmann, J. Corander, Bayesian optimization for likelihood-free inference of simulator-based statistical models. *J. Mach. Learn. Res.* **17**, 1–47 (2016).
38. A. L. Traud, P. J. Mucha, M. A. Porter, Social structure of Facebook networks. *Phys. A* **391**, 4165–4180 (2012).
39. J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaghiotto, W. Van den Broeck, C. Régis, B. Lina, P. Vanhems, High-resolution measurements of face-to-face contact patterns in a primary school. *PLOS ONE* **6**, e23176 (2011).



40. L. A. Adamic, N. Glance, The political blogosphere and the 2004 u.s. election: Divided they blog, in *Proceedings of the 3rd International Workshop on Link Discovery*, series and number LinkKDD '05 (ACM, New York, NY, USA, 2005) pp. 36–43.
  41. C. Seierstad, T. Opsahl, For the few not the many? The effects of affirmative action on presence, prominence, and social capital of women directors in norway. *Scand. J. Manag.* **27**, 44–54 (2011).
  42. J. M. McPherson, J. R. Ranger-Moore, Evolution on a dancing landscape: Organizations and networks in dynamic blau space. *Soc. Forces* **70**, 19–42 (1991).
  43. S. Chen, A. Mira, J.-P. Onnela, Flexible model selection for mechanistic network models via super learner. *J. Complex Networks* **2019**, cnz024 (2019).
  44. D. V. Foster, J. G. Foster, P. Grassberger, M. Paczuski, Clustering drives assortativity and community structure in ensembles of networks. *Phys. Rev. E* **84**, 066117 (2011).
  45. T.-Q. Peng, Assortative mixing, preferential attachment, and triadic closure: A longitudinal study of tie-generative mechanisms in journal citation networks. *J. Informet.* **9**, 250–262 (2015).
  46. A. Mele, A structural model of homophily and clustering in social networks, (September 3, 2017), note available at SSRN: <https://ssrn.com/abstract=3031489> or <http://dx.doi.org/10.2139/ssrn.3031489>.
  47. A. Mele, A structural model of dense network formation. *Econometrica* **85**, 825–850 (2017).
  48. T. A. B. Snijders, P. E. Pattison, G. L. Robins, M. S. Handcock, New specifications for exponential random graph models. *Sociol. Methodol.* **36**, 99–153 (2006).
  49. T. Gross, B. Blasius, Adaptive coevolutionary networks: A review. *J. R. Soc. Interface* **5**, 259–271 (2008).
  50. T. Gross, H. Sayama, *Adaptive Networks: Theory, Models and Applications* (Springer, 2009).
  51. G. Zschaler, Adaptive-network models of collective dynamics. *Eur. Phys. J. Special Topics* **211**, 1–101 (2012).
  52. J. Ito, K. Kaneko, Spontaneous structure formation in a network of chaotic units with variable connection strengths. *Phys. Rev. Lett.* **88**, 028701 (2001).
  53. S. Bornholdt, T. Röhl, Self-organized critical neural networks. *Phys. Rev. E* **67**, 066118 (2003).
  54. P. Holme, M. E. J. Newman, Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys. Rev. E* **74**, 056108 (2006).
  55. F. Vazquez, J. C. González-Avella, V. M. Eguiluz, M. S. Miguel, Time-scale competition leading to fragmentation and recombination transitions in the coevolution of network and states. *Phys. Rev. E* **76**, 046120 (2007).
  56. G. Iñiguez, J. Kertész, K. K. Kaski, R. A. Barrio, Opinion and community formation in coevolving networks. *Phys. Rev. E* **80**, 066119 (2009).
  57. P. Klimek, M. Diakonova, V. M. Eguiluz, M. S. Miguel, S. Thurner, Dynamical origins of the community structure of an online multi-layer society. *New J. Phys.* **18**, 083045 (2016).
  58. T. Raducha, B. Min, M. S. Miguel, Coevolving nonlinear voter model with triadic closure. *EPL* **124**, 30001 (2018).
  59. R. Axelrod, The dissemination of culture: A model with local convergence and global polarization. *J. Conflict Resolut.* **41**, 203–226 (1997).
  60. J. S. Coleman, Relational analysis: The study of social organizations with survey methods. *Hum. Organ.* **17**, 28–36 (1958).
  61. Wolfram Research, Inc. Mathematica, version 11.1, 2017.
  62. G. Timár, S. N. Dorogovtsev, J. F. F. Mendes, Scale-free networks with exponent one. *Phys. Rev. E* **94**, 022302 (2016).
- Acknowledgments:** Numerical simulations were performed using computer resources within the Aalto University School of Science “Science-IT” project. J.U.-C. and M.K. thank the support from the Academy of Finland through the ECANET-project (No. 32779). We thank R. Dunbar, J. Saramäki, R. Hari, and M. San Miguel for helpful feedback. **Author contributions:** All authors contributed to the writing of the article and interpretation of the results. A.A., G.J., and M.K. contributed to the simulation and analysis software and deriving the mathematical results. A.A., J.U.-C., and M.K. prepared and analyzed the data. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The Last.fm network dataset used in this study has been uploaded to Zenodo and is available under the accession number 3726824. The other data sets have been downloaded from publicly available data repositories. Key parts of the code used in this study are available from M.K. upon reasonable request.
- Submitted 17 April 2019  
Accepted 19 February 2020  
Published 8 May 2020  
10.1126/sciadv.aax7310
- Citation:** A. Asikainen, G. Iñiguez, J. Ureña-Carrión, K. Kaski, M. Kivelä, Cumulative effects of triadic closure and homophily in social networks. *Sci. Adv.* **6**, eaax7310 (2020).

## Cumulative effects of triadic closure and homophily in social networks

Aili Asikainen, Gerardo Iñiguez, Javier Ureña-Carrión, Kimmo Kaski and Mikko Kivela

*Sci Adv* 6 (19), eaax7310.  
DOI: 10.1126/sciadv.aax7310

### ARTICLE TOOLS

<http://advances.sciencemag.org/content/6/19/eaax7310>

### SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2020/05/04/6.19.eaax7310.DC1>

### REFERENCES

This article cites 55 articles, 6 of which you can access for free  
<http://advances.sciencemag.org/content/6/19/eaax7310#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).